Universiteit Leiden

Traces of 'Pre-Indo-Iranian'

Chronological Layers and Structural Characteristics of Early Indo-Iranian Loanwords

Axel Palmér

S2091135

Thesis submitted in partial
fulfillment for the degree of
Master of Arts – Linguistics (Research)

Leiden University Centre for Linguistics

Faculty of Humanities

Leiden University

July 2019

Supervisor: Prof.dr. A.M. Lubotsky

Second Reader: Dr. A. Kloekhorst

Contents

Acknowledgements

First of all, I would like to thank Prof.dr. Sasha Lubtosky, who supervised this thesis and always found time for our (sometimes lengthy) meetings in his busy schedule. I would also like to thank Prof.dr. Guus Kroonen, who taught me crucial background knowledge relevant to this thesis and offered me so many opportunities to realize my potential. I am grateful to Dr. Alwin Kloekhorst, who supported my academic development during the ResMa programme in many ways. Sincere thanks go to Dr. Tijmen Pronk, whose many excellent classes in Indo-European linguistics I attended with pleasure.

I would also like to thank all the staff and students of LUCL, several of whom I consider among my closest friends. Especially Oscar Billing, with whom I spent almost every day of every week studying in the MPhil room, and enjoyed countless discussions about our work and (thankfully!) other matters, has my sincere gratitude. I thank my parents Anne and Kjell for their loving support throughout the years. Finally, I would like to express my loving thanks to my partner Linde for moving with me to Leiden and building our lives together here.

Abstract

In this thesis, I study loanwords of unknown origin in Proto-Indo-Iranian and early Post-Proto-Indo-Iranian. According to the Central Asian Substrate Hypothesis, Indo-Iranian speakers migrated to Central Asia around 2000 BCE and came into contact with the agricultural BMAC civilization, which resulted in a body of loanwords into Proto-Indo-Iranian, borrowed from the language of the BMAC people. Following a methodology for identifying non-Indo-European vocabulary in Indo-European languages, I argue that 74 out of 103 previously suggested loanwords can plausibly be analyzed as loanwords (chapter 3). Only a handful of these may have been borrowed from known languages. After establishing the relative chronology of Proto-Indo-Iranian sound changes (chapter 2), I divide the 74 early Indo-Iranian loanwords into chronological layers based on when they were borrowed (chapter 3-4). I argue that 21 words were borrowed after the disintegration of Proto-Indo-Iranian. Moreover, I argue that many of the remaining 53 loanwords that are reconstructable to Proto-Indo-Iranian were borrowed towards the end of this stage. Finally, I integrate the chronological layers into my analysis of structural characteristics of early Indo-Iranian loanwords and describe two new phonological patterns of loanwords (chapter 5). The fact that many loanwords are shown to have been borrowed in late PII or Post-PII, i.e. after Indo-Iranian speakers migrated to Central Asia, is consistent with the timeline of the Central Asian Substrate Hypothesis. Second, the newly discovered phonological characteristics provide additional support for the Central Asian Substrate Hypothesis, since they increase the likelihood that most loanwords originate in the same language.

Abbreviations

Languages

A. = Ashkun

Akk. = Akkadian

Arab. = Arabic

Av. = Avestan

Bactr. = Bactrian

Bal. = Balochi

Baxt. = Baxtiari

Bur. = Burušaski

Elam. = Elamite

Finn. = Finnish

GAv. = Gathic Avestan

Germ. = German

Goth = Gothic

Gr. = Ancient Greek

Hitt. = Hittite

Hung. = Hungarian

Ind. = Indic

Ir. = Iranian

Ke. = Ket dialect (Selkup)

Khot. = Khotanese

Km. = Kamviri

Kt. = Kataviri

Lat. = Latin

LAv. = Late Avestan

Lith. = Lithuanian

Mar. = Mari

MHG = Middle High German

MI = Middle Indic

MiP = Middle Persian

MIr. = Middle Irish

MoP = Modern Persian

N = Narym dialect (Selkup)

Nur. = Nuristani

OAv. = Old Avestan

OBur. = Old Burušaski

OCS = Old Church Slavonic

OHG = Old High German

OIr. = Old Irish

ON = Old Norse

OP = Old Persian

Orm. = Ormuri

Oss. = Ossetic

Par. = Parāčī

Parth. = Parthian

PCelt. = Proto-Celtic

PFU = Proto-Finno-Ugric

PFV = Proto-Finno-Volgaic

PGm. = Proto-Germanic

Phl. = Pahlavi

PIE = Proto-Indo-European

PII = Proto-Indo-Iranian

PInd. = Proto-Indic

PIr. = Proto-Iranian

Pkt. = Prakrit

Pre-II = Source language(s) of loanwords in Indo-Iranian

Pre-PII = Pre-Proto-Indo-Iranian

Pto. = Pashto

PToch. = Proto-Tocharian

PU = Proto-Uralic

PUg. = Proto-Ugric

Rom. = Romani

Ru. = Russian

Sa. = Sami

Sam. = Samoyedic

SCr. = Serbo-Croatian

Selk. = Selkup

Skt. = Sanskrit

Sogd. = Sogdian

Šu. = Šughni

ToA = Tocharian A

ToB = Tocharian B

Ty. = Tym dialect (Selkup)

W. = Waigali

V. = Vasi-vari

Wakh. = Wakhi

Veps. = Vepsian

Yagh. = Yaghnōbī

Yazg. = Yazgulyami

Y-M = Yidgha-Munji

aor. = aorist

AV = Atharvaveda

BL = Bartholomae's Law

BMAC = Bactria-Margiana Archaeological Complex

BrL = Brugmann's Law

dat. = dative

dial. = dialectal

gen. = genitive

GL = Grassmann's Law

inj. = injunctive

Irr. = Irregular

Lim. = Limited

nom. = nominative

Other abbreviations

pl. = plural

pres. = present

Rem. = Remarkable

RV = Ṛgveda

sg. = singular

Symbols

> = regularly develops into

< = regularly derives from

>> = develops by analogy into, borrowed into

<< = developed by analogy from, borrowed from

: = corresponds to

1. Introduction

The topic of this thesis is loanwords of unknown origin in early Indo-Iranian. In other words, the thesis treats early Indo-Iranian words that are neither inherited from Proto-Indo-European (PIE), nor innovated within Indo-Iranian based on inherited roots, but borrowed from languages with which Indo-Iranian came into contact in prehistory. I use Pre-Indo-Iranian (Pre-II) as a cover term for the unknown donor language(s) of early Indo-Iranian loanwords. Included within the scope of "early Indo-Iranian" vocabulary is that of Proto-Indo-Iranian (PII). However, the term also includes words shared between Indic[1] and Iranian (and Nuristani) that cannot be reconstructed to PII, but nevertheless must have entered the Indo-Iranian languages at an early date, shortly after the disintegration of PII.

1.1. The goal of the thesis

The thesis has three main goals. The first goal is to establish which early Indo-Iranian words are loanwords rather than inherited from PIE. With a few exceptions, all previously suggested early Indo-Iranian loanwords are disputed, and alternative Indo-European (IE) etymologies have been proposed. Therefore, an essential step of this study is to evaluate the proposals of previous literature, to determine for each proposed loanword whether IE origin can be excluded or not.

The second goal is to classify early Indo-Iranian loanwords into chronological layers. The purpose is to determine how diverse vs. uniform the early Indo-Iranian loanwords are in terms of relative time of borrowing. Based on established regular phonological correspondences, it can be determined whether possible cognates in Indic and Iranian go back to PII or not. This allows loanwords to be classified as PII or Post-PII.

However, the goal of the thesis is also to determine whether different chronological layers of loanwords exist within PII. Based on the relative chronology of PII sound changes, I will investigate whether, on the one hand, some words must have undergone certain PII sound changes, and therefore must have been borrowed *before* these occurred, or, on the other hand, some words cannot have undergone certain PII sound changes, and therefore must have been borrowed *after* these occurred.

The third goal is to describe patterns in the phonology and morphology of early Indo-Iranian loanwords. The purpose is to increase our understanding of the Pre-II language(s) with which early Indo-Iranian came into contact. Besides being an intriguing question in itself, this is a crucial step in the methodology of studying loanwords of unknown origin. If phonological

[1] I use "Indic" instead of the more traditional term "Indo-Aryan".

and morphological patterns in the loanword corpus are found, this in itself lends additional support in favor of postulating an unknown source language. Since every language has a phoneme inventory and a phonotactic system, phonological patterns among loanwords, which cannot be explained by the phonology of the recipient language, imply that a foreign linguistic system is fossilized behind them. Similarly, recurring morphological traits such as foreign suffixes imply an underlying morphological system.

Below, previous literature on Indo-Iranian loanwords and related topics will be discussed, followed by a more detailed formulation of research questions.

1.2. The Indo-Iranian languages

Indic and Iranian are the two major sub-branches of Indo-Iranian. For historical linguistic purposes, the Old Indo-Iranian languages (Vedic) Sanskrit, Avestan and Old Persian are the most important sources. However, since the Old Iranian corpus is limited, Middle and Modern Iranian languages also play a crucial part in Indo-Iranian historical linguistics. Evidence from Middle and Modern Indic languages is less commonly seen in the literature, but sometimes preserves archaic features that Vedic Sanskrit had lost.

The Nuristani languages of Afghanistan are commonly considered to form a third sub-branch of Indo-Iranian. However, the internal relationship between Indic, Iranian and Nuristani remains unclear. Scholars have argued that Nuristani forms an intermediate subgroup with Iranian (Mayrhofer, 1983) or Indic (Blažek & Hegedűs, 2012, p. 43), or that Nuristani was the first branch to split off from PII, and that Indic and Iranian constitute a subgroup (Hegedűs, 2012, p. 145). As there is no general consensus, I will assume, for the purposes of this thesis, that Nuristani is equally closely related to Indic as it is to Iranian.

1.3. Methodology and hypotheses of previous literature

1.3.1. Non-IE elements in IE languages

A series of publications have developed a methodology for identifying and systematically studying non-IE vocabulary of unknown origin in ancient IE languages (cf. Kuiper, 1991, 1995; Beekes, 1996, 2010; Schrijver, 1997; Lubotsky, 2001b). The methodology for identifying prehistoric loanwords is based on five criteria:

1) Limited geographical distribution
2) Irregular phonological correspondences
3) Remarkable morphology
4) Remarkable phonology
5) Specific semantics

The first criterion applies if a word is restricted to one branch of IE or several branches that are (or were in prehistory) spoken close to one another. This criterion stands out, since it is in most cases a prerequisite for postulating borrowing in the first place. While in theory straightforward, it is often the case that an IE etymology has been suggested, but that its validity is disputed. Therefore, careful etymological analysis is always necessary.

The second criterion applies if a word is attested in two or more IE languages, but does not show regular sound correspondences based on what we know from the inherited vocabulary, and is therefore not reconstructable to PIE.

The third criterion applies if a word shows a derivational pattern or a suffix that is marginal or absent in the inherited vocabulary. It is important to remember that loanwords eventually adapt to the native morphology, generally following a productive pattern. Thus, a nominal suffix *-*bso*- would be a clear indication of a loanword, in spite of the fact that it is thematic, since *-*bs*- is clearly non-IE. In the case of verbs, loanwords are expected to belong to a productive class, e.g. thematic rather than athematic.

The fourth criterion applies if a word contains phonemes or phonemic sequences that are marginal or absent in the inherited vocabulary, e.g. two mediae in the root, the vowel **a* (depending on one's views on PIE phonology).

The fifth criterion applies if a word is particularly "borrowable" due to its semantics. For example, words for cultural phenomena as well as flora and fauna are more easily borrowed than "basic" vocabulary (Tadmor et al., 2010).

As pointed out by Schrijver (1997, p. 296), none of these criteria is in itself decisive when it comes to identifying loanwords. Limited geographical distribution may be accidental, irregular phonological correspondences may be the result of analogy, remarkable morphology and phonology may represent hitherto unknown inherited features, and words with specific semantics may of course be inherited. Therefore, loanwords should ideally be identified based on two or more of these criteria.

Besides the five criteria above, a crucial methodological principle of identifying and studying loanwords of unknown origin is the notion of recurring irregularities and structural characteristics (Schrijver, 1997, p. 296). In isolation, phonological and morphological features can be used to identify loanwords, but when the same irregularity or foreign-looking structural characteristic is found in several words, it drastically increases the plausibility that they are loanwords. Recurring irregularities and structural characteristics can also indicate which loanwords originate in the same language.

An aspect that set the studies cited above aside from other studies of lexical borrowing is that the donor language, or "substratum" language as it is often called, is unrecorded and unknown, and is thus only preserved in the loanwords themselves. The crucial step forward that these studies represent, therefore, is the acknowledgement that such unknown prehistoric languages can, and should, be studied in historical linguistics. To some, the notion of postulating linguistic entities based on loanwords has seemed too methodologically problematic to be taken seriously. Indeed, there is a certain risk that a new substrate language is used as a *magic wand* each time a scholar is unable to explain an irregular correspondence or an obscure lexeme. This would be similar to postulating an additional phoneme to explain a single cognate set. However, this criticism is only valid when borrowing is used as an *ad hoc* explanation to a particular problem. When, on the other hand, the methodology outline above is followed, the situation changes, because if recurring irregularities and structural patterns are observed in loanwords, postulating one substrate language can provide a solution to many unrelated problems at once.

1.3.2. Non-IE elements in Indo-Iranian

Kuiper (1991) studied non-IE elements in Vedic Sanskrit. He identified hundreds of loanwords, along with various morphosyntactic features which Sanskrit acquired in contact with non-IE languages in South Asia. One of the most salient types of loanwords in Old Indic is the so-called *CVCV̄CV* type, cf. Skt. *caṣála-* 'knob', trisyllabic words with a medial long vowel or diphthong. This structure is rare in IE words, since these normally consist of a root and a suffix, both usually monosyllabic.

In a series of publications, Witzel (1995; 1999a; 1999b; 2003; 2006; 2009) investigated loanwords in Vedic Sanskrit, Indo-Iranian, and the linguistic (pre-)history of South Asia in general. The main contribution of Witzel's work lies in the early Indo-Iranian loanwords that he proposes, as well as his discussion of some structural characteristics of these words. Moreover, Witzel (2003) puts early Indo-Iranian loanwords in a broader perspective, incorporating possible shared borrowings in other languages, such as Burušaski, Dravidian, Anatolian, Greek, and languages of the Caucasus. A recurring irregularity in early Indo-Iranian loanwords proposed by Witzel (2003, p. 45) is an *r*/*n*-alternation, argued to reflect dialectal variation in the substrate language(s).

Although the significance of Witzel's work should not be underestimated, it suffers from the occasional inclusion of words with clear or likely PIE origin (e.g. PII *$*mad^hu$-* 'honey',

2003, p. 13) as well as the lack of methodological stringency in postulating common origins of loanwords (cf. chapter 3 on **ganTuma-*).

Lubotsky (2001b) systematically investigated vocabulary that is shared between Indic and Iranian, but not found in other IE languages, i.e. words that fulfill the first criterion of the methodology outlined above. In this material, he identified 55 loanwords. Most show regular correspondences and can be reconstructed to PII, whereas others show irregular correspondences. Additionally, 23 verbal roots isolated to Indo-Iranian were listed as possible loanwords, although Lubotsky deemed it impossible to distinguish between inherited and borrowed verbs (2001b, p. 310).

Lubotsky realized that several structural characteristics of PII loanwords are identical to those of specifically Indic loanwords, as described by Kuiper (1991). These features include the *CVCV̄CV* type, voiceless aspirates, frequent palatal stops, frequent clusters with **-s-*, the cluster **-rṷ-*, and the suffixes *-ig-*, *-pa-*, and *-h-* (Lubotsky, 2001b, p. 305). Based on this similarity, Lubotsky proposed that PII and Indic loanwords originate in the same language or related languages, spoken in Central Asia on the one hand, and in the Punjab on the other. This hypothesis will be further discussed below.

Furthermore, Lubotsky (2001b, p. 306) argued that loanwords with the irregular correspondence Indic *s* : Iranian *s* were first borrowed into Indic and then transmitted to Iranian.

Kümmel (2017) collected Indo-Iranian vocabulary related to animal husbandry and agriculture. He found that most terms for domesticated animals are inherited, whereas several terms for cereals and other domesticated plants are not. Words in the latter group are potential early Indo-Iranian loanwords.

As the literature review shows, non-IE vocabulary in Indo-Iranian has received some attention from previous scholarship. However, it is not yet fully integrated into Indo-Iranian lexicography, as is evident from the *Etymologisches Wörterbuch des Altindoarischen* (*EWAia*). Although it sometimes acknowledges the possibility of borrowing, *EWAia* does not take into account the systematic study of loanwords of unknown origin. Partly, this may be because some of the aforementioned studies were not yet available at the time of publication, but the dictionary also shows skepticism towards such proposals. This is expressed by the employment of *ad hoc* explanations, such as how the *s* of Skt. *sūcī́* 'needle' is said to be analogical from *sīv-* 'to sew', in order to explain the irregular correspondence to Ir. **ćūkā-* / **ćūčī-* (*EWAia* II, p. 739). In this case, assuming borrowing is preferable, since the Indo-Iranian word for 'needle' fulfills three of five criteria of a loanword: limited geographical distribution, irregular

phonological correspondences, and specific semantics. In other cases, *EWAia* simply dismisses proposed borrowings as "unnecessary" (II, 241) or "implausible" (II, p. 151).

EWAia only considers borrowing as a possibility when a known source language exists. For some words, an Austroasiatic, Dravidian or Uralic source has been suggested. However, since these languages are known to have borrowed from Indo-Iranian, the direction of borrowing is often difficult to prove.

1.3.3. Indo-Iranian origins: homeland and migration

Studying prehistoric language contact is one of the main pieces of linguistic evidence for prehistoric migrations and language spread. A loanword from one language to another suggests that speakers of the donor language and recipient language were in contact, which usually[2] presupposes geographical proximity of the speaker communities. However, when it comes to loanwords of unknown origin, the situation is somewhat reversed: the prehistoric location of the recipient language delimits the possible locations of the donor language(s). Therefore, a short review of the current views on the origin of the Indo-Iranian languages and their speakers is due.

The Indo-Iranian branch originates in PIE. The question of when and where PIE was spoken has generated two fundamentally different hypotheses. The *Steppe Hypothesis* places the IE homeland in the nomadic Yamnaya culture on the Pontic-Caspian Steppe around 3500-3000 BCE (Mallory, 1989; Anthony, 2007). This view has been rivalled by the *Anatolian Hypothesis* (Renfrew, 1987), which claims that Proto-Indo-European dispersed with the spread of agriculture from Anatolia around 7000 BCE.

Recently, strong evidence for large scale migrations from Yamnaya steppe populations into Europe and Asia was offered by geneticists (Haak et al., 2015), favoring the Steppe Hypothesis. The Steppe Hypothesis is also favored by the linguistic evidence, since PIE had terminology for wheeled vehicles, which were invented after 4000 BCE, consistent with the chronology of the Yamnaya culture (Anthony & Ringe, 2015). From the IE homeland on the Pontic-Caspian Steppe, Indo-Iranian speakers eventually migrated all the way to South and Western Asia, as evidenced by the high degree of Steppe Ancestry in the DNA of modern Indo-Iranian speaking populations (Damgaard et al., 2018).

Kuz'mina (2007) approached the question of the Indo-Iranian migration and homeland from an archaeological perspective, incorporating linguistic and anthropological evidence to

[2] In some cases, words are more mobile, as it were, than the speakers who use them; *Wanderwörter* can spread from one community to the other, without the original source language being in contact with all subsequent recipients.

some extent. She argued that prehistoric Indo-Iranian speakers inhabited the Sintashta (2100-1800 BCE) and Andronovo (2000-900 BCE) cultures. By retracing cultural development in the archaeological record, Kuz'mina (2007, p. 205) found that the Economic and Cultural Type (ECT) of the Indo-Iranian speaking Sauromatians and Saka cultures descends directly from the Andronovo cultures, which in turn succeeded the older Sintashta culture. The pastoral Sintastha culture, situated to the south-east of the Ural Mountains, is thus a plausible Indo-Iranian homeland. This hypothesis is also supported by the many PII loanwords into (Proto-)Uralic (Koivulehto, 2001). An archaeolinguistic argument is that chariotry, for which several terms are reconstructable to PII (Witzel, 2001), originated in the Sintastha culture (Kuznetsov, 2006).

From their homeland in the Sintashta culture around 2000 BCE, Indo-Iranian speakers spread southwards to the areas in which Indo-Iranian languages are still spoken today. Moreover, Indo-Iranian languages continued to be spoken in Central Asia, with some groups (e.g. the Alans) spreading westwards to Europe.

1.3.4. The BMAC culture and the "Central Asian Substrate Hypothesis"

The Bactria-Margiana Archaeological Complex (BMAC) denotes a Central Asian Bronze Age civilization east of the Caspian Sea, to the south of the Andronovo horizon. With its origins in the first half of the 3rd millennium, the BMAC civilization was at its peak ca. 2400-1700 BCE (Francfort, 2005, p. 260). Around its fortified settlements, the BMAC people practiced irrigation farming, cultivating wheat, barley, lentil, pea, grass pea, chick pea, grape, apple and flax (Spengler et al., 2014).[3] Domesticated animals include cattle, sheep, camels, pigs and donkeys (Witzel, 2000, p. 4). Especially interesting is the archaeological evidence of groups of mobile pastoralists, who lived outside of the fortified settlements, and whose animals may have grazed the fields of the farmers after the harvest (Spengler et al., 2014, p. 808, 816). According to Spengler et al. (ibid.), the fact that animal dung was used as fuel by the farmers indicates non-hostile contacts between the groups. Since no written documents have been excavated from the BMAC civilization, the identity of its language(s) is unknown.

Witzel (2003) and Lubotsky (2001b) have elaborated the hypothesis that most loanwords of unknown origin in PII originate in an unknown language of the BMAC civilization. I refer to this as the "Central Asian Substrate Hypothesis". The hypothesis combines archaeological and linguistic arguments into a plausible scenario.

[3] Whether millet was cultivated is uncertain. Spengler et al. (2014, p. 817) did not find evidence for millet in southern Central Asia earlier than the Iron Age, but notes that this could be accidental.

First, if the Sintashta culture is accepted as the PII homeland, the early Indo-Iranians must have come into contact with BMAC groups as they spread southwards, their presence being attested in the Near East in the 16-15th centuries BCE (Mallory, 1989, p. 38), perhaps even as early as the 18th century (Kroonen et al., 2018, p. 12). Archaeological evidence for steppe influence is seen in BMAC pottery (Witzel, 2000, p. 7). Moreover, as noted above, there is evidence for temporary settlements of pastoralists in the BMAC area (Salvatori, 2008, p. 64), which could very well have belonged to Indo-Iranian speakers from the pastoral Andronovo cultures. Furthermore, the spread of Indo-Iranian to South and Western Asia has been connected to the BMAC cultural influence in these areas in the second half of the 2nd millennium (Witzel, 2000, p. 8). According to Mallory (1998), the Indo-Iranians eventually assimilated to the culture of the BMAC and transmitted it to the south, which would explain why there is little direct cultural influence from the steppe south of the BMAC. Language contact between Indo-Iranian speaking steppe populations and BMAC farmers is thus likely.

Second, the semantics of some early Indo-Iranian loanwords make a BMAC origin likely. Witzel (2003, p. 25) mentions **Hustra-* 'camel', **kHara-* 'donkey' and **išt(i)-* 'brick' as the clearest cases, since the camel and donkey were present in the BMAC culture, but not in the steppe where Indo-Iranian originates, and since BMAC settlements are built with bricks. Lubotsky (2001b, p. 307) mentions **i̯au̯ī̯ā-* 'canal', which can be connected to the irrigation farming technique of the BMAC, as well as several other terms referring to building technology.

Thus, the Central Asian Substrate Hypothesis is supported both by the archaeological evidence for contact between steppe populations and BMAC populations, and by the fact that several loanwords seem to reflect BMAC material culture. However, many of the previously proposed early Indo-Iranian loanwords cannot be directly compared to the BMAC culture, since they denote abstract concepts or other notions not visible in the archaeological record. These are instead hypothesized to originate in the BMAC language(s) simply because they are PII loanwords without a known source.

1.4. Research Questions

As we have seen, previous research has identified a group of early Indo-Iranian loanwords, discovered a number of structural characteristics of this group of words, and put them in an archaeolinguistic context by proposing a plausible language contact scenario. The linguistic, rather than archaeological perspective, is the main focus of this thesis, i.e. the loanwords themselves and the language(s) they may have come from.

1.4.1. Identifying loanwords

The words recognized as early Indo-Iranian loanwords differ depending on the author, and it is therefore necessary to reevaluate previous proposals. This serves two purposes: firstly, insights from Lubotsky (2001b), Witzel (2003) and Kümmel (2017) will be synthesized to provide a more complete picture of the material. Secondly, since some works are inexplicit with regards to methodology, it is unclear whether all proposed loanwords have been analyzed under the same criteria. By applying the methodology outlined in 1.3.1 to all previously proposed loanwords, the analysis will become more explicit and the results more uniform.

Unlike in previous literature, verbs will also be taken into account. In principal, verbs can be analyzed according to the same methodology as nouns. However, as the extensive Indo-Iranian verbal morphology requires verbs to be analyzable as monosyllabic roots, borrowed verbs often require more adaptation to the native system than nouns, and are thus more difficult to differentiate from the inherited vocabulary. The presence of archaic derivations, e.g. nasal infix present, strongly suggest IE origin. Borrowing will only be considered when archaic derivations are absent.

In some cases, a known source language of a loanword has been proposed. Such proposals will be evaluated, since a plausible known source would be a strong argument for postulating borrowing.

1.4.2. Chronological layers

An aspect of early Indo-Iranian loanwords that has not been systematically taken into account in earlier literature is the time of borrowing. The development of Indo-Iranian can be divided into chronological stages following its separation from the rest of IE: Pre-PII, PII, and Post-PII.[4] Generally, the Central Asian Substrate is identified with loanwords in PII (Lubotsky, 2001b, p. 301). Yet, Witzel (2003) also assigns Post-PII loanwords, some of which may be very late borrowings, to the Central Asian Substrate. This is problematic. Chronological stages reflect temporal development, but indirectly often reflect geographical movement, since a principal cause for the disintegration of PII must have been the geographical separation of speaker communities that would later become Indic, Iranian and Nuristani. Therefore, loanwords in different chronological layers should not *a priori* be lumped together.

Lubotsky reconstructs loanwords with irregular correspondences to PII, arguing that the proto-language was a continuum of differentiated dialects that nonetheless underwent shared

[4] After this, of course, follows Proto-Iranian and Proto-Indic and their respective historical development until the modern period, but this goes beyond the scope of this discussion.

innovations (2001b, p. 302). On the one hand, it is true that the notion of clearly definable linguistic entities such as "PII" does not fully capture the complexity of language development. On the other hand, by not assuming a uniform PII, methodological stringency is decreased, since the importance of regular sound change is downplayed. The possibility of variation within PII should not be excluded, but irregular correspondences in loanwords nevertheless suggest that, at the time of borrowing, the linguistic community was disintegrating, implying greater distance between speaker groups and their dialects. Therefore, regular vs. irregular correspondences is a relevant point of division when it comes to loanwords, and is one of the main research questions of the thesis.

More specifically, I will investigate which loanwords can be reconstructed to PII, and which cannot, based on current knowledge about the regular correspondences between the branches of Indo-Iranian. Moreover, I will explore whether Post-PII loanwords were borrowed independently from the same source, transmitted from one branch of Indo-Iranian to the other, or borrowed from different sources. This will help to determine the likelihood of a Central Asian Substrate origin of early Indo-Iranian loanwords.

Another question that will be addressed is whether there are different chronological layers of loanwords within PII. The inherited vocabulary of Indo-Iranian has undergone all PII sound changes, but this is not necessarily true for PII borrowings. If the structure of a loanword is such that it cannot go back to Pre-PII, it must have been borrowed at a later stage. Besides allowing for a more precise chronological stratification of early Indo-Iranian loanwords, showing that a loanword cannot go back to Pre-PII would also provide a strong argument in favor of a non-IE origin. Conversely, if a word must have gone through certain PII sound changes, it must have been borrowed at an earlier stage.

The investigation of chronological layers hinges on establishing a relative chronology of PII sound changes. Lubotsky (2018) proposes a chronology, but since several aspects of Indo-Iranian historical phonology are debated, key points must be reviewed and revised.

A final purpose of dividing early Indo-Iranian loanwords into chronological layers is to improve the analysis of structural characteristics of borrowed vocabulary (cf. below).

1.4.3. Structural characteristics

Previous literature has proposed several phonological and morphological characteristics of early Indo-Iranian loanwords, which corroborate the hypothesis that many words originate in the Central Asian Substrate. In this thesis, I will reevaluate previously proposed structural characteristics, and examine the material for additional patterns. If more characteristics are observed, it would strengthen the Central Asian Substrate Hypothesis.

Furthermore, chronological layers of early Indo-Iranian loanwords will be incorporated into the study of structural characteristics. I will attempt to determine whether structural characteristics hold for words within the same layer, and then compare the layers to each other. Structural differences between the layers would point to multiple source languages, whereas similarities would point to contact with the same language (or related languages) over an extended period of time. This process serves to further test the Central Asian Substrate Hypothesis, and could provide new insights into the Pre-Indo-Iranian linguistic landscape of Central, South and Western Asia.

It is important to kept in mind that structural differences between chronological layers may be due to changes in the recipient language (Indo-Iranian) rather than differences in the donor language(s). Whether this is likely or not depends on the feature in question. For example, the *CVCV̄CV* type looks equally foreign in PII as it does in Indic. On the other hand, voiceless aspirates are less 'marked' in Indic than in PII, since such stops are fully integrated in the phonology of the former but not of the latter.

A caveat is that loanwords attested in a single branch may still be inherited from PII. Unless there is a clear argument against this (i.e. the attested phonological structure cannot develop regularly from PII), it is difficult to disprove. The basic methodological principle, however, is to only reconstruct loanwords with cognates in Indic and Iranian to PII.

Another caveat is that loanwords are only indirect attestations of the source language, which have undergone adaptation to the structure of the recipient language. Accordingly, the phonology and morphology of loanwords must have undergone some kind of change as the loanwords are "nativized", i.e. integrated into the Indo-Iranian linguistic system (Hock, 1991, p. 390). However, structural patterns of the source language may still be carried over to the recipient language in a (more or less) regular way (Hock, 1991, p. 394).

1.5. Organization of the thesis

The thesis is structured as follows. Chapter 2 describes the most important PII sound changes and arranges them in a relative chronology. In chapter 3, previously proposed early Indo-Iranian loanwords are analyzed etymologically, taking the relative chronology established in chapter 2 into account. Chapter 4 summarizes and discusses the findings regarding the chronological layers of early Indo-Iranian loanwords. In chapter 5, previously proposed structural characteristics of Indo-Iranian loanwords are discussed, and new patterns are presented. Chapter 6 summarizes and concludes the thesis.

2. Proto-Indo-Iranian historical phonology

In this chapter, the historical development of PII phonology will be discussed. The goal is to describe the sound changes from PIE to PII, focusing on their relative chronology. When possible, the phonetic realization of PII phonemes will be described. This treatment will serve as a basis for establishing chronological layers and analyzing structural characteristics of early Indo-Iranian loanwords. Below, the PIE and PII phoneme inventories are given for reference, as reconstructed by Beekes (2011, p. 119) and Lubotsky (2018, p. 1875), respectively.

2.1. Proto-Indo-European phoneme inventory

	labial	dental	palatal	velar	labiovelar
stops	*p*	*t*	*ḱ*	*k*	*kʷ*
	(*b*)	*d*	*ǵ*	*g*	*gʷ*
	bʰ	*dʰ*	*ǵʰ*	*gʰ*	*gʷʰ*
fricative		*s*			
laryngeals			*h₁*	*h₂*	*h₃*
liquids		*l* *r*			
nasals	*m*	*n*			
semivowels	*u*		*i*		
vowels[5]	*e* *o*				
	ē *ō*				

[5] Many scholars reconstruct a third vowel **ă*, but there are many arguments against this (Lubotsky, 1989).

2.2. Proto-Indo-Iranian phoneme inventory

	labial	dental	palatal	velar
stops	*p*	*t*		*k*
	b	*d*		*g*
	b^{h}	*d*^{h}		*g*^{h}
affricate			*ć*	*č*
			ǰ	*ǰ*
			ǰ^{h}	*ǰ*^{h}
fricative		*s*		
laryngeal				*H*
liquid		*r*		
nasals	*m*	*n*		
semivowels	*u*		*i*	
vowels	*a* *ā*			

Common allophones:

/s/ = [*s*, *š*, *z*, *ž*]

/i/ = [*i*, *i̯*]

/u/ = [*u*, *u̯*]

/r/ = [*r*, *r̥*, *l*, *l̥*]

2.3. Sound changes from PIE to PII

2.3.1. Vowels

The PII vowel system was reduced in comparison to its PIE predecessor, showing a merger of non-high vowels as **a* and **ā*. However, before the PII vowel merger, two important sound changes must have occurred: Brugmann's Law (BrL), i.e. lengthening of **o* in open syllables, and palatalization of velars.[6] The chain of vowel developments can be described as follows:

[6] In reality, the palatalization of velars was allophonic until it was phonologized as a result of the vowel merger.

Table 1. Vowel changes from PIE to PII

PIE →	**Pre-PII →**	**PII: BrL →**	**PII: vowel merger**
**e*	**e*	**e*	**a*
**ē*	**ē*	**ē*	**ā*
**o*	**oC.*	**o*	**a*
	**o.C*	**ō*	**ā*
**ō*	**ō*	**ō*	
**$h_{2\text{-}3}$e*	**$h_{2\text{-}3}$a*	**$h_{2\text{-}3}$a*	**Ha*
**e$h_{2\text{-}3}$*	**a$h_{2\text{-}3}$*	**a$h_{2\text{-}3}$*	**aH*

In Pre-PII, laryngeals **$h_{2\text{-}3}$* phonetically color an underlying **e*. The resulting vowel must be assumed to have been different from both **e* and **o* for two reasons: 1) **h_3e* did not undergo BrL (Lubotsky, 1990), i.e. it was not phonetically identical to **o*,[7] and 2) **e$h_{2\text{-}3}$* did not palatalize a preceding velar (Ollet, 2014). Although the evidence that **e$h_{2\text{-}3}$* was a non-palatalizing context is scarce, there are no secure cases where **e$h_{2\text{-}3}$* does palatalize a preceding velar (cf. 2.3.3.4.). Thus, it seems likely that phonetic coloring did take place in Pre-PII. Ollet (2014, p. 163) argues that this contradicts Lubotsky's claim that **h_3e* did not undergo BrL, but this is a false dilemma. It is perfectly possible that **h_3e* was phonetically different from **o*, e.g. **[h_3a]. Thus, there is no contradiction between acknowledging laryngeal coloring as a phonetic rule in Pre-PII and accepting that **h_3e* did not undergo BrL.

In spite of the above, evidence for phonological coloring of **$h_{2\text{-}3}$e* > **Ha* will come from the relative chronology of PII sound changes, which will be discussed in section 2.4.[8]

Note that under the scenario above, **h_2* and **h_3* appear not to have been phonemically distinct in Pre-PII. One may object to the idea that **h_2* and **h_3* colored **e* to the same vowel **a*, despite being phonemically distinct and giving different coloring effects (**h_2a* vs. **h_3o*) elsewhere in Indo-European. Indeed, the only reason for arguing that **h_3e* yielded the same vowel as **h_2e* in Pre-PII is to explain why **h_3e* does not undergo BrL, while at the same time being distinct from **e*. However, if Kloekhorst (2018, p. 89) is right that **h_3* was the labialized variant of **h_2*, an additional argument could be that the loss of distinction between **h_3* and **h_2*

[7] One of Lubotsky's examples, Skt. *ávi-* 'sheep' < PIE **$h_3$eu̯i-* is challenged by ToB *$\bar{a}^u$wi*, which seems to point to **h_2e/ou̯i-* (Kim, 2000). However, other examples like Skt. *ánas-* 'cart' ~ Lat. *onus-* 'burden' < **$h_3$enos-* and Skt. *ápas-* ~ Lat. *opus* 'work' < **$h_3$epos-* remain convincing.

[8] In short, there is evidence that the merger of the laryngeals preceded BrL. Therefore, the colored vowel is phonologized as **a* before the general PII vowel merger.

is parallel to the loss of labialization of *K^w, which in Indo-Iranian merges with its non-labial counterpart *K.

In the context *$\check{V}HC$, the vowel underwent compensatory lengthening when the laryngeal was lost, cf. Skt. *mā́tar-* ~ Av. *mātar-* 'mother' < PII **maHtar-* < PIE **meh₂ter-*.

2.3.2. Laryngeals

In this section, the development of the PIE laryngeals in PII will be described. The effect of laryngeals on adjacent vowels and consonants, however, are discussed in the sections treating those phonemes.

2.3.2.1. Consonantal laryngeals

The PIE laryngeals eventually merge into PII *H. This is deduced by the fact that the laryngeals give identical reflexes in terms of vocalization, deglottalization of mediae, Lubotsky's law, and Indic aspiration.[9] Due to its interaction with the mediae (cf. 2.3.3.3.), it is likely that PII *H was a glottal stop. Since consonantal laryngeals seem to be preserved in some positions in Iranian (Kümmel, 2018), phonemic *H must have been retained throughout PII.

2.3.2.2. Laryngeal vocalization

Laryngeal vocalization (LV) changed interconsonantal laryngeals to *i. While LV eventually affects most interconsonantal laryngeals in Indic, for PII it is only securely reconstructable for final syllables, i.e. *H > *i / C_(C)# (Lubotsky, 2018, p. 1882). Among the examples are Skt. *jáni-* ~ OAv. *jaini-* 'wife' < PII **ǰani-* < **genH-* and Skt. *sádhiṣ-* ~ LAv. *hadiš*, OP *hadiš* 'seat, residence' < PII **sad^{h}is-* < **sedHs-*.

Some examples of LV in initial syllables can be reconstructed for PII, viz. Skt. *pitár-*, OP *pitar-*, OAv. dat.sg. *piθrē* 'father' < PII **pitar-* < **pHtar-*, but OAv. nom.sg. *ptā-* does not reflect a vocalized laryngeal. Another example is Skt. *aśiṣat*, OAv. *sīšōiṯ* 'to instruct, command' < PII **ćiša-* < **ćHsa-* < **ḱh₂s-* (Lubotsky, 2018, p. 1883).

In middle syllable position, Indic (and perhaps Nuristani) shows LV, but Iranian does not (Ravnaes, 1981, p. 261). A reasonable hypothesis is thus that medial laryngeals remained consonantal in PII, were lost in Proto-Iranian, but vocalized in Proto-Indic. However, there is one case of PII LV in a medial syllable: Skt. *duhitár-* ~ Nur. Prasun *lüšt* 'daughter' < PII

[9] Skt. *piba-* 'to drink' < PIE **pi-ph₃-e-* shows potential evidence that *h_3 did not merge with the other laryngeals in PII (Kümmel, 2018, p. 163). However, since *ph_3 > *b seems to be a PIE development, reflected also in Lat. *bibō* 'to drink' and OIr. *ebait* 'they drink', the laryngeal may have been lost at an early stage. An alternative explanation of **pi-bh₃-e-* is that the Pre-PIE root was **beh₃-*, and that *b > *p in initial position (Kortlandt, 1996). In that case, the development to **pi-b^{h}He-* may have been prohibited by the constraint against tautosyllabic tenues and aspiratae. In general, the lack of clear examples of aspiration/deglottalization by *h_3 may be due to the rarity of the phoneme in PIE.

**dʰujʰitar-* < **dʰugHtar-* < PIE **dʰugh₂ter-* (Lubotsky, 2018, p. 1883). PII **i* is reconstructed to account for the palatalization of **gʰ*. Conversely, OAv. *dugədar-* and LAv. *duxtar* show no trace of LV. To avoid projecting LV back to before PII palatalization, Kümmel (2016b, p. 220) suggests that Skt. *duhitár* derives directly from **dugʰitar-* with debuccalization of **gʰ* > *h*, and that the palatalization in Prasun *lüšt* is secondary, which is *ad hoc*.

Previous theories on Indo-Iranian LV can be divided into two types: those advocating a single LV process and those operating with two separate LV processes.

The former type of theory assumes that interconsonantal laryngeals were vocalized once in PII. To explain the lack of vocalic reflexes in medial syllables in Iranian, the proponents of this theory introduce various phonological rules and analogical processes. Schmidt (1973, p. 54) proposed that laryngeals were lost in the sequence **CHCC*, perhaps already in PIE. This sound change would have given rise to paradigmatic alternations like **dʰugh₂ter-* / **dʰugtr-* and **ph₂ter-* / **ptr-*, which according to Schmidt produced the attested Indo-Iranian paradigms through levelling of different stems in Indic and Iranian. A reversed version of this theory, whereby a laryngeal was vocalized in the sequence **CHCC* is advocated by Beekes (1981, p. 285).

Lipp (2009, p. 356) argued that LV only affected pretonic laryngeals, **-CHC-´*. Since laryngeals in initial syllables are by definition pretonic, an alternation between pretonic LV and post-tonic non-LV would only be visible in medial and final syllables. It is immediately clear that Lipp's accent rule did not operate in final syllables, where LV is firmly attested, since such laryngeals are by default post-tonic. Thus, the rule could in fact only predict the outcome of laryngeals in medial syllables (i.e. *CV́CHCV* > *CV́CCV*, but *CVCHCV́* > *CVCiCV́*).

Lipp's accent rule would explain cases like Skt. *déva-tta-* 'God-given' < **dai̯u̯á-dH-ta-* (no LV) vs. Skt. *duhitár-* < **dʰugHtár-* (LV). However, the expected regular outcome of PII **-dH-ta-* 'given' would be Skt. ***-ddha-*, since **d* would be deglottalized to **dʰ* by the laryngeal,[10] with subsequent progressive assimilation by Bartholomae's Law. Therefore, Skt. *-tta-* could be a secondary formation from the root *dā*, rather than a regular outcome. In the latter case, Skt. *duhitár-* seems to follow the accent rule, but to explain the lack of LV in OAv. *dugədar-* etc., Schmidt's rule of laryngeal loss **CHCC* > **CCC* must be invoked. Another problematic case is LAv. and OP *Vištā*(*spa*)- < **û̯i-sH-ta-* ~ Skt. *víṣita-* < **û̯í-sḭ-ta-*. The Skt. form shows LV of a post-tonic laryngeal. It might of course be explained away as secondary, but at face value, it suggests that accent did not influence vocalization. Ultimately, Lipp's

[10] Cf. section 2.3.3.3.

accent rule cannot replace any of the other phonological rules, and does not explain enough material to be considered likely. In fact, there are so many exceptions that the accent rule can just as well be reversed with Beekes (1981, p. 285), who assumed that only post-tonic **H* was vocalized.

All theories that operate with a single PII laryngeal vocalization must assume extensive processes of analogy to explain the lack of LV in Iranian middle syllables. In the case of Lipp (2009), unexpected reflexes of LV in Indic must be attributed to accent shifts or analogical extensions. While it is not unexpected that paradigmatic alternations in PII would have been levelled in the separate branches, the fact that Iranian always shows the forms without LV suggests a phonological conditioning rather than analogical levelling.

The second type of theory, advocated by Kuiper (1976, p. 243) and Kümmel (2016b, p. 221), operates with two distinct processes of LV, one in PII that only affects laryngeals in initial and final syllables, and another in Indic that affected remaining interconsonantal laryngeals. This theory better accounts for the lack of LV in medial syllables in Iranian, and for the fact that **H̥* can yield either *ī* or *i* in Indic, whereas only *i* is found in Iranian. The one clear counterexample, Skt. *duhitár-*, however, is not easy to explain away. Thus, the theory of two LVs has the advantage of being capable of explaining the divergent treatment of laryngeals in middle syllables. The disadvantage is that the laryngeal in **dʰugHtar-* must have been affected by the first LV, despite being in a medial syllable.

In addition to the above, there are different opinions regarding the phonetics of LV. Most scholars, including Mayrhofer (1986), Schmidt (1973), Werba (2005), Lipp (2009) and Kümmel (2016b) have assumed that LV was realized as an anaptyctic vowel adjacent to the laryngeal, i.e. **Hi* or **iH*. Although the 'anaptyxis theory' is reasonable at first glance, since laryngeals, being obstruents, cannot be vocalized in the same way as resonants, it faces methodological problems and cannot explain the attested material.

The first problem regards the assumption that **H̥* = **Hi*. The vowel is posited after the laryngeal to explain the 'double reflex' of **H̥*, yielding both aspiration of a preceding stop and the vowel *i*, e.g. **dʰugʰitar-* < **dʰugHitar-* < **dʰugHtar-*. However, this cannot explain why PII **pH̥tar-* > **pHitar-* 'father' becomes Skt. *pitár-* ~ OP *pitā-* and not ***phitár-* and ***fitā-*, respectively. Mayrhofer (1986, p. 138) therefore assumed that **H̥* > **iH* in initial syllables. However, this does not explain why **iH* did not yield long **ī*, like other **VHC* sequences. Byrd (2015, p. 33) argued that the anaptyctic vowel in **iH* was a non-moraic vowel that was not lengthened by laryngeals like other vowels. However, since **H̥* elsewhere merges with PII **i*, there is no independent reason to assume that **iH* was an "extremely short vowel" (Byrd, 2015,

p. 31). Kümmel (2016b, p. 223) instead assumes laryngeal loss in initial syllables, arguing that the anaptyxis of **i* in these cases is a separate process that affected clusters of three consonants, e.g. **pHtr-* > **ptr-* > **pitr-*. However, in the case of Skt. *aśiṣat* ~ OAv. *sīšōiṯ* < PII **ćHsa-*, there would be no phonetic motivation for anaptyxis after laryngeal loss, since **ćsa-* is a licensed cluster in PII. Thus, in Kümmel's scenario, core evidence like **ćisa-* < **ćHsa-* must be explained away as analogical.

Secondly, it is problematic to assume that a single phoneme **H̥* yielded aspiration and a vowel **i*, e.g. a 'double reflex'. Moreover, the data does not require it. In the case of voiceless aspirates, there are words where a laryngeal appears to yield both aspiration and *i*, cf. Skt. *pṛthivī́-* 'earth' < **pr̥tHu̯ī-* and *pathíbhiḥ* 'with/from the road' < **pn̥tH-bhis*. Yet, in all such cases, the aspiration may have been introduced by analogy from related words where the laryngeal was not interconsonantal, e.g. Skt. *pṛthú-* < **pr̥tHu-*. In the latter case, levelling of *-th-* must have occurred in nom.sg. *pánthās* anyway, since LAv. *paṇtå̊* preserves the regular reflex of PII **pantaH-s*. Thus, there is no reason to believe that the aspiration in Skt. *pathíbhiḥ* is old. To my knowledge, the only case where analogical extension of aspiration is impossible is Skt. *átithi-* 'guest' < **atH̥tHi-*, and here aspiration is conspicuously absent.[11]

In the case of secondary voiced aspirates, Lubotsky (2018, p. 1882) has offered an alternative explanation: in a cluster **DH*, like in PII **dʰugHtar-*, mediae were not aspirated, but deglottalized and thus merged with **Dʰ* (cf. 2.3.3.3.). As this is a dissimilatory process, it does not preclude that the laryngeal was later vocalized to **i*. When LV occurred, the media had already lost its glottalic feature.

Thus, in terms of phonology, LV should be viewed as **H* > **i*, without an intermediate biphonemic stage. While the exact conditioning and chronology of Indo-Iranian LV remain unclear, it is evident that it occurred in initial and final syllables in PII, as well as in **dʰugHtar-*.

2.3.2.3. Laryngeal metathesis

In the sequence *CHiC* and *CHuC*, the laryngeal and semivowel underwent metathesis, cf. Skt. *bhūtá-* ~ LAv. *būta-* 'become' < **bʰuHta-* < **bʰh₂uto-* and Skt. *pītá-* 'drunk' < **piHta-* < **ph₃ito-* (Lubotsky, 2018, p. 1884). That the laryngeal originally preceded the semivowel is shown by Ru. (dial.) *bávit'* 'to linger', Goth. *bauan* 'to live, dwell' < **bʰeh₂u-* (Kortlandt, 1986, p. 90) and Skt. *pāyáya-* 'cause to drink' < **paH̯i-a̯ia-* < **poh₃i̯-e̯ie-*.[12]

[11] Unless the unaspirated *t* is explained by Grassmann's Law; in that case, Skt. *átithi-* cannot be used as an argument.

[12] The semivowel in **bʰeh₂u-*, **peh₃i-* may be a fossilized suffix, if these roots derive from PIE **bʰeh₂-* 'to shine' and **peh₃-* 'to drink'.

In addition to Indo-Iranian, examples of laryngeal metathesis exist in several IE branches. Gr. *πῖθι* 'drink!' < **pih₃-dʰi* seems to show the same development as in Skt. *pītá-*. Next to Goth. *bauan*, ON *búa* 'to dwell' probably continues a metathesized zero-grade **bʰuh₂-* (Kroonen, 2013, p. 71). Lat. *grūs* 'crane' < **ǵruh₂-* most likely shows metathesis (Vaan, 2008, p. 274), since SCr. *žȅrãv* shows that the word was originally a *u*-stem (Kortlandt, 1985, p. 120). The Anatolian evidence is scarce, but Hitt. *šuḫḫa-ⁱ* / *šuḫḫ-* 'to scatter' < **suh₂-*, if from the same root as *išḫuu̯ai-ⁱ* / *išḫui-* 'to throw, scatter, pour' < **sh₂u-oi-* (Kloekhorst, 2008, p. 892), seems to show a metathesized variant of **seh₂u-*.

Thus, laryngeal metathesis seems to have been a PIE, in any case Pre-PII, development.

2.3.2.4. Laryngeal accent shift

Lubotsky (1988, p. 50) observed that *i*- and *u*-stems derived from roots with laryngeals after the root vowel (**Ce*(*C*)*H*(*C*)-) are generally oxytone in Sanskrit. By contrast, roots without laryngeals are barytone or oxytone depending on their ablaut pattern. To explain this, Lubotsky postulated a laryngeal accent shift from the root to the suffix in PII, i.e. **Cé*(*C*)*H*(*C*)-*U*- > **Ce*(*C*)*H*(*C*)-*Ú*-.

Two exceptions to this rule, Skt. *íṣṭi-* 'sacrifice' and *yájyu-* 'eager to sacrifice' < PIE **Hieh₂ǵ-*, derive from a root where the laryngeal was lost before **-ǵC-* with Lubotsky's Law (cf. 2.3.3.3.). Therefore, it seems that the laryngeal accent shift was posterior to Lubotsky's Law.

Two other counterexamples, Skt. *bhū́mi-* 'earth' < **bʰúH-mi-* and Skt. *bhū́ri-* 'much' < **bʰuH-ri-*, were explained by assuming that the laryngeal accent shift preceded laryngeal metathesis (Lubotsky, 1988, p. 53). Since the laryngeal originally preceded the vowel in the root **bʰeh₂u-*, this would explain the barytonesis of Skt. *bhū́mi-* and *bhū́ri-*. However, as seen in the previous section, laryngeal metathesis may have been a PIE development, whereas the laryngeal accent shift clearly is not. Therefore, the accentuation of *bhū́mi-* and *bhū́ri-* must be explained otherwise. Although no good alternative is available at present, *bhū́mi-* and *bhū́ri-* should be treated as exceptional, since the laryngeal accent shift explains a clear majority of the available evidence.

2.3.2.5. Loss of intervocalic laryngeals

In intervocalic position, laryngeals were lost in PII (Lubotsky, 1995, p. 229). However, whenever the sequence *-VHV-* was separated by a morpheme boundary, i.e. *VH-* + *-V* or *V-* + *-HV-*, the laryngeal could be restored by analogy, often yielding a disyllabic long vowel or diphthong in Vedic Sanskrit. As argued by Lubotsky (1995, p. 220), BrL must be anterior to

the loss of intervocalic laryngeals, cf. Skt. *dāyi* ‘was given’ < **dāHi* < **doh₃i*. Moreover, loss of intervocalic laryngeals must be anterior to **N̥ > *a*, since **-aHn̥-* yields a disyllabic long vowel, even when there is no model for restoration of the laryngeal, cf. GAv. dat.sg. *vātāi* ‘wind’ /*va’atai*/ ~ Skt. *vā́ta-* ‘wind’ /*va’ata*/ < PIE **h₂ueh₁nto-* (Lubotsky, 1995, p. 230).

2.3.3. Stops

In PII, the PIE stops developed into stops and affricates. The labial and dental series were retained, whereas the labiovelar and velar series merged.[13] Secondary affricates from palatalized velars emerge from the merger of **e* and **o*. In this section, I discuss special developments of stops in PII, as well as their phonetic interpretation. I leave out some early changes like the depalatalization of **Ḱ* before **r* (Kloekhorst, 2011), which are not strictly speaking Indo-Iranian but Post-PIE developments shared by several branches.

2.3.3.1. Phonetics of the stops

For the three series of PIE stops and their descendants in PII I use the terms tenuis/voiceless for **T*, media/glottalic for **D*, and media aspirata/voiced (aspirate) for **D^h*. The first set of terms (tenuis/media/aspirata) is used as a cover term without phonetic implications. The second set (voiceless/glottalic/voiced) is used when the phonetic interpretation of the stops is significant. The cover symbols **T/*D/*D^h* reflect the traditional (non-glottalic) phonetic interpretation of the stops.

Although not yet commonly accepted by scholars of Indo-European linguistics, the Glottalic Theory offers an explanation to a number of unrelated features of the stops in PIE and in the daughter languages. The version of the Glottalic Theory employed here states that the PIE mediae **D* were pre-glottalized stops (Kortlandt, 2018), i.e. stops with an inherent glottalic feature realized before the occlusion, [ˀt] etc.

In addition to offering a plausible phonetic explanation to the root constraint against the type **De(R)D-*, the scarcity of PIE **b* and the make-up of the PIE stop system in general (Kümmel, 2012, p. 299), the Glottalic Theory is supported by comparative evidence. For example, Winter’s Law in Balto-Slavic (Winter, 1978), Lachmann’s Law in Latin (Kortlandt, 1985), Lubotsky’s Law in Indo-Iranian (1981), the Kortlandt effect (1983), and lengthening before mediae in Anatolian (Kloekhorst, 2014, p. 230ff) all suggest that PIE mediae were not plain voiced stops, but pre-glottalized.

[13] Some scholars contend that labiovelars were preserved in PII, but this cannot be correct (cf. 2.3.5.2.).

A consequence of the Glottalic Theory is that the contrastive features of the tenues and mediae aspiratae must be reinterpreted:[14] since the mediae are not plain voiced stops but glottalic, there is no reason to assume that the aspiratae contrasted with the tenues in both voice and aspiration. In fact, voiced aspirates are only attested in Indic, where they are more properly described as 'breathy voiced' (Kloekhorst, 2016, p. 234). The aspiration feature of Indic voiced aspirates is not phonetically identical to the aspiration of voiceless stops (Kobayashi, 2017, p. 331). Thus, the aspiration feature in Greek voiceless aspirates need not be historically related to the breathy voice feature of Indic. The PIE mediae aspiratae are therefore best interpreted as plain voiced stops.

The stops and affricates in PII generally continue the PIE situation with tenues = voiceless, mediae = pre-glottalized (cf. 2.3.3.3.), and aspiratae = voiced. In Iranian, the mediae and aspiratae merged, probably as a plain voiced stop (Hoffmann & Forssman, 1996, p. 95). In Indic, a new series of voiceless aspirated stops emerged from clusters of tenues + laryngeal (Kuryłowicz, 1935, p. 46), and the aspiratae became voiced aspirated stops.

2.3.3.2. Phonetics of the palatals

In PIE, the palatals were most likely stops since they became velar stops in the *centum* languages. However, in Indo-Iranian languages they emerge as affricates or fricatives. The goal of this section is to approximate the phonetic quality of PII primary palatals by comparing the reflexes in Indo-Iranian languages.

2.3.3.2.1. Indic

The outcome of PII **ć* is Skt. *ś*. Synchronically, *ś* is a sibilant that does not only reflect **ć* but also **s* and **š* in external sandhi. Its place of articulation is either alveopalatal [ɕ] or palato-alveolar [ʃ] (Kobayashi, 2004, p. 55). The voiced counterpart of Skt. *ś* is *j*, which continues PII **ǰ* as well as **ǰ*. Synchronically, *j* is a palato-alveolar affricate [ʤ] (Kobayashi, 2004, p. 74). The outcome of **ǰʰ* is Skt. *h*, which is phonetically a voiced glottal fricative [ɦ] (Kobayashi, 2017, p. 331). However, forms like Skt. *jáhāti* 'leaves' < **ǰʰa-ǰʰaH-ti* show that *h* must have been an affricate before the application of Grassmann's Law.

In sum, the reflex of voiceless **ć* and voiced aspirate **ǰʰ* are fricatives. While the latter must have been an affricate in the prehistory of Indic, the reflex of **ǰ* is synchronically an affricate. All are alveopalatal or palato-alveolar.

[14] The contrast between **T* and **Dʰ* may have been realized as a length opposition in Proto-Indo-Anatolian, which later developed into a voice opposition in 'classical' PIE (Kloekhorst, 2016).

2.3.3.2.2. Iranian

Traditionally, the PII palatals **ć*, **ǰ* and **ǰʰ* were thought to yield dental fricatives **s*, **z* in PIr. However, this reconstruction only accounts for the Avestan reflexes *s*/*z*, leaving crucial evidence from other Iranian languages out of consideration.

In OP, **ć* generally yields a labiodental fricative *θ*, whereas **ǰ(ʰ)* merges with the dental stop *d*, perhaps pronounced as a voiced fricative [ð] (Cantera, 2017, p. 492). While *θ* could have developed from a dental fricative [s] or affricate [ʦ], *d* can hardly go back to [z], but most likely reflects an earlier dental affricate [ʣ]. Lubotsky (2001a, p. 49) argued that OP *θ* went through an intermediate stage **s*, since PII **sć* > OP *-s-* in medial position but *θ-* in initial position. His idea is that **sć* yielded **ss*, which was simplified to **s-* in initial position, becoming *θ*, but not in internal position, since a geminate was tolerated here. This seems reasonable, since if *θ* derives directly from [ʦ], we would expect **sć* to be pronounced [sʦ] or [ttʦ], which could have yielded *-st-* or *-θ-*, but hardly *-s-*.

Beyond Avestan and OP, the evidence speaks more clearly against reconstructing PIr. dental fricatives. In Khotanese the reflex of **-ću̯-* is *-śś-*, cf. *aśśä* 'horse' < PII **Haću̯a-*, which is to be interpreted as an alveo-palatal fricative [ɕ]. Elsewhere **ć* > *s*. According to Kümmel (2019, p. 15), dental **s* could have been secondarily retracted to [ɕ] before **u̯*. Sims-Williams (1998, p. 136), on the other hand, argues that the pre-form of *-śś-* cannot have been dental, but must have been pronounced further back, viz. alveo-palatal.

As for the manner of articulation, Khot. *dasta*, Parth. *dst*, Sogd. *δst* 'hand' < PII **ǰʰasta-* all underwent dissimilation of **ǰʰ* > **d* due to the following **s*, which shows that **ǰʰ* was an affricate in Proto-Iranian. PII **ǰʰ* otherwise becomes a voiced dental fricative in these languages, cf. Khot *aysu* [azu] 'I' < PII **HaǰʰHam*. Moreover, early Iranian loanwords into Tocharian, e.g. ToB *etswe* 'mule' << **aʦwa*, show evidence of an affricate (Peyrot, 2018).

In conclusion, the Iranian evidence suggests that PII **ć*, **ǰ* and **ǰʰ* became PIr. dental affricates [ʦ], [ʣ] or alveo-palatal affricates [ʨ] [ʥ].

2.3.3.2.3. Nuristani

The Nuristani reflexes of PII **ć*, **ǰ* and **ǰʰ* are *c*, *j*, pronounced as dental affricates [ʦ], [ʣ] (Blažek & Hegedűs, 2012, p. 46).

2.3.3.2.4. Proto-Indo-Iranian

Thus, all branches of Indo-Iranian suggest that the PII palatals were affricates. Moreover, in Indic and perhaps Khotanese, reflexes of the palatals are alveo-palatal, whereas the remaining languages have dentals. As an unconditioned change from dental/alveolar > alveo-palatal

affricate is barely attested cross-linguistically (Kümmel, 2007, p. 232), it seems highly probable that the alveo-palatal pronunciation is original. The PII palatals may thus with some confidence be interpreted as alveo-palatal affricates: **ć* = [tɕ] (voiceless), **ǰ* = [ʔdʑ] (glottalic) and [dʑ] (voiced).

2.3.3.3. Stops in contact with laryngeals

Indic voiceless aspirates *ph, th, kh* derive from clusters of tenues + laryngeal, e.g. *tistha-* 'to stand' < **ti-stH-a-*, dat.sg. *sákhye* 'companion' < **sakHi̯ai̯* (Kuryłowicz, 1935, p. 46). In Iranian, tenues become fricatives before any consonant, including laryngeals, so there is no reason to assume an intermediate stage of voiceless aspirates (*pace* Cantera, 2017, p. 21).

Nevertheless, in both Indic and Iranian, clusters of media + laryngeal merge with voiced aspirates. This is most clearly seen in Skt. *máhi* < **majH* < PIE **meǵh₂*, Skt. *sadhiṣ-* < **sadHs-* < PIE **sedh₁s-* and Skt. *duhitár* < **dʰugʰHtar-* **dʰugHtar-* < PIE **dʰugh₂ter-*. Since OAv. *dugədar* was affected by BL, it must go back to **dʰugʰtar-* < **dʰugHtar-* as well.

In the Sanskrit forms above, the laryngeal appears to show a double reflex, yielding both aspiration of a preceding stop and **i*. To explain this development, scholars have assumed that **H̥ > *Hi*. However, this hypothesis cannot explain the attested material (cf. 2.3.2.2.). An alternative solution was offered by Lubotsky (2018, p. 1882). Under the assumption that the mediae were pre-glottalized and the laryngeals had merged as [ʔ], the change **DH > *DʰH* may be understood as a dissimilation of the glottalic element of the stop. The dissimilated **D* merged with **Dʰ*, which was probably a plain voiced stop, later aspirated in Indic. As it is a dissimilatory process, laryngeal deglottalization does not preclude that the laryngeal was subsequently vocalized to **i*.

The deglottalization process is phonetically paralleled by Lubotsky's Law (1981), which constitutes dissimilatory loss of laryngeals in the sequence **-HDC- > *-DC-*, cf. Skt. *pắjrá-* 'firm' < **paHjra-*. Due to their phonetic similarity, it is likely that the two sound changes occurred simultaneously.

2.3.3.4. Palatalization of velars

The outcome of the Pre-PII velars **k, *g, *gʰ* depends on the phonological environment. Before **e* and **i*, the velars are palatalized to affricates **č, *ǰ, *ǰʰ*. As evidenced by Skt. *duhitár-* and Prasun *lüšt* 'daughter' < **dʰuǰʰitar-* < **dʰugHtar-*, secondary **i* < **H̥* also caused palatalization. This change is linked to the PII vowel merger, since the change **K > *Č* was phonologized when the conditioning factor, **e* vs. **o*, disappeared. Due to ablaut (e.g. **e ~ *o* or **oi ~ *i*), alternations between velar stops and palatal affricates became a frequent phenomenon in verbs,

cf. Skt. 3sg.pres. *hánti* ‘slays’ < Pre-PII **g^{h}enti* but 3pl.pres. *ghnánti* ‘they slay’ < **g^{h}nenti*. The original distribution is often blurred by analogy in the daughter languages, cf. Skt. *gácchati* ~ LAv. *jasaiti* ‘goes’ < **gm̥sćati*.

Lubotsky (2001a) argued that **sk* and **sḱ* were not phonemically distinct in PIE. Although the outcome of PIE **sk* has three separate reflexes in Indo-Iranian, **sć*, **sč* and **sk*, they stand in complementary distribution. In a non-palatalizing context, **sk* is retained. Before **e* and **i*, **sk* is palatalized to **sč* as expected, but **sč* later becomes **sć* unless it is preceded by an obstruent (Lubotsky, 2001a, p. 53).

A special problem is whether **eh$_{2\text{-}3}$* caused palatalization of a preceding **K* or not. As shown by Ollet (2014), there is no good evidence for palatalization of velars before **eh$_{2\text{-}3}$*. The difficulty lies in the fact that paradigmatic alternations were often levelled out, and so the absence of palatalization before **eh$_{2\text{-}3}$* does not necessarily reflect the original situation. However, although Avestan usually generalizes the palatalized variant in paradigms, OAv. 3sg.aor. *gāṯ* ‘went’ < **g^{w}eh$_{2}$-t* shows no palatalization. I conclude that **eh$_{2\text{-}3}$* most likely did not palatalize a preceding velar, because the vowel was lowered by laryngeal coloring.

2.3.3.5. Bartholomae’s Law

Bartholomae’s Law (BL) describes a progressive voicing assimilation which affected aspiratae in Indo-Iranian. Clear examples are Skt. *buddhá-* ‘awoken’ < **b^{h}ud^{h}-ta-* and OAv. 3sg.inj.med. *aogədā* ‘to announce’ < **HaHug^{h}-ta*.

In Iranian, clusters of the shape **D^{h}s* were also affected, cf. OAv. (*pairii-*)*aoγža* < **HaHug^{h}-sa* and *diβža-* < **d^{h}i-(d^{h})b-z^{h}a-*. In Indic, however, **D^{h}s* becomes **Ts*, cf. Skt. *dipsa-* < **d^{h}i-(d^{h})b-z^{h}a-*[15] (*LIV*, p. 133). The *LIV* explains the Sanskrit outcome as secondary restoration of *-s-*. The reason for assuming that *-ps-* replaced earlier **-bz^{h}-* is that the initial *d-* of Skt. *dipsa-* according to *LIV* has been deaspirated by Grassmann’s Law (GL). However, other **D^{h}s* clusters such as Skt. aor.inj. *dhukṣa-* ‘to milk’ < **d^{h}ug^{h}-sa-*, which can hardly be analogical,[16] appear to remain unaffected by GL. It seems more likely that Skt. *dabh-* originally had an initial media, PIE **deb^{h}-*. The only reason why *LIV* assumes PIE **d^{h}eb^{h}-* is Umbr. dat.sg. *fefure* ‘damage’ < **d^{h}eb^{h}-os-ei̯*. However, as Schirmer (1998, p. 64), who proposed the Umbrian etymology, shows, there are several other possible interpretations of Umbr. *fefure*, and the assumption of an *s*-stem requires analogical restoration of *-u-* < **-o-* in the suffix, which does not normally occur in the language.

[15] The loss of root-initial **d^{h}* is regular due to cluster simplification, cf. Skt. *nadbhyas* ‘grandson’ < **naptbhyas* < **napt-* (Kobayashi, 2017, p. 334)
[16] Since other derivations of **d^{h}ug^{h}-* did undergo GL, cf. Skt. 3sg.pres. *dogdhi* < **d^{h}au̯g^{h}-ti*.

If Skt. *dipsa-* instead goes back to PII **di-(d)b^h-sa-*, the development of **D^{h}s* may be explained differently. Within Indic, the PII plain voiced stops (= aspiratae) became aspirated. However, the aspiration could not be realized phonetically in **D^{h}s* clusters, since **[ʔdibz^{h}a] was impossible. Therefore, the stop instead merged with the tenues, yielding [ʔdipsa]. This also explains why aspiration was maintained in Skt. *dhukṣa-* < **d^{h}ug^h-sa-*, since **g^h* was never aspirated here. Since Iranian never developed aspiration, the voicing of **D^{h}s* = [Dz] in e.g. *diβža-* was retained.

Since there are slight differences between Indic and Iranian in the operation of BL, it is likely that the process was phonetic in PII and not phonologized until Post-PII.

2.3.4. The sibilant **s*

PIE had a single sibilant phoneme **s*. In PII, the phoneme **s* has voiced [z] and palatalized [š] allophones.

Already in Pre-PII **s > š / i, u, r, K_*. Known as the *RUKI*-rule, this sound change is shared by the *satəm* languages and was thus most likely active at a very early date. On the other hand, the rule remained active for a long period, since secondary instances of *i, u, r, K* (e.g. **i* < **H̥*) could trigger it (Lubotsky, 2018, p. 1881). Yet, there is one case of PII **š* which is not conditioned by *RUKI*, namely **šu̯ećš* 'six' < PIE **sueḱs*.[17] Moreover, the clusters **tć* and **ćs* may have merged as [tš] already in PII, although Khotanese evidence suggests that they were kept separate (Cantera, 2017, p. 25). Thus, certain instances of **š* could be analyzed as phonological, but the *RUKI*-rule remained active until Post-PII.

2.3.5. Liquids

2.3.5.1. General development

The outcome of the PIE liquids **r* and **l* in PII is complicated. In Iranian languages **r* and **l* merge (Cantera, 2017, p. 15). In Sanskrit, *r* and *l* are synchronically phonemically distinct, but the distribution does not always match the etymological origins (Lubotsky, 2018, p. 1878). The irregular distribution of *r* and *l* in Sanskrit is often explained as dialectal variation that was adopted into the literary language (Burrow, 1973, p. 84). As a consequence, Skt. *r* and *l* are not decisive for etymological analysis. For the purposes of this study, therefore, I will operate with a single PII liquid **r*.

PII **r* has two main allomorphs conditioned by the phonotactic environment where it occurs. Whenever **r* is next to a vowel or precedes a vocalic resonant, it is consonantal

[17] The initial **s-* may be secondary, perhaps by contamination by **septm* 'seven' (Kroonen, 2013, p. 431).

(Schindler, 1977). Conversely, whenever **r* is interconsonantal, it is vocalic **ṛ*. Vocalic **ṛ* is retained in Sanskrit, and is generally rendered as *ər* in Avestan (Hoffmann & Forssman, 1996, p. 91).

2.3.5.2. Liquid + laryngeal clusters

The cluster **ṛH* shows a special development in Indo-Iranian. In Sanskrit, the reflexes are *ir̆* and *ur̆*. The outcome is conditioned by three factors: 1) the following phoneme, 2) the accent, and 3) whether the environment is labial or not. When **ṛH* precedes a vowel, the anaptyctic vowel is short, cf. Skt. *hiraṇya-* 'gold' < **ǰʰṛHaṇi̯a-* and *tirá-* 'to cross' < **tṛHá-*. The same is true when **ṛH* precedes **i̯V* or **u̯V*, cf. Skt. *turyā́ma* 'we shall conquer' < **tṛHi̯ā́ma* and *bhurváṇi-* 'victorious' < **bʰṛHu̯áni-*. However, if the liquid is accented, the vowel is long, cf. Skt. *tū́rva-* 'to cross' < **tṛ́Hu̯a-* and *-śī́rya* 'having smashed' < **ćṛ́Hi̯a-* (Lubotsky, 1997). Elsewhere (**rHC* where *C* is not **i̯/u̯*), the vowel is long, regardless of the accent, cf. Skt. *dīrghá-* 'long' < **dṛHgʰa-* and *ī́rṣyant-* 'being envious' < **ṛ́Hsi̯ant-*. Finally, the quality of the vowel is *ur̆* if **ṛH* is preceded by a labial consonant (*p*, *b*, *bh*, *m*, *u̯*) or followed by **Cu̯*. Exceptions to this rule (mainly verbal forms) can be explained as analogical to other forms in the paradigm where a conditioning labial exists (Lubotsky, 1997, p. 139). Elsewhere, the vowel is *ir̆*, cf. the examples above.

Burrow (1957), and more recently Clayton (2018), argued that **ṛH* > *ur̆* also when preceded by a PIE labiovelar. However, counterexamples like Skt. *girí-* 'mountain' < **gʷrH-i-* and *gīrṇá-* 'swallowed' < **gʷrh₃-no-*, as well as the unlikelihood of preserved labiovelars in a *satəm* language, render this analysis difficult.

The Indic development of **ṛH* may be expressed by the following set of rules:

**ṛH* > *ur* / $C_{+labial}$ _*V*, _*u̯V*
**ṛH* > *ūr* / $C_{+labial}$ _*C*, _*Cu̯*, _*ú̯V*
**ṛH* > *ir* / $C_{-labial}$ _*V*, _*i̯V*
**ṛH* > *īr* / $C_{-labial}$ _*C*, _C, _*í̯V*

The Iranian reflexes of **ṛH* also vary depending on the phonological environment. In most Iranian languages, labial and non-labial contexts are differentiated, showing different anaptyctic vowels (Clayton, 2018). In Avestan, the reflexes are *ar*, *ər* and *ruu*. The outcome is dependent on 1) the accent, 2) the following phoneme, and 3) whether the environment is labial or not (Cantera, 2001).

Under the accent, **ṛH* becomes *ar*, cf. OAv. *pauruua-* 'first' ~ Skt. *pū́rva-* < **pṛ́Hu̯a-*. The same is true when **ṛH* precedes a vowel, independent of the accent, cf. OAv. *tarə̄*

‘sideways’ < **tr̥Has* and *pouru-*[18] (< **paru-*) ‘much, many’ < **pr̥Hú-* (Skt. *purú-*). In unaccented position before consonants, the development of **r̥H* depends on the context. Before **u̯*, unaccented **r̥H* yields *ruu*, cf. Av. *zruuan-* ‘life-time’ < **jr̥H-u̯án-* (Lubotsky, 1997, p. 144). In non-labial contexts, unaccented **r̥H* becomes *ar*, cf. OAv. *darəga-* ‘long’ < **dr̥Hg^{h}á-*. In labial contexts, unaccented **r̥H* becomes *ər*, cf. OAv. *pərəna-* ‘full’ ~ Skt. *pūrṇá-* < **pr̥Hná-*. A labial context is whenever **r̥H* is preceded by a labial consonant (*p*, *b*, *bh*, *m*, *u̯*) or followed by **Cu̯*.

The Iranian development of **r̥H* may be expressed by the following set of rules:

**r̥H* > *ar* / *_́*, *_V*, *C*$_{\text{-labial}}$ *_C*,

**r̥H* > *ər* / *C*$_{\text{+labial}}$ *_C*, *_Cu̯*

**r̥H* > *ruu* / *_u̯*[19]

Cantera (2001, p. 25) argues that initially, **r̥H* > *ər* in all contexts, but that *ər* later became *ar* unless blocked by a labial environment. If so, there is no actual labialization, so much as lowering of the vowel in non-labial contexts. Moreover, any *ər* that does not become *ar* eventually merges with the reflex of **r̥* (without laryngeal). In other Iranian languages, the anaptyctic vowels of original **r̥H* and **r̥* are clearly labial in the appropriate contexts, cf. Phl. *purr* ‘full’ < **pr̥Hná-*, Pto. *murγə́* ‘bird’, MoP *murw* ‘bird’ < **mr̥gá-* and Y-M *purs* ‘to ask’ < **pr̥sća-*.

Although both Indic and Iranian show processes of labialization, they also show significant differences that preclude projecting the vocalization of **r̥H* to PII.

2.3.6. Nasals

2.3.6.1. General development

Like the liquids, the PIE nasals **n* and **m*, retained in PII, had consonantal and vocalic allophones. The consonantal nasals are generally preserved in Indo-Iranian languages. Between consonants and at word boundaries, the nasals are vocalic **n̥* and **m̥*. They later merge with **a* in most positions (Lubotsky, 2018, p. 1876). An exception is **m* in word-initial position before resonants, which remains consonantal and yields **b* before **r* in Indic, cf. LAv. *mraoiti* ‘says’ ~ Skt. *brávīti* ‘says’ < **mrau̯H-ti*.

[18] Here, *ar* was secondarily rounded (Hoffmann & Forssman, 1996, p. 90)
[19] Lubotsky (1997, p. 144) did not find examples with **i̯*.

2.3.6.2. Nasal + laryngeal and nasal + resonant clusters

In some contexts, the outcome of the vocalic nasals is **an*, **am*. This occurs in the contexts **Ṇ̥RV* and **Ṇ̥HV*, cf. Skt. *mányate* ~ OAv. *maṅiiete* 'thinks' < PII **manḭa-* < PIE **mn̥̂-i̯e-*,[20] Skt. *namrá-* 'loyal' ~ LAv. *namra-* 'respectful' < **namrá* < **nm̥ró-*,[21] Skt. *-tama-* ~ Av. *-təma-* < **-tamHa-* < **-tm̥Ho-*,[22] Skt. *hanmás* 'we slay' < **jʰn̥̂-mas*. The sequence **Ṇ̥HC* instead became **aHC* > **āC*, cf. Skt. *jātá-* ~ Av. *zāta-* 'born' < **ǵn̥h₁-tó-*. A sequence ***Ṇ̥RC* never developed regularly, since the second resonant would have been vocalized instead of the nasal (Schindler, 1977).

There is one major exception to the above. When **n̥̂* precedes consonantal **n*, the outcome is **an-*, cf. Skt. *tanóti* 'stretches' < **tn̥̂-neu̯-ti*. Most likely, the expected geminate nasal **-nn-* was simplified to **n* (Kümmel, 2005, p. 322). In general, the vocalization of nasals in nasal present formations is aberrant (Schindler, 1977, p. 56). For example, the plural **tn-nu̯-énti* gives Skt. *tanvánti* 'they stretch' instead of ***tnanvánti* < ***tn-n̥̂u̯-énti*, perhaps due to paradigmatic levelling.

2.3.7. Semivowels

The PIE semivowels **i* and **u* are retained in PII and have consonantal and vocalic allophones. Next to a vowel a semivowel is consonantal, whereas an interconsonantal semivowel is vocalic, cf. Skt. 3pl. *vidúr* ~ LAv. *vīδarə* 'they know' < **u̯id-*. Although PII **u* in principle always derives from PIE **u*, PII **i* also arose secondarily via laryngeal vocalization (cf. 2.3.2.2.).

The sequences **iHC* and **uHC* eventually yield long vowels **īC* and **ūC*, when the laryngeal is lost with compensatory lengthening, cf. Skt. *bhūtá-* ~ LAv. *būta-* 'become' < PII **bʰuHta-*. However, this change is probably quite late. In Nuristani, the sibilant in **uHs* is not affected by *RUKI*, cf. Kati *mussā*, Prasun *mūs'ū* 'mouse' < PII **muHs-*, indicating that the *RUKI*-rule was phonologized before **uHs* > **ūs*. As seen in 2.3.4., the *RUKI*-rule was probably not phonologized in PII. Since PII **muHs-* > Skt. *mūṣ-* and Av. *mūš-* show the effect of *RUKI*, the developments **iH* > **ī* and **uH* > **ū* form an isogloss between Indic and Iranian, excluding Nuristani.

[20] The zero-grade is paralleled by Gr. *μαίνομαι* 'to be furious'.
[21] *-ró-* derivatives normally take zero-grade of the root (Kümmel, 2005).
[22] Cf. Lat. *in-timus* 'inner' (Lubotsky, 2018, p. 1876).

2.4. Relative chronology of Indo-Iranian sound changes

In this section, the relative chronology of the Indo-Iranian sound changes described in 2.3. is discussed.

Lubotsky (2018, p. 1877) has argued that LV in final syllables is one of the earliest PII sound changes, preceding BrL. The basis for this argument is the long vowel of Skt. (*yúva-*)*jāni-*[23] '(having a young) wife' < **-g^{w}onh$_2$-*, in contrast to the simplex Skt. *jáni-* < **g^{w}enh$_2$-*. If the laryngeal had been consonantal when BrL operated, it would have closed the syllable and prevented lengthening of **o*. While the *o*-grade is expected in a compound, not all Vedic compounds follow this pattern, cf. *pr̥ṣṇī-mātara-* 'having P. as mother' < **-meh$_2$ter-o-*. However, it is less likely that the long vowel of *-jāni-* is secondary than the short vowel of *-mātara-*, since the latter conforms to the synchronic simplex form *mātar-*. In any case, LV must precede the PII palatalization, since **i* < **H̥* also caused palatalization (Skt. *duhitár-* ~ Prasun *lüšt* < **d^{h}uj^{ʰ}itar-* < **d^{h}ugHtar-*.

LV is not the earliest in the relative chronology of PII sound changes. Forms like Skt. *sadhiṣ-* ~ Av. *hadiš-* 'seat' < **sad^{h}is-* < **sad^{h}H̥s-* < **sadHs-* < PIE **sedh$_1$s-*, Skt. *máhi* 'great' < **maj^{ʰ}i* < **maj^{ʰ}H̥* < **majH* < PIE **meǵh$_2$* and Skt. *duhitár-* 'daughter' < **d^{h}uj^{ʰ}itar-* < **d^{h}ug^{h}H̥tar-*< **d^{h}ugHtar-* < PIE **d^{h}ugh$_2$ter-* show that laryngeal deglottalization must precede LV.[24] The mediae must have been deglottalized already when LV occurred, since laryngeals at that point merge with **i*, losing the glottalic feature that caused deglottalization. Lubotsky's Law is most likely contemporary with laryngeal deglottalization.

Furthermore, by the time of laryngeal deglottalization, the three PIE laryngeals had already merged into a glottal stop **H* = [ʔ]. However, the difference in vowel quality must have remained after the laryngeal merger, as **eH* < **eh$_1$* later caused palatalization, whereas **aH* < **eh$_{2\text{-}3}$* did not. At this point, the phonetic coloring of **e* was phonologized, since the conditioning factor (**h$_1$* ≠ **h$_{2\text{-}3}$*) was lost. In 2.3.1., I argued that laryngeal coloring could have been sub-phonemic until the PII vowel merger, but since the laryngeals merged before the PII vowel merger, laryngeal coloring must have been phonologized earlier than previously thought. Thus, this conclusion is not based on direct evidence, but follows from considerations on relative chronology.

[23] The palatal *j-* is assumed to be analogical to the simplex form.
[24] Only Skt. *duhitár-* is decisive, since in the paradigms of *sadhiṣ-* and *máhi*, the laryngeal would only have been vocalized in certain case forms. Therefore, if the original nom./acc.sg.n was **maǰi* < **majH̥*, **j^{ʰ}* could have been levelled throughout the paradigm by analogy to the gen.sg. **maj^{ʰ}Hás* < **majHás*. In the case of **sad^{h}is-*, only nom./acc.sg. is attested, but the original genitive could have been **sadH-as-as*.

After LV, BrL caused lengthening of **o* > **ō* in open syllables. At this stage, PII had three phonemic vowels **ĕ*, **ŏ* and **ă*, of which the first (together with **i*) caused phonetic palatalization of preceding velars (/*ke*/ = [*če*]). Subsequently, the vowels merged as **ă*, causing the secondary palatals to become phonemic.

As discussed in 2.3.2.4., the laryngeal accent shift must have been posterior to Lubotsky's Law. The loss of intervocalic laryngeals is posterior to BrL. The vocalization of **N̥* > **a*(*N*) is posterior to the loss of intervocalic laryngeals.

The change *V̆HC* > *V̄C* is posterior to the vocalization of nasals, since **N̥HC* > **āC*, most likely through an intermediate stage **aHC*. As Nuristani preserved **i*/*uHC* sequences, the change *V̆HC* > *V̄C* seems to be an Indic-Iranian isogloss. In the strict sense, it is Post-PII, but could be regarded as a shared innovation between Indic and Iranian, if Nuristani was the first to split off from PII.

With these considerations, I arrive at the following relative chronology:

Table 2. Relative chronology of Proto-Indo-Iranian sound changes

<table>
<tr><th>Phase</th><th colspan="3">Sound change</th></tr>
<tr><td>1 (Pre-PII)</td><td>Phonetic coloring of
*$h_{2\text{-}3}e$ > *$h_{2\text{-}3}a$</td><td>Laryngeal metathesis
*CHUC > *CUHC</td><td>RUKI-rule</td></tr>
<tr><td>2 (PII)</td><td colspan="3">Laryngeal merger to glottal stop *H, phonologization of *a</td></tr>
<tr><td>3</td><td colspan="3">Lubotsky's Law (*ʔ$^{?}$DC > *$^{?}$DC), laryngeal deglottalization (*$^{?}$Dʔ > *Dʔ)</td></tr>
<tr><td>4</td><td>Laryngeal vocalization
*H > *i / C_(C)#</td><td colspan="2" rowspan="2">Laryngeal accent shift *Cé(C)H(C)-U- >
*Ce(C)H(C)-Ú-</td></tr>
<tr><td>5</td><td>Brugmann's Law
*o > *ō / _$C.</td></tr>
<tr><td>6</td><td>Loss of intervocalic laryngeals</td><td rowspan="2">Vowel merger
*ĕ, *ŏ, *ă > *ă̆</td><td rowspan="2">Phonologization of palatalized velars *č, *ǰ, *ǰʰ</td></tr>
<tr><td>7</td><td>*N̥ > *a(N)</td></tr>
<tr><td>8 (Post-PII)</td><td colspan="2">laryngeal loss with compensatory lengthening *VH > V̄</td><td>Vocalization of *r̥H clusters</td></tr>
</table>

3. Etymological analysis of proposed loanwords

In this chapter, previously proposed loanwords into Proto-Indo-Iranian are analyzed etymologically according to the methodology outlined in chapter 1. Each entry includes a discussion of potential etymologies, why a word can or cannot be considered a loanword (applicable criteria are shaded), and when it was borrowed. The words are divided into two main categories: loanwords and non-loanwords. The loanwords (sections 3.1.-3.4.) are divided into chronological layers, which are further discussed in chapter 4. The non-loanwords (section 3.5.) are either inherited, i.e. have plausible or possible IE etymologies, or simply lack evidence for borrowing.

3.1. Loanwords I: Pre-Proto-Indo-Iranian or early Proto-Indo-Iranian

1. PII **ućig-* 'sacrificing priest'

Ind. Skt. *uśíj-*

Ir. OAv. *usig-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (I, p. 235) argues that it is derived from *vaś-* 'to wish'. However, from a morphological perspective, the word looks non-IE, given the suffix **-ig-* (almost unique for this word, cf. *AiGr*. II, 2, p. 321).

Another word from the religious sphere, Skt. *ṛtvíj-* 'priest', is often analyzed as a compound of *ṛtu-* 'season' + *-ij-* 'sacrificing' (*EWAia* I, p. 258). However, while the root *yaj-* 'to sacrifice' reflects a palatal stop (PIE **Hieh$_2$ǵ-*), the *-k* of nom.sg. *ṛtvík* reflects a plain velar (Lubotsky, 2008). For this reason, it is likely that PII **ućig-* and Skt. *ṛtvíj-* contain the same suffix **-ig-* and may have been borrowed from the same source. The same is likely for Skt. *vaṇíj-* 'merchant' and *bhuríj-* '?'.

The suffix **-ig-* is palatalized in forms like the genitive Skt. *uśíj-as* (< **ućig-es*). The non-palatalized form is reflected in Skt. nom.sg. *uśik,* instr.pl. *uśigbhyas* and OAv. nom.sg. *usixš*. While it is not unthinkable that the paradigm could have been secondarily adapted to fit the pattern of inherited palatalized velars, the most straightforward explanation is that **ućig-* was borrowed before the PII palatalization of velars.

3.2. Loanwords 0: Proto-Indo-Iranian, but no further indication of date of borrowing

2. PII **aka-*

Ind. Skt. *áka-* ‘pain’

Ir. Av. *aka-* ’bad’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (I, p. 39) takes it as a derivative of *añc-* ‘to bend’ < PIE **h₂enk-*, which is semantically possible. However, this analysis is morphologically problematic, since Sanskrit would reflect an accented zero-grade **Hn̥k-o-*. The “suffix” **-ka-*, found in several loanwords, may also indicate non-IE origin.

3. PII **anću-* ‘Soma plant’

Ind. Skt. *aṃśú-*

Ir. Av. *ąsu-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). No IE etymology (*EWAia* I, p. 37). Witzel (2003, p. 37) identifies the Soma plant as ephedra, native to mountainous regions of Central and South Asia. The word may share a common origin with ToA *añcwaṣi* ‘made of iron’, assuming that the color of iron was associated with the color of ephedra (Pinault, 2003b).

4. PII **atHarṷan-* ‘priest’

Ind. Skt. *átharvan-*

Ir. LAv. *āθrauuan-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). The word had no IE etymology (*EWAia* I, p. 60), but is sometimes connected to Av. *ātar-* ‘fire’, also of unknown origin. In reality, the connection between the words is probably folk-etymological. This could explain the irregular correspondence Indic *ar* : Iranian *ra*, if earlier **aθarṷan* was remodeled to *āθrauuan-* based on the oblique stem *āθr-* ‘fire’. The “suffix” **-arṷa-* is found in other proposed loanwords.

5. PII **atka-* ‘cloak’

Ind. Skt. *átka-*

Ir. LAv. *aδka-*, *aṯ.ka-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky, since the suffix **-ka-* is normally denominal (2001b, p. 304). Attempts at IE etymology (e.g. PIE **tek-* ‘weave’, *EWAia* I, p. 58) are unconvincing. LAv.

aʇ.ka- with ‘implosive’ *ʇ*, instead of ***aθka-*, as well as the variant *aδka-* could point to original **adka-*, cf. LAv. *ʇbi-* ‘twice’ < **dṷi-*.

6. PII **(H)āni-*

Ind. Skt. *āṇí-* ‘linchpin, hip’

Ir. –

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Kuiper (1991, p. 89), Pinault (2003a, p. 132) and Witzel (2003, p. 33). *EWAia* (I, p. 161) is agnostic as to the exact etymology of this word. The proposed connection to Gr. *ὠλένη* ‘elbow’, Lat. *ulna* ‘forearm’, PGm. **alīnō-* ‘forearm’ < PIE **Heh₃l-én-eh₂-*, which is not semantically clear, would require the assumption of Fortunatov’s Law.

According to Pinault (2003a), the word is found in the compound *kalyāṇī-* lit. ‘with beautiful hips’. The connection between Skt. *kaly-* and Gr. *καλός* is difficult, since this requires the reconstruction of PIE ***kal-*. Furthermore, Skt. *-ṇ-* remains unexplained, unless one assumes that a dialectal variant **karyāṇi-*, where *-r-* would cause the **n* to become retroflex, influenced the simplex form.

Lastly, the lack of an Iranian cognate puts it into question whether the word can be reconstructed for PII. ToB *oñi* ‘hip’ cannot have been borrowed from Indic, since Indic *ā* is adapted to Tocharian *a* in later borrowings (Pinault, 2003a, p. 131). Thus, if *oñi* was borrowed from Indo-Iranian, it was most likely at a very early stage, perhaps PII. However, it is difficult to exclude that Tocharian borrowed the word independently from a Central Asian source.

7. PII **bʰarṷ-* ‘to chew, eat’

Ind. Skt. *bharv-* ‘to chew’

Ir. LAv. *aš.baouruua-* ‘where there is much to eat’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). No IE comparanda are available (*EWAia* II, p. 253). The cluster **-rṷ-* is found in several PII loanwords but not in secure PIE roots. In 11 out of 12 cases of root final **-u* in *LIV*, a laryngeal precedes **-u*.[25] The roots **melh₂u-* ‘to grind’ and **deh₃u-* ‘to give’ are clearly secondary roots from **melh₂-* ‘to grind’ and **deh₃-* ‘to give’,

[25] The exception is **bʰeru-* ‘to boil’, which is semantically clearly unrelated to PII **bʰarṷ-*. Since it is isolated to Italo-Celtic, it need not go back to PIE. It is strange that on the one hand, root final **-u-* seems to be a fossilized suffix, as evidenced by **melh₂u-* ‘to grind’ and **deh₃u-* ‘to give’, but on the other hand, it appears to correlate phonologically with a preceding laryngeal. This suggests that the suffix **-u-* was only reanalyzed as part of the root in a specific phonological context, perhaps related to the loss of laryngeals in various IE branches.

perhaps originally *u*-stems. However, this is an unlikely origin of **b^{h}ar̥u̯-* ‘to chew’, since **b^{h}ar-* ‘to bear’ is semantically unrelated.

8. PII **b^{h}iš-aj-* ‘healer’

Ind. Skt. *bhiṣáj-* ‘physician’

Ir. LAv. *bišaziia-* ‘to cure’, *-biš-* ‘healing’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). IE origin, as argued by *EWAia* (II, p. 264), is very unlikely, especially considering the foreign-looking suffix **-aj-* (with agentive function?). The Sanskrit verbal form *bhiṣákti* ‘heals’ with *-kt-* (< **gt*) instead of expected *-ṣṭ-* (< **jt*) suggests that the verbal forms are secondary.

9. PII **bīja-* ‘seed, semen’

Ind. Skt. *bī́ja-*

Ir. Sogd. *byz’k*, Par. *bīz*, Khot. *bījä* < **bīzya-* (Bailey, 1979, p. 280)

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Witzel (2003, p. 33). IE origin is highly improbable as the root contains two mediae, one of which is **b*. Kent’s (1950, p. 29) attempt to connect OP (*Bagā-*)*bigna-* is rightly rejected by *EWAia* (II, p. 227), since OP *g* cannot reflect PII **j*. As for the semantics, either of the reconstructable meanings could have developed from the other, so it is not certain that the word originally belonged to agricultural terminology.

10. PII **ćar̥u̯a-* ‘Name of a deity’

Ind. Skt. *śarvá-*

Ir. LAv. *sauruua-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b) and Witzel (2003). *EWAia* (II, p. 621) presents no plausible IE etymology. Pinault (2003b) connects **ćar̥u̯a-* to ToB *śerwe*, ToA *śaru* ‘hunter’, arguing that Toacharian borrowed at an early stage from Iranian. Since the words are probably non-IE, it is unclear whether this is the correct direction of borrowing (Adams, 2013, p. 695). The “suffix” **-ar̥u̯a-* is potentially also found in **atHar̥u̯an-* etc.

11. PII **ćyā-* ‘to freeze, congeal’

Ind. Skt. *śyā-*

Ir. Oss. *syjyn*, *sujun*, Yagh. *ši-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). No IE comparanda adduced by *EWAia* (II, p. 660) or *LIV* (p. 360). There are no other examples of anlaut **ḱi̯-* in *LIV*. The only PIE root with a similar cluster is **ǵi̯eu̯H-* ‘to chew’ (*LIV*, p. 168).[26] In view of this, it is more likely that **ćyā-* is a loanword than an isolated inherited root.

12. PII **gadā-* ‘club’

Ind. Skt. (Su+) *gadā-*

Ir. LAv. *gaδā-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

As pointed out by Lubotsky (2001b, p. 303), **gadā-* cannot be IE since it would reflect a root with two mediae. *EWAia* (I, p. 460) does not propose an IE etymology.

13. PII **gr̥da-* ‘penis’

Ind. Skt. *gr̥dá-* ‘penis’

Ir. LAv. *gərəδō.kərəta-* ‘cutting of the genitals’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

As pointed out by Lubotsky (2001b, p. 303), **gr̥da-* cannot be IE since it would reflect a root with two mediae. *EWAia* (I, p. 494) mentions some previously suggested IE etymologies (e.g. **geR-d-*), which are implausible.

14. PII **Hustra-* ‘camel’

Ind. Skt. *úṣṭra-*, *uṣṭár-*

Ir. Av. *uštra-*, OP *uša-*(*bāri-*), Sogd. *xwštr-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

EWAia (I, p. 237) discusses some attempts at IE etymology but admits that the word looks non-IE (the suffix **-tro-* is not normally used for animates). Moreover, it is unsurprising that a word for ‘camel’ would be borrowed by Indo-European speakers, since the camel was domesticated only in the mid-3rd millennium in Iran (Heide, 2011, p. 367). Initial laryngeal is strongly suggested by Av. *Zaraθuštra-* < **ǰarat-Hustra-* ‘having aging camels’. Its presence either means that the source language had an initial consonant that was adopted as PII **H-*, or that at

[26] The only other case in *LIV*, **ǵi̯eH-* ‘to bereave’, is isolated to Indo-Iranian and need not go back to PIE.

the time of borrowing, any vowel-initial word would automatically be pronounced with an initial laryngeal, like in PIE.

15. PII **indra-* ‘name of a God’

Ind. Skt. *índra-*

Ir. LAv. *iṇdra-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

As noted by Lubotsky (2001b, p. 311), if the word was IE, **n* should have vocalized, giving ***i̯adra-*. *EWAia* (I, p. 192) discusses several unlikely etymologies. Parpola (2015, p. 66) suggests that **indra-* was borrowed from PU **ilmar* / **inmar* ‘thunder god’, derived from PU **ilma-* ‘sky’, reflected in Finn. *Ilmarinen* and Udmurt *Inmar*.[27] However, since the suffix **-ri-* is a Finnic innovation and cannot be reconstructed for PU (Frog, 2012, pp. 215-6), this is unlikely.

16. PII **išt*(*i*)- ‘brick’

Ind. Skt. *iṣṭakā* (VS, Br +), *iṣṭikā* (Sū+)

Ir. LAv. *ištiia-*, OP *išti-*, MiP *xišt-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b) and Witzel (1999b, p. 54). *EWAia* (I, p. 201) discusses possible connections with the PIE **ies-* ‘to boil’. This is semantically problematic, since a brick is burnt, not boiled. Moreover, a *-ti-* derivation should yield an abstract noun. Lastly, if Persian *x-* reflects a laryngeal (Kümmel, 2016a, p. 83), the word cannot reflect PIE **is-ti-*.

Indic has both *-akā-* and *-ikā-* suffixes, whereas Iranian has *-i-* and *-i̯a-*. For PII, an original *i*-stem seems the most likely, but variation may have existed already at this stage. A related word is ToB *iścem* ‘clay’, which may have been borrowed from Indo-Iranian (Witzel, 2003, p. 30).

[27] Probably borrowed from Finnic, cf. Frog (2012).

17. PII **ǰʰarmi̯a-* ‘house’

Ind. Skt. *harmyá-*

Ir. Av. *zairimiiāuuant-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (II, p. 807) suggests an IE etymology from **ǵʰer-* ‘to cover’, which does not explain the PII suffix **-mi̯a-*.[28] Semantically, the word fits together with other loanwords pertaining to permanent structures (e.g. **išt(i)-* ‘brick’).

18. PII **kāća-* ‘grass’

Ind. Skt. *kā́śa-*

Ir. MoP *kāh*, MoP *kašk*, Munji *kosk* < PIr. **kaćaka-*

Nur. Km. *kaċo*, Kt. *kċo*, W. *kac*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Kümmel (2017, p. 284). The connection between the Indic and Iranian forms is dismissed by *EWAia* (I, p. 345) for unclear reasons. The variant with suffix *-ka-* in Iranian is aberrant in that the root vowel is short. In any case, **kāća-* can be reconstructed to PII based on Skt. *kā́śa-* and MoP *kāh*.

19. PII **kaći̯apa-* ‘tortoise’

Ind. Skt. *kaśyápa-*

Ir. LAv. *kasiiapa-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b) and Witzel (2003, p. 35). *EWAia* (I, p. 331) provides no IE etymology. The “suffix” *-pa-* is non-IE and is found in other loanwords, e.g. **pāpa-*, **stupa-*. If borrowed from a Central Asian language, **kaći̯apa-* may have referred to the Russian tortoise, *Testudo horsfieldii*, native to the area of the BMAC culture.[29]

[28] Rather, one would have to assume a root structure **ǵʰerm-*, which is unparalleled in PIE (*LIV*, p. 708)
[29] Cf. *The Reptile Database* (http://reptile-database.reptarium.cz/species?genus=Testudo&species=horsfieldii)

20. PII **kadru-* ‘brown’

Ind. Skt. *kádru-* ‘reddish brown’

Ir. MoP *kahar* ‘light brown’, LAv. *kadruua.aspa* ‘with brown horses’ (name of a mountain)

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b) and Witzel (2003, p. 33), who notes that words for secondary colors are often borrowed. *EWAia* (I, p. 295) rightly rejects previous attempts at connecting the Indo-Iranian forms to Gr. *Κόδροϛ*.

21. PII **kapāra-* ‘bowl’

Ind. Skt. *kapā́la-* ‘dish, bowl’

Ir. MiP *kabārag*, MoP *kabāra* ‘vessel’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (I, p. 300) suggests a connection to Lat. *capiō* ‘to take’, *caput* ‘head’ and OE *hafola*. For the Latin words, de Vaan (2008, p. 90) reconstructs a root PIE **kh₂p-*, but argues that substrate origin in **kap-* is equally likely. Kroonen reconstructs PGm. **hafe/alan-* < **kap-ola-* for the OE form and **ha(u)bu/eda-* ‘head’ based on other Germanic cognates (2013, p. 215). Beekes (1996, p. 220) considers these to originate in a European substrate language, but does not discuss **kapāra-*. The main problem with connecting the Indo-Iranian and European material is that the only way to explain why **k-* remains non-palatalized in Indo-Iranian is to reconstruct PIE **a.*[30] The problem is not solved by assuming a laryngeal in the root, because **kh₂ep-* would have given Skt. *kh-*, Ir. *x-*, and **keh₂p-* a long vowel.

A possible scenario, if the word is non-IE, is to assume parallel Post-PIE borrowing by Italic, Germanic and Indo-Iranian from a common European substrate language.[31] This scenario presupposes that the word entered Indo-Iranian after **a* had been phonologized. It also presupposes a relative geographical proximity of Indo-Iranian, Italic and Germanic speakers. However, since **kapāra-* has the *CVCV̄CV* structure, characteristic of Sanskrit loanwords, it more likely derives from an Asian language. Therefore, I will treat the similarity between PII **kapār/la-* and OE *hafola*, Latin *caput* etc. as a chance similarity.[32]

[30] The only other possibility for Indo-Iranian is PIE **knp-* / **kmp-*, but this is not supported by the European words.

[31] This scenario is based on the possibility that speakers of Indo-Iranian first migrated to Europe, before turning eastwards to Asia.

[32] Alternatively, the word diffused as a Wanderwort from Europe to Asia or the other way around, but as it is not a prototypical ‘culture word’ prone to spreading over large areas (like e.g. ‘wheat’), this seems unlikely.

22. PII **kapaṷ̯ta-* ‘pigeon’

Ind. Skt. *kapóta-* ‘pigeon’

Ir. OP *kapautaka-* ‘blue’, MiP *kabōd* ‘grey-blue, pigeon’, Khot. *kavūt* ‘pigeon’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b) as one of the trisyllabic *CVCV̄CV* words in PII. *EWAia* (I, p. 303) considers IE origin likely based on the suffix *-ta-* which is often found in colors, but here the meaning ‘pigeon’ is probably primary. The *-ka-* suffix in OP is a trivial innovation. According to Berger (1959, p. 58), the source of the word is Austroasiatic, e.g. Santali *potam*, Mundari *pudām* ‘dove’, with *ka-* as a prefix (also Kuiper, 1991, p. 42).

PII **kapaṷ̯ta-* cannot with certainty be assigned to late PII, since the non-palatalized PII *kap-* could go back to Pre-PII **kN̥p*.

23. PII **kapHa-* ‘phlegm, mucus’

Ind. Skt. *kapha-* ’phlegm’

Ir. LAv. *kafa-*, Khot. *khavȧ* ‘mucus’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (I, p. 303) offers no IE etymology, but suggests on the basis of Khot. *khavȧ*, which seems to reflect PIr. **χafa-*, that it could be borrowed or represent a ‘Kraftwort’. Burrow (1973, p. 26) suggests borrowing from Uralic, cf. Hung. *hăb* ‘foam, froth’, Veps. *kobé* ‘wave, foam’, and Sam. (Kam.) *khòwü* ‘foam’. However, since these words go back to PU **kompa-* ‘wave’ (Sammallahti, 1988, p. 537), it is unclear why the **m* would not be reflected in Indo-Iranian.[33] Moreover, the Uralic word does not refer to bodily fluids like in Indo-Iranian but rather to frothy water. In view of the formal and semantic problems, a Uralic origin must be regarded as speculative.

24. PII **kHā-* ‘well, source’

Ind. Skt. *khā́-*

Ir. LAv. *xā-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b), due to its semantics and phonology. *EWAia* (I, p. 451) offers no IE etymology, and denies a connection to the verb Skt. *khan-*, LAv. *kan-* ‘to dig’. *LIV* (p. 344), however, reconstructs PIE **k$^{(u)}$eh$_2$-* ‘to dig’ with a nasal present **k$^{(u)}$-né/n̥-h$_2$-*, which allegedly yielded a secondary root **kanH-* and the PII word for ‘well’. The aspirated *kh-*

[33] It is of course possible that PII borrowed from a Uralic language where **m* had been lost.

in Sanskrit is explained as secondary from the zero-grade **kh*(*i*)-, but this is not attested anywhere. Even if the connection between 'well' and 'to dig' is maintained, the problem remains that both are isolated to Indo-Iranian.

25. PII **kHara*- 'donkey'

Ind. Skt. (AVP+) *khara*-

Ir. LAv. *xara*-

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b), Witzel (2003, p. 35), and hesitantly *EWAia* (I p. 447). A relation to Akk. (Mari) *ḫāru* (*ḫârum*, *ayarum*) 'donkey foal', as proposed by Eilers (1959, p. 467), is plausible, but it is difficult to prove that Akkadian was the direct source of the PII word. The Akkadian word, in turn, is likely borrowed from West Semitic *ʿaḫr* 'foal' (*CAD* H, p. 118), which is semantically and formally further removed from **kHara*-.

Witzel (2003, p. 29) believes PII **kHara*- to be connected to Skt. *garda-bhá*- 'donkey' and ToB *kercapo* 'donkey' < PToch. **kercäpā*-. The Tocharian word is likely borrowed from Indo-Iranian, since *-$b^h a$- is a common "animal-suffix" in Indo-Iranian.[34] Tocharian probably borrowed the word before the PII vowel merger as *$gord^{(h)}eb^h o$-, since *$d^{(h)}$ was palatalized to **c* within Tocharian (Adams, 2013, p. 210). If *$gord^{(h)}e(b^h o)$- was borrowed into early PII, it is possible that **kHara*- was borrowed into PII at a later stage. Possibly, PII **g*- (= [g^{ʔ}] / [k^{ʔ}]) and *kH*- (= [kʔ] or [kh]) represent different adaptations of the same sound in the source language, since both contain a glottalic element.

26. PII **kšīra*- 'milk'

Ind. Skt. *kṣīrá*-

Ir. MiP *šīr*, Y-M *xšīra*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b), partly due to the cluster **kš*-. *EWAia* (I, p. 433) does not offer an IE etymology. Kroonen (2013, pp. 261-2) argues for a connection to PGm. **hwaja* ~ **huja*- 'whey', which he derives from an *i*-stem adjective *tk^w-*ōi*-. According to him, PII **kšīra*- would derive from *$tk^w ih_2$-*ro*-, a *-*ro*- derivation of a *h_2-collective of *tk^w-*ōi*-. A similar derivation would be Alb. *hirrë* 'whey' < *$tk^w iH$-*r*-*neh*$_2$-, although this may be a loanword from Hungarian (Szemerényi, 1958, p. 171). To explain the semantics, Kroonen argues that *tk^w-*ōi*- has the same root as PGm. **Þinhla*- 'curdled milk' and Skt. *takrá*-

[34] Cf. Skt. *vr̥ṣa-bhá*- 'bull'

‘buttermilk’ < *t(*e*)*mk-lo-*. This is difficult, however, since the root **temk-* ‘to thicken’ has a nasal which is absent in **tk^{w}-ōi-*. Furthermore, it is implausible that a word for raw milk would be derived from a root meaning ‘to thicken’, which is clearly associated to processed milk. Based on these considerations, Kroonen’s etymology of PII **kšīra-* seems unlikely.

27. PII **kućsi-* ‘round side of the body’

Ind. Skt. *kukṣí-* ‘cheek’

Ir. Sogd. *qwšy-* ‘side of the body’

Nur. W. *küc* ‘belly’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). The morphology is not easily explainable from an IE perspective. *EWAia* (I, p. 360) argues for a derivation based on an unattested *s*-stem **kuć-as-*. However, the root of **kuć-as-* itself is unknown: the proposed root cognates LAv. *kusra-* ‘arching’, Skt. *kuśá-* ‘grass’ or Skt. *kuśī́* ‘?’ also lack IE etymologies and are semantically dissimilar.

28. PII **matsi̯a-* ‘fish’

Ind. Skt. *mátsya-*

Ir. LAv. *masiia-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (II, p. 298) supports a connection to PGm. **mati-* ‘food’. According to this etymology, **matsi̯a-* is a **-i̯o-*derivative of an unattested *s*-stem **med-s-*. However, such a derivational chain is improbable and it is unclear why the word would mean ‘fish’.

PU **maća* ‘fish net’, reflected in Mar. *mača* ‘fish net’, Sam. Selk. Ke. *maazeng*, N *mā́šek*, Ty. *mə̑̄sa* ‘net, trawl’ (*UEW*, p. 263), is a possible source of the Indo-Iranian word, in which case the meaning ‘fish’ could have developed metonymically from ‘fish net’.[35] However, *UEW* concedes that the reconstruction is uncertain due to the scarcity of cognates. Indeed, neither Sammallahti (1988) nor Aikio (2015) reconstruct PU **maća*. According to them, Mari *-č-* regularly corresponds to Selk. *-tc-*, *-c-* or *-tč-*, cf. PU **ceca* ‘uncle’ > Mar. *čəčə* ~ Selk. Ke. *citca*, *cica*, N *četčeka*. It is thus unlikely that PII **matsi̯a-* was borrowed from Uralic.

[35] I thank Niels Schoubben for bringing this possibility to my attention. Décsy (1990, p. 89) reconstructs **matja*, where *tj* is simply an alternative way to write *ć*.

29. PII **mr̥ga-*

Ind. Skt. *mr̥gá-* ‘wild forest animal’

Ir. LAv. *mərəγa-* ‘bird’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (II, p. 371) mentions a proposed connection to Gr. *μάργος* ‘mad, furious’. However, this word shows so much irregular variation in within Greek that it is most likely Pre-Greek in origin (Beekes, 2010, p. 905). As the semantics relate to flora and fauna, a loanword seems likely.

30. PII **muska-* ‘testicle’

Ind. Skt. *muṣká-* ‘testicle’

Ir. MiP *mušk* ‘musk’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a *Wanderwort* by Lubotsky (2001b), who believes that MiP *mušk* was borrowed from Indic. However, the MiP form could also have been inherited from PII, especially since the suffix **-ka-* is characteristic of other proposed PII loanwords. *EWAia* (II, p. 363) argues that Skt. *muṣká-* ‘testicle’ is derived from PIE **mṓs-* ‘mouse’, evolving from a literal meaning ‘little mouse’, similar to Lat. *mūs-culus* ‘muscle’. As *muṣká-* has a short *ŭ*, this would require the long **ū* of **mṓs-* to be explained by monosyllabic lengthening. However, the acute accent of SCr. *mȉš* ‘mouse’ < **muHs-*, ToB *maścitse* < **mu̯H̥s-* (Beekes, 2010, p. 985) and Nur. Prasun *mṍ̃s’ū* ‘mouse’ (without *RUKI*) all point to PIE **muHs-*. Therefore, Skt. *muṣká-* cannot be derived from ‘mouse’. Gr. *μόσχος* ‘musk’ is likely borrowed from an Iranian source, and further spread to Lat. *muscus* ‘musk’.

31. PII **nagna-*

Ind. Skt. (AVP+) *nagnáhu-* ‘yeast’

Ir. Sogd *nγny*, Pashto *naγan*, Bal. *nagan*, *naγan*, MiP *nān* ‘bread’ < PIr. **nagna-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (II, p. 6) assumes that the Sanskrit word was borrowed as **nagna-hvā-* from Iranian **nagna-xᵛada-*, cf. MoP *nānxᵛāh-* ‘bread spices’. This is rather *ad hoc*, since the change from **nagna-hvā-* > **nagna-hu-* in Sanskrit is left unexplained.

Bailey (1979, p. 179) derives **nagna-* ultimately from PIr. **ni-kana-* lit. ‘put down (into the ashes)’. Semantically, this is acceptable, but the change from **ni-* > **na-* is irregular. More likely, this is a loanword.

32. PII **pāpa-* ‘bad’

Ind. Skt. *pāpá-*

Ir. LAv. *pāpa-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (II, p. 120) offers no IE etymology, rejecting a connection to Skt. *pāmán-* ‘skin disease’ and Gr. *πῆμα* ‘disaster, sorrow’. Morphologically, **pāpa-* looks non-IE due to the suffix **-pa-*, found in other proposed loanwords. Assuming reduplication of a root **peH-* (**pe-pH-o-*) cannot explain **pāpa-*.

33. PII **parsa-* ‘sheaf’

Ind. Skt. *parṣá-*

Ir. LAv. *parša-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Witzel (2003, p. 33). *EWAia* (II, p. 101) gives no IE etymology. The suffix **-sa-* is perhaps found in another loanword, **pīi̯ūša-*, although in **parsa-*, the **s* could also be analyzed as part of the root. The fact that ‘sheaf’ is an agricultural term increases the likelihood that **parsa-* is a loanword.

34. PII **pau̯asta-*

Ind. Skt. *pavásta-* ‘cover, garment’

Ir. OP *pavastā-* ‘clay envelope’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (II, p. 105) dismisses the idea that the Iranian word is borrowed from Indic and concludes that the word has no etymology. The long final *-ā* in OP is regular. Since the morphology (suffix **-asta-*?) is inexplainable from an IE perspective, **pau̯asta-* is likely a loanword.

35. PII **rāći-*

Ind. Skt. *rāśí-* ‘heap, mass’

Ir. Pashto *ryāša-* ‘heap (of grain)’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a possible loanword by Lubotsky (2001b), who does not exclude a connection to Skt. *raśmí-* ‘reins, rope’. This etymology is hesitantly advocated in *EWAia* (II, p. 449), who postulates a root PII **rać-* ‘to bind’ < PIE **laḱ-*, with a supposed cognate in Lat. *laqueus* ‘loop of rope’ and *lacio* ‘to entice’. However, as de Vaan (2008, p. 327) points out, Lat. *laqueus*

reflects a labiovelar *k^w rather than a palatal *$ḱ$. As for *lacio*, it is likely connected to Lat. *lacer* ‘mutilated’ and Gr. *ἀπέληκα* ‘I have torn off’ and *λακίς* ‘tatters of clothes < PIE *$l(e)h_2k$- ‘to tear’ (Beekes, 2010, p. 826), in which case the *a*-vocalism in Latin reflects a laryngeal[36], incompatible with Skt. *raśmí-*. Formally, PII *$rāći$- ‘heap’ could derive from *$leh_2ḱ$-*i*- but the semantics are difficult to explain. The word may have a connection to agricultural terminology.

36. PII *$ringa$- ‘mark’

Ind. Skt. *liṅga-*

Ir. LAv. (*haptō-*)*iringa-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Witzel (2003, p. 33). *EWAia* (II, p. 479) argues that the word is related to Lith. *lýgus* ‘alike’ < PIE *$leig$-, also reflected in PGm. *$līka$- ‘alike’. The foremost problem of this etymology is the *-*n*- in Indo-Iranian, resembling the IE nasal infix, which is entirely unexpected in a nominal form. Comparative evidence for this nasal has been argued to exist in Gr. *ἐναλίγκιος* ‘like’ (Pisani, 1981, p. 207), but this is impossible due to the irregular correspondence Gr. *-nk-* : PII *-ng-*. A root of the structure *$Reing$-/*$Rieng$- is unparalleled in IE (cf. *LIV*).

37. PII *$ṛsi$- ‘seer’

Ind. Skt. *ṛ́ṣi-*

Ir. Av. *ərəši-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b), due to the abnormal initial accent in Sanskrit. As shown by Lubotsky (1988, p. 54), Sanskrit. *i*-stems with zero-grade in the root are generally oxytone.

38. PII *$sćāga$- / *$sćaga$- ‘male goat’

Ind. Skt. *chā́ga-*, *chagalá-*

Ir. Oss. *sæğ*/*sæğæ*, Wakh. *čəy*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (I, p. 558) argues that the long root vowel in Sanskrit is a back formation from the feminine *chā́gā-* ‘female goat’, itself a *vṛddhi*-derivation of *$sćaga$-. However, the variation may also be an indication of borrowing.

[36] And not PIE *o > Lat. *a* / *l*_

A connection to PGm. **skēpa-* 'sheep' has been suggested. However, that the Germanic **-p-* would be the result of dissimilation from earlier **skēka-* (cf. *EWAia* I, p. 559) is *ad hoc*. One could postulate a parallel borrowing of a preform **skĕ́g^{w}o-* into Germanic and Indo-Iranian, but it is improbable that Germanic would nativize **g^{w}* as **b*, since Germanic retains the phoneme **g^{w}*. Moreover, if the long vowel of Skt. *chága-* is secondary, the PII form **sćága-* looks less similar to PGm. **skēpa-*.

39. PII **spāra-* 'ploughshare'

Ind. Skt. *phála-*

Ir. MoP *supār*, Išk. *uspir*, Wakh. *spūndr*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b), since the anlaut Skt. *ph-* : Ir. *sp-* is irregular.[37] Semantically, a loanword is not unexpected, since 'ploughshare' is an agricultural term. *EWAia* (II, p. 204) argues that Skt. *phála-* < **spāla-* and that PII **spāla-* < PIE **spelH-* 'to split' can be reconstructed. However, the proposed cognate OCS *plěvǫ* 'to weed, separate the husk from the grain" (LIV, p. 577) is semantically unconvincing and requires the assumption of *s*-mobile, which is not synchronically attested in either Slavic or Indo-Iranian.

Another problem is that **sp-* > *ph-* is a Middle Indic development, which requires the assumption that Skt. *phála-* is a 'prakritism', i.e. a colloquial form which had undergone a 'Middle Indic' sound change already in Vedic times. However, this hypothesis may be supported by the context in which Skt. *phála-* is attested. The word occurs in RV X.117 (the 'Praise of Generosity'), which, according to Jamison & Brereton is "unusual in both subject matter and tone" and makes "no mention of divinities […], an almost unique situation in the Ṛgveda" (2014, p. 1586). Importantly, the style is "colloquial and conversational" (ibid.). In this context, a colloquial word is not unexpected. Since Vedic is not equal to Proto-Indic, it is not impossible that certain lower-prestige dialects parallel to Vedic had already undergone a sound change **sp-* > *ph-* at this time. Based on these considerations, **spāra-* will be treated as a PII loanword.

[37] The chronology of the development of PII *sP*-clusters in Indic is complicated. As pointed out by Kobayashi (2004, p. 72), the change of PII **sć* > Skt. (*c*)*ch* precedes the change of Skt. *sp-* > MI (*p*)*ph* by hundreds of years.

40. PII **stuka-* / **stupa-* ‘tuft of hair’

Ind. Skt. *stúkā-*, *stupá-* ‘hair’, *stū́pa-* ‘hair, top beam of house’

Ir. Oss. *styg*/*stug*, Y-M *stūγ* ‘long hair’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b), due to the suffixal variation *-ka-*/*-pa-* in Sanskrit. Although *EWAia* (II, p. 760) agrees that Skt. *stúkā-* and *stupá-* are connected, it suggests that the variation can be ascribed to dissimilation or different extensions of an unattested root **stu-*, neither of which seems likely. Moreover, the suffixes **-ka-*/**-pa-* are attested in other proposed loanwords.

41. PII **u̯āćī-*

Ind. Skt. *vā́śī-* ‘axe’

Ir. LAv. *vāsī-* ‘pointed knife’, Oss. *was* ‘axe’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (II, p. 548) assumes IE origin in connection to OHG *wahs* ‘sharp’, or as an Indo-Iranian innovation based on the root Skt. *vāś-* ‘to roar’ (cf. Germ. *Klinge* ‘blade’ ~ *klingen* ‘to sound’). The latter explanation is semantically highly uncertain and further unhelpful since *vāś-* lacks IE comparanda. In the AV, Sanskrit has a parallel form *vāsī-* with an unexpected dental *s*, which could indicate non-IE origin.

Parpola (2012, p. 161) and Kümmel (2019) argue that PII **u̯āćī-* is borrowed from PU **weŋći* ‘knife’, reflected in Finn. *veitsi* ‘knife’, *veitsä* ‘to cut’, Hung. *vés* ‘to chisel’ (cf. *UEW*, p. 565). However, the reconstruction of PU **weŋći* is uncertain, since the regular outcome of **ŋć*[38] is Hung. *gy* (Pystynen, 2014). The word is not reconstructed by either Sammallahti (1988), Aikio (2015) or Zhivlov (2014). Until a solution for the problematic Uralic correspondences is found, the connection to Indo-Iranian remains speculative.

42. PII **u̯and*(*H*)- ‘to praise’

Ind. Skt. *vand*^{i}-

Ir. Av. *vaṇd-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (II, p. 502) and LIV (p. 681) project the Indo-Iranian forms to PIE **u̯end-*. However, the seṭ-character of the Sanskrit root points to **u̯endH-*, which we would expect to give Skt. ***vandh-*. This increases the likelihood that PII

[38] As Pystynen (2014) points out, the proper reconstruction is **ńć*.

*$\underset{\sim}{\mu}$and(*H*)- is a loanword. The root has no archaic derivations that would indicate that it is old (Macdonell, 1916, p. 416). As it is often used in religious contexts, *$\underset{\sim}{\mu}$and(*H*)- could have been borrowed along with deity names such as **indra*- etc.

43. PII *$\underset{\sim}{\mu}$arāj^{h}a- ‘wild boar’

Ind. Skt. *varāhá-*

Ir. LAv. *varāza-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b), due to the trisyllabic *CVCV̄CV* structure. *EWAia* (II, p. 514) mentions a proposed connection to PCelt. *ǵ^{h}or$\underset{\sim}{\mu}$o- with irregular “Konsonantenvertauschung”, which can hardly be correct. Although PFV **orase* ‘boar’ looks related, it is probably borrowed from an Indo-Iranian language (Rédei, 1986, p. 54) .

44. PII **umā*-(*kā*)- ‘flax, linseed’

Ind. Skt. *úmā-* ‘flax’

Ir. Y *imoγō*, *ümoγō*, M *yimaγå* ‘linseed’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a Post-PII *Wanderwort* by Lubotsky (2001b). *EWAia* (I, p. 225) dismisses non-IE origin but offers no alternative etymology. The initial vowels of the Yidgha-Munji words derive from **u*- with umlaut (Morgenstierne, 1938, p. 96). Since the suffix *-kā-* is productive in Iranian, there is nothing against assuming that this word entered the language in PII times. Skt. *kṣumā-* ‘flax’ may be indirectly related. It is noteworthy that flax was cultivated in the BMAC culture (Spengler et al., 2014).

45. PII *$\underset{\sim}{\mu}$rīj^{h}i- ‘rice’

Ind. Skt. *vrīhí-*

Ir. Pto. *wr’iže*, OP **vrīzi*- < PIr. *$\underset{\sim}{\mu}$rīǰi-, Khot. *rrīysū* < PIr. *$\underset{\sim}{\mu}$rīǰuka*, Orm. *rízan* < PIr. *$\underset{\sim}{\mu}$rīǰana-*, Sogd. *rysk* < PIr. *$\underset{\sim}{\mu}$rīǰaka-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a *Wanderwort* by *EWAia* (II, p. 598). Kümmel (2017, p. 283) reconstructs PII *$\underset{\sim}{\mu}$rīj^{h}i- based on Skt. *vrīhí-*, Pto. *wr’iže* and OP **vrīzi-*, the latter being reconstructed based on Elam. *mi-ri-zi-iš* ‘rice’. The Elamite form is probably borrowed from “Median”, where PIr. *ǰ$^{(h)}$ > *z*.

The other Iranian forms show various (productive) suffixes *-uka-*, *-aka-*, and *-ana-*, neither of which precludes that they were inherited from PII and later reshaped.

It is often assumed that Gr. *ὄρυζα* ‘rice’ was borrowed from an Iranian language (Beekes, 2010, p. 1112). If so, Greek probably adopted the word when *υ* = [i] and *ζ* = [z], as the ancient Greek pronunciation would not reflect the Iranian phonology.

The fact that rice was not cultivated in the BMAC culture (Spengler et al., 2014) reduces the likelihood that **u̯rīj^{h}i-* originated in the Central Asian Substrate, even though it is reconstructable to PII.

In addition to the above, Iranian has another similar word for rice, reflected by MiP *blnc*, *brynz*, Sogd. *βrync* ‘rice’, reconstructable as PIr. **brinǰa-* (Kümmel, 2017, p. 283). A descendant of PIr. **brinǰa-* is probably the source of Gr. *ὀρίνδης* ‘rice flour bread’ (Beekes, 2010, p. 1102). Furthermore, Arm. *brinj* ‘rice’ seems to be borrowed from a Middle Iranian source. Since it only exists in Iranian, **brinǰa-* was most likely borrowed at a later stage than PII **u̯rīj^{h}i-*. However, PIr. **brinǰa-* contains an **-n-* which is absent in the older word. Since **ī* is unlikely to develop into **-in-*, this indicates that **brinǰa-* was borrowed from a different source language than **u̯rīj^{h}i-*, rather than the a later historical stage of the same language.

46. PII **u̯r̥tka-* ‘kidney’

Ind. Skt. *vr̥kká-*

Ir. OAv. *vərəδka-*, MiP *gurdag*, Khot. *bilga-*, Y-M *wulγa*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (II, p. 571) argues that it is derived from **vart-* ‘to turn’. However, since the suffix *-ka-* is denominal, whereas *vr̥t-* is a plain root, this is improbable. Also, the etymology is not semantically convincing.

3.3. Loanwords II: Proto-Indo-Iranian, borrowed after certain sound changes

47. PII **čāt(u̯āla)-* ‘pit, well’

Ind. Skt. (Br.) *cā́tvāla-*

Ir. LAv. *cāt-*, Sogd. *čʾt*, Bactr. *σαδο*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). No IE etymology has been suggested (*EWAia* I, p. 539). The Sanskrit word is not directly comparable to the Iranian forms due to the “suffix” *-vāla-*, itself of unknown origin. Skt. *cā́tvāla-* has the trisyllabic *CVCV̄CV* structure, which is unparalleled in Iranian. Bactr. *σαδο* reflects either **čātă̆-* or **čāti-* (Davary, 1982, pp. 137, 264). The morphological variation between Indic and Iranian suggests a more recent time of borrowing.

48. PII **i̯au̯ī̯ā-* ‘canal’

Ind. Skt. **yavīyáˊ-* (metrically restored), *yavyáˊ-*

Ir. OP *yauviyā-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b), on account of its *CVCV̄CV* structure. *EWAia* (II, p. 405) does not offer an IE etymology. According to Witzel (2003, p. 32), the Sanskrit and OP words cannot go back to the same proto-form: this is incorrect, however, because *a* > *au* /_*v* occurs elsewhere in OP, and the length of *i* cannot be deduced from the script, cf. *tauviyah-* ~ Skt. *távyas-* / *távīyas-* ‘stronger’. For the inclusion of **i̯au̯ī̯ā-* in layer II, I refer to section 4.3.

49. PII **ǰaǰʱa/ukă̆-* ‘hedgehog’

Ind. Skt. *jáhakā-* (YV+), Lahnda *jahā-*

Ir. LAv. *dužaka-*, Bal. *ǰaǰuk*, *dužux*, MoP *žūža-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (I, p. 582) mentions no convincing IE etymology. Besides, it has reduplication, which is morphologically remarkable, and words for animals are easily borrowed.

The initial *d-* of LAv. *dužaka-* and Bal. *dužux* (instead of expected **j-*) could be explained as dissimilation of *ǰ…ǰ* > *d…ǰ*[39], although OAv. *jījišəṇtī* ‘they conquer (repeatedly)’ < PII **ǰai-* (Cheung, 2007, p. 222) renders this unlikely. Another question is whether the *u*-vowel in LAv. *dužaka-* is old. *AiWB* (p. 755) analyses *dužaka-* as a compound *duž-* + *-aka-* ‘having bad hooks’. Given the related Indic words, this cannot be the actual etymology (Av. *duž-* ‘bad’ : Skt. *duḥ-* < **dus-*), but folk-etymologically this analysis is conceivable. Thus, it is possible that Iranian speakers remodelled **ǰažaka-* (with regular *ž* < **ǰ*) as **duž-aka-* ‘having bad hooks’.

Although the *-u-* in the initial syllable of some Iranian forms may be secondary, Bal. *ǰaǰuk* points to PIr. **ǰaǰuka-*, whereas the Indic forms point to **ǰaǰʱakā-*. The suffixes *-uka-* and *-aka-* could be secondary innovations within Iranian and Indic, in which case Lahnda *jahā-* preserves an older form. However, the morphological and phonological variation suggests that the word was borrowed in late PII. Since it occurs before **u* and **ā*, **ǰʱ* was most likely not palatalized within PII but borrowed as such.

[39] Cf. Khot. *dasta-* ‘hand’ < **ǰasta-* with dissimilation of the fricative element of **ǰ*.

50. PII **kárus-* ‘damaged’

Ind. Skt. *kárū-ḍatin-* ‘with damaged teeth’

Ir. Sogd. *krwʾ* ‘gap’, *krw δntʾk* ‘with damaged teeth’, MoP *karve* ‘decayed teeth’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (I, p. 313) gives no IE etymology. MoP *karve* is part of a larger group of borrowings from Sogdian (Henning, 1939, p. 96) and thus does not inform on the PII situation.

Sogd. *krw δntʾk* is read /*karw dandāk*/. Importantly, the first part of the compound is also attested as simplex *karw’* /*karwā*/ (Gharib, 1995, p. 194), for which reason *EWAia* (I, p. 313) reconstructs PIr. **karu̯a-* ‘gapped, damaged’. This reconstruction is homophone to the preform of LAv. *kaurva-*[40] ‘thin-haired’ < PIr. **karu̯a-* ‘thin-haired’, cognate to Skt. *kṷ̌lva-* ‘thin-haired’ < PII **kĺ̥H-u̯o-*, and further related to Lat. *calvus* ‘bald’.[41] Henning (1939, p. 96) proposed an etymological relation between PIr. **karu̯a-* ‘thin-haired’ and PIr. **karu̯a-* ‘gapped, damaged’. However, if the latter is linked to Skt. *kárū-ḍatin-*, this is impossible, since the retroflex *-ḍ-* and long *-ū-* show that Skt. *kárū-* goes back to **karuž-* < **karuš-*. The suffix **-us-* (*s*-stem derived from *u*-stem?) is rare but not unparalleled in Indo-Iranian (*AiGr*. II, 2, p. 477).

The fact that Skt. *kárū-* goes back to **karuž-* precludes a scenario where Skt. *kárū-ḍatin-* was borrowed from (proto-)Sogdian, as Sogdian preserves *-z-* in this position, e.g. *ʾztyw* < *uzdahyu-* (Gershevitch, 1961, p. 44). The reverse scenario, that Sogd. *krw δntʾk* (with alternative reading as /*karu dandāk*/) was borrowed from Sanskrit, would explain the lack of *-z* in Sogdian, but cannot explain the existence of Sogdian simplex *krw’* /*karwā*/ ‘gap’. To save the latter scenario, one would have to separate Sogd. *krw’* ‘gap’ from *krw δntʾk* ‘with damaged teeth’, and instead connect it to LAv. *kaurva-* ‘thin-haired’ and Skt. *kṷ̌lva-* ‘id.’, which seems rather *ad hoc*.

As both possible directions of borrowing seem impossible, the remaining possibilities are to assume that Sogd. *krw δntʾk* was remade based on a synchronic stem **krw* which had somehow lost its *-s*, or that Sogdian and Sanskrit borrowed from slightly different sources. As

[40] For the meaning ‘thin-haired’, cf. Lubotsky (1997, p. 42). Naturally, the reading of LAv. *kaurva-* as ‘thin-haired’ and not ‘damaged’ hinges partly on the etymological identification with Skt. *kṷ̌lva-* ‘thin-haired’. For the compounds *kaurvō.gaoša-* ‘with … ears’, *kaurvō.dūma-* ‘with … tail’ and *kaurvō.barəša-* ‘with … neck’ (*AiWB*, p. 456), all said of horses, *kaurvō-* could theoretically be read as either ‘damaged’ or ‘thin-haired’, although the latter seems slightly more likely in the context.

[41] Cf. de Vaan (2008, p. 85). To account for the *a*-vowel of Lat. *calvus*, we must reconstruct a thematicized weak *u*-stem **klH-eu-o-*. In Indo-Iranian, the word was thematicized as **kĺ̥H-u̯o-*, with accented zero-grade in the root. This explains the long **ū* in Skt. *kūlva-* and non-labialized PIr. **karu̯a-*. EWAia (I, p. 449) suggests that Skt. *kharvá-* ‘mutilated’ should be equated to PIr. **karu̯a-* ‘damaged’, but this is impossible since Skt. *kh-* < **kH-* and corresponds to Ir. *x-*.

the compound with the IE word for ‘tooth’ is quite specific, it seems better to assume the former scenario, i.e. a single borrowing event with a subsequent loss of *-s* in Sogdian.

The structure of PII **kárus-* is difficult to derive from Pre-PII. With its non-palatalized anlaut, it could only reflect Pre-PII **kórHus-*.[42] However, this is difficult, since Pre-PII **kórHus-* should have undergone the laryngeal accent shift (> ***korHús-*) and laryngeal metathesis (> ***koruHs-*). Accordingly, as **k-* is not palatalized, PII **kárus-* must either have been borrowed with an **a* (after the phonologization of **a*), with an **ŏ* (after BrL), or after the palatalization of velars.

51. PII **maḭūkHa-* ‘peg’

Ind. Skt. *mayū́kha-* ‘peg for stretching the woof’

Ir. OP *mayūxa-* ‘doorknob’, Sogd. *myγk* ‘peg’, MiP and MoP *mēx* ‘peg, nail’, Oss. *mīx*/*mex* ‘stake’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b), who argues against a derivation from the root *may-* ‘to erect, build’ (*EWAia* II, p. 317), since the “suffix” **-ūkHa-* is left unexplained. For the inclusion of **maḭūkHa-* in layer II, I refer to section 4.3.

52. PII **pīḭūša-* ‘beestings’

Ind. Skt. *pīyū́ṣa-*

Ir. Wakh. *pyix̌*, Munji *fəyū*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

EWAia (II, p. 138) considers this word to be related to Skt. *páyas-* ‘milk’ and *pay-* ‘to swell’. Lubotsky (2001b, p. 303) rejects this due to the unusual derivation (suffix *-ūsa-* / *-sa-*?) and the unexpected long *ī*, and argues that the word is a characteristic loanword with the structure *CVCV̄CV*. The idea that Skt. *pīyū́ṣa-* is a compound of *pī-* ‘to swell’ + *yū́ṣa-* ‘broth’ (*AiGr*. II, 2, p. 500) is semantically unlikely and formally problematic, since *pī-* is a bare root. For the inclusion of **pīḭūša-* in layer II, I refer to section 4.3.

[42] The laryngeal closes the syllable, accounting for the lack of BrL. Other input forms are impossible: Pre-PII **kerus-* > ***čarus-*, **korus-* > ***kārus-*, **kN̥rus-* > ***kaNrus-*, **keh$_{2\text{-}3}$rus-* > ***kaHrus-*, and **kHerus-* > Ind. ***kharus-* ~ Ir. ***xarus-*.

53. PII **pusća-* ‘tail’

Ind. Skt. *púccha-* ‘tail’

Ir. LAv. *pusa-* ‘head dress’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (II, p. 140) maintains the possibility that the word is related to PGm. **fuhsa-* ‘fox’. This etymology is supported by Kroonen (2013, p. 158), who reconstructs **puk-so-* for the Germanic word and **puḱ-sk-o-* for Indo-Iranian. However, the fact that Indo-Iranian and Germanic do not reflect a single PIE form, and the abnormal suffixes **-so-* / **-sko-* weakens the plausibility of the etymology. Furthermore, the correspondence Skt. *-cch-* : Av. *-s-* reflects PII **-sć-* < **-sč-*. Yet the paradigm of **puḱ-sk-o-* (> **pu-sk-o-*[43]) would not provide a palatalizing context for the **-sk-* cluster, except in the vocative **puḱ-sk-e*, which, for a word ‘tail’ is a highly unlikely model of analogy. Thus, PII **pusća-* must have been borrowed after the PII phonologization of palatalized **sč*.

A possible source of **pusća-* is PFU **ponci* ‘tail’.[44] However, since it is unclear why the PFU **-n-* is not reflected in PII, the connection remains speculative.

3.4. Loanwords III: Post-Proto-Indo-Iranian

54. Pre-II **?

Ind. Skt. *áṇu-*

Ir. MiP *ʾrzn*, MoP *arzan*, Pto. *zdən*, Wakh. *yirzn* < PIr. **(H)arǰaná-*, MiP *ʾlwm*, Baxt. *halum*

Nur. V. *üǰʾũ̃*, A. *azʾũ* , W. *ə̃zü* < **(H)arǰaná-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Kümmel (2017, p. 283) proposes that the Iranian and Nuristani words are loanwords, and adds Skt. *áṇu-* ‘millet’ as a possible cognate. This connection presupposes PII **-rǰn-* > Indic *-ṇ-*, for which I know of no other examples. Skt. *áṇu-* is thus unlikely to be related to the Iranian and Nuristani words. MiP *ʾlwm* and Baxt. *halum* have been influenced by **ganTuma-* ‘wheat’.

[43] **-ḱsk-* was simplified to *-sk-* probably already in PIE, cf. **prḱ-ske-* > **pr-ske-* ‘to ask’.
[44] For the reconstruction cf. Sammallahti (1988, p. 547).

55. Pre-II **banγα-* ‘hemp’

Ind. Skt. *bhaṅgá-* ‘hemp’

Ir. LAv. *baŋha-* ‘a plant, narcotic’, MoP *bang* ‘hemp’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Witzel (2003, p. 34). *EWAia* (II, p. 241) argues that MoP *bang* is borrowed from Sanskrit. For Skt. *bhaṅgá-*, a semantically unconvincing connection to *bhañj-* ‘to break’ is suggested. LAv. *baŋha-* (< (virtual) **b^{h}ansa-* / **b^{h}asa-*) is semantically and formally close but does not correspond regularly to Skt. *bhaṅgá-*. Kümmel (2019) reconstructs PII **b^{h}anga-*, which he interprets as a borrowing from PU **pɐŋka-* ‘mushroom’. However, the reconstruction is based only on Sanskrit. Given the synchronically close form and meaning of Skt. *bhaṅga-* and LAv. *baŋha-*, it seems more plausible that they represent parallel Post-PII borrowings, with different adaptations of a Pre-II **banγα-* (vel sim.).

56. Pre-II **ćika*(*tā*)-

Ind. Skt. *síkatā-* ‘sand, gravel’

Ir. OP *θikā-* ‘gravel’, Sogd. *šykth*, Khot. *siyatā-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b) and *EWAia* (II, p. 728) due to the irregular correspondence Ind. **s-* : Ir. **ć-* and the absence of an IE etymology. Lubotsky (2001b, p. 306) argues for an Indic >> Iranian direction of borrowing, but this need not be the case. Since OP and Sogdian seem to reflect PIr. **ćika*(*tā*)-, where **ć* is still an affricate, it was probably not borrowed from Indic, which has initial **s-*. Khot *siyatā-* reflects **ćiɻatā-*.

To assume an independent parallel borrowing by Indic and Iranian is complicated by the difficulty of determining a plausible common source form of the word. If the source language had an initial dental affricate [ʦ], we would expect PIr. **ć-* but Skt. *ts-* or possibly *kṣ-*. Reconstructing a palatal affricate [ʨ] might be compatible with PIr. **ć-* but can hardly explain Skt. *s-*. Ultimately, however, parallel borrowing cannot be excluded, since irregular adaptations to the native linguistic structure may have occurred.

The other direction of borrowing, Iranian >> Indic, is possible if one assumes that Skt. *síkatā-* was borrowed from an Iranian language where PIr. **ć-* > **s-*. Given the difficulty of the two other scenarios, this might be the most plausible one.

57. PII **dŭ̆rćа- / *dr̥ćа-* ‘(goat’s) wool, hair’

Ind. Skt. *dūrśá-* ‘(large) garment’

Ir. Wakh. *δirs*/*δırs*/*δürs*/*dərs* ‘wool of goat/yak’, Šu. *δox̌c* ‘body hair, course cloth’, Y-M *lirs*/*līrs*/*lurs* ‘goat’s hair’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

EWAia (I, p. 740) reconstructs PIE **dr̥H-ḱo-*, but the non-IE-looking suffix **-ḱo-* is not reflected in any of the proposed root cognates. Lubotsky (2001b, p. 311) takes it as a loanword based on the irregular correspondence between Indic and Iranian, reconstructing PII **dr̥Hća-* / **dr̥ćа-*. However, PII **-r̥H-* > Skt. *-ūr-* only regularly occurs in labial contexts, i.e. $C_{+labial}$ _ or _$Cu̯$, which means that Skt. *dūrśá-* rather reflects PII **dūrćа-*.

Wakh. *δirs*/*δırs*/*δürs*/*dərs* reflect dialectal variants (Morgenstierne, 1938, p. 481). PII **dr̥ćа-* is a possible ancestor of Wakh. *dərs* (ibid.). However, *-ə-* may also reflect **-u-*, e.g. Wakh. *dəγd* ‘daughter’ < PIr. **dugdar-*. The other forms, Wakh. *δirs*/*δırs*/*δürs*, do not show the normal outcome of **r̥*, nor the outcome of **r̥* in a labialized context (e.g. Wakh. *pʊrs* ‘to ask’ < PII **pr̥sćа-*). Instead, the vowels of Wakh. *δirs*/*δırs*/*δürs* seem to reflect delabialized **-u-* (Morgenstierne, 1938, p. 480). Thus, **durćа-* is a more likely origin of the Wakhi forms.

The same is true for Y-M *lirs*/*līrs*/*lurs*, where the vocalism of the attested dialectal variants could go back to **ŭ̆* (Morgenstierne, 1938, pp. 96-7), thus being compatible with a reconstruction PIr. **durćа-* or **dūrćа-*.

In Šughni, *-o-* can be the outcome of **r̥*, but only when a long **ā* in the following syllable causes *a*-umlaut (**r̥* > **ūr* > **ār* > *ox̌*, cf. Sokolova (1967, pp. 56, 58)). Moreover, **r̥* > **ūr* occurs in stressed position, whereas unstressed **r̥* becomes **ir* or **ar* (Sokolova, 1967, p. 61). Since Skt. *dūrśá-* has oxytone accentuation, this would mean that either Indic or Iranian underwent an accent shift. On the other hand, the normal outcome of PIr. **ŭ̆* is Šughni *u*, unless affected by *i*- or *a*-umlaut, becoming *i* or *a* (Sokolova, 1967, p. 49). It is thus difficult to connect Šu. *δoxc* the other Indo-Iranian forms by regular sound changes.

In conclusion, the attested words variously point to PII **dūrćа-*, **dŭ̆rćа-*, or **dr̥ćа-*. This suggests that the word was borrowed after the disintegration of PII.

58. Pre-II **ganDərṷa-* ‘a mythical being’

Ind. Skt. *gandharvá-*

Ir. LAv. *gaṇdərəβa-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (I, p. 462) offers no IE etymology but mentions the irregular correspondence with Gr. *Κένταυροι*. As pointed out by Lubotsky (2001b, p. 303), the Sanskrit and Avestan forms are irregular correspondences, since Skt. *-arvá-* would reflect PII **-arṷa-* whereas LAv. *-ərəβa-* would reflect PII **-r̥b$^{(h)}$a-*. The most likely explanation is that LAv. *gaṇdərəβa-* was borrowed after the sound change **b > β / V_V*. This indicates that the source language had a fricative sound, which was adapted as *-β-* in Avestan, since PIr. **ṷ* was still a glide. In Indic, however, the fricative was adopted as Skt. *-v-*. Skt. *gandh-* may reflect earlier **gand^{h}-* or **g^{h}and^{h}-* with Grassmann’s Law.

59. Pre-II *ganTi-* ‘smell’

Ind. Skt. *gandhá-* / *-gandhi-* ‘smell’

Ir. LAv. *gaiṇti-* ‘bad smell’, OP *gasta-* ‘evil, repugnant’, Khot. *ggañu* ‘stench’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (I, p. 461) concedes that the origin is uncertain but does not consider non-IE origin. Bailey (1979, p. 79) postulates a root **gan-* ‘to smell’, which would be the basis of the Iranian forms (e.g. *-ti-* derivation in Av.), with the variant **gan-d-* in Indic. However, the dental stop must have been part of the root, as shown by OP *gasta-* < **gn̥t-ta-* and other Iranian forms, which all show a root-final stop (Cheung, 2007, p. 103). However, Khot. acc.sg. *ggañu* seems to point to **d* rather than **t* (Bailey, 1979, p. 79). Skt. *gandhá-* has a variant *-gandhi-* in compounds, which bridges the gap to LAv. *gaiṇti-*.

Due to the irregular correspondence Indic *dh* : Iranian *t*, the word was most likely borrowed in Post-PII times. Especially noteworthy is that the same irregular correspondence is found in **ganTuma-*.

60. Pre-II **ganTuma-*

Ind. Skt. *godhū́ma-* ‘wheat’

Ir. LAv. *gaṇtuma-* ‘wheat’, Pto. *γan|əm*, Parth. *gndm*, Wakh. *ǰ̌ədim*, MiP *gnm*, Khot. *ganama*, Bal. *gandūm*, Yazg. *γ^{w}ont*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by *EWAia* (I, p. 498) and Lubotsky (2001b). Skt. *godhū́ma-* has probably been affected by folk etymology, reanalyzed as a compound *go-dhū́ma-* lit. ‘cow-smoke’

(*EWAia* I, p. 498; Kümmel, 2017, p. 282). Thus, it could very well reflect earlier **gandhu̍ma-*. The many Iranian cognates variously point to either PIr. **t* or **d* and long or short **ŭ̄* (Kümmel, 2017, p. 281). Yazg. *γʷont* < **gantu-* lacks the "*-m*-suffix" seen elsewhere in Indo-Iranian. The variation clearly points to a parallel Post-PII borrowing.

Conversely, Witzel (2003, p. 31) reconstructs a single PII form **gantuma-*, from which, according to him, the forms with *d* and *ū* developed via folk etymology. However, there is good reason for reconstructing a second preform **gandŭ̄ma-*: firstly, forms reflecting **d* are also found in Iranian languages that otherwise show no apparent trace of folk-etymological restructuring (since they preserve short *ŭ* and anlaut *gan-*); secondly, the same irregular correspondence Indic *dh* : Iranian *t* is independently attested in Skt. *gandhá-* / *gandhi-* 'smell' ~ LAv. *gaiṇti-* 'bad smell'.

Similar words for 'wheat' appear outside of Indo-Iranian. *EWAia* (I, p. 499) mentions Burušaski *gur*, pl. *guri*/*eŋ* 'wheat' < **γorum*, as well as Hitt. *kant-* 'wheat' and Arab. *ḥiṇtatun* 'wheat' < **hnt-*, to which Gr. *χόνδρος*[45] 'grain' may be added. ToB *kanti* 'bread' probably belongs here as well (Adams, 2013, p. 146).

According to Berger (1970, pp. 40-42), Burušaski is the source of all Indo-Iranian forms, since the *-m*-suffix is native to this language. However, according to Berger's data, a suffix *-m-* never occurs as a separate morpheme in Burušaski, but always as part of a plural morpheme *-miŋ*, e.g. *ǰi* 'soul' ~ *ǰimiŋ* 'souls' (1970, p. 37). Additionally, *-miŋ* is not directly attested for *gur* in Burušaski, but is according to Berger indirectly attested in Rom. *kharmin* 'wheat' << OBur. **γor-miŋ*. Berger furthermore argues that the *r* in Burušaski would have developed in the plural **γor-miŋ* < **γun-miŋ* < **γund-miŋ* and spread analogically to the singular. The Indo-Iranian forms would have been borrowed from the earliest stage when the dental stop was still preserved. The main problem with the whole scenario is the many unverified steps of Burušaski historical development. It cannot be excluded, for example, that Bur. *gur* was borrowed from an Indic language as **godum-*, and then lenited *-d-* to *-r-* within Burušaski.

Berger (1970, p. 40) argues that Rom. *kharmin* must have been borrowed from a form with initial **γ-*, and finds such a form in compounds like Bur. *sauriŋ* 'ration' < **sa-γuriŋ* lit. 'wheat of the day'. However, the fact that **-γ-* is only found in compounds means that it could represent an original **-g-* that was lenited in intervocalic position.

[45] However, the *-ρ-* remains unexplained. Also, Gr. *χόνδρος* could reflect earlier **χόνρος* with epenthetic *-δ-*.

Witzel (2003, p. 31) further adduces Basque *gari* ‘wheat’, but this is most likely a chance similarity since *gari* may come from **wari* (Berger, 1970, p. 43).

In conclusion, the direction of borrowing of **ganTuma-* is difficult to determine. The *-m*-suffix more likely originates in an unknown source language than in Burušaski. Due to the irregular correspondences, the word will be treated as a Post-PII borrowing.

61. Pre-II **Kaḭća-* ‘hair’

Ind. Skt. *kéśa-* ‘head hair’, *keśavá-* ‘with long hair’

Ir. LAv. *gaēsa-* ‘curly hair’, *gaēsu-* ‘with curly hair’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). To explain the irregular correspondence Skt. *k-* : LAv. *g-*, *EWAia* (I, p. 401) postulates a contamination of original **geśa-* with the semantically close Skt. *késara-* ‘hair, mane’, yielding *kéśa-* by analogy. However, this scenario fails to explain why *ś* was retained, while **g* was substituted, even though the model *késara-* has a dental *s*. Furthermore, Skt. *késara-* itself may be a loanword, given the absence of the *RUKI-* rule in this word (*EWAia* I, p. 401). Perhaps Skt. *kéśa-*, *késara-* and LAv. *gaēsa-* are indirectly connected as parallel Post-PII loanwords.

62. Pre-II **?*

Ind. Skt. *khaḍgá-* (JB) ‘rhinoceros’

Ir. MoP *karkadān* ‘rhinoceros’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Witzel (2003) and *EWAia* (I, p. 443). Arab. *karkaddan* ‘rhinoceros’ and Gr. *καρτάζωνος* ‘rhinoceros’ also belong here, probably as a borrowing from Persian. Kuiper (1948, p. 137) adds Akk. *kurkizānu*, which is formally similar but means ‘pig, piglet’ (*CAD* K, p. 561). In the more western languages, the words contain the “suffix” **-d*(*z*)*an*, whereas Sanskrit lacks this element. Kuiper (ibid.) identifies the prefix *kar-* as evidence for Proto-Munda origin. In that case, Skt. *khaḍ-* reflects **khar-*. In any case, the attested Indo-Iranian words are so dissimilar that they must be classified as Post-PII borrowings, probably from different sources.

63. Pre-II *?

Ind. Skt. *masū́ra-* ‘lentil’

Ir. MiP *mycwk*/*myšwk* < PIr. **mižuka-* ‘lentil’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

EWAia (II, p. 335) gives no IE etymology. Kümmel (2017, p. 284) suggests that the Indic and Iranian forms are Post-PII borrowings from the same source. Indeed, they share some features that could support this hypothesis (**mVsu-*), but differ in the suffix. In Sanskrit, unaccented *-ra-* can form denominal adjectives (*AiGr*. II, 2, p. 849-858). However, there is no indication that *masū́ra-* was originally a denominal adjective. It was more likely borrowed as a trisyllabic word with the structure *CVCV̄CV*. On the other hand, Iranian **mižuka-* could be analyzed as **miž-uka-*, as *-uka-* is a common denominal suffix. This decreases the likelihood that **mižuka-* and *masū́ra-* have the same source. Both words were likely borrowed Post-PII.

64. Pre-II **mVša-* ‘bean’

Ind. Skt. *mā́ṣa-*

Ir. MiP *māš*, Šu. *maẍ*, Sogd. *mwškh*, Yagh. *mušk* < **mušakā-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a *Wanderwort* by Lubotsky (2001b). *EWAia* (II, p. 352) concludes that there is no satisfactory IE etymology. The word is clearly non-IE since there is no regular source for *š* after *ā*. The variant **māša-* has a wider distribution than **muša-*, occurring in both Indic and Iranian, as well as in ToB *māśak* ‘mung bean’ (Adams, 2013, p. 483) and Arabic *māš*, and may have spread westwards from India. In that case, MiP *māš* could also be borrowed from Sanskrit. However, Ir. **muša*(*-kā*)- is more likely a parallel Post-PII borrowing (Kümmel, 2017, p. 284).

65. Pre-II **naḭiT*^{s}(*a*)- ‘skewer’

Ind. Skt. *nikṣ-* ‘to pierce’, *nī́kṣaṇa-* / *nékṣaṇa-* ‘skewer, fork’

Ir. LAv. *naēza-* ‘sharp point of needle’, MiP *nēzag* ‘lance’, MoP *neš* ‘skewer’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (II, p. 41) considers IE origin likely, but argues against explaining the root final **-s* as a remnant of an old desiderative. Indeed, a desiderative seems semantically unmotivated. Thus, it seems more likely that the attested variation is due to parallel Post-PII borrowing. An affricate in the source language may have been adopted as PIr. **ǰ* [ʤ], but PInd. **tš*[46] (Skt. > *kṣ*). MoP *neš* (< **naḭitš-*?) looks to be closer

[46] Although we might have expected PInd. **ǰ* = [ʤ] (< PII * **ǰ* and **ǰ*) (Kobayashi, 2017, p. 331).

to the Indic form. The long vowel of Skt. *nī́kṣaṇa-* is unexpected and could be an additional argument for non-IE origin.

66. Pre-II **pərd-a(n)k-* 'leopard, panther'

Ind. Skt. *pṛ́dāku-* 'snake'

Ir. MoP *palang* < PIr. **pard-*, Sogd. *pwrδnk*, Pto. *pṝng* 'leopard'

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Witzel (2003, p. 35). *EWAia* (II, p. 163) offers no IE etymology and doubts the connection between the Indic and Iranian forms on semantic grounds. However, the semantics of Skt. *pṛ́dāku-* 'snake' may very well be secondary, in which case Skt. (JüS) *pṛ́dāku-* 'tiger, panther' probably is closer to the original meaning. Furthermore, the structure of Skt. *pṛ́dāku-* may have been influenced by *sṛ́dāku-* 'lizard' (Witzel, 1999a, p. 44), which potentially hides the original form of the word. The Iranian forms are generally connected to Gr. *πάρδαλις* 'panther' (Beekes, 2010, p. 1152), although they seem to go back to **pərd-ank-*. The "suffix" **-ank-* is reminiscent but not identical to the Sanskrit desinence *-āku-*.

Additionally, Witzel (2003, p. 35) argues that Gr. *πάνθηρ* 'panther' ultimately has the same origin as Gr. *πάρδαλις* and the Indo-Iranian words. While alternations in voicing/aspiration of stops is common in early Greek loanwords, the alternation *r*/*n* is not. However, it is possible that earlier **πάρθηρ* >> Gr. *πάνθηρ* due to folk etymology (*πᾶν* 'all' *θηράω* 'to hunt', cf. Beekes, 2010, p. 1150), or by irregular dissimilation.

67. Pre-II **pinda-*

Ind. Skt. *píṇḍa-* 'lump'

Ir. Khot. *piṇḍaa*[47]

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Witzel (2003, p. 33). *EWAia* (II, p. 128) considers the possibility that Skt. *píṇḍa-* is a loanword. In any case, it seems likely that it existed in Iranian, since it was probably borrowed as Arm. *pind* 'firm, dense'.[48] Given the distribution of the word and the unexplained retroflex in Sanskrit, it is possible that *píṇḍa-* was borrowed into Sanskrit and then spread to Armenian via Khotanese, or that Indic and Iranian borrowed the word independently.

[47] Bailey (1979) does not include Khot. *piṇḍaa* in his dictionary.

[48] Martirosyan (2010, p. 552) considers Arm. *pind* to be IE, derived from PIE **b^{h}end^{h}-* 'to bind'. However, since the expected outcome of this root would have been ***bind*, the etymology requires the assumption of Grassmann's Law in Armenian, for which there are no further examples.

68. Pre-II *?

Ind. Skt. *śāli-* ‘unhusked rice’

Ir. Pto. *šole*, Orm. *šōl*, Par. *šēl*, YM *šālē*

Nur. Km. *šāl’i-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

EWAia (II, p. 632) presents no IE etymology. Due to the initial *š-*, Kümmel (2017, p. 283) assumes that the Iranian forms are borrowed from Nuristani or Sanskrit. Also, since *l* is secondary in Iranian languages, this seems likely. Neither Nur. *š-* corresponds regularly to Skt. *ś-*. The word was likely borrowed into Indic in Post-PII times and later diffused to other Indo-Iranian languages.

69. Pre-II *?

Ind. Skt. *śaṇá-* ‘hemp’

Ir. MiP *šan*, MoP *kanab*, Khot. *kaṃha* ‘hemp’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Witzel (2003, p. 34) and *EWAia* (II, p. 605). This old word for hemp is likely connected to Gr. *κάνναβις* ‘hemp’ (which was borrowed into many European languages), Sumerian *kunibu* ‘hemp’ (Beekes, 2010, p. 636) and Akk. *qunnabu* ‘an aromatic’ (*CAD* Q, p. 306). According to Bailey (1979, p. 52), Khot. *kaṃha* goes back to **kanfa-* < **kanaba-*, which would bring it very close to the Gr. and Near Eastern words. Somewhat confusing is the relationship between the *k*-initial words and Skt. *śaṇa-*, which from an IE perspective looks like a *centum-satəm* distribution. However, given the retroflex *-ṇ-*, *śaṇa-* looks like a more recent borrowing. Moreover, an alternation of *k*/*ś* is found elsewhere in Sanskrit loanwords (Witzel, 1999a, p. 34). In light of this, the words for ‘hemp’ most likely reflect Post-PII loanwords from different sources.

70. Pre-II *?

Ind. Skt. *sarṣapa-* ‘mustard’

Ir. Khot. *śśaśv*(*a*)*-āna-*, Sogd. *šywšp-δn*, MiP *span-dān*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a *Wanderwort* by Lubotsky (2001b) and *EWAia* (II, p. 712). Gr. *σίνᾱπι* ‘mustard’ probably belongs here as well. The Iranian words reflect a compound with *dāna-* ‘seed’, but the first part of the compound presents irregular correspondences. Khotanese and Sogdian show disyllabic words whereas MiP *span-* is monosyllabic. Khot. *śś-* and MiP *sp-* could be taken as regular reflexes of **ćṷ-*, but given its position in the word MiP *sp-* seems to correspond to Khot

-śv-, which is irregular. The extra initial syllable of Khot. *śśa-* and Sogd. *šy-*, which Persian lacks, is comparable to Skt. *sa-* and Gr. *σί-*. However, Skt. *sarṣapa-* is quite dissimilar from the Iranian forms. This suggests that the Indic and Iranian words were borrowed from different source languages.

71. Pre-II *?

Ind. Skt. *siṁhá-* 'lion'

Ir. Parth. *šarg*, Khot. *sarau* 'lion'

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Witzel (2003, p. 45). Perhaps belonging here is Arm. *inc/j* 'lion'. Semantically, it is not surprising that IE languages from the Eurasian steppe would borrow a word for 'lion'.

The words show variation between **-r-* and **-n-*. Moreover, the anlaut correspondence Skt. *si-* : Khot. *sa-* is irregular. Parth. *šarg* points to **čį-* or **kš-* (Kümmel, 2019). The *-h-* in Skt. *siṁhá-* reflects a primary or secondary palatal **ǰ^{h}/ǰ^{h}*, whereas Iranian points to **g$^{(h)}$*. Due to the variation, the words were clearly borrowed Post-PII.

72. Pre-II **šu̯aį̯pa-* 'tail'

Ind. Skt. *śépa-*, Pkt. *cheppā-* 'tail, penis'

Ir. LAv. *xšuuaēpā-* 'tail'

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (II, p. 654) gives no IE etymology, but rejects non-IE origin. Moreover, Pkt. *cheppā-* is explained as a generalized sandhi variant (°*c chépam*) of Skt. *śépa-*. However, since the anlaut of Pkt. *cheppā-* (**chepyā-*) corresponds to LAv. *xšuuaēpā-*, it need not be secondary. The correspondence is paralleled by Pkt. *cha* 'six' ~ Av. *xšuuaš* 'six' < **šu̯aćs* (Lubotsky, 2000). Skt. *śépa-* could reflect earlier **śvepa-* with dissimilation of *-v-* before a labial consonant, cf. Skt. *śiti-pád-* 'with white feet' < **śviti-* < **ḱu̯iti-*.

The origin of the anlaut cluster is complicated. Lubotsky (2000, p. 260) tentatively suggests PII **pću̯aį̯pa-* (< PIE **pḱu-* 'cattle' + **u̯eį̯p-* 'to swing'?), but this is unlikely since Avestan preserves initial **pć-* as *fš-*, cf. Av. *fšūmant-* 'having cattle' < **pću-mant-*. Furthermore, **pću̯aį̯pa-* could hardly yield Skt. **śvepa-* in view of *kṣumánt-* 'having cattle'.

It appears that two variants must be reconstructed: **šu̯aį̯pa-*, continued by Iranian and Prakrit, and **ću̯aį̯pa-* continued by Sanskrit. The anlaut of **šu̯aį̯pa-* cannot be secondary from earlier **su̯-*, because unlike **šu̯aćs* 'six', where original **s* >> **š* by assimilation to **ćs* (= [tš]),

**šu̯aḭpa-* does not provide any conditioning for a similar assimilation.[49] Rather, **šu̯aḭpa-* was borrowed with anlaut **š-*. In Sanskrit, the source word was adapted as **śvepa-*, indicating that **ć* had probably become a fricative in Indic at this point. The irregular correspondences indicate that the word was borrowed Post-PII.

73. Pre-II **(t)sūkV̄-* 'needle' >> Indo-Iranian **ćūkā-* / **sūčī-* / **ćuči* / **ćaučaṇ̣ia-*

Ind. Skt. *sūcī́-* 'needle'

Ir. LAv. *sūkā-*, Wakh. *sic* 'needle', MiP *sozan*, Oss. *sūʒīn*/*soʒīnæ*, Khot. *suṃjsañu* 'needle'

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b), due to the irregular correspondence Ind. **s* : Ir. *ć-*. *EWAia* (II, p. 739) instead assumes that earlier Skt. **śūcī́-* >> *sūcī́-* by analogy to *sīv-* 'to sew'. As this is rather *ad hoc*, one might entertain Lubotsky's (2001b, p. 306) hypothesis that Indic borrowed the predecessor of Skt. *sūcī́-* from an unknown source, which was then borrowed into Iranian. However, since PII **ć* remained an affricate in PIr., Skt. *sūcī́-* is an unlikely source of PIr. **ćūkā-* etc. Moreover, the fact that Sanskrit has *-c-*, never *-k-*, makes it improbable that LAv. *sūkā-* was borrowed from Indic.

While Wakh. *sic* < **ćuči* resembles Skt. *sūcī́-*, the MiP, Ossetic and Khotanese words reflect the rather divergent form **saučaṇ̣ia-* (Bailey, 1979, p. 427). This much variation points to a Post-PII borrowing.

However, as the palatalization of velars was a PII development, this requires the assumption of two source forms, one with palatalized **-c-* and one with **-k-*. Yet, the alternation of **-c-* and **-k-* seems to correlate with the quality of the following vowel (*-cī-* vs. *-kā-*, except in **ćaučaṇ̣ia-*), indicating regular palatalization of **kī* > **čī*. This correspondence gives the impression of an old borrowing, whereas the irregular correspondence Indic *s-* : Iranian *s-* points to the opposite. This paradox has no easy solution, but one may speculate that the variation *č*/*k* arose as the loanwords for needle were adapted to the native linguistic structure, where velars were normally palatal before **ī*.

[49] Cf. Skt. *svápna-* ~ Av. *xᵛafna-* < PII **su̯apna-*.

74. Pre-II **u̯īna-* ‘lute’

Ind. Skt. *vī́ṇā-*

Ir. Khot. *bīna*, Sogd. *wyn’*, MiP *win*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword into PII by Witzel (2003, p. 33). *EWAia* (II, p. 568) gives no etymology but states that the word may have diffused amongst the Indo-Iranian languages, although the direction of borrowing is not clear. The irregular retroflex *ṇ* in Sanskrit indicates that the word was not inherited from PII. It may have been borrowed into Sanskrit and then spread to Iranian and eventually also to Arm. *vin*, or into Indic and Iranian independently.

3.5. Words with IE etymologies or insufficient evidence for borrowing

75. PII **āćā-* / **aćas-*

Ind. Skt. *ā́śā-* f. ‘space’

Ir. LAv. *asah-* n. ‘region’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (I, p. 178) offers no IE etymology. The irregular correspondence Indic *ā* : Iranian *ă* could be taken as an argument for non-IE origin. However, morphologically, the alternation between Skt. *ā́śā-* (< **Hóḱ-eh₂-*?) and LAv. *asah-* (< **Heḱ-os*?) looks old.

76. PII **ćan-* ‘to ascend’

Ind. Skt. adv. *śanaiḥ* ‘gradually’

Ir. LAv. *san-*, Khot. *san-* / *sata-* ‘to rise’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (II, p. 608) and *LIV* (p. 324) assume an IE verbal root **ḱen-* also reflected in Arm. *snanim* ‘to be raised, grow’. Since both Iranian and Armenian have *nu*-presents, IE origin is likely.

77. PII **ćṷitra-* ‘white’

Ind. Skt. *sítsa-* (AV) ‘lead’

Ir. SW-Iranian **siça-* ‘white’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

EWAia (II, p. 734) argues that Skt. *sítsa-* is borrowed from an Iranian word **siça-* ‘white’, which cognate to Skt. *śvitrá-* ‘white’ < PIE **ḱṷit-*. Against this, Witzel (2003, p. 33) argues that a borrowing from SW-Iranian, does not fit with the fact that Skt. *sítsa-* is attested already in the

AV. Yet, this counterargument seems to be built on a misunderstanding, as Witzel (ibid.) notes that "[t]he Persians moved into the Persis and Anšan from NW Iran only after c. 700 BCE". However, the (pre-)historical location of the Persian speech community is not crucial for the hypothesis that Skt. *sḱsa-* was borrowed from a SW-Iranian dialect: the only relevant assumption is that the sound change PIr. **ćy̯* > **s* and **θr* > **ç* had already occurred in this language. Borrowing of an originally IE word from SW-Iranian thus remains a possible explanation of Skt. *sḱsa-*. In any case, Skt. *sḱsa-* must be a late borrowing, as it does not show the effect of the *RUKI*-rule.

78. PII **daćā-*

Ind. Skt. *daśā-* 'hem'

Ir. Khot. *dasa-*, Bal. *dasag-* 'thread'

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (I, p. 710) mentions possible connections to PGm. **tagla-* 'hair' and OIr. *dual* 'tuft, plait' < PIE **doḱ-lo-*, to which Matasović (2009, p. 102) adduces SCr. *dlàka* 'single hair' < PSl. **daḱlā-* (with depalatalization). Kroonen (2013, p. 504) points out that **tagla-* could be an inner-Germanic formation from a different root, however. PII **daćā-* (< PIE **deḱ-eh₂-*) is derived differently from the proposed cognates. If the original meaning was 'hair', an **-eh₂-*derivation could yield 'single hair' >> 'thread'. It is possible that the word is inherited.

79. PII **dʰu̯aǰ-* 'to flutter'

Ind. Skt. *dhvajá-* 'banner'

Ir. LAv. *dβōža-* 'to flutter', Sogd. *wy-δβys-* 'to bloom'

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (I, p. 800) derives it from the root PII **dʰeu̯H-* 'hin und herbewegen' with a suffix **-eg-*. However, **dʰu̯aǰ-* does not reflect a laryngeal, and the supposed suffix does not look IE. Since the original distribution of palatalized and plain velars in verbs is often distorted, the existence of **ǰ* before non-palatalizing vowels (e.g. Skt. *dhvajá-*) does not prove non-IE origin.

80. PII *g^h*as*- ‘to devour’

Ind. Skt. *ghas*-

Ir. LAv. *gah*-

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). No IE comparanda given by *EWAia* (I, p. 514). However, it may be connected to PGm. **gamman*- ‘stall, hut’ and Arm. *gom* ‘fold (for cattle)’ < *g^h*os-mo*- (Kroonen, 2013, p. 166), although the semantic connection is weak. IE origin cannot be excluded.

81. PII *g^h*au̯s*- ‘to make sound, hear’

Ind. Skt. *ghoṣ*-

Ir. Av. *gaoš*-

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (I, p. 518) mentions a possible connection to PGm. *gauma*- ‘heed, attention’ (< *g^h*ou-mo*-), in which case the PII verb would be an original *s*-present/desiderative of the same root. In this scenario, the PIE *s*-stem would originally have meant ‘to wish to be heard’, contrasting with the *s*-less stem ‘to be heard’.[50] Kroonen (2013, p. 171) proposes a different etymology for the Germanic words, which he connects to Skt. *gū́hati* ‘to hide’. This is formally possible, but semantically less attractive than the above. In any case, there is not enough evidence to postulate substrate origin.

82. PII **Hat*-

Ind. Skt. *at*- ‘to wander’

Ir. LAv. *xvāθra*- ‘well-being’ (< **su̯-at-ra*), *a-pairi.āθra*- ‘unavoidable’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (I, p. 56) suggests a connection to Lat. *annus* ‘year’ (< **atno*- ‘which goes’), reconstructing PIE *h_2*et*- ‘to wander’ (cf. *LIV*, p. 273). Although this may not be correct, there is no further indication that PII **Hat*- is a loanword.

[50] A similar situation would be reflected in Skt. *śru*- ‘to hear’ but *śroṣa*- ‘to be obedient’ (lit. ‘to wish to hear’)

83. PII **H̯uap-* ‘to strew, scatter’

Ind. Skt. *vap-*

Ir. OAv. (*vī-*)*uuāpaṯ*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (II, p. 503) gives no convincing IE cognates. Kloekhorst (2008, pp. 430-1) argues that Hitt. *ḫuu̯app-*^{i} / *ḫupp-* ‘to throw, be hostile towards’ is related to the Indo-Iranian forms and reconstructs PIE **h₂uóph₁-ei*, with a final laryngeal to explain the Hittite geminate *-pp-* (since *-p-* would otherwise have been lenited). The initial laryngeal reflex in Hittite fits with the lengthened *vī-* in OAv. A root final laryngeal, however, is not compatible with the Indo-Iranian evidence, since that should have given Skt. **vaph-*, Av. **vaf-*. A solution is to reconstruct PIE **h₂uep-*, and assume that the geminate *-pp-* in Hittite was levelled from the 3pl.pret. *ḫuppēr*, where *-p-* would have escaped lenition since it was preceded by a short vowel. In conclusion, this verb is likely of IE origin.

84. PII **H̯uap-* ‘to shave, shear’

Ind. Skt. *vap-*

Ir. Khot. *patävutta-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (II, p. 504) assumes that the verb is etymologically identical to **H̯uap-* ‘to strew, scatter’. As a semantic development from ‘to scatter (hair) > ‘to shear’ seems plausible, the verb may be treated as IE.

85. PII **H̯u̯id^{h}H-* ‘to split in two’

Ind. Skt. *vyadh-* ‘to wound, hurt’

Ir. LAv. *vīδ-* ‘to pierce’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested as a possible loanword by Lubotsky (2001b). Conversely, *EWAia* (II, p. 591) and *LIV* (p. 294) reconstruct PIE verbal root **h₂u̯ied^{h}-* based on Indo-Iranian. An initial **h₂-* is postulated based on Gr. *ἠίθεος* / *ἄθεος* ‘unmarried youth’ (Tichy, 1993, p. 15), but since *ἄθεος* is likely a hyperdorism, Greek could also reflect **h₁-* (Beekes, 1992, p. 172). Lubotsky (1994, p. 204) explains Gr. *ἠίθεος* and Skt. *vidhávā-* ’widow’ from PIE **du̯i-d^{h}h₁-u-* ‘widow(er)’ (lit. ‘bereft of its half’, with Kortlandt effect **d* > **h₁*), and connects this compound to Skt. *vidh-* ‘to allot, apportion’.

In my opinion, Skt. *vyadh-* belongs to the same etymon[51] as a full grade of the secondary root PII **H̯id^{h}H-* ‘to split in two’. In PII, this root underwent semantic change from ‘to split in two’ >> ‘to divide, allot’ on the one hand, and >> ‘to pierce’ on the other, and finally, within Sanskrit >> ‘to wound’. This analysis is supported by the fact that a) Skt. *vidh-* ‘to allot, apportion’ never takes the full grade (*EWAia* II, p. 555) and b) the Iranian evidence. In LAv., we find a root *vīδ-* with the present stem *viθiia-* ‘to pierce’ (Kellens, 1995, p. 55). The same root likely underlies *vaēδa-* ‘Wurfgeschoss, Name einer bestimmten Angriffswaffe’ (*AiWB*, p. 1320) and *a-šəmnō-vīd-* ‘das Ziel nicht erreichend, verfehlend’ (*AiWB*, p. 257). It is also found in MiP *wistan* ‘to shoot, throw’, Pto. *wīštəl* ‘to shoot, hit’, Šu. *wēδ-d* ‘to throw’ (*EWAia* II, p. 592). The Iranian forms show a different full grade (~ **H̯uaid^{h}H-*) than Indic (~ **H̯id^{h}H-*), indicating the secondary nature of this formation. In sum, I regard this verb as IE in origin.

86. PII **i̯ātu-* ‘black magic’

Ind. Skt. *yātú-*

Ir. LAv. *yātu-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Witzel (2003, p. 38). *EWAia* (II, p. 411) suggests connections to either Skt. ^{1}*yā-* ‘to travel’, ^{2}*yā-* ‘to request’ or ^{3}*yā-* ‘to attack’, the first two of which have IE cognates. Lubotsky (1988, p. 47) holds ^{3}*yā-* as the most likely base of **i̯ātu-* on semantic grounds. Perhaps ^{3}*yā-* is etymologically identical to ^{2}*yā-*, if a semantic development ‘to pursue’ >> ‘to attack’ is assumed. In any case, IE origin seems likely.

87. PII **ǰ^{h}ai-* ‘to incite’

Ind. Skt. *hi-*

Ir. LAv. *frazaiiaiiāmi*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be loanword by Lubotsky (2001b). *EWAia* (II, p. 802) suggests connection to PGm. **gaiza-* ‘spear, tip’ and OIr. *gae* ‘spear’, but these reflect a root **ǵ^{h}eis-* (Kroonen, 2013, p. 164), which might be comparable to Skt. *héṣas-* ‘weapon’ and *heṣ-* ‘to damage’, but not to Skt. *hi-*. Even if isolated, the *nu*-present derivation in Sanskrit suggests IE origin.

[51] Cf., less explicitly, Melchert (1977, p. 113).

88. PII **ǰʰas-* ‘to laugh’

Ind. Skt. *has-*

Ir. LAv. *jahī, jahikā-* ‘prostitute’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). No IE comparanda given by *EWAia* (II, p. 811). Given the formal similarity to PII **gʰas-* ‘to devour’, one might consider an etymological connection between the two. Starting from **gʰes-* with an original meaning ‘to open the jaws’, it is possible that the palatalized variant **ǰʰas-* < **gʰes-* was lexicalized as ‘to laugh’, whereas the non-palatalized **gʰas-* < **gʰ(o)s-* was lexicalized as ‘to devour’. Whether they are related or not, there is not enough evidence to postulate substrate origin.

89. PII **kuč-* ‘to crook, bend’

Ind. Skt. *kuc-*

Ir. MiP *n-gwč-*, Khot. *us-kuj-* ‘to rise up’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (I, p. 361) regards the proposed connection to OIr. *cúar* ‘bent’ uncertain. However, Matasović (2009, p. 228) has shown that MIr. *cúar* ‘curved’ may derive from **kukro-* or **koukro-* (with regular loss of **k* before **r*). This has further cognates in BSL, cf. Lith. *kaũkas* ‘lump’ and PSl. **kùka-*[52] ‘hook’ (Derksen, 2008, p. 256). In view of the formal and semantic correspondences, it is in my opinion likely that PII **kuč-* is IE.

90. PII **magʰa-* ‘gift, offering, sacrifice’

Ind. Skt. *maghá-*

Ir. Av. *maga-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b) due to its connection to ritual practices. *EWAia* (II, p. 289) assumes IE origin in connection with Goth. *magan* ‘to be able’ and OCS *mogǫ* ‘id.’. While the semantic match is not perfect, it is not farfetched enough to exclude IE origin.

[52] The Slavic acute is analogical to **kl̥ùka-* (Derksen, 2008, p. 256).

91. PII **marǰʰa-* ‘udder’

Ind. Skt. *malhá-* ‘with hanging belly/udder’

Ir. LAv. *mərəzāna-* ‘belly’, gen.sg. *maršuiiå̊-* ‘paunch’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

According to Lubotsky (2001b, p. 312), the proposed connection to Lith. *mìlžtis* ‘to swell up’ (*EWAia* II, p. 334) is impossible since the Baltic acute reflects PIE **ǵ*, whereas PII **ǰʰ* must go back to **ǵʰ*. However, since PII **ǰʰ* could also go back to PIE **ǵH* with PII deglottalization, the etymology is possible if PIE **(h₂)melǵH-* ‘to swell’ is reconstructed.

LAv. *mərəzāna-* seems to reflect **mḷ̊ǵH-ono-*. The gen.sg. *maršuiiå̊-* looks related, but *š* < **ǰ* only regularly occurs before **n*.

Lubotsky (2001b, p. 312) furthermore suggests a connection to Skt. *bárjaha-* ‘udder’, also entertained in *EWAia* (II, p. 211, 334). Skt. *bárjaha-* cannot reflect the same PII form as Skt. *malhá-*, since *m* > *b* only occurs before consonantal *r* (cf. Skt. *bravīti* < **mreu̯H-ti*). Neither is there an analogical model for initial *b-*. The only way to connect the words would be to assume parallel borrowings **malǰʰ-* ~ **barǰ(ʰ)-* from a non-IE source.

92. PII **monH-i-*

Ind. Skt. *maṇí-* ‘necklace’

Ir. LAv. (*zarənu-*)*maini-* ‘with golden neck-jewel’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Witzel (2003, p. 33). Due to the irregular retroflex in Sanskrit, the word cannot readily be reconstructed for PII. *EWAia* (II, p. 293), however, argues for IE origin in **monh₂-i-*, further reflected in PGm. **manja-* ‘necklace’ and derived from **mon-eh₂-* ‘neck’, reflected in PGm. **manō-*. Note that the Indo-Iranian forms could derive from **mn̥h₂-i-* or **monh₂-i-*. Although the retroflex in Sanskrit remains unexplained, IE origin cannot be excluded.

93. PII **nard-* ‘to hum, complain’

Ind. Skt. *nṛ̥d-*

Ir. Sogd. *nrδ-*, MiP *nāl*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (II, p. 22) proposes no IE etymology but suggests that the word could be onomatopoeic. This is difficult to disprove, but neither a particularly compelling hypothesis.

94. PII **raǰʰ-*

Ind. Skt. *rah-* ‘to be abandoned’

Ir. MiP *rāz* ‘mystery’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (II, p. 442) concludes that there is no IE etymology.

95. PII **sagʰ-*

Ind. Skt. *sagh-* ‘to be able to bear’

Ir. LAv. *azgatō* ‘unbearable’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (II, p. 686) assumes origin in PIE **segʷʰ-* on the basis of Gr. *σθένος* ‘strength’. However, due to the unclear suffix **-eno-*, *σθένος* may not be IE (Beekes, 2010, p. 1326).

Alternatively, **sagh-* could derive from the same PIE root as Skt. *sáhate* ‘conquers’, *sáhas-* ‘victory’, Av. *hazah-* ‘victory’ < PIE **segǵʰ-* ‘to hold’, cf. Gr. *ἔχω* ‘to have, hold’. In the zero-grade **sgǵʰ-*, **s* would depalatalize **ǵʰ* > **gʰ*. From the zero-grade, a secondary root **segʰ-* was formed, continued in PII **sagh-*. The lexical split would only have occurred in *satəm* languages, where **ǵʰ* and **gʰ* were distinct. This analysis allows LAv. *azgatō* to be connected to Gr. *ἄσχετος* ‘irresistible’ (*AiWB*, p. 228).[53]

96. PII **srans-*

Ind. Skt. *sraṃs-* ‘to fall apart’

Ir. OAv. *rå̊ŋhaiiən* ‘they make fall away’

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be loanword by Lubotsky (2001b). *EWAia* (II, p. 783) offers no IE etymology.

97. PII **stHūna-* ‘pillar’

Ind. Skt. *sthūṇā-*, *sthūnā-*

Ir. LAv. *stunā-*, OP *stūnā-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Witzel (2003). The irregular Sanskrit retroflex *-ṇ-* is unexplained, but may be secondary given the non-retroflex variant form. *EWAia* (II, p. 768), argues that the word

[53] However, since **n-sgʰ-eto-* should have yielded **n-zǰʱ-eto-* > **a-zǰʰata-*, the velar of LAv. *azgatō* must be secondary.

is related to Skt. *sthūrá-* 'big, strong' and Gr. *στῦλος* 'column, pillar', all from PIE **sth₂u-* (itself from **steh₂-* 'to stand' with a *u*-extension). The long *ū* in Skt. *sthūrá-* and *sthū́ṇā-* presupposes a proto-form **stuH-*, presumably from metathesis of **stHu-*, in which case the aspirate *-th-* must be analogical from some other form of the same root, e.g. **stHeu-*. IE origin is possible.

98. PII **sṷag-* 'to embrace'

Ind. Skt. *svaj-*

Ir. LAv. *pairiš.xᵛaxta*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (II, p. 788) and *LIV* (p. 610) reconstruct PIE **sṷeng-* (**n* reflected in Skt. *pári-ṣvañjalya-*), a root also reflected in MHG *swanc* 'movable' and OIr. *seng* 'skinny'.

99. PII **u̯i̯ak-* 'to encompass'

Ind. Skt. *vyac-*

Ir. MoP *gunǰidan*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (II, p. 590) argues for an IE origin in relation to Lat. *vinciō* 'to bind'. *LIV* (p. 696) reconstructs **u̯i̯ekʷ-* 'to encompass' and adds Gr. (Thess.) *ἴμψας* 'having yoked' as evidence for the labiovelar. However, there is reason to doubt that *ἴμψας* belongs here. Firstly, the *-s-* in the Greek word is also found in *ἴμψιος* '(of the) yoke' (Beekes, 2010, p. 591), implying that it is part of the root. Secondly, in this scenario, the *-m-* would be a relic of the nasal present, which should be absent in the aorist. Furthermore, Lat. *vinciō* 'to bind' could very well be related to Lat. *vincō* 'to conquer' (Vaan, 2008, p. 679), which does not reflect a labiovelar but goes back to PIE **u̯ei̯k-* (*LIV*, p. 670). PII **u̯i̯ak-*, cannot derive from PIE **u̯ei̯k-* since the root structures differ.

Despite the lack of convincing cognates outside of Indo-Iranian, the fact that PII **u̯i̯ak-* formed a nasal present indicates IE origin.

100. PII **u̯i̯atH-*

Ind. Skt. *vyath-* 'to be unsteady'

Ir. OAv. *a-iβiθura-* 'unshakeable'

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (II, p. 591) takes it as a secondary root from **ui-* + *-eth₂*. This analysis is problematic since there is no semantically fitting root

*teH- to explain the second part of the root (cf. *LIV*, p. 616: *teh_2- ‘to steal’, *teh_2- ‘to thaw’. However, as PII *u̯i̯atH- may be cognate to PGm. *witt/dōn- ‘to tremble’, it could be IE.

101. PII *u̯ik- ‘to separate’

Ind. Skt. *vic-*

Ir. LAv. *vic-*, MiP *wēxtan*/*wēz-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (II, p. 577) proposes an etymological connection to Hitt. *ḫuek-^{zi}* ‘to slaughter’, which is formally impossible (Kloekhorst, 2008, p. 407). However, other possible cognates are Lat. *victima* ‘sacrificial animal’, lit. ‘the separated one’ (Vaan, 2008, p. 675), and PGm. *wīha-* ‘holy’. For this reason, IE origin cannot be excluded.

102. PII *u̯r̥ćsa- ‘tree’

Ind. Skt. *vr̥kṣá-*

Ir. LAv. *varəša-*

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Taken as a loanword by Lubotsky (2001b). *EWAia* (II, p. 572) hesitantly supports a connection to Skt. *válśa-* ‘twig’, LAv. *varəsa-* ‘hair’, OCS *vlasъ* ‘hair’ < PIE *uolḱ-o-. This explanation is possible under the assumption that PII *u̯r̥ćsa- derives from an *s*-stem *uelḱ-es- ‘twig’, from which a thematic possessive derivative *ulḱ-s-ó- ‘having twigs’ >> ‘tree’ was formed. While this scenario requires the assumption of an unattested s-stem, the connection between Skt. *vr̥kṣá-* ‘tree’ and *válśa-* ‘twig’ is semantically likely, and the postulated derivational chain explains their semantic relationship in a morphologically convincing way.[54] As such, PII *u̯r̥ćsa- ‘tree’ could be of IE origin.

103. PII *u̯riH- ‘to oppress, collapse’

Ind. Skt. *vlī-*

Ir. LAv. *uruuīnaitīš* (acc.pl.)

Lim. distribution	Irr. correspondences	Rem. morphology	Rem. phonology	Specific semantics

Suggested to be a loanword by Lubotsky (2001b). *EWAia* (II, p. 598) derives it from an extended root *u̯R-eiH- for unclear reasons. Although the verb is isolated to Indo-Iranian, archaic-looking derivations like Skt. *vlināti* < *uli-ne-H-ti* means that IE origin cannot be excluded.

[54] Cf. Skt. *vatsá-* ‘yearling’ < *uet-s-ó-.

4. Chronological layers in early Indo-Iranian loanwords

In chapter 3, early Indo-Iranian loanwords were classified into chronological layers. In this chapter, each layer is discussed as a whole. The goal is to summarize the findings from chapter 3 and to discuss potential implications of the chronological layers for the Central Asian Substrate Hypothesis.

4.1. Layer I: Pre-PII or early PII

Layer I consists of loanwords borrowed into Pre-PII or into early PII before the operation of certain PII sound changes. A single word (**ućig-*) was argued to belong to this layer.

The descendants of PII **ućig-* 'sacrificing priest' show a regular alternation between velar *-k-* / *-g-* and palatalized **-ǰ-* in the paradigm. The most likely reason for this alternation is that **ućig-* underwent palatalization before the PII vowel merger. However, it cannot be excluded that it was secondarily adapted by analogy to inherited stems with a similar alternation, as these were common in the morphological structure of early Indo-Iranian.

Thus, the evidence for Pre-PII borrowings is scarce or non-existent. However, it must be kept in mind that any loanword in layer 0 may have been borrowed at the Pre-PII or early PII stage as well. Accordingly, the scarcity of words in Layer 1 does not in itself prove that Indo-Iranian borrowed less during the Pre-PII or early PII period.

4.2. Layer 0: PII (unspecified)

Layer 0 refers to 45 loanwords that are reconstructable for PII, but for which there is no further indication of time of borrowing. Theoretically, these words may have been borrowed at any time after the disintegration of PIE up until the split of PII. However, certain words in layer 0 have features that suggest a more precise classification.

Four words[55] in layer 0 contain voiceless aspirates (**p^h*, **t^h*, **k^h*). I have chosen to consistently analyze these as clusters of voiceless stops + laryngeal (**pH*, **tH*, **kH*), since phonemic voiceless aspirates are an Indic innovation. However, given that **H* was phonetically a glottal stop,[56] it seems likely that **PH*-clusters in loanwords represent adaptations of monophonemic stops or fricatives in the source language(s), rather than clusters. Stops in tautosyllabic clusters with glottals automatically adopt the laryngeal feature of the glottal, since this becomes a feature of the cluster as a whole (Kehrein, 2002). Since **H* was in the process of being lost as a segmental phoneme in late PII (having been vocalized, lost intervocalically

[55] **atHarvan-*, **kapHa-*, **kHā-*, **kHara-*.
[56] Or, according to Kümmel (2018), a glottal fricative [h].

etc.), the sequence **PH* was probably phonetically equivalent to **P^h*. Therefore, it is likely, but not provable, that loanwords in layer 0 with voiceless aspirates were borrowed in late PII.

Another four words[57] in layer 0 contain long vocalic semivowels **ī* and **ū*. In inherited words, these long vowels arose regularly from the sequences **iH* and **uH*. This development is shared by Indic and Iranian but not Nuristani (cf. chapter 2). Although it is impossible to prove, it seems unlikely that loanwords containing **ī* or **ū* were borrowed as **iH* and **uH*. These sequences are connected to the ablauting morphophonology characteristic of Indo-European, and none of these loanwords show ablaut. It is more likely that loanwords containing **ī* or **ū* were borrowed at a time when **iH* and **uH* had already become long vowels, i.e. late PII. However, this should be regarded as a tentative conclusion.

Lastly, the semantics of some words provide additional indications of their time of borrowing. PII **Hustra-* 'camel' and **kHara-* 'donkey' both denote animals that are absent from the Indo-European homeland and rather associated with Central Asian cultures (Witzel, 2000, p. 4). PII **parsa-* 'sheaf', **spāra-* 'ploughshare' and **u̯rīj́^{h}i-* 'rice' all relate to agriculture, which was not practiced in the Sintashta culture (Judd et al., 2018) or on the Eurasian steppe until after 2000 BCE (Anthony, 2007). Therefore, it is unlikely that any of these words were borrowed into Pre-PII, before Indo-Iranian speakers had migrated to Central Asia.

Thus, out of 45 loanwords in layer 0, 13 show indications of being borrowed towards the end of Indo-Iranian linguistic unity.

4.3. Layer II: late PII

Layer II consists of 7 loanwords borrowed into late PII, i.e. after certain PII sound changes had already occurred. Below, the rationale for including words in layer II based on the relative chronology of PII sound changes will be discussed.

PII **pusća-* 'tail' and **ǰaj́^{h}a/ukắ-* 'hedgehog' are included in layer II, since they contain palatalized velars before non-palatalizing vowels. Therefore, they are unlikely to have been borrowed before the PII palatalization of velars. In the case of **ǰaj́^{h}a/ukắ-*, the morphological variation among the attested Indo-Iranian forms also suggests a late time of borrowing. The non-palatalized **ka-* in PII **karus-* also indicates borrowing after the palatalization of velars, but could also be because it was borrowed after the phonologization of **a*. In any case, it cannot be Pre-PII.

[57] **bīǰa-*, **kšīra-*, **u̯āćī-*, **u̯rīj́^{h}i-*. In the case of **u̯āćī-*, the long **ī* is synchronically a suffix, and could theoretically be secondary due to adaptation to the native morphology.

PII **čāt(u̯āla)-* ‘pit, well’ is included in layer II due to the morphological variation in the attested forms, which suggests that the Indo-Iranian speaker communities were disintegrating at the time of borrowing. It is thus likely that the word was borrowed after the palatalization of velars, i.e. initial **č-* was borrowed as such and not palatalized within PII.

The inclusion of **maį̯ūkHa-* ‘peg’, **pīį̯ūša-* ‘beestings’ and **į̯au̯īį̯ā-* ‘canal’ in layer II is based on indirect evidence. The long vowels **ī* and **ū* go back to **iH* and **uH*. Thus, if e.g. **pīį̯ūša-* goes back to Pre-PII, it must be reconstructed as **piHį̯uHsa-*. There is no direct counterevidence against reconstructing **piHį̯uHsa-* etc., but the fact that the attested words seem to belong to the *CVCV̄CV* type makes it highly likely that they were borrowed with a medial long vowel. Since the sound change **i/uH > *ī/ū / _C* is the latest change shared by Indic and Iranian, it is likely that **maį̯ūkHa-*, **pīį̯ūša-* and **į̯au̯īį̯ā-* were borrowed quite late. Strictly speaking, this is circular reasoning, since the analysis of **pīį̯ūša-* as a *CVCV̄CV* word and the rejection of the reconstruction **piHį̯uHsa-* are logically co-dependent. However, the *CVCV̄CV* type is so pervasive in Indo-Iranian loanwords that I consider it very likely that loanwords attested with this structure belong to the same group.

4.4. Layer III: Post-PII borrowings

Layer III consists of 21 loanwords that are attested in both Indic and Iranian, but cannot be reconstructed for PII due to irregular sound correspondences that cannot be explained by secondary processes. However, the formal and semantic similarity of 19 of these loanwords puts it beyond reasonable doubt that they share a common origin.

In chapter 3, I argued that these 19 loanwords have been borrowed along four different paths: 1) source language >> Indic *and* Iranian, independently, 2) source language >> Iranian >> Indic, 3) source language >> Indic >> Iranian, and 4) source language *A* >> Indic, source language *B* >> Iranian.

Group 1 (parallel borrowings) consists of 13 words: **banγa-* ‘hemp, narcotic’, **dŭ̆rća-* ‘(goat’s) wool, hair’, **ganDarų̯a-* ‘a mythical being’, **ganTi-* ‘smell’, **ganTuma-* ‘wheat’, **Kaį̯ća-* ‘hair’, **mVša-* ‘bean’, **naį̯Tˢ(a)-* ‘skewer’, **pinda-* ‘lump’, **pərd-a(n)k-* ‘panther’, **šų̯aį̯pa-* ‘tail’, **(t)sūkV̄-* ‘needle’, and **ų̯īna-* ‘lute’. For justification of each case, I refer to chapter 3.

As an example of group 1, consider Skt. *kéśa-* ‘head hair’ (**kaį̯ća-*) next to LAv. *gaēsa-* ‘curly hair’ (< **gaį̯ća-*). The irregular initial stop correspondence can hardly be explained by assuming that Sanskrit or Avestan borrowed the word from the other: Indic **kaį̯ća-* would most likely have been adopted as Iranian ***kaį̯ća-* since voiceless /k/ exists in the Iranian phoneme

inventory. Similarly, Iranian **gaⱬća-* would most likely have been adopted as Indic **gaⱬća-*. The most plausible scenario is therefore parallel borrowing from a Pre-II source that I write as **Kaⱬća-*. This "reconstruction" is not necessarily accurate, but merely an approximation of the source form. Since the Sanskrit and Avestan words are so similar, it is likely that they were borrowed only shortly after the split of PII, during a period of dialectal differentiation.

A single word constitutes group 2: **ćikatā-* 'sand, gravel' may have been borrowed into Iranian and then spread to Indic. However, parallel borrowing is also possible.

Similarly, a single word constitutes group 3: Skt. *śāli-* 'unhusked rice' likely spread from Indic into Iranian and Nuristani.

Finally, four words constitute group 4: *khaḍgá-*/**karkadan-* 'rhinoceros', *sarṣapa-* (etc.) 'mustard', *śaṇa-*/*kanaba-* 'hemp', and *siṁhá-*/*sarau* 'lion'. All words refer to animals and plants common in South Asia and the Middle East. They may ultimately derive from a common source,[58] but the Indic and Iranian reflexes are so dissimilar that their immediate source languages must have been different. This does not exclude the possibility of a Central Asian Substrate origin, but it implies that at least one of the Indo-Iranian variants was borrowed from another language.

The two remaining words in layer III are *(*H*)*arǰaná-*/*áṇu-* 'millet' and *masū́ra-*/**mižuka-* 'lentil'. In these cases, the proposed Indic and Iranian cognates are too phonologically divergent to plausibly share a common origin. However, the structure of Skt. *masū́ra-* indicates that it belongs to the group of Indic *CVCV̄CV* words.

4.5. Implications of chronological analysis

The division of early Indo-Iranian loanwords into chronological layers has consequences for the analysis of the loanword corpus and the Central Asian Substrate Hypothesis.

As stated above, the small size of layer I does not in itself prove that most early loanwords entered Indo-Iranian at a later stage, since words from layer 0 could in theory be very old. However, considering the 7 words of layer II, 21 of layer III, and the fact that 13 words in layer 0 show features that suggest a late time of borrowing, the general trend is clear: most loanwords were not borrowed into Pre-PII, but into the later stages of PII and Post-PII. This result supports the Central Asian Substrate Hypothesis, since the time of borrowing coincides with the period when PII is believed to have been spoken in the Sintastha and Andronovo cultures.

[58] Either because the source languages were distantly related or because the source languages, in turn, had borrowed the word from a common source.

Another consequence is that loanwords in different layers cannot *a priori* be considered to originate in the same language(s). This is especially true for Post-PII loanwords (layer III) vs. PII loanwords (layers I, II and 0). Instead, they will (at least initially) be treated as originating in different languages. Crucially, however, this does not necessarily imply "different languages" in the phylogenetic sense of mutually unintelligible linguistic communities. Rather, "different languages" may represent different chronological stages of the same language, i.e. "Old Pre-II", "Middle Pre-II" etc.

Loanwords belonging to the same layer may in theory originate in different languages or be separated by hundreds of years of linguistic development. In other words, a layer may be more diverse, in terms of absolute chronology, than is discernible by the available historical linguistic methodology. However, with the current methodology, it is these layers that must be the basic units of analysis.

The next step in our analysis is to determine whether the two situations (borrowing from phylogenetically different languages vs. borrowing from different chronological stages of the same language) can be differentiated based on the data itself. As discussed in chapter 1, Lubotsky argued, based on structural similarities between PII loanwords and Indic loanwords, that "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" (2001b, p. 306). In other words, he found similar structural characteristics in separate chronological layers, which demonstrate a link between the source languages of both layers. Now, structural characteristics of the layers proposed in this study can be analyzed in a similar fashion.

5. Structural characteristics of Indo-Iranian loanwords

This chapter discusses the phonological and morphological structure of Indo-Iranian loanwords, including characteristic word structures, recurring sequences of phonemes, and recurring irregular correspondences. The goal is to show how the structure of loanwords differs from inherited words and to determine what the structure of loanwords reveals regarding the structure of the substrate language(s). In accordance with the results of chapter 4, the chronological layers 0-III will be kept apart in the analysis, in order to determine whether different layers represent borrowings from the same language(s) or different language(s).

5.1. The *CVCV̄CV*-type

The recurrence of non-IE trisyllabic words with a medial long vowel or diphthong in Sanskrit was described by Kuiper (1991). The existence of *CVCV̄CV* words in PII was demonstrated by Lubotsky (2001b, p. 306), who put forward the hypothesis that the source of these words in PII and Sanskrit was the same language, or at least related languages. Therefore, the treatment of *CVCV̄CV* words is crucial for understanding the Pre-II linguistic landscape of Central and South Asia.

Table 3. CVCV̄CV words in early Indo-Iranian loanwords

	CVCV̄CV
Layer I (Pre-/early PII)	
Layer 0 (PII)	**kapāra-* **kapaṷta-* **ṷarāj^{h}a-*
Layer II (late PII)	**i̯aṷīi̯ā-* **mai̯ūkHa-* **pīi̯ūša-* **čāt(ṷāla)-*
Layer III (Post-PII)	
Total: 7	

Evidently, four *CVCV̄CV* words belong to the late PII layer, although **čāt(ṷāla)-* is uncertain since the trisyllabic structure is only attested in Indic. The three remaining *CVCV̄CV* words belong to layer 0, since it could not be demonstrated that they have or have not undergone certain PII sound changes. However, due to their structural similarity, it is probable that all PII *CVCV̄CV* words were borrowed from the same language. The fact that other *CVCV̄CV* words

were borrowed in late PII indicates that **kapāra-*, **kapau̯ta-* and **u̯arāǰʰa-* too were borrowed towards the end of Indo-Iranian linguistic unity. This would also explain why the initial **k-* of **kapāra-* and **kapau̯ta-* is not palatalized, without requiring the reconstructions Pre-PII **kN̥pāra-* vs. **kN̥pau̯ta-*.

While 6 *CVCV̄CV* words are securely reconstructable for PII, Kuiper's (1991, pp. 90-93) list of non-IE words in Vedic includes 63 *CVCV̄CV* words.

In the PII group, all *CVCV̄CV* words are thematic *a*-stems. Conversely, in the Indic group, 23 of 63 words belong to other stem types, e.g. 9 *u*-stems.[59] Witzel (2003, p. 33) considers all thematic loanwords as original consonant stems (e.g. **kapau̯t-*) that were thematicized within PII. If this is true, PII *CVCV̄CV* words, which are all thematic, could derive from disyllabic words. On the other hand, it is unlikely that the many athematic stems among the Indic *CVCV̄CV* words are secondary. Therefore, under Witzel's analysis, the link between PII and Indic *CVCV̄CV* words is challenged. However, since there are consonant stems among the PII loanwords (**ućig-*, **bʰišaǰ-*), there is no reason to assume that all thematic loanwords were athematic in the source language(s) and subsequently thematicized within PII. Thus, the link between PII and Indic *CVCV̄CV* words should be maintained.

Yet, the morphological difference between the groups could originate in a morphological difference in the source languages. It is therefore more likely that *CVCV̄CV* words in PII and Indic originate in slightly different, though related, languages, than in the same language.

5.2. *r*/*n*-alternation

Witzel (2003, p. 45) proposed that the Central Asian Substrate had a dialectal variation of *r*/*n*, reflected in Indo-Iranian loanwords as well as other words in languages of Asia Minor, the Middle East and the Caucasus. The evidence consists of the following words:

Table 4. Evidence for r/n-alternation in loanwords

meaning	***r*-variants**	***n*-variants**
leopard, panther	Skt. *pŕ̥dāku-*, PIr. *pard-*, Gr. *πάρδαλις*	Gr. *πάνθηρ*
lion	Parth. *šarg*, Khot. *sarau*	Skt. *siṁhá-*, Arm. *inc/j*
wheat	Bur. *gur*, Basque *gari*	LAv. *gaṇtuma-*
water, river	Bur. *hur*, Macro-Caucasian **(t)sir-*	Skt. *sindhu-*, LAv. *həṇdu-*
mustard	Skt. *sarṣapa-*	MiP *span-dān*, Gr. *σίνᾱπι*

[59] Namely: *ikṣvākú-*, *jábāru-*, *jarā́yu-*, *kiyā́mbu-*, *kúṇāru-*, *pŕ̥dāku-*, *urvārú-*, *viṣṇāpú-*, to which *palāṇḍu-* may be added.

First, that there would be any etymological relation between Burušaski *hur* and PII **sind^{h}u-* is by no means certain. As shown in chapter 3, it is impossible to ascertain that the *r* of Bur. *gur* did not develop secondarily from **γund-* or **γud-*. Likewise, the *n* of Gr. *πάνθηρ* could be secondary.

Secondly, one may question whether "*r*/*n*-alternation" is a proper way to describe the variation. In 'wheat' and 'water, river', *-r-* corresponds to the cluster *-nd$^{(h)}$-*. In 'leopard, panther', one the other hand, both variants show a cluster, *-rd-* vs. *-nd^{h}-*. The word for 'mustard' has *-rs-* in Indic but *-n-* in Iranian, although Khot. and Sogdian have neither *-n-* nor *-r*(*s*)*-*. Only in the word for 'lion', does *-r-* in Iranian seem to correspond to *-n-* in Indic and Armenian, but here the root vowels are also different. Thus, the evidence for *r*/*n*-alternation is more heterogeneous than has previously been acknowledged.

Of the Indo-Iranian evidence, only Skt. *sindhu-*, Av. *həṇdu-* can be reconstructed for PII, whereas the rest show irregular correspondences pointing to Post-PII borrowing. Indic and Iranian have the same *r*/*n*-variant for 'leopard, panther', 'wheat', 'water, river', but different variants for 'lion' and 'mustard'. As for the geographical distribution, the *r*-variants are found in both Indic and Iranian, as well as Greek, Burušaski and Caucasian languages. The *n*-variants are absent from Burušaski and Caucasian languages, but seen in Indic, Iranian, Greek and Armenian, the latter being geographically close to Caucasian languages. If the *r*/*n*-alternation originates in dialectal variation, we would expect a clearer geographical distribution.

Thus, the words used as evidence show more variable elements than *r* vs. *n*, i.e. the *r*/*n*-alternation is not their lowest common denominator. As such, it is methodologically hazardous to use the concept of *r*/*n*-alternation to equate words that in reality are very different. Secondly, their geographical and chronological distribution offers no reason to assume that the variation would originate specifically in a Central Asian language.

5.3. The irregular correspondence Indic dh : Iranian t

A recurring irregular correspondence in Post-PII loanwords can be observed based on the dental stops in Skt. *godhū́ma-* 'wheat' ~ LAv. *gaṇtuma-*, Khot. *ganama*, Bal. *gandūm* 'wheat' and Skt. *gandhá-* 'smell' ~ LAv. *gaiṇti-* 'bad smell', Khot. *ggañu* 'stench'. The Sanskrit, Khotanese, and Balochi words point to PII **d$^{(h)}$*, whereas LAv. points to **t*. Although Witzel (2003, p. 31) reconstructs a single PII form of 'wheat', I argue that the irregular variation cannot be explained by secondary developments (cf. chapter 3).

It is interesting to note that the first syllable of **ganTuma-* and **ganTi-* is identical.[60] This similarity led Witzel (2003, p. 31) to assume a folk-etymological relationship between the words, viz. Skt. *godhū́ma-* << **gandha-dhū́ma-* 'perfume smell'. However, other than their formal similarity, there is no indication that such a folk-etymological influence would have taken place, since Skt. *godhū́ma-* synchronically looks like a compound of *go-* 'cow' and *dhū́ma-* 'smoke'. A folk-etymological reanalysis of **ganTuma-* based on **ganTi-* and **umā-* 'flax' is not unthinkable for Iranian, but in that case the expected outcome would be ***ganT̰uma-*. Finally, if the similarity is due to folk etymology, original **t* must have been generalized in Avestan, but original **d*^{h} in Indic, which is unnecessarily complicated.[61] It seems preferable to project the reason behind the phonological similarity of the words to their source language.

The recurring irregular correspondence *dh* : *t* allows for some interesting observations. First, the fact that the same irregular correspondence is found twice makes it highly likely that these words were borrowed from the same language.

Second, in the source language, the words contained a sound which was adapted as **d*^{h} in Sanskrit and some Iranian languages, but **t* in LAv. Based on the reconstructed phonology of Indic and Iranian, it is difficult to find a plausible explanation for this irregularity. Both Proto-Indic and Proto-Iranian have voiceless and voiced stops, implying that if the source had [t] or [d], they should have been adapted as such. The source words might have contained a breathy stop, close to Indic *dh*, but that would most likely have been adapted as Iranian *d*, not *t* as in LAv. Thus, it is possible that the source language had a different sound altogether, or, more precisely, a sound, alien to the synchronic Indic and Iranian phonologies, that was interpreted as *t* by some Iranian speakers but *dh* (which may still have been [d] at this point) by Indic speakers. Alternatively, one may assume that Indic and Iranian borrowed at different points in time, and that the source language underwent a change from **t* > **d* (vel sim.) in the meantime, but this is unlikely, given the relatively short time span from the disintegration of PII until the attestation of the separate branches.

5.4. Non-initial mediae in clusters with **r or *n*

A previously unnoticed feature of Indo-Iranian loanwords is the tendency for non-initial[62] mediae to co-occur with *n*, *r* (sometime both) or *r̥*. Here is the evidence:

[60] A similar anlaut is found in Skt. *gandharvá-* ~ LAv. *gaṇdərəβa-*, although here Indic and Iranian both point to voiced **d*$^{(h)}$.
[61] If, for example, originally **ganti-* vs. **ganduma-*, then the influence must have gone in opposite directions in Sanskrit and Avestan.
[62] Initial mediae occur in: **dŭ́rća-*/**dr̥ća*, **gadā-*, **bīja-*, **gr̥da-*, LAv. *gaēsa-*, *gaṇtuma-*, *gaiṇti-*.

Table 5. Non-initial mediae in early Indo-Iranian loanwords

	Elsewhere	***r*-clusters**	***n*-clusters**
Layer I (Pre-/early PII)	**ućig-*		
Layer 0 (PII)	**bīǰa-* **bhiš-aǰ-* **gadā-* **sćāga-*	**gr̥da-* **mr̥ga-* **kadru-*	**u̯and(H)-* **ringa-* **indra-* **nagna-*
Layer II (late PII)			
Layer III (Post-PII)		**(H)arǰaná* *khaḍgá-/*karkadan* **pərd-a(n)k-*	**banγa-* **ganTuma-* **pinda-*
Total: 18	5	6	7

Table 5 shows that most (13/18) word-internal mediae co-occur with **n*, **r* or **r̥*. Since the words for 'rhinoceros' may be borrowed from different sources, they are best left out of this discussion. Likewise, **(H)arǰaná-* has no Indic equivalent.

In most cases, the media occupies the coda position of the cluster, but in **kadru-* and **nagna-*, the media occupies the onset. The mediae are mostly dental or velar, rarely palatal, and never bilabial.

In layer 0, 7/11 words follow the pattern. Of the 4 words with mediae that do not, two (**bīǰa-*, **bhiš-aǰ-*) contain palatals. By contrast, mediae in clusters with **r* and **n* are never palatals in PII loanwords. This potentially indicates that palatals stops, i.e. affricates, adhered to different phonotactic rules than stops in the source language(s). PII **gadā-* may in principle go back to **gn̥dā-*, in which case it would fit into the pattern.

The high frequency of mediae in clusters with **n*, **r* or **r̥* is contrasted by the low frequency of other stops in these positions in PII loanwords. Three cases are attested: **ańću-*, **Hustra-* and **u̯r̥tka-*. Since palatal stops do not seem to be part of the pattern, **ańću-* may be disregarded. PII **Hustra-* has a variant **Hustar-*, implying that the **r* may originally not have been part of the cluster. The **t* in **u̯r̥tka-* could in principle have been a media **d* originally, since it would have been devoiced by sandhi anyway.

In Indo-Iranian inherited vocabulary, all three stop series occur in clusters with **n*, **r* or **r̥*, cf. Skt. *vártate* ~ LAv. *varətata* 'to turn', Skt. *pardate* ~ LAv. *pərədən* 'to fart', Skt. *spárdhate* 'to contest', Skt. *pánthā-* ~ Av. *paṇtā* 'way', Skt. *skándati* 'jumps', Skt. *bandhaya-*

~ LAv. *baṇdaiie-* ‘to bind’, etc. It is thus likely that the co-occurrence of dental and velar mediae with **n*, **r* or **r̥* in loanwords reflects a feature of the source language(s). In other words, it is likely that the source language(s) of these PII borrowings only allowed one type of stop in clusters with **n*, **r* or **r̥*, which was nativized as PII mediae.

Depending on the phonetic interpretation of PII mediae, the above feature appears more or less salient. If the mediae were pre-glottalized, the correlation with **n* and **r* is quite salient. On the other hand, if the mediae were plain voiced stops at the time of borrowing, the co-occurrence with nasals is quite trivial, since voicing of stops after nasals is very common cross-linguistically (Kümmel, 2007, p. 53). By implication, the feature would be less likely to reflect the phonological system of a single substrate language. However, the same does not apply to the co-occurrence of mediae and **r*.

In layer III, **pərd-a*(*n*)*k-*, **pinda-* and **banγα-* follow the pattern. The Iranian reflexes of **ganTi-* and **ganTuma-* show either voiceless or voiced stops after **n* (cf. chapter 3). The Indic equivalents have voiced aspirates. One possibility is that the variation is caused by the changes in the stop systems of Iranian and Indic. Another is that **ganTi-* and **ganTuma-* were borrowed from a different language than the loanwords with mediae in clusters with **n*, **r* or **r̥*. A third possibility is that the correlation of mediae with **n* and **r* only holds for the PII layers. With so few examples in Post-PII, the correct scenario cannot be determined with any degree of certainty.

5.5. Correlation between **i* and affricates

Lubotsky (2001b, p. 304) observed the high frequency of palatal stops and clusters containing **s* in Indo-Iranian loanwords. Another tendency is the correlation between **i* and affricates. The evidence is presented below:

*Table 6. Loanwords containing *i, as well as loanwords containing palatals*

	Palatal + *i	**Other obstruents + *i**	***i elsewhere**	**Other palatals**
Layer I (Pre-/early PII)	**ućig-*			
Layer 0 (PII)	**u̯rīj^{h}i-* **kaćia̯pa-* **rāći-* **u̯āćī-* **ćyā-*	**b^{h}iš-aj-* **kućsi-* **išt(i)-* **kšīra-* **matsia̯-* **r̥si-* **bīja-*	**j^{h}armiia̯-* **āni-* **ringa-* **indra-*	**anću-* **bīja-* **ćaru̯a-* **kāća-* **b^{h}iš-aj-* **j^{h}armiia̯-* **sćāga-* **u̯arāj^{h}a-*
Layer II (late PII)		**pīia̯ūša-*	**ia̯u̯īia̯ā-* **maia̯ūkHa-*	**pusća-*
Layer III (Post-PII)	**ćika(tā)-*	**(t)sūkV̄-* (**sūčī-*) **ganTi-* **pinda-*	**mižuka-* **naia̯T^{s}(a)-* *śāli-* (Skt.) **šu̯aia̯pa-* **u̯īna-*	**dŭrća-* **Kaia̯ća-* *śaṇá-/kanaba-*
Total: 41	7	11	11	12

Focusing on the PII words (layer I, 0, II) where **i* follows an obstruent (n=13), we see that 5 co-occur with a primary palatal. Of the 8 remaining cases, three (**kućsi-*, **kšīra-*, **matsia̯-*) show clusters with **s*, another (**r̥si-*) a simple **s*. Note that the cluster of **kućsi-* could go back to **tć*. Three cases (**pīia̯ūša-*, **b^{h}iš-aj-*, **bīja-*) show labial stops preceding **i*. Only **išt*(*i*)- has a dental stop before **i*, but the original derivational suffix is not certain for this word (cf. chapter 3).

Thus, non-labial obstruents before **i* are predominantly palatal affricates or clusters with **-s*, which are phonetically close to affricates. There is no inner-Indo-Iranian explanation for this phenomenon, since the PII palatalization before **i* only affects velars and does not produce primary palatals, except in the case of PIE **ske*/*i* > PII **ść*. Given this distribution, a reasonable hypothesis is that non-labial stops were affricated before **i* in the source language(s). The original stops may have been dental or velar.

That said, not all primary palatals in PII loanwords can be analyzed in this way: 9 loanwords contain palatals that are not followed by **i*. Cases like **sćāga-*, **ǰʰarmi̯a-*, and **ćaru̯a-* could reflect palatals before a high vowel Pre-PII **e*, which could be assumed to have caused palatalization in the source language(s), but this is mere speculation. Moreover, the palatal in **ańću-* appears in a distinctly non-palatal context. Thus, we must assume that the source language(s) of these words also possessed phonemic palatals (that were not conditioned by a following **i*), or at least, a phoneme that was nativized as PII palatals.

In Post-PII loanwords, **ćika*(*tā*)- seems to follow the above pattern. In the case of **sūčī-*, the **č* may be secondary. On the other hand, **ganTi-* contradicts the distribution. The evidence is too scarce to allow for a clear analysis.

5.5.1. Dental stops

Above, the high frequency of palatal stops and clusters with **s* before **i* in PII loanwords was explained by postulating a process of affrication of dental or velar stops in the source language(s). If this hypothesis is correct, we would expect dental and velar stops in loanwords to occur in non-palatalizing context. The evidence for dental stops is given below:

Table 7. Loanwords containing dental stops

	Before *u	**Before *r**	**Before thematic vowel or *-ā-**	**Elsewhere**
Layer I (Pre-/early PII)				
Layer 0 (PII)	**stuka-*	**Hustra-* **kadru-* **indra-*	**gadā-* **gr̥da-* **kapau̯ta-* **pau̯asta-*	**atHarvan-* **atka-* **matsi̯a-* **išt*(*i*)- **u̯and*(*H*)- **u̯r̥tka-*
Layer II (Late PII)	**čāt*(*u̯āla*)-			
Layer III (Post-PII)	**dŭrća-* **ganTuma-*		**ćikatā-* **pinda-* **pərd-a*(*n*)*k-*	**ganTi-* **ganDəru̯a-*
Total: 22	4	3	7	8

In the PII layers (0, II), dental stops almost exclusively occur in non-palatalizing contexts. This is most clearly seen in the 5 cases were a dental stop precedes **u* or **r*. In 4 cases, a dental stop precedes the suffixes *-a-* (< **-o-*) or *-ā-* (< **-eh₂-*), which were both non-palatalizing in PII. Of course, the suffix vowels may have been added after the words were borrowed, but in any case, there is no indication that the stops were followed by an **i* or another palatalizing vowel in the source language. Of the 6 remaining cases, **atHarvan-*, **atka-* and **u̯r̥tka-* show **t* in clusters with **H* and **k*, which may be treated as non-palatalizing contexts. Conversely, **matsi̯a-* belongs to the affricated group described in the previous section. For **išt*(*i*)*-*, the original derivation may not have been an *i*-stem. In the case of **u̯and*(*H*)*-*, the dental occurs in different contexts depending on the derivation.

The fact that dental stops occur in non-palatalizing contexts provides indirect support for the hypothesis that PII palatals and clusters before **i* reflect affricated stops in the source language(s) of PII loanwords. The source language(s) seems to show a complementary distribution of **t, *d* / _**u, *r, *ă* and **ć, *j*ʰ*, *tć, *ts* / _**i*, which most likely reflects a historical process.

In Post-PII loanwords, there is one case (**ganTi-*) of a dental stop in a palatalizing context. This suggests that the above analysis only holds for the PII layers, or that **ganTi-* was borrowed from a different source language.

5.5.2. Velar stops

As stated above, affricates before **i* could in principle also reflect original velar stops, in which case we would expect velar stops in to occur in non-palatalizing contexts. The evidence is given below:

Table 8. Loanwords containing velar stops

	Before consonant	**Before thematic vowel**	**Before other vowels**	**Word-final position**
Layer I (Pre-/early PII)				**ućig-*
Layer 0 (PII)	**kHā-* **kHara-* **kšīra-* **nagna-*	**aka-* **atka-* **mr̥ga-* **muska-* **ringa-* **sćāga-* **stuka-* **u̯r̥tka-*	**gadā-* **gr̥da-* **kaći̯apa-* **kāća-* **kadru-* **kapāra-* **kapau̯ta-* **kapHa-* **kućsi-*	
Layer II (Late PII)	**mai̯ūkHa-*	**ǰajʰa/ukă̄-*	**kárus-* **čāt(u̯āla)-*	
Layer III (Post-PII)		**banγα-* *khaḍgá-/*karkadan siṁhá-* (Skt.)	**ćikatā-* **ganDəru̯a-* **ganTi-* **ganTuma-* **Kai̯ća-* **kanaba-* **(t)sūkV̄-*	**pərd-a(n)k-*
Total: 36	5	12	18	2

Pre-vocalic velar stops in PII and Post-PII loanwords mostly occur before **ă̄*, **u* or **r̥*, which are non-palatalizing contexts. Here, PII **ă̄* cannot go back to a high vowel Pre-PII **ĕ̄*, since the velar would have been palatalized within PII. Some velars occur before **H* or **n*, which are also non-palatalizing contexts. PII **kšīra-* is part of the affricated group described above.

There are five cases of palatalized velars: three in PII (**ǰaǰʰa/ukă̆-*, **čāt(u̯āla)-*, **ućig-*) and two in Post-PII (**(t)sūkV̄-*, Skt. *sim̐há-*). Only **ućig-* (with palatalized **ǰ* in the gen.sg.) is likely to have undergone PII palatalization. The palatalized velar of Skt. *sūcī́-* (<< **(t)sūkV̄-*) may have arisen secondary within Indo-Iranian. In the remaining cases, the palatalized velar could reflect the original form of the source language.

As with the dentals, the absence of velar stops before **i* is compatible with the hypothesis that PII palatals and clusters before **i* reflect affricated stops in the source language(s). However, the existence of palatalized velars in PII loanwords could indicate that affricated velars in the source language(s) were adapted as PII **Č*. In that case, dental stops are the most likely origin of affricates + **i*. Another indication of this is that **matsi̯a-* and **kućsi-* (if < **kutći-*) contain dental clusters.

5.6. The sequence **-ru̯-*

The recurring sequence **-ru̯-* in PII and Sanskrit loanwords was observed by Lubotsky (2001b, p. 304). Among early Indo-Iranian loanwords, the evidence consists of **atHaru̯an-*, **ćaru̯a-*, and **ganDəru̯a-*, to which I have added **bʰaru̯-*. Although **-ru̯-* is not absent in the inherited vocabulary, the sounds are generally separated by a morpheme boundary, e.g. Skt. *sárva-* ‘all’ ~ Av. *hauruua-* ‘whole’ < PIE **solh₂-u̯o-*, which is probably a thematicized *u*-stem (Pronk, 2011, p. 189). The loanwords with **-ru̯-* could in principle contain a morpheme boundary as well, but there is no indication that that is the case.

The word **ganDəru̯a-* is Post-PII, whereas the three remaining words are PII. Although the number of words is quite small, it is noteworthy that the **-ru̯-* cluster is found in both chronological layers.

6. Conclusions

In this chapter, the results of the thesis will be summarized and discussed.

6.1. Summary of main results

The 103 words of possible non-IE origin that have been analyzed in this study may be divided into two groups. 29 words do not fulfill the necessary criteria to make substrate origin likely. In the majority of cases, this is due to the existence of possible or plausible IE comparanda. For 8 words, however, there is no IE etymology, but no other criteria make a non-IE origin likely.[63] Although borrowing is in principle as likely as inheritance in such cases, they were left out of the loanword corpus to avoid interference with the results.

The remaining 74 words can be considered as loanwords according to the applied methodology. Besides lacking IE etymologies, these words show structural peculiarities that separate them from the inherited lexicon and/or specific semantics that make them particularly liable to borrowing.

It has furthermore been demonstrated that the 74 early Indo-Iranian loanwords cannot be ascribed to the same chronological layer in the history of Indo-Iranian. The majority, 53 words, are reconstructable to PII (layers 0, I, II). Within this group, only one word (**ućig-*) shows evidence of an early Pre-PII time of borrowing, although this is not absolutely certain. On the other hand, 7 words (layer II) were borrowed in late PII, after the operation of various sound changes, such as the phonologization of **a*, the palatalization of velars, and the lengthening of short vowels preceding laryngeals (**V̌H > *V̄*). The remaining 45 words (layer 0) could theoretically have been borrowed at any point during Pre-PII or PII, but 13 words have features that indicate a late PII time of borrowing.

21 loanwords showing irregular correspondences were classified as Post-PII. 19 of these loanwords are most likely related, and 13 of those are so similar that they reflect parallel borrowings by Indic and Iranian from the same source language. The two remaining words are too dissimilar to have any etymological relation.

The analysis of structural characteristics of early Indo-Iranian loanwords generally supports the conclusions of previous literature. However, the evidence for Witzel's *r*/*n*-alternation is very scarce. Even if it is accepted, the wide geographical distribution of the *r*/*n*-variants do not support the idea that the words originated in the Central Asian Substrate.

[63] These are **āćā-*/**ać̣as-*, **dʰu̯aǰ-*, **ǰʰai-*, **nard-*, **raǰʰ-*, **srans-*, **u̯riH-*, **u̯i̯ak-*.

Two new structural characteristics were proposed in chapter 5. In PII loanwords, velar and dental stops are absent before **i*. In contrast to this, non-labial obstruents before **i* are palatals (phonetically affricates) or clusters with **s* (that are phonetically close to affricates). Thus, obstruents seem to be complementarily distributed depending on the following vowel. I argued that this reflects a feature of the source language, probably affrication of dental stops before **i*.

Another characteristic of PII loanwords is that non-initial dental and velar mediae co-occur with **n* and **r*. Conversely, tenues and aspiratae almost never appear in this position. Since Indo-Iranian allows all series of stops in clusters with **n* or **r*, I argued that this reflects a phonotactic feature of the source language of the loanwords.

Neither of the new structural characteristics seem to hold for the Post-PII layer. Interestingly, for both characteristics, the counterexamples are the words **ganTi-* 'smell' and **ganTuma-* 'wheat', which consequently appear increasingly isolated from the rest of the loanword corpus. In fact, other Post-PII loanwords are generally in line with the PII pattern. The **-rṷ-* cluster of Post-PII **ganDərṷa-* also shows a link between the layers. It may thus be best to view **ganTi-* and **ganTuma-* as outliers, perhaps originating in a different language than the rest.

Together with the patterns proposed in previous literature, the evidence for affricates + **i* and **n*/*r* + mediae lend additional evidence to the hypothesis that most early Indo-Iranian loanwords originate in the same unknown substrate language.

6.2. Identity of source languages

A minority of early Indo-Iranian loanwords have been proposed to originate in known languages.

The possibility of a Uralic origin has been discussed for six words:

1) Skt. *bhaṅgá-* 'hemp' << PU **pe̮ŋka-* 'mushroom'
2) **indra-* << PU **ilmar* / **inmar* 'thunder god'
3) **matsi̯a-* 'fish' << PU **maća* 'fish net'
4) **u̯āćī-* 'axe' << PFU *wäŋći* 'knife'
5) **kapHa-* 'phlegm' << PU **kompa* 'wave'
6) **pusća-* 'tail' << PFU **ponci* 'tail'.

The first word is not reconstructable to PII, and since it refers to a domesticated plant, a BMAC origin seems more likely. Moreover, the cluster *-ng-* has been shown to be characteristic of PII loanwords of unknown origin. Words 2, 3, and 4 are not reconstructable to Proto-Uralic,

which is the stage at which contact between Uralic and PII otherwise seems to have taken place. In 5, the semantics of the proposed source does not match the Indo-Iranian word. While semantic change may be considered, a Uralic origin of **kapHa-* remains speculative. The 6th case is semantically plausible, but the correspondence Uralic **n* : PII **Ø* presents a formal problem. Interestingly, words 4 and 5 also show the correspondence Uralic **n* : PII **Ø*, but since they present other problems, it is unlikely that this reflects a regular adaptation strategy of Uralic loanwords into Indo-Iranian.

Ultimately, it seems unlikely that Uralic was a major donor language of early Indo-Iranian loanwords.

A Near Eastern source has been proposed for **kHara-* 'donkey', **ganTuma-* 'wheat', **ganDərṷa-* 'a mythical being', and **kanaba-* 'hemp', based on similar words in languages of the Near East and, to some extent, Greek. While the ultimate source could be a known language of the Near East (e.g. Sumerian for **kanaba-*), a direct source of borrowing cannot be determined for any of these words. Therefore, an intermediary language located in Central Asia, transmitting Near Eastern (agricultural) vocabulary, remains equally likely.

Thus, to account for the 74 early Indo-Iranian loanwords treated here, it remains necessary to assume unknown donor language(s).

6.3. Implications for the Central Asian Substrate Hypothesis

The Central Asian Substrate Hypothesis places the unknown donor language of early Indo-Iranian loanwords in the BMAC culture. As we have seen, a core argument of the hypothesis is the semantics of certain PII loanwords, which can be connected to the material culture of the BMAC. However, many loanwords cannot be linked to material culture. Using the structural characteristics advanced in this study, words *without* reference to material culture can be connected to words *with* reference to material culture. For example, since **kaćiapa-* 'tortoise' and **ṷāćī-* 'axe, knife' can be connected to the BMAC (Lubotsky, 2001b, p. 307), it becomes increasingly likely that other loanwords with affricate + **i*, like **ćyā-* 'to freeze, congeal', **ućig-* 'sacrificing priest', and **matsia-* 'fish', also originate in the Central Asian Substrate. In this way, the study offers new ways to bridge the gap between linguistics and archaeology in support of the Central Asian Substrate Hypothesis. However, a complicating factor in this particular case is that **ṷrīj^{h}i-* 'rice', also with affricate + **i*, cannot easily be connected with the BMAC culture, since rice was not cultivated here.

Another contribution of the present study to the Central Asian Substrate Hypothesis is that the time of borrowing of many loanwords has been shown to be late PII or shortly Post-

PII, rather than Pre-PII. This supports the hypothesis, since the language contact between Indo-Iranian and the Central Asian Substrate is believed to have occurred after the founding of the Sintashta culture, where PII was probably spoken, at a time when Indo-Iranian speakers, identified with the Andronovo cultures, spread over a larger area in Central Asia.

6.4. Directions for future research

Several questions remain open for future research. The division of loanwords into chronological layers could be supplemented by a detailed integration of archaeological data, to investigate whether chronological layers of linguistic development can be connected to archaeological layers of cultural development. On the linguistic side, one could investigate to what extent the newly proposed structural characteristics of loanwords hold for the corpus of Indic loanwords proposed by Kuiper (1991). Lastly, future research would benefit from incorporating potentially crucial evidence from Middle and Modern Iranian, Indic and Nuristani languages to a greater extent.

Bibliographical abbreviations

AiGr. = Debrunner, A. & Wackernagel, J. (1954). *Altindische Grammatik Band II, 2: Die Nominalsuffixe*. Göttingen: Vandenhoeck & Ruprecht.

AiWB = Bartholomae, C. (1904). *Altiranisches Wörterbuch*. Berlin: Walter De Gruyter & Co.

CAD = Gelb, I. J., et al. (1956-2010). *The Assyrian Dictionary of the Oriental Institute of the University of Chicago. 21 vols.* Chicago: The Oriental Institute at the University of Chicago.

EWAia = Mayrhofer, M. (1992-2001). *Etymologisches Wörterbuch des Altindoarischen I-III.* Heidelberg: Universitätsverlag Carl Winter.

LIV = Rix, H. & Kümmel, M. (2001). *Lexikon der indogermanischen Verben.* Wiesbaden: Dr. Ludwig Reichert Verlag.

UEW = Rédei, K. (1988). *Uralisches Etymologisches Wörterbuch*. Budapest: Akadémiai Kiadó

References

Adams, D. Q. (2013). *A Dictionary of Tocharian B.* Amsterdam/New York: Rodopi.

Aikio, A. (2015). The Finnic 'Secondary e-stems' and Proto-Uralic Vocalism. *Suomalais-Ugrilaisen Seuran Aikakauskirja 95*, 25-66.

Anthony, D. W. (2007). *The Horse, the Wheel, and Language: How Bronze Age Riders from the Eurasian Steppe Shaped the Modern World.* Princeton/Oxford: Princeton University Press.

Anthony, D. W. & Ringe, D. (2015). The Indo-European Homeland from Linguistic and Archaeological Perspectives. *Annual Review of Linguistics 1(1)*, 199-219.

Bailey, H. W. (1979). *Dictionary of Khotan Saka.* Cambridge: Cambridge University Press.

Beekes, R. S. (1981). The Neuter Plural and the Vocalization of the Laryngeals in Avestan. *Indo-Iranian Journal 23*, 275-287.

Beekes, R. S. P. (1992). Widow. *Historische Sprachforschung 105 (2)*, 171-188.

Beekes, R. S. P. (1996). Ancient European Loanwords. *Historische Sprachforschung 109 (2)*, 215-236.

Beekes, R. S. P. (2010). *Etymological Dictionary of Greek.* Leiden: Brill.

Beekes, R. S. P. (2011). *Comparative Indo-European Linguistics: An Introduction (2nd ed.).* Amsterdam: John Benjamins Publishing Company.

Berger, H. (1959). Deutung einiger alter Stammesnamen der Bhil aus der vorarischen Mythologie des Epos und der Purana. *Wiener Zeitschrift für die Kunde Süd- und Ostasiens und Archiv für indische Philosophie III*, 34-82.

Berger, H. (1970). Die Burušaski-Lehnwörter in der Zigeunersprache. *Indo-Iranian Journal 3*, 17-43.

Blažek, V. & Hegedűs, I. (2012). On the position of Nuristani within Indo-Iranian. In R. Sukač & O. Šefčík, *The Sound of Indo-European 2* (pp. 40-66). München: Lincom.

Burrow, T. (1957). Sanskrit "*gṝ-/gur-*" 'to Welcome'. *Bulletin of the School of Oriental and African Studies 20(1),* 133-144.

Burrow, T. (1973). *The Sanskrit Language.* London: Faber and Faber.

Byrd, A. M. (2015). *The Indo-European Syllable.* Leiden: Brill.

Cantera, A. (2001). Die Behandlung der idg. Lautfolge (*C*)*RHC*- im Iranischen. *Münchener Studien zur Sprachwissenschaft 61*, 7-27.

Cantera, A. (2017). The Phonology of Iranian. In J. Klein, B. Joseph & M. Fritz, *Handbook of Comparative and Historical Indo-European Linguistics* (pp. 481-503). Berlin: De Gruyter Mouton.

Cheung, J. (2007). *Etymological Dictionary of the Iranian Verb.* Leiden: Brill.

Clayton, J. (2018). *Rounding of Indo-Iranian *R(H).* Retrieved from UCLA: Program in Indo-European Studies: https://pies.ucla.edu/IEConference/IEChandouts/clayton_j_2018h.pdf

Damgaard et al. (2018). The First Horse Herders and the Impact of Early Bronze Age Steppe Expansions into Asia. *Science 360 (6396).*

Davary, G. D. (1982). *Baktrisch: ein Wörterbuch auf Grund der Inschriften, Handschriften, Münzen und Siegelsteine.* Heidelberg: Groos.

Décsy, G. (1990). *The Uralic Proto-Language: A Comprehensive Reconstruction.* Bloomington: Eurolingua.

Derksen, R. (2008). *Etymological Dictionary of the Slavic Inherited Lexicon.* Leiden: Brill.

Eilers, W. (1959). Akkad. kapsum "Silber, Geld" und Sinnverwandtes. *Die Welt des Orients 2*, 322-337 + 465-469.

Francfort, H.-P. (2005). La civilisation de l'Oxus et les Indo-Iraniens et Indo-Aryens en Asie Centrale. In G. Fussman, J. Kellens, H.-P. Francfort, & X. Tremblay, *Aryas, Aryens et Iraniens en Asie Centrale* (pp. 253-328). Paris: Collège de France.

Frog. (2012). Confluence, Continuity and Change in the Evolution of Mythology: The Case of the Finno-Karelian Sampo-Cycle. In A.-L. Siikala & E. Stepanova, *Mythic Discourses:*

Studies in Uralic Traditions. Studia Fennica Folkloristica 20. (pp. 205-254). Helinki: Finnish Literature Society.

Gershevitch, I. (1961). *A Grammar of Manichean Sogdian.* London: Basil Blackwell.

Gharib, B. (1995). *Sogdian Dictionary.* Tehran: Farhangan Publications.

Haak et al. (2015). Massive Migrations from the Steppe were a Source for Indo-European Languages in Europe. *Nature 522*, 207-220.

Hegedűs, I. (2012). The RUKI-rule in Nuristani. In B. Nielsen Whitehead, T. Olander & B. A. Olsen, *The Sound of Indo-European 1* (pp. 145-167). Copenhagen: Museum Tusculanum Press.

Heide, M. (2011). The Domestication of the Camel. *Ugarit-Forschungen 42*, 331-382.

Henning, W. B. (1939). Sogdian Loan-Words in New Persian. *Bulletin of the School of Oriental Studies 10*, 93-106.

Hock, H. H. (1991). *Principles of Historical Linguistics.* Berlin: Mouton de Gruyter.

Hoffmann, K. & Forssman, B. (1996). *Avestische Laut- und Flexionslehre.* Innsbruck: Institut der Sprachwissenschaft der Universität Innsbruck.

Jamison, S. W. & Brereton, J. B. (2014). *The Rigveda: The Earliest Religious Poetry of India.* Oxford: Oxford University Press.

Judd et al. (2018). Live in the Fast Lane: Settled Pastoralism in the Central Eurasian Steppe during the Middle Bronze Age. *American Journal of Human Biology 30*, 1-23.

Kehrein, W. (2002). *Phonological Representation and Phonological Phasing.* Tübingen: Niemeyer.

Kellens, J. (1995). *Liste du verbe avestique.* Wiesbaden: Dr. Ludwig Reichert Verlag.

Kent, R. G. (1950). *Old Persian: Grammar, Texts, Lexicon.* New Eaven: American Oriental Society.

Kim, R. (2000). Reexamining the History of Tocharian B ‘ewe’. *Tocharian and Indo-European Studies 9*, 37-43.

Kloekhorst, A. (2008). *The Hittite Inherited Lexicon.* Leiden: Brill .

Kloekhorst, A. (2011). Weise’s Law: Depalatalization of Palatovelars before **r* in Sanskrit. *Indogermanistik und Linguistik im Dialog: Akten der XIII. Fachtagung der Indogermanischen Gesellschaft vom 21. bis 27. September 2008 in Salzburg* (pp. 261-270). Wiesbaden: Reichert Verlag.

Kloekhorst, A. (2014). *Accent in Hittite: A Study in Plene Spelling, Consonant Gradation, Clitics, and Metrics.* Wiesbaden: Harrassowitz.

Kloekhorst, A. (2016). The Anatolian Stop System and the Indo-Hittite Hypothesis. *Indogermanische Forschungen 121*, 213-247.

Kloekhorst, A. (2018). Anatolian Evidence Suggests that the Indo-European Laryngeals *h_2 and *h_3 were Uvular Stops. *Indo-European Linguistics 6*, 69-94.

Kobayashi, M. (2004). *Historical Phonology of Old Indo-Aryan Consonants.* Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Kobayashi, M. (2017). The phonology of Indic. In J. Klein, B. Joseph & M. Fritz, *Handbook of Comparative and Historical Indo-European Linguistics* (pp. 325-344). Berlin: De Gruyter Mouton.

Koivulehto, J. (2001). The Earliest Contacts between Indo-European and Uralic Speakers in the Light of Lexical Loans. In C. Carpelan, A. Parpola & P. Koskikallio, *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations* (pp. 235-263). Helsinki: Suomalais-Ugrilainen Seura.

Kortlandt, F. H. H. (1983). Greek Numerals and PIE Glottalic Consonants. *Münchener Studien zur Sprachwissenschaft 42*, 97-104.

Kortlandt, F. H. H. (1985). Long Vowels in Balto-Slavic. *Baltistica 21(2)*, 112-124.

Kortlandt, F. H. H. (1985). Proto-Indo-European Glottalic Stops: the Comparative Evidence. *Folia Linguistica Historica 6(2)*, 183-201.

Kortlandt, F. H. H. (1986). Posttonic *w in Old Irish. *Éiru 37*, 89-92.

Kortlandt, F. H. H. (1996). The High German Consonant Shift. *Amsterdamer Beiträge zur älteren Germanistik 46*, 53-57.

Kortlandt, F. H. H. (2018). Proto-Indo-European Glottalic Stops: The Evidence Revisited. *Münchener Studien zur Sprachwissenschaft 71(1)*, 147-153.

Kroonen, G. (2013). *Etymological Dictionary of Proto-Germanic.* Leiden : Brill.

Kroonen, G., Barjamovic, G. & Peyrot, M. (2018). Linguistic Supplement to Damgaard et al. 2018: Early Indo-European Languages, Anatolian, Tocharian, Indo-Iranian. *https://doi.org/10.5281/zenodo.1240524*.

Kuiper, F. B. J. (1948). *Proto-Munda Words in Sanskrit.* Amsterdam: Noord-Hollandsche Uitgevers Maatschapij.

Kuiper, F. B. J. (1976). Old East Iranian Dialects. *Indo-Iranian Journal 18*, 241-253.

Kuiper, F. B. J. (1991). *Aryans in the Rigveda.* Amsterdam: Rodopi.

Kuiper, F. B. J. (1995). Gothic bagms and Old Icelandic ylgr. *NOWELE 25*, 63-88.

Kuryłowicz, J. (1935). *Études Indoeuropéennes I.* Krakow: Polska Akademija Umiejetnosci.

Kuz'mina, E. E. (2007). *The Origins of the Indo-Iranians.* Leiden: Brill.

Kuznetsov, P. F. (2006). The Emergence of Bronze Age Chariots in Eastern Europe. *Antiquity 80,* 638-45.

Kümmel, M. J. (2005). Vedisch *tand-* und ein neues indoiranisches Lautgesetz. In G. Schweiger, *Indogermanica: Festschrift Gert Klingenschmitt* (pp. 321-332). Taimering: Schweiger VWT-Verlag.

Kümmel, M. J. (2007). *Konsonantelwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion.* Wiesbaden: Dr. Ludwig Reichert Verlag.

Kümmel, M. J. (2012). Typology and Reconstruction. In B. Nielsen Whitehead, T. Olander & B. A. Olsen, *The Sound of Indo-European: Phonetics, Phonemics, and Morphophonemics* (pp. 291-329). Copenhagen: Museum Tusculanum Press.

Kümmel, M. J. (2016a). Is ancient old and modern new? Fallacies of attestation and reconstruction (with special focus on Indo-Iranian). In D. M. Goldstein, S. W. Jamison & B. Vine, *Proceedings of the 27th Annual UCLA Indo-European Conference, Los Angeles, October 23rd ad 24th, 2015* (pp. 79-96). Bremen: Hempen Verlag.

Kümmel, M. J. (2016b). Zur 'Vokalisierung' der Laryngale im Indoiranischen. In D. Gunkel, J. T. Katz, B. Vine & M. Weiss, *Sahasram Ati Srajas: Indo-Iranian and Indo-European Studies in Honor of Stephanie W. Jamison* (pp. 216-226). Ann Arbor/New York: Beech Stave Press.

Kümmel, M. J. (2017). Agricultural Terms in Indo-Iranian. In M. Robbeets & A. Savelyev, *Language Dispersal Beyond Farming* (pp. 275-290). Amsterdam: John Benjamins Publishing Company.

Kümmel, M. J. (2018). The Survival of Laryngeals in Iranian. In L. Beek, A. Kloekhorst, G. Kroonen, M. Peyrot & T. Pronk, *Farnah: Indo-Iranian and Indo-European Studies in Honor of Sasha Lubotsky* (pp. 162-172). Ann Arbor/New York: Beech Stave Press.

Kümmel, M. J. (2019). Early Indo-Iranic Loans in Uralic: Sounds and Strata. *Contacts: Archaeology, genetics, languages joining forces to shed light on early contacts (4000 BC – 1000 AD) between Indo-European and Uralic speakers.* Sveaborg.

Kümmel, M. J. (2019). *Einführung ins Ostmitteliranische.* URL: https://www.academia.edu/30130317/Einführung_ins_Ostmitteliranische_Introduction_to_Eastern_Middle_Iranian_.

Lipp, R. (2009). *Die indogermanischen und einzelsprachlichen Palatale im Indoiranischen: Band II, Thorn-Problem, indo-iranische Laryngalvokalisation.* Heidelberg: Universitätsverlag Winter.

Lubotsky, A. M. (1981). Gr. *pḗgnumi* : Skt. *pajrá-* and loss of laryngeals before mediae in Indo-Iranian. *Münchener Studien zur Sprachwissenschaft 40*, 133-138.

Lubotsky, A. M. (1988). *The System of Nominal Accentuation in Sanskrit and Proto-Indo-European.* Leiden: Brill.

Lubotsky, A. M. (1989). Against a Proto-Indo-European Phoneme **a*. In T. Venneman, *The New Sound of Indo-European* (pp. 53-66). Berlin: Walter de Gruyter.

Lubotsky, A. M. (1990). La loi de Brugmann et **H₃e*-. *La reconstruction des laryngeales (Bibliothèque de la Faculté de Philosophie et Lettres de l'Université de Liège 253)*, 129-136.

Lubotsky, A. M. (1994). RV. *ávidhat*. In G. E. Dunkel, G. Meyer, S. Scarlata & C. Seidl, *Früh-, Mittel-, Spätindogermanisch. Akten der IX. Fachtagung der Indogermanischen Gesellschaft vom 5. bis 9. Oktober 1992 in Zürich* (pp. 201-206). Wiesbaden: Dr. Ludwig Reichert Verlag.

Lubotsky, A. M. (1995). Reflexes of Intervocalic Laryngeals in Sanskrit. In W. Smoczýnski, *Kurylowicz Memorial Volume. Part One* (pp. 213-233). Krakow: Universitas.

Lubotsky, A. M. (1997). The Indo-Iranian reflexes of PIE **CRHUV*. In A. M. Lubotsky, *Sound Law and Analogy: Papers in honor of Robert S.P. Beekes on the occasion of his 60th birthday* (pp. 139-154). Amsterdam: Rodopi.

Lubotsky, A. M. (2000). Indo-Aryan 'six'. In M. Ofitsch & C. Zinko, *125 Jahre Indogermanistik in Graz* (pp. 255-261). Graz: Leykam.

Lubotsky, A. M. (2001a). Reflexes of Proto-Indo-European **sk* in Indo-Iranian. *Incontri linguistici 24*, 25-57.

Lubotsky, A. M. (2001b). The Indo-Iranian Substratum. *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* (pp. 301-317). Helsinki: Mémoires de la Société Finno-ougrienne.

Lubotsky, A. M. (2008). Vedic 'ox' and 'sacrificial cake'. In A. M. Lubotsky, J. Schaeken & J. Wiedenhof, *Evidence and Counter-evidence. Essays in Honor of Frederik Kortlandt. Volume 1: Balto-Slavic and Indo-European Linguistics* (pp. 351-360). Amsterdam: Rodopi.

Lubotsky, A. M. (2018). Proto-Indo-Iranian Phonology. In J. Klein, M. Fritz & M. Fritz, *Handbook of Comparative and Historical Indo-European Linguistics* (pp. 1877-1888). Berlin: De Gruyter Mouton.

Macdonell, A. A. (1916). *A Sanskrit Grammar for Students.* New Delhi: D.K. Printworld.

Mallory, J. P. (1989). *In Search of the Indo-Europeans: Langauge, Archaeology and Myth.* New York, N.Y: Thames and Hudson.

Mallory, J. P. (1998). A European Perspective on Indo-Europeans in Asia. In V. Mair, *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia* (pp. 175-201). Washington/Philadephia: The University of Pennsylvania Museum Publications.

Martirosyan, H. (2010). *Etymological Dictionary of the Armenian Inherited Lexicon.* Leiden: Brill.

Matasovic, R. (2009). *Etymological Dictionary of Proto-Celtic.* Leiden: Brill.

Mayrhofer, M. (1983). Lassen sich die Vorstufen des Uriranischen nachweisen? *Akten der Österreichischen Akademie der Wissenschaften*, 249-255.

Mayrhofer, M. (1986). *Indogermanische Grammatik. 2. Halbband: Lautlehre.* Heidelberg: Carl Winter Universitätsverlag.

Melchert, H. C. (1977). Tocharian Verb Stems in *-tk-*. *Zeitschrift für vergleichende Sprachforschung 91*, 93-130.

Morgenstierne, G. (1938). *Indo-Iranian Frontier Languages Vol. II: Iranian Pamir Languages (Yidgha-Munji, Sanglechi-Ishkashmi and Wakhi.* Oslo: H. Aschehoug & Co.

Ollet, A. (2014). Evidence of Laryngeal Coloring in Proto-Indo-Iranian. *Historische Sprachforschung 127*, 150-165.

Parpola, A. (2012). Formation of the Indo-European and Uralic (Finno-Ugric) Language Families in the Light of Archaeology: Revised and Integrated 'Total' Correlations. *A Linguistic Map of Prehistoric Northern Europe. Suomalais-Ugrilaisen Seura Toimituksia 266*, 119-184.

Parpola, A. (2015). *The Roots of Hinduism: The Early Aryans and the Indus Civilization.* Oxford: Oxford University Press.

Peyrot, M. (2018). Tocharian B etswe 'mule' and Eastern East Iranian. In L. Beek, *Farnah: Indo-Iranian and Indo-European Studies in Honor of Sasha Lubotsky* (pp. 270-283). Ann Arbor/New York: Beech Stave Press.

Pinault, G.-J. (2003a). Skt *kalyāṇa-* interprété à la lumière des contacts en asie centrale. *Bulletin de la société de linguistique de Paris 98*, 123-161.

Pinault, G.-J. (2003b). Futher links between the Indo-Iranian substratum and the BMAC language. *Paper given at the 12th World Sanskrit Conference.* Helsinki.

Pisani, V. (1981). Greco ἐναλίγκιος ed ἄντην. *Indogermanische Forschungen 86*, 206-211.

Pronk, T. (2011). The Saussure Effect in Indo-European Languages other than Greek. *Journal of Indo-European Studies 39*, 176-193.

Pystynen, J. (2014). Palatal Unpacking in Finnic. *Papers presented at the 19th Conference of the Finno-Ugric Association of Canada.*

Ravnaes, E. (1981). The Development of *ə*/Interconsonantal Laryngeal in Iranian. *Indo-Iranian Journal 23*, 247-273.

Rédei, K. (1986). *Zu den indogermanisch-uralischen Sprachkontakten.* Wien: Österreichische Akademie der Wissenschaften.

Renfrew, C. (1987). *Archaeology and Language, the Puzzle of Indo-European Origins.* London: Jonathan Cape.

Salvatori, S. (2008). The Margiana Settlement Pattern from the Middle Bronze Age to the Parthian–Sasanian: a Contribution to the Study of Complexity. In S. Salvatori, M. Tosi & B. Cerasetti, *The Bronze Age and Early Iron Age in the Margiana Lowlands: Facts and Methodological Proposals for a Redefinition of the Research Strategies* (pp. 57-74). Oxford: Archaeopress.

Sammallahti, P. (1988). Historical Phonology of the Uralic Languages with Special Reference to Samoyed, Ugric and Permic. In D. Sinor, *The Uralic Languages: Description, History, and Foreign Influences* (pp. 478-554). Leiden: Brill.

Schindler, J. (1977). Notizen zum Sieversschen Gesetz. *Die Sprache 22*, 56-65.

Schirmer, B. (1998). *Studien zum Wortschatz der Iguvinischen Tafeln: Die Verben des Betens und Sprechens.* Frankfurt am Main: Peter Lang Europäischer Verlag der Wissenschaften.

Schmidt, G. (1973). Die iranischen Wörter für “Tochter” und “Vater” und die Reflexe des interkonsonantischen *H* (*ə*) in den idg. Sprachen. *Zeitschrift für vergleichende Sprachforschung 87 (1)*, 36-83.

Schrijver, P. (1997). Animal, Vegetable, and Mineral: Some Western European Substratum Words. In A. M. Lubotsky, *Sound Law and Analogy* (pp. 293-316). Amsterdam: Rodopi.

Sims-Williams, N. (1998). The Iranian Languages. In A. Giacalone Ramat & P. Ramat, *The Indo-European Languages* (pp. 125-153). London: Routledge.

Sokolova, V. S. (1967). *Genetičeskie otnošenija jazguljamskogo jazyka i šugnanskoj jazykovoj gruppy*. Leningrad: Izd-vo "Nauka," Leningradskoe otd-nie.

Spengler, R. N., Cerasetti, B., Tengberg, M., Cattani, M. & Rouse, L. M. (2014). Agriculturalists and Pasoralists: Bronze Age Economy of the Murghab Alluvial Fan, Southern Central Asia. *Vegetation History and Archaeobotany 23*, 805-820.

Szemerényi, O. (1958). Greek *γάλα* and the Indo-European term for „milk". *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen 75*, 170-190.

Tadmor, U., Haspelmath, M. & Taylor, B. (2010). Borrowability and the Notion of Basic Vocabulary. *Diachronica 27:2*, 226-246.

Tichy, E. (1993). Kollektiva, Genus femininum und relative Chronologie im Indogermanischen. *Historische Sprachforschung 106 (1)*, 1-19.

Vaan, M. (2008). *Etymological Dictionary of Latin and the other Italic Languages.* Leiden: Brill.

Werba, C. H. (2005). Sanskrit *duhitár*- und ihre (indo-)iranischen verwandten. In G. Schweiger, *Indogermanica: Festschrift Gert Klingenschmitt* (pp. 699-732). Taimering: Schweiger VWT-Verlag.

Winter, W. (1978). The Distribution of Long and Short Vowels in Stems of the Type Lith. *ésti* : *vèsti* : *mèsti* and OCS *jasti* : *vesti* : *mesti* in Baltic and Slavic Languages. *Recent Developments in Historical Phonology*, 431-446.

Witzel, M. (1995). Early Indian History: Linguistic and Textual Parameters. In G. Erdosy, *Language, Material Culture and ethnicity. The Indo-Aryans of Ancient South Asia* (pp. 85-125). Berlin: De Gruyter.

Witzel, M. (1999a). Early Sources for South Asian Substrate Languages. *Mother Tongue (special issue)*, 1-76.

Witzel, M. (1999b). Substrate Languages in Old Indo-Aryan (Rgvedic, Middle and Late Vedic). *Electronic Journal of Vedic Studies 5-1*, 1-67.

Witzel, M. (2000). The Home of the Aryans. In A. Hintze & E. Tichy, *Anusantatyai. Festschrift für Johanna Narten* (pp. 283-338). Dettelbach: J. H. Röll.

Witzel, M. (2001). Autochtonous Aryans? The Evidence from Old Indian and Iranian Texts. *Electronic Journal of Vedic Studies 7(3)*, 1-118.

Witzel, M. (2003). Linguistic Evidence for Cultural Exchange in Prehistoric Western Central Asia. *Sino-Platonic Papers 129*, 1-70.

Witzel, M. (2006). Early Loan Words in Western Central Asia: Indicators of Substrate Populations, Migrations, and Trade Relations. In V. H. Mair, *Contact and Exchange in the Ancient World* (pp. 158-190). Honolulu: University of Hawai'i Press.

Witzel, M. (2009). The Linguistic History of Some Indian Domestic Plants. *Journal of BioSciences 34(6)*, 829-833.

Zhivlov, M. (2014). Studies in Uralic Vocalism III. *Journal of Language Relationship 12*, 113-148.

Appendix: Reference list of analyzed vocabulary

Layer	Indo-Iranian	Meaning	#
PII I	**ućig-*	sacrificing priest	1
PII 0	**aka-*	bad	2
	**anću-*	Soma plant	3
	**atHarṷan-*	priest	4
	**atka-*	cloak	5
	**(H)āni-*	linchpin, hip	6
	**bʰarṷ-*	to chew	7
	**bʰiš-aj-*	healer	8
	**bīja-*	seed, semen	9
	**ćarṷa-*	Name of a deity	10
	**ćyā-*	to freeze, congeal	11
	**gadā-*	club	12
	**gr̥da-*	penis	13
	**Hustra-*	camel	14
	**indra-*	name of a God	15
	**išt(i)-*	brick	16
	**ǰʰarmiḭa-*	house	17
	**kāća-*	grass	18
	**kaćḭapa-*	tortoise	19
	**kadru-*	reddish brown	20
	**kapāra-*	dish, bowl	21
	**kapaṷta-*	pigeon	22
	**kapHa-*	phlegm	23
	**kHā-*	well, source	24
	**kHara-*	donkey	25
	**kšīra-*	milk	26
	**kućsi-*	~ side of the body	27
	**matsḭa-*	fish	28
	**mr̥ga-*	wild animal	29
	**muska-*	testicle	30

Layer	Indo-Iranian	Meaning	#
	**nagna-*	bread	31
	**pāpa-*	bad	32
	**parsa-*	sheaf	33
	**paṷasta-*	cover	34
	**rāći-*	rope	35
	**ringa-*	mark	36
	**ṛ̊si-*	seer	37
	**sćāga-* / **sćaga-*	goat	38
	**spāra-*	ploughshare	39
	**stuka-* / **stupa-*	tuft of hair	40
	**ṷāćī-*	axe	41
	**ṷand*(*H*)*-*	to praise	42
	**ṷarāj́*^{h}*a-*	boar	43
	**umā-*(*kā*)*-*	flax	44
	**ṷrīj́*^{h}*i-*	rice	45
	**ṷṛ̊tka-*	kidney	46
PII II	**čāt*(*ṷāla*)*-*	pit, well	47
	**i̯aṷīi̯ā-*	canal	48
	**ǰaǰ*^{h}*a*/*ukă̆-*	hedgehog	49
	**kárus-*	damaged	50
	**mai̯ūkHa-*	peg	51
	**pīi̯ūša-*	beestings	52
	**pusća-*	tail	53
Post-PII	***(*H*)*arǰaná-*/*áṇu-*	millet	54
	**banγα-*	hemp	55
	**ćika*(*tā*)*-*	sand, gravel	56
	**dŭ̆rća-*	(goat's) wool, hair	57
	**ganDərṷa-*	a mythical being	58
	**ganTi-*	smell	59
	**ganTuma-*	wheat	60
	**Kai̯ća-*	hair	61
	khaḍgá-/**karkadan*	rhinoceros	62

Layer	Indo-Iranian	Meaning	#
	*masu̍ra-/*mižuka-*	lentil	63
	**mVša-*	bean	64
	**naḭT^{s}(a)-*	skewer	65
	**pərd-a(n)k-*	leopard, panther	66
	**pinda-*	lump	67
	śāli- (Skt.)	unhusked rice	68
	*śaṇá-/*kanaba-*	hemp	69
	sarṣapa- (Skt.) etc.	mustard	70
	siṁhá- (Skt.) etc.	lion	71
	**šu̯aḭpa-*	tail	72
	**(t)sūkV̄-*	needle	73
	**u̯īna-*	lute	74
Inherited	**āćā-* / **aćas-*	space, region	75
	**ćan-*	to ascend	76
	**ćṷitra-*	white	77
	**daćā-*	thread, hem	78
	**d^{h}u̯aǰ-*	to flutter	79
	**g^{h}as-*	to devour	80
	**g^{h}au̯s-*	to make sound, hear	81
	**Hat-*	to wander	82
	**Hu̯ap-*	to strew, scatter	83
	**Hu̯ap-*	to shave, shear	84
	**Hu̯id^{h}H-*	to split in two	85
	**i̯ātu-*	black magic	86
	**ǰ^{h}ai-*	to incite	87
	**ǰ^{h}as-*	to laugh	88
	**kuč-*	to crook, bend	89
	**mag^{h}a-*	gift, offering, sacrifice	90
	**marǰ^{h}a-*	udder	91
	**monH-i-*	necklace	92
	**nard-*	to hum, complain	93
	**raǰ^{h}-*	to be abandoned	94

Layer	Indo-Iranian	Meaning	#
	**sag^{h}-*	to be able to bear’	95
	**srans-*	to fall away/apart	96
	**stHūna-*	pillar	97
	**sṷag-*	to embrace	98
	**ṷi̯ak-*	to encompass	99
	**ṷi̯atH-*	to be unsteady	100
	**ṷik-*	to separate	101
	**ṷr̥ćsa-*	tree	102
	**ṷriH-*	to oppress, collapse	103