

## **Outer and Inner Indo-Aryan, and northern India as an ancient linguistic area**

Claus Peter Zoller  
University of Oslo

### **Abstract**

The article presents a new approach to the old controversy concerning the veracity of a distinction between Outer and Inner Languages in Indo-Aryan. A number of arguments and data are presented which substantiate the reality of this distinction. This new approach combines this issue with a new interpretation of the history of Indo-Iranian and with the linguistic prehistory of northern India. Data are presented to show that prehistorical northern India was dominated by Munda/Austro-Asiatic languages.

Keywords: Indo-Aryan, Indo-Iranian, Nuristani, Munda/Austro-Asiatic history and prehistory.

### **Introduction**

This article gives a summary of the most important arguments contained in my forthcoming book on Outer and Inner languages before and after the arrival of Indo-Aryan in South Asia. The

traditional version of the hypothesis of Outer and Inner Indo-Aryan purports the idea that the Indo-Aryan Language immigration<sup>1</sup> was not a singular event. Yet, even though it is known that the actual historical movements and processes in connection with this immigration were remarkably complex, the concerns of the hypothesis are not to reconstruct the details of these events but merely to show that the original non-singular immigrations have left revealing linguistic traces in the modern Indo-Aryan languages. Actually, this task is challenging enough, as the long-lasting controversy shows.<sup>2</sup> Previous and present proponents of the hypothesis have tried to fix the difference between Outer and Inner Languages in terms of language geography (one graphical attempt as an example is shown below p. 106) which, in turn, was explained in terms of different immigration routes (northern vs. southern) and/or in terms of earlier vs. later immigrations.<sup>3</sup> But all such conjectures – also the well-founded ones – fail to explain what differentiates Outer Languages *essentially* from Inner Languages. My answer is that the historical development of the Outer Languages following the Proto-Indo-Iranian stage differed to some extent from the development that led from Proto-Indo-Iranian to Old Indo-Aryan (and, almost needless to say, from the development to Old Iranian). This is a claim otherwise associated only with Nuristani. Nuristani is seen by many, though not by all, as constituting a third branch within Indo-Iranian. I will argue below that this is only partially correct: There is much evidence to show that there continued to exist a linkage

---

<sup>1</sup> Here and in the forthcoming book I mostly speak of *language* immigration. I thereby bypass the question whether this immigration was primarily due to population movements or primarily due to language adaptation. This question is not essential for my arguments. I will also not argue here with the proponents of the Indigenous Aryans Theory since their arguments have nothing to do with the model proposed here.

<sup>2</sup> It needs to be understood that the opponents of the Outer-Inner Language theory (or the agnostics) usually do not doubt that the Indo-Aryan immigration was a non-singular and fairly long process. Meanwhile much evidence has been accumulated to show this (see in the Literature e.g. quoted publications by Michael Witzel or Asko Parpola). There is also evidence that speakers of Indo-Aryan were already in South Asia before the arrival of the Vedic Aryans as has been shown e.g. by Rainer Stuhmann (2016). But neither Witzel nor Parpola nor others are concerned with modern linguistic reflexes.

<sup>3</sup> I will argue below that only the latter paradigm makes sense because today it is impossible to draw clear-cut borders between presumed Outer and Inner Languages.

of languages/dialects<sup>4</sup> even after the separation of Proto-Indo-Aryan from Proto-Iranian.

The article consists of four main parts and several sections: Part I recapitulates briefly the scholarly history of the debate from its beginnings in the 19<sup>th</sup> Century till today. Part II discusses succinctly some linguistic terms and facts concerning the history of Indo-Aryan which have been ignored by the proponents and opponents of the hypothesis but which are crucial for a clear understanding of the model proposed here. I will argue in Part III that from among the many previously proposed arguments, only the one pertaining to the contested historical origin of the Middle IndoAryan (MIA) *-alla/illa/ulla-* suffix is of diagnostic value for supporting the hypothesis.<sup>5</sup> In Part IV the old hypothesis of Outer and Inner Languages is integrated into a new theory.<sup>6</sup> The core of this new theory consists of two propositions the fulfilment or non-fulfilment of which decides on the validity or invalidity of the theory. Here is a summary of the two propositions:

- Especially – but not only – peripheral New Indo-Aryan languages<sup>7</sup> have to show evidence for OIA and PIE features neither found in Vedic nor in Classical Sanskrit.<sup>8</sup>

---

<sup>4</sup> Following the terminology of Malcolm Ross, I will henceforth use the term ‘lect’ as a cover term for both language and dialect. A linkage of lects arises after lectal differentiation of a language. An example of a linkage of lects are the language varieties found in the Hindi Belt.

<sup>5</sup> This suffix is of Proto-Indo-European (PIE) origin, but it is unknown in Old Indo-Aryan (OIA) and only treated by Indian grammarians of MIA from around the 4<sup>th</sup> or 5<sup>th</sup> Century CE onwards. On pp. 93ff. I will show that this is a striking example for the gradual penetration of Outer Language (OL) features into the area of the Inner Languages (IL).

<sup>6</sup> I use the term not in the Anglo-Saxon but in the traditional German way: A theory is a system of propositions, which is used to describe or explain aspects of reality and make predictions about potential/possible (future) observations.

<sup>7</sup> That means especially (but not only) Outer Languages which are by definition peripheral both in geographical and in cultural-linguistic terms, i.e. typically non-written languages spoken by non-dominant populations that have been linguistically influenced only little or only moderately by the Indo-Aryan koinés. I will explain in Part II my understanding of the term koiné. With “especially but not only Outer Languages” I mean that e.g. Braj Bhasha, even though it is the sister language of Hindi and also located in the *Madhya deśa*, the ancient center of Vedic language and culture, contains significantly more Outer Language features than Hindi. I explain this

- South Asia has long since been recognized as constituting a linguistic area (*Sprachbund*). Its formation is usually ascribed to interactions between Indo-Aryan and Dravidian. Influence through Munda is regarded as less important and even less important is the influence through Tibeto-Burman. This situation makes it safe to assume that before the advent of Indo-Aryan (and Dravidian?<sup>9</sup>) northern India was characterized by a substantially different type of linguistic landscape. This is standing to reason vis-à-vis the very long history of human habitation in South Asia. Hence, the second proposition predicts that the earlier immigration – namely of the ancestor of the Outer Languages – must have experienced a *strong* impact from linguistic features of the prehistoric linguistic area, and which has left clear traces in the modern Outer Languages, whereas the later Vedic language immigration led to a *weak* impact on Vedic and its follow-up languages.

My contention is that if linguistic data can indeed be presented to support both propositions, this is a definite confirmation of the theory of a distinction between Outer and Inner Languages. Hence, Part III will present (a) some diverse linguistic examples which support the first proposition (i.e. inherited linguistic data not found in OIA but only at later historical stages), and (b) will present other linguistic examples of a completely different nature because they are non-inherited. These latter data are presently predominantly found in

---

fact by presuming that Proto-Braj Bhasha was stronger influenced by Outer Language features than Proto-Hindi.

<sup>8</sup> This proposition entails the prediction that if there were (at least) two different immigrations of two different OIA languages/dialects into South Asia (one of them the Vedic language) then their mutual mingling must have required many centuries before the first Outer Language features surfaced in the MIA phase of the koinés.

<sup>9</sup> Whereas some Dravidologists, like Bhadiraju Krishnamurti (2003), are agnostic on the question of the origin of Dravidian, I may also mention G. Uma Maheshwar Rao (University of Hyderabad) who pursues Dravidian and Mongolian comparative studies, and the publication of Jaroslav Vacek (1989). But in any case, this question is not of central relevance for the argumentation of this article because it is most unlikely that early Dravidian was once spoken all over the same area where now Indo-Aryan is spoken.

north-western Indo-Aryan Outer Languages,<sup>10</sup> in Tibeto-Burman West Himalayish and in its related, but meanwhile extinct Zhang Zhung language,<sup>11</sup> and in the language isolate Burushaski.<sup>12</sup>

The examples of (b) presented below, which are non-inherited (i.e. not of Indo-Aryan pedigree), show clearly that the linguistic area of northern India, including large stretches of the mountains between Himalayas and Hindu Kush, was in prehistoric times (i.e. before the arrival of Indo-Aryan) deeply influenced by Munda/Austro-Asiatic languages and their typical linguistic features.<sup>13</sup> The fact that these Munda/Austro-Asiatic words and features are practically absent from Vedic and Classical Sanskrit, that they are still rare in MIA *deśya* dialects (see below) but found in impressive number in the just-mentioned modern languages is a clear proof for the correctness of the above-formulated second proposition.<sup>14</sup>

---

<sup>10</sup> This means mainly Nuristani, Dardic and Pahārī languages spoken roughly between Uttarakhand and eastern Afghanistan.

<sup>11</sup> West Himalayish is a group of small Tibeto-Burman languages scattered through the high valleys of the Himalayas between Nepal and Kashmir. The Zhang Zhung language was spoken in large parts of Upper Tibet. It died out as a written language in the 8<sup>th</sup> Century CE. Specialists find the closest relationship between Zhang Zhung language and West Himalayish, forming together a separate branch (different from Tibetan) within the Tibeto-Burman language family.

<sup>12</sup> Spoken in the Northern Areas of Pakistan.

<sup>13</sup> The expression Munda/Austro-Asiatic is a kind of kludge. The Austro-Asiatic language family comprises around 160 languages which are spoken from eastern India throughout large parts of Southeast Asia and even in some areas of southeast China. They are usually divided into two main branches: Munda and Mon-Khmer. The Munda languages are spoken in Bangladesh and India; in India in West Bengal, Odisha, Jharkhand, Chattisgarh, Uttar Pradesh, Bihar and Andhra Pradesh. A western offshoot is Korku spoken in Madhya Pradesh and Maharashtra. The linguistic history of Austro-Asiatic is till today much less well understood than the history of Indo-Aryan. Despite the undisputable linguistic kinship between Munda and Mon-Khmer, both branches also differ impressively from each other. There seems to be a tendency among specialists to suggest that Munda is intrusive from Southeast Asia into India, but there have also always been others who have argued the other way round. Due to the many open questions I use the term Munda/Austro-Asiatic (in accordance with Parpola) as a kludge. The conclusions I draw from the findings of clear Munda/Austro-Asiatic languages (language features) mainly in the north-west of South Asia are found towards the end of this article. But already here I want to make clear that I will not make a definite statement with regard to the homeland of Austro-Asiatic.

<sup>14</sup> We will see below that nevertheless Munda/Austro-Asiatic was not the only language family of prehistoric northern India.

### Part I A short history of the Outer and Inner Indo-Aryan controversy

The hypothesis actually goes back to the 19<sup>th</sup> Century missionary and linguist August Friedrich Rudolf Hoernle who formulated it the first time in his 1880 publication. However, it only came into prominence through Abraham Grierson (*Linguistic Survey of India* 1.1: 116-118 and his article 1931-33), but was resolutely rejected by Suniti Kumar Chatterji (1926: 150-169).<sup>15</sup> This related to the following arguments regarding the following preservations in northwestern (and eastern) Indo-Aryan: preservation of final *-i*, *-e* (and *-u*); epenthesis; *i > e*, *u > i*, *ai* and *au > e* and *o*; *c, j > ċ, dz*; phonological status of *ñ* and *ñ̃*; *l > r* and *ḍ > r*; *d ~ ḍ*; *d > j*; *-mb- > -m-*; *-r- > -ṛ-*; *-s- > -h-*; *ś, ṣ, s > ś*; tendency for loss of word-medial aspiration; non-occurrence of compensatory vowel-lengthening; feminine *-ī*; ablative postposition; synthetic declension; pronominal affixation; *-l-* past and adjectival *-l-*; lexical evidence. Chatterji rejects all Grierson arguments that the quoted features would reflect a difference between Outer and Inner Languages. I fully agree with Chatterji<sup>16</sup> with the exception of three features: the alternation *d ~ ḍ*,<sup>17</sup> the historical process *c, j > ċ, (d)z* and the historical origin of the *-l(l)-* past which I discuss below. Chatterji's rejection of the hypothesis brought the discussion to an effective standstill until it was revived almost hundred years later by Franklin Southworth (2005a).

Besides the '*-l(l)-* past argument', Southworth suggests in addition the following features as characteristic for Outer Languages as against Inner Languages (2005a: 136ff.): modern reflexes of OIA gerundive *-(i)tavya-*; unequal geographical distribution of reflexes of OIA *r*; lack of length contrast in *i* and *u*; word accent; change *l → n*; lexical evidence. The six arguments are unequally persuasive and all in all not really convincing. For instance, quoting a few parallel

<sup>15</sup> Chatterji discusses here Grierson's article from 1920 on Indo-Aryan vernaculars, published in *BSOAS* I,III.

<sup>16</sup> It is impossible to recapitulate here all of Chatterji's linguistic arguments in detail. The interested reader is advised to read the quoted passages.

<sup>17</sup> This is actually a special case of the more general alternation dental ~ palatal ~ retroflex, which is indeed an Outer Language feature because the most likely cause for it is the Munda/Austro-Asiatic linguistic area (it is a Munda, not an Austro-Asiatic feature). It will not be discussed in this article.

lexemes in Marāṭhī, Bengali and Chakma (p. 145f.) is not enough evidence for showing that several thousand years ago two distinct IA immigrations into South Asia took place. But for a discussion of the different values and problems concerning the six arguments I have again to refer to my forthcoming publication. However, I do want to mention here that George Cardona arrived at the following conclusion on Southworth's attempt (Cardona and Jain 2003: 19).<sup>18</sup> “[I]t [is] fair to say that these conclusions are not sufficiently backed up by detailed facts about the chronology of changes to merit their being accepted as established.”

Does this unsatisfying situation mean that the hypothesis of Outer Languages and Inner Languages is passé? Certainly not! As a matter of fact, the weaknesses do not lie in the conceptions of the hypothesis but in the arguments proposed so far to support it. Above I formulated two propositions, but besides them more background information needs to be introduced.

## **Part II Further background information**

### *Koinés, lingua francas and ‘village dialects’*

There exist various definitions for the terms koiné and lingua franca which quite often either differ from each other or simply hold that both terms mean the same. Indeed, the terms seem to overlap; still, it is possible, and necessary for our purpose, to differentiate them from one another even though it is not feasible here to treat this topic in a comprehensive way. It is also obvious that there probably exist only gradual differences between certain real koinés and certain real lingua francas.

I first refer to the definitions given by the German *Duden* editors: a koiné is ‘a language created by the leveling of dialect differences’ whereas a lingua franca is ‘an interlanguage of a larger multilingual space’. This means, koinés and lingua francas have different forms of genesis. Whereas a koiné is the outcome of a dialect selected, standardized and canonized (through standard grammar, canon of literature, etc.) within a community of speakers of closely

---

<sup>18</sup> Cardona had occasion to read Southworth's manuscript before its publication in 2005. For a more general critique of the hypothesis see also Masica (1991: Appendix II).

related dialects, a lingua franca must and usually does not undergo this process of selection, standardization and canonization. On the contrary, speakers of a language A can agree with speakers of a language B to use any language C for their communication needs. Whereas a koiné has typically a touch of ‘elite’, like Sanskrit, there are among lingua francas – even though the term is used independently of the history and structure of such a language – also pidgins and creoles with inferior status like Pidgin English spoken in the Solomon Islands and Vanuatu. But koinés and lingua francas share the feature that both can detach themselves from their original contexts. For instance, Sanskrit was and partly still is a koiné in the Dravidian language area. This leads to one more characteristic to be mentioned here: neither a koiné nor a lingua franca must be someone’s mother tongue. But what is about the very many ‘village dialects’ which are so characteristic for South Asia (and many other areas)? I suggest defining them negatively: they are neither koinés nor are they lingua francas, they are – languages. In this function, and under this perspective, a ‘village dialect’ from the Karakorum is, despite lack of detailed knowledge of its history, not necessarily less important for the reconstruction of the history of Indo-Aryan than Sanskrit is. But now I turn to some more concrete facts related to South Asian koinés and ‘village dialects’.

In the textbooks dealing with the history of Indo-Aryan, a linguistic family tree is drawn with OIA on top and ever new ramifications down to the modern languages, even though every specialist can tell that the different MIA languages (e.g. Pali) are not direct descendants of the Vedic corpus. Thomas Oberlies states about Pali and other MIA languages (1990: 39): “These languages are by no means straightforward continuants of the Old Indo-Aryan (= OIA) of the Vedic corpus.” He refers to a complex and much discussed problem; however, I think, one important factor why this is so has not been sufficiently taken into account, namely the continuous and long-lasting formation of ever new koinés. Every ‘normal’ language is characterized by so-called emblematic features which bestow it a regional identity. Koinés are devoid of emblematic features – they have been cleansed of such traits – and it is therefore difficult to identify them with a specific region (see Ross 1997 for more details). The largest part of the history of Indo-Aryan is only known through the succession of one koiné after the other. Without claim for



completeness: Vedic → Sanskrit (“the purified one” as the prototypical koiné) → Pali → Jaina Apabhraṃśa → Sādhukkarī → Modern Standard Hindi, Fiji Hindi, Modern Standard Bengali etc. All these koinés came into prominence due to political, cultural and religious factors. Thus, their written documents present a skewed and one-sided picture of the many forms of Indo-Aryan at their time. Even though Sheldon Pollock has observed a movement for vernacularization in the Indian Middle Ages – in short, a slowly growing interest of the elites in demotic language use – he also concludes correctly (2006: 287) that “... nowhere ... has literature been coeval with its language, not even with its written form. The histories of vernacular languages in South Asia demonstrate this unequivocally, not least by the temporal gap mentioned earlier that separates the moment of literization,<sup>19</sup> or the attainment of literacy, from the moment of literarization, or the attainment of literature – a gap that is often chronologically appreciable and always historically significant.”

Besides these koiné-internal fractures mentioned by Pollock, there must have existed also linguistic incongruences between successive koinés (remember Oberlies’ observation), and there existed also always an asymmetrical relationship between the respective dominant koinés – wearing the ‘emblem’ of literacy – and the many more non-dominant ‘village dialects’, which never achieved literization or literarization and whose emblems consisted of regional peculiarities which also included specific linguistic innovations and archaisms. If we provisionally equate the dominant koinés with the Inner Languages and the non-dominant ‘village dialects’ with the Outer Languages, and if we accept that every koiné was in the beginning not a koiné but a ‘village dialect’, then we can expect that the interfaces between the successive koinés were not only joinings but also breakages where Outer Language features could easily sneak in. Out of a number of examples I give here two to show what I mean:

- It has been known for a long time that MIA Pali and Prakrit (*i)dha* ‘here’ and Aśokan (*hi*)*da* continue PIE *\*h<sub>1</sub>idh<sub>a</sub>* ‘here’ (Mallory and Adams 2006: 418) and are thus more archaic than Vedic *ihá* ‘here’ because of the Vedic loss of *-d-*. This is

---

<sup>19</sup> Pollock here means literization in the sense of acquiring the ability to write.

just one of around 50 examples discussed by Oberlies (1999) who concludes his article with these words (p. 48): “Some of these forms and words – such as *idha* ... are phonetically older than even Vedic, while some must be the continuations of certain dialectical variations within Old Indo-Aryan.” We should realize that Pali and Prakrit were in use more than a millennium after Vedic and that some of the Pali and Prakrit forms go even back to pre-Vedic even though both koinés got purified of regional emblems similar like Sanskrit.

- Western etymological dictionaries of Old Indo-Aryan contain also quite many words which are tagged with the term ‘lexicographic’. That is, they are not documented in genuinely old Sanskrit texts but only found in lexica written in India.<sup>20</sup> In publications dealing with IA etymological matters they are always treated with utmost prudence because of their late attestation and they are frequently suspected to be free concoctions of the lexicographers. But what to do if such words, and not few of them, are actually found in modern, sometimes peripheral, small languages without written heritage? Here a few examples out of many:<sup>21</sup> P. *kathal* ‘a plough; also the main shaft of a plough’ < OIA lex. *kuntala-* ‘plough’; Bng. *kiṇḍūrɔ* ‘strong (as a man)’ with metathesis < OIA lex. *kunḍīra-* ‘strong, powerful’; Garh. *kujeri, kujyaru* ‘mist, haze; fog’ < OIA lex. *kujjhaṭi-* ‘a fog or mist’; Pr. *čō* ‘bribe’ < OIA lex. *chāya-* ‘bribe’; Kṭg. *gəriśtu* ‘small quantity of cow dung’ < OIA lex. *gopurīṣa-* ‘cow-dung’ plus diminutive suffix, etc. Since it is very unlikely that, at least in peripheral languages like Nuristani<sup>22</sup> Prasun and West Pahārī Bangāñī,<sup>23</sup> these words were borrowed from Sanskrit thesauri, one possible explanation is that these are words of real Old Indo-Aryan origin even though they are not found in Vedic and Classical Sanskrit. This is clearly so in case of Kṭg. *gəriśtu* which is a compound with second component <

<sup>20</sup> The oldest extant thesaurus of this type, the *Amarakośa*, is dated ca. 400 CE, but Sanskrit lexica were produced until the time of the Mughals.

<sup>21</sup> Many other examples will be found in my forthcoming publication.

<sup>22</sup> The Nuristani languages are spoken in East Afghanistan and North Pakistan.

<sup>23</sup> The many varieties of West Pahārī are spoken in Himachal Pradesh and parts of Uttarakhand.

documented OIA *púrīṣa-* ‘ordure’. In other examples mentioned in this paragraph the origin is less clear, and in still other cases the lexicographers simply erred: S. *alu* ‘young crow’ is not a Sanskrit word deriving < OIA lex. *ali-* ‘crow’ (Monier-Williams) but both are borrowings from Munda, cf. Sora and Kharia *ol-* ‘to crow’.<sup>24</sup>

Despite these complexities, there is no doubt that at the time of the immigration of Old Indo-Aryan into South Asia a whole bunch of Indo-Aryan dialects/variants existed. Parpola arrives at the same conclusion even for a much earlier period of Aryan because he argues (2002: 79): “According to the testimony of the numerous and partly very early Aryan loanwords in the Uralic (Finno-Ugric) languages spoken in the forest zone of eastern Europe ... the Aryan proto-language was dialectally differentiated from the start.” However, this is not enough to propose a fundamental difference between Outer and Inner Languages because the above-quoted data can still be accommodated within a model of prolonged language immigration, naturally entailing a whole bunch of dialects/variants. In order to get a clearer and more conclusive point of view, it is helpful to introduce here some principles of the so-called Social Network Model (Ross 1997) which was designed to get to grips with the linguistic history of the Austronesian family of languages. This family is spread over an enormously large part of the Pacific Ocean side of our globe and one of its distinguishing features is its lack of any ancient tradition of written literature.

#### *Language fissure and lectal<sup>25</sup> differentiation*

Ross suggests the following definition (1997: 212): “Fissure is reflected in discrete bunches of innovation, lectal differentiation is reflected in overlapping (bunches of) innovations. These two patterns reflect different SCEs.<sup>26</sup> Language fissure is usually the result of a single event which divides one group of speakers into two, whilst

<sup>24</sup> What it means to find a Munda word in Sindhī and in a Sanskrit thesaurus but not in Old and Middle Indo-Aryan will be explained below in the last part.

<sup>25</sup> As pointed out above in footnote 4, Ross uses the noun ‘lect’ and the adjective ‘lectal’ in order to circumvent the distinction between ‘language’ and ‘dialect’ which indeed is hardly tenable from a linguistic point of view.

<sup>26</sup> Speech community events.

lectal differentiation entails the (usually gradual) geographic spread of a group of speakers.” An example for a fissure thus is the division between Old Iranian and Old Indo-Aryan which is characterized by discrete bunches of innovations. An example for a lectal differentiation is the history from Old to New Indo-Aryan. However, I want to direct attention to the fact that sometimes innovations that have come up in one language (area) through a fissure, nevertheless can go across the new border and spread into the other language (area). Below I will give an example for this. But here I want to add that Ross of course follows the well-founded communis opinio that language fissures and differentiations are always the result of innovations and not of preservations.

The language fissure which separated Old Iranian (OIr.) and Old Indo-Aryan had this form (here only a selection of the processes): Proto-Indo-Iranian (PII)  $*k̑, *g̑, *g̑^h$ <sup>27</sup> > OIA  $*ć > ś, j$  and  $*j^h > h$ ; PII > OIr. >  $*ć > *ć$ <sup>28</sup> (preserved in Nuristani) >  $s$  (Avestan),  $*j$  and  $*j^h > (d)z$  (note Iranian loss of aspiration). We see that Proto-Iranian and Proto-Aryan shared for some time the same process of PIE  $*k̑ > *ć$ . But whereas in Proto-Iranian this was followed (a) by depalatalization of  $*ć > *ć$  (= [tʃ] as in Zoller) and then (b) by deaffricatization of  $*ć > s$ , in Proto-Aryan the  $*ć$  changed directly into the palatal sibilant  $ś$  without undergoing deaffricatization. Nuristani preserved the stage of the depalatalized affricate  $ć$ . Thus we get the following equation for ‘10’: Avestan *dasa* – modern Nuristani *duc* – Vedic *dāśa*- (modern Dardic *daš*, Hindi *das*). This Nuristani archaism (and several other linguistic peculiarities) led Georg Morgenstierne to the postulation of a third branch within Indo-Iranian.<sup>29</sup> It also led him to postulate the widely accepted dictum (1961: 139): “There is not a single common feature distinguishing Dardic, as a whole, from the rest of the Indo-Aryan languages ... Dardic is simply a convenient cover term to

<sup>27</sup> These are so-called palatovelars.

<sup>28</sup> A depalatalization change from “tsh” to “ts”.

<sup>29</sup> Actually, this idea was also suggested before him, e.g. by Abraham Grierson. However, Morgenstierne’s predecessors allocated the wrong languages to the wrong branches. It was the achievement of Morgenstierne to correct this.

denote a bundle of aberrant Indo-Aryan hill languages.”<sup>30</sup> I show now that both claims are questionable.

*Nuristani, Dardic and other Outer Languages*

Nuristani has changed Proto-Indo-Iranian \**ć* into dental *ć* whereas Old Indo-Aryan changed \**ć* into *ś*. It seems self-evident that the change \**ć* > *ć* occurred before the change \**ć* > *ś* because both in Proto-Nuristani and in Old Iranian no deaffricatization of \**ć* took place. On the other hand, in Old Indo-Aryan no depalatalization of the fricative took place (OIA *ś* is still preserved in Dardic and West Pahārī). And the Proto-Indo-Iranian palatal affricates going back to Proto-Indo-European labiovelars (e.g. \**ć* < PIE \**kʷ*) were preserved in Old and Middle Indo-Aryan and in many New Indo-Aryan languages, as well as in Iranian as palatal affricates (cf. e.g. Vedic *catvāras* ‘4’, Hindi *cār*, New Persian *čahār* all < PIE \**kʷétuor-es*). On the other hand, Old Indo-Aryan *ś* and *c, j* are relatively frequently in Nuristani and, in case of *c, j*, very frequently in Dardic reflected with the dental affricates *ć, dz*: Nuristani Kāmdeshi *čāč’am* ‘large lizard’ (connected with OIA *śimśumāra-* ‘crocodile’), Dardic Indus Kohistani *ćλur* ‘4’ (but Hindi *cār*), Kāmdeshi *zā-* and Indus Kohistani *zāv* both ‘give birth’ (OIA *JAN* and Hindi *jannā* ‘to be born’). Even though there is no direct evidence for the antiquity of depalatalization in IA languages, it must be old. MIA Gāndhārī<sup>31</sup> had *ć* and (*d*)*z* sounds, even though their phonological status is unclear, and depalatalization is found in many IA languages between Dardic and Assamese.<sup>32</sup> It is sometimes claimed that depalatalization occurred early in Nuristani and late in Dardic. But this cannot be true because e.g. Nuristani Waigālī *čatk’a* ‘sharp; clever’ is apparently a fairly recent borrowing from Urdu *caṭak* ‘quickness; brightness’ which shows that we are dealing here with a long-lasting phonological process. Whereas OIA, and most MIA and NIA have only one series of (palatal) affricates, Nuristani

<sup>30</sup> The approximately 27 Dardic languages are spoken in North Pakistan. Kashmiri is usually also counted as a Dardic language but in my opinion it is an interlink between Dardic and West Pahārī.

<sup>31</sup> Was spoken in northwestern South Asia and in the Oasis towns of Central Asia.

<sup>32</sup> Chatterji (1926: 154f.) tries to explain depalatalization in dialects of Bengali, Assamese and Oriyā as result of Tibeto-Burman or Dravidian influence. Given the very wide spread of the phenomenon, this is more than unlikely and anyway does not work in case of Nuristani and Dardic.

and Dardic have mostly three (e.g.  $\check{c}$ ,  $\zeta$ ,  $\acute{c}$ )<sup>33</sup> and West Pahārī two (e.g.  $\acute{c}$ ,  $c$ ). This parallels their respective sibilant sub-systems where most Dardic languages have three ( $\check{s}$ ,  $\zeta$ ,  $s$ ), and West Pahārī and the Chittagong dialect of Bengali have preserved two ( $\acute{s}$ ,  $s$ ). These are more natural subsystems than the OIA sub-system with three sibilants but only one order of affricates, namely palatals. Therefore, where this standard OIA system prevailed, it later-on swiftly led to a reduction of the three OIA sibilants to only one already before Ashoka. We thus see that the archaism in Nuristani is simply due to the fact that depalatalization occurred *before* deaffricatization<sup>34</sup> whereas it occurred in Proto-Dardic *after* deaffricatization. This difference in the relative temporal sequence of two sound changes is hardly sufficient for postulating a separate language branch. If my thoughts are correct, the following conclusion is obvious: the speakers of Vedic must have lost direct contact with speakers of Old Iranian before that contact was lost by the speakers of the Aryan ancestor of the Outer Languages. In fact, there is another sound change discussed below which further supports this assumption.

Note also Cardona's assessment of the relative position of Nuristani. He writes (in Cardona and Jain 2003) p. 22: "Given that Nuristani lacks spirants  $f$ ,  $\theta$ , and  $x$  typical of Iranian, which deaspirated voiced aspirates ... it is reasonable to conclude that the deaspiration took place independently in Iranian and Nuristani ..."<sup>35</sup> We have seen above the Nuristani and Iranian change of  $*j$  and  $*j^h > (d)z$  whereas aspirated  $*j^h$  is reflected in Old Indo-Aryan as  $h$ : Proto-Indo-Iranian  $*j^h_1rd-$  'heart' > OIA  $h_1'd-$ , Nuristani Prasun *zir*, Old Avestan *zərəd-*. Cardona comments on this theme by discussing some

<sup>33</sup> That is, palatal, retroflex, dental.

<sup>34</sup> Of course, there are also many words in Nuristani where depalatalized affricates got subsequently deaffricatized.

<sup>35</sup> It is usually claimed that another important difference between Nuristani and Indo-Aryan is that Nuristani has completely (and early) lost all aspiration. I will not deal with this topic here as it is not of relevance for the discussion of the theory, but mention just two facts. First: Nuristani has not totally lost aspiration, at least in some languages it continues in a similar way as the automatic aspiration of unvoiced stops in most Germanic languages. Second: Rachel Lehr writes about the Darrai Nur dialect of Dardic Pashai (2014: 12): "The loss of aspiration is a feature of some Dardic languages, to differing degrees. Pashai shows no evidence of an aspiration contrast." Even though this appears to be a rather late development, it shows again the close correlation of Nuristani and Dardic phonological processes.

relevant details concerning Grassmann's Law which would be too complex to reproduce here. But Cardona clearly shows that a Proto-Indo-Aryan *\*jh* phoneme can be reconstructed which is reflected both in OIA *\*jh* (> *h*) and Proto-Nuristani *\*j/ž*, and he arrives at the following conclusion (p. 25): “Thus, the apparently great difference in the treatment of PIE *\*k* and so on cannot serve to demonstrate that Nuristani languages are definitely to be considered a branch separate from Indo-Aryan.”

Above-quoted Nuristani *duč* displays the vowel change *a > u* which is extremely common in Nuristani and Dardic languages as well as in many other Outer Languages. It is also found in Iranian. Here just a few examples from among many dozen collected by me: Avestan *puxða-* ‘fifth’ and Nuristani Ashkun *punc* ‘five’, Dardic Kalasha *šulá* ‘wood’ < OIA *śalākā-* ‘any small stake or stick’ (12349),<sup>36</sup> West Pahārī High Rudhārī *peṭabhurāi* ‘pregnant’ with first component < OIA *\*peṭṭa-* ‘belly’ (8376) and second < OIA *bhārin-* ‘bearing a load’ (9466), Bengali *usti* ‘bone’ either borrowed or deriving from OIA *ásthi-* ‘bone’ (quoted in Bodding 1936 v: 690). The process seems to have started in Avestan (see Hoffmann and Forssman 1996: 64); at least it is there where it is documented for the first time. Thus it is also very common in Middle and New East Iranian languages as in Khotanese *mūra-* ‘bird’ < Avestan *mərəga-* and Bartangi *δus* ‘10’. If the process started in East Iranian, it crossed a language fissure and entered OL Indo-Aryan. But it is practically absent from Old Indo-Aryan, and Georg Pischel (§ 111) has only a few Prakrit examples for context free *a > u*. Also this is a fairly strong argument for saying that the speakers of Vedic must have lost direct contact with speakers of Old Iranian before that contact was lost by the speakers of the Aryan ancestor of the Outer Languages. Both depalatalization and context-free change of *a > u* are already good arguments for supporting the old hypothesis: Besides some inconspicuous dialect variations within Vedic language,<sup>37</sup> a presence of other Old Indo-Aryan lects – separated from Vedic through fissures that may turn out to be more consequential – is already likely and will get support from compelling arguments introduced below.

<sup>36</sup> Numbers in parenthesis refer to lemmata in the CDIAL.

<sup>37</sup> None of these variations (see e.g. publications of Witzel in the Literature) has left traces in the New Indo-Aryan languages.

Here I would like to recall again the linguistic truism that language fissures occur due to innovations whereas preservations/ archaisms are not diagnostic for the identification of branches. This means that the question which *innovations* distinguish Nuristani from Indo-Aryan (and Iranian?) has not been answered. Richard Strand tries to explain this in an article for the *Encyclopedia Iranica* with the title “Nurestāni languages.”<sup>38</sup> From around the second half of the article he discusses the evolution of these languages. Strand divides this evolution into six phases; 1. Aryan phase, 2. Early Iranian phase, 3. Traditional phase, 4. Indo-Aryan phase, 5. Nurestān phase, 6. Afghan-Islamic phase. I have again to stave off the reader to my forthcoming publication in which I deal with all linguistic features suggested by Strand as characteristic for Nuristani. But I can present already here my summary: There is not a single Nuristani innovation – apart from innovations that have led to inner-Nuristani lectal differentiations – which does not have parallels either in Iranian (a few) or Indo-Aryan (many, i.e. most in Dardic but quite a number also in other IA languages). Here just two randomly selected examples. Within the “Indo-Aryan phase” Strand mentions “Anticipation of *r*.” This is elsewhere called “Dardic metathesis” (e.g. Kalasha *krum* ‘work’ < OIA *kárman-* ‘work’ [also with *a* > *u*]) and is widespread in Dardic and West Pahārī, and occasionally also found in other western and north-western languages. Within the “Nurestān phase” he mentions “Initial spirants assimilate following occlusion to become affricates.” An example is \*š<sub>g</sub>čil’a ‘slack’ (< OIA \*š<sub>r</sub>thilá-) > Nuristani Prasun čič’il. This is actually a so-called coronal consonant harmony (see Arsenault 2012) and is also found (frequently with aspiration fronting) e.g. in Dardic Indus Kohistani čhič<sub>h</sub>v̄ ‘to learn’ < OIA śikṣate ‘learns’ or West Pahārī Bangāñī čhič<sub>h</sub> ‘lime’ < OIA śvitrá- ‘white’, etc. Already these few examples should make clear that it is useful not only to study the uniqueness of Nuristani but also its deep interconnections with surrounding language families. Thus I suggest concentration on obvious features that show fluid transitions – and not sharply differentiated branches – from East Iranian to Nuristani to Dardic (and some other Outer Languages). Nuristani shares e.g. the following features with its wider surroundings, several of which are likely to be shibboleths of Outer Languages:

<sup>38</sup> <http://www.iranicaonline.org/articles/nurestani-languages>



- In Nuristani, or at least in the Waigalī variety of Nišigrām, a short *a* is pronounced quite back as [ɑ] as in Pashto and other Iranian languages, but not as in Indo-Aryan where a short *a* is typically pronounced more centrally like [ʌ]. This is not an Outer Languages feature but simply an example for the influence of Iranian phonology on Nuristani.<sup>39</sup>
- At least in the Nišigrām variety of Waigalī and in Prasun there exists a pitch accent at the word level. The accent appears usually, but not always, on the last syllable. Whether or not it has a distinctive function is not quite clear, but apparently it can shift its position in a word (Degener 1998: 36ff.). Thus this Nuristani accent system is very similar to accent systems in East Iranian languages like Pashto (see Morgenstierne 1973). This contrasts with many Dard languages which have preserved the older stage of flexible accent positions in the words. Examples (only with initial accent): OIA *nīla-* ‘dark blue’ — Dardic Kalasha *nīla* but Nuristani Kāmdeshi *niłá* ‘black’; OIA *bhrātṛ-* ‘brother’ — Dardic Kalasha *báya* but Nuristani Prasun *əv’ə* (if same derivation); OIA *ásru-* ‘tear’ — Dardic Shina *āśo* but Nuristani Prasun *ūč’ū*.
- In the Nišigrām variety of Waigalī vowel nasalization is contrastive; the language shares this feature with many IA languages, but in Iranian, e.g. in Pashto, vowel nasalization is not contrastive.
- Josef Elfenbein states about Pashto and IA “Lahndā”<sup>40</sup> (1997: 745): “There is also, as in Lhd, a spontaneous change of *ŋ* to *ṽr*, e.g. *rūṽr* < *rūnr* < *\*rūxšna-*.” Exactly the same phonetic phenomenon of shift of [+nasal] from a consonant to a preceding vowel – thus VN > ṽC – is known from Nuristani, from Dardic languages like Indus Kohistani, and quite frequently it is also found in other Outer Languages like poetic Ṭihriyālī dialect of Garhvālī *ḍrē lege* (Hindi *āne lag gaye*) ‘(they) started to come’, Braj-Awadhī as in *kāṛī* ‘a long

<sup>39</sup> I heard this [ɑ] many times both when working with a Waigalī language consultant in Oslo and from my Indus Kohistani language consultants in Pakistan whenever they started talking in Pashto.

<sup>40</sup> This is actually an outdated term for Hindko and Siraiiki.

deep basket' < OIA *kāraṇḍa-* 'basket', Nepālī, Bihārī, Bhojpurī, Bengali as in *rāṅ* 'widow' < OIA *raṇḍa-* 'maimed', in Assamese as e.g. in *kār* (*kāṅ*) 'arrow' < OIA *kāṇḍa-* 'arrow', and, significantly, in (North) Munda languages as e.g. in Kharia *hāriya* ~ *haṅiya* ~ *haṇḍiya* 'pots for holding rice beer' which has a parallel in Bengali *hāṅrī* 'pot'.<sup>41</sup> The Kharia and Bengali words are connected with OIA *\*hāṇḍa-* 'pot' (14050) and Chatterji wonders (1926: 553) about its possible relationship with OIA *bhāṇḍa-* 'pot' (9440), but loss of word-initial consonants or of word-initial closure is found in various Outer Languages<sup>42</sup> and in Munda, but it is almost unknown in OIA. In MIA it is not found in the standard Prakrits but only in *Deśya* Prakrit, which is significant. This will be discussed below.

- A velar nasal consonant is articulated in two different ways in northern South Asia: [ŋ] (*-ṅ(-)*) or [ŋg] (*-ṅg(-)*). The latter pronunciation is found in Old Indo-Aryan and in an Inner Language like Hindi, e.g. OIA *āṅga-* and Hindi *aṅg* 'limb' (114). The former pronunciation is found in Nuristani, in some Dardic languages, in Burushaski and at the other end of the IA world, e.g. in Assamese. The former pronunciation is also a typical characteristic of Munda/Austro-Asiatic. Since the phenomenon is found at the western and the eastern fringes of IA (the pronunciation may also be found in some IA languages in-between, but at the moment I am not aware of one), it is most likely an example of the *strong* impact of Munda/Austro-Asiatic on the first wave of OIA language immigration. Examples: Nuristani *aṅust'a* 'finger-ring' but OIA *aṅguṣṭhya-* 'pertaining to thumb or big toe' (138), Dardic Pashai *aṅ* 'arm' < OIA *āṅga-* 'limb' (114), Burushaski *aṅáro* 'Tuesday' borrowed < OIA *aṅāraka-* 'the planet Mars' (126), Assamese *āṅuli* 'finger' < OIA *aṅūli-* 'finger' (135), Munda Santali *baṅ baṅ* 'gaping hole' and *aṅ* 'body' (borrowing of 114).

<sup>41</sup> In the Chittagong dialect of Bengali the parallel is *āri* 'earthen saucepan', i.e. here [+nasal] has completely disappeared and initial *h-* has changed into a tone.

<sup>42</sup> Perhaps the most notorious case is Nuristani Prasun.

*Word and syllable languages, syncope and sesquisyllables*

The two topics of ‘word and syllable languages’ and ‘syncope and sesquisyllables’ have normally nothing to do with each other. But in Part IV we will see that a kind of phonological conspiracy appears to have taken place in languages of north-western South Asia: Inherited morphonological word language features characteristic of Indo-Iranian collaborated, so to say, with sesquisyllabic word structures which are one of the most important distinguishing features of the Austro-Asiatic family of languages.

For more details distinguishing word and syllable languages see Auer 2001. Here follows a small selection of important differences:

<b>classification parameters</b>	<b>word languages</b>	<b>syllable languages</b>
1) onsets/codas	complex	simple
2) accent	word level	phrase level
3) sonority hierarchy	little importance	important

- 1) Two examples for complex onsets are given in the next paragraph (Khowar, Prasun).
- 2) For instance, Hindi has only sentence intonation whereas e.g. Indus Kohistani uses pitch accents at the word level (see examples above).
- 3) Nuristani Kāmdeshi *sk'io* ‘fat’ has the more sonorant *s-* preceding the less sonorant *-k-* which is not possible in a syllable language. Even more extreme is the situation in Mon-Khmer Khasi (spoken in Meghalaya, more on it below) which not only allows a whole gamut of initial consonant clusters but also clusters of the type *bt-*, *bth-*, *bs-* which seem to violate Greenberg’s universal that in clusters it is the voiced stops which tend to be closer to the vocalic syllable nucleus (Jenny and Sidwell 2014: 284f.).

Syllable languages like Spanish, Munda, Hindi and Dardic Kalam Kohistani are more or less easier pronounceable than word languages like High German, Danish or Old Iranian. Syllable languages are therefore speaker-friendly, but make it more difficult to recognize word and morpheme boundaries, whereas word languages are hearer-friendly, i.e. they facilitate the decoding of morphological structures

and thus of information units. Even though it is clear that the development from OIA to NIA is basically one from a word to syllable languages,<sup>43</sup> one needs to be aware that frequently there exist also profound typological differences between genetically closely related languages (e.g. among the varieties of West Pahārī). But for us especially interesting is the following observation by Peter Auer:<sup>44</sup> “While the transition from syllable-language to word-language is unmarked in language change, the transition from word-language to syllable-language only occurs – top/down processes of language planning and standardization excluded – as a consequence of genetic non-transmission/difficult communication (extensive migration, second language acquisition, etc.)” This observation can be applied to the early history of Indo-Aryan: a main cause for the *marked* shift (i.e. not naturally language-internal) from the OIA word language to the NIA syllable languages was that in Vedic times speakers of local non-IA languages began in increasing numbers to learn and speak OIA as an L2. This hypothesis is confirmed by Kuiper who states (1991: 96): “The contact [of local people] with the community of Indo-Aryan speakers must primarily have been maintained by bilinguals, particularly among the lower strata of artisans and peasants (an aspect often overlooked by Vedists) and these must have been the essential factor in conforming the Vedic language to foreign patterns of the Indian linguistic area.” And Martin Kümmel (2014: 204) observes: “Most modern Indo-Aryan languages are often considered prototypical syllable languages. In contrast to that, Old and Middle Iranian allow many more consonant clusters and generally show more word-related features, especially in the East: Sogdian, Khwarezmian or modern Pashto may be characterized as quite typical word languages ...” Note, however, that in north-western South Asia, even though the area is mixed with regard to word and syllable languages, there are several clear word languages like Dardic Khowar and Nuristani Prasun, cf. Khowar *gr'isp* ‘summer’ and Prasun *çn(e)*- ‘to

---

<sup>43</sup> According to Martin Kümmel, quoted right below, OIA had already syllable language characteristics which only intensified in later stages of IA. Thus it is actually more correct to say that the process from OIA to NIA was basically one from a language with yet few syllable language features towards languages with more and more syllable language features.

<sup>44</sup> [http://www.frias.uni-freiburg.de/de/das-institut/archiv-frias/school-of-lili/veranstaltungen/Dateien\\_Veranstaltungen/a1](http://www.frias.uni-freiburg.de/de/das-institut/archiv-frias/school-of-lili/veranstaltungen/Dateien_Veranstaltungen/a1) — see there powerpoint page 21.

sneeze'. West Pahārī Bangānī is a syllable language allowing only few initial clusters but in nearby varieties of West Pahārī one finds e.g. *rgāṇu* 'to colour' (cf. H. *raṅgānā*) or *cmār* 'cobbler (Hindi *camār*)' (both forms showing syncope on which more below). We come upon a comparable situation with regard to Munda/Austro-Asiatic in eastern India. Whereas almost all Munda languages are marked syllable languages,<sup>45</sup> Mon-Khmer Khasi is known for its enormous amount of consonant clusters (Jenny and Sidwell [2014: 248] mention 127 different two-consonant clusters). Many of them go back to Proto-Mon-Khmer.<sup>46</sup> The difference in syllable structure between Mon-Khmer (e.g. Khasic) and Munda is seen in the following table:

PAA	Mon-Khmer		Munda	
* <i>blu:ʔ</i> 'thigh' <sup>47</sup>	Palaung	<i>blu</i>	Sora	<i>bulu-</i>
			Kharia	<i>bhulu</i>
* <i>kmu:ʔ</i> 'dirty'	Khmer	<i>khmau</i>	Mundari	<i>humu</i> <sup>48</sup>
	Kuy	<i>kmau</i>	Kurku	<i>kumu</i>
* <i>priəl</i> 'hail'	Bahnar	<i>prəl</i>	Gtaʔ	<i>bire:l</i>
	Khasi	<i>phria</i>	Gorum	<i>are:l</i>
* <i>ba:r</i> 'two'	Bahnar	<i>ba:r</i>	Mundari	<i>bar</i>

It is interesting to see that we have two comparable situations in the north-west and in the east of South Asia: more or less marked word languages are found in the north-west (e.g. Nuristani, Dardic, West Pahārī, partly Panjabi and Sindhī), whereas in the east only Khasic is

<sup>45</sup> An apparent exception is Gtaʔ with many initial clusters. But according to Anderson (2008), this is an innovation and an isolated case.

<sup>46</sup> Indo-Aryan (including Nuristani) and Iranian words are transliterated and transcribed according the common usage of Indologists and Iranianists. Burushaski has been transcribed in different ways by different authors. I follow here the system of Hermann Berger. These practices contrast with those of the Austro-Asianists (including those who work on Munda) who, in the majority of cases, present their data according to the International Phonetic Alphabet (IPA). I follow here this tradition although this entails sometimes small complications. Thus IA *j* corresponds roughly with AA *j* and *y* with *j*. Paul Olav Bodding in his Santali dictionary follows his own system. I have adapted data from him to the majority practice. In a very few cases my graphic presentations deviate slightly from the original because of font problems.

<sup>47</sup> For a fuller presentation of this lemma see Sidwell (2010: 124).

<sup>48</sup> Here and with Gorum *are:l* we have a case of deletion or weakening of an initial consonant. This is a characteristic of Munda which has affected Outer Languages to some extent. It will be discussed in more detail in part IV.

found like a rock in turbulent waters. Indeed, Sidwell and Rau write (in Jenny and Sidwell 2014: 283): "...Standard Khasi is important for comparative purposes; it shows a rich inventory of consonant clusters and vestiges of morphology. Also, being isolated geographically from SE Asia the branch has not been under the same areal pressures to restructure phonologically as many other AA groups." What is true for many other AA groups is also true for Munda and for many IA languages. The massive trend towards syllable languages is not 'natural' but, as pointed out above, is linguistically marked. Thus the question arises: were Indo-Aryan and Munda pushed in this direction by a third language family? Without being able to go into detail here, it is known that the MIA and NIA syllable and word structures display many similarities with Dravidian. Yet, I hesitate to accept that the push of Indo-Aryan and Munda in this direction could have been caused by Dravidian (alone).

### **Part III The first proposition**

I have stated above (p. 76) that from among the bundle of distinctive features suggested by Grierson and Southworth as characteristic for the Outer Languages, I regard only three as convincing:  $d \sim \bar{d}$ , the historical process  $c, j > \bar{c}$ ,  $dz$  and the historical origin of the  $-l(l)$ - past. The first feature will not be discussed in this article, the second has been dealt with above, and the third is the main topic of Part III here. This will be supplemented by a few examples of words from Outer Languages which are of Proto-Indo-European origin but not found in Vedic and Classical Sanskrit. I repeat here the assertion of the first proposition: the theory has to be accepted as accurate when linguistic data of Proto-Indo-European provenance only surface in Middle Indo-Aryan or later. This must be so because it must have taken centuries of mutual interpenetration of the originally separate Outer and Inner Languages until Outer Language features were registered. This is clearly the case with the  $-l(l)$ - past. The dispute about its origin and its cogency of proof stood in the center of the whole controversy. I therefore discuss it in detail.

*The historical origin of MIA -alla/illa/ulla-*

According to Cardona, the state of facts is as follows (in Cardona and Jain 2003: 18). He begins with a truism: “That modern Indo-Aryan languages are divisible into affiliated subgroups is beyond doubt. Thus, it is reasonable to say that there are eastern, northwestern, southwestern, and midlands groups. On the other hand, the precise manner in which a family tree is to be drawn up as well as the exact affiliation of particular languages ... are issues which have been not fully settled... Grierson ... divided Indo-Aryan into what he termed outer, mediate and inner sub-branches... A major criterion for recognizing an outer sub-branch was the occurrence of *-l-* in past participle forms ... (Grierson 1927: 140) as in Sindhī *māryō* or *māē-l* both meaning ‘beaten.’ Grierson did not, however, establish how the formation in question could be a common innovation of all the languages concerned, and Chatterji (1926: 167) was without doubt justified when he denied that an *-l-* past was a valid criterion for establishing an outer group, noting that it functioned as a past marker only in Eastern languages and Marathi... Moreover, no historical evidence is cited to demonstrate how the use of such participle forms developed in late Prākṛit or the earlier stages of languages like Gujarati, for which we have early sources.” — Cardona is not correct here: *l-* past markers and *l-* participles are also found in Dardic and West and Central Pahārī. For instance Dardic Brokskad *gālo* ‘(he) went’ and West Pahārī Sirājī of Ḍōḍā *kērī-lō* ‘was made up’. It seems also to be found in Nuristani, however not in past function but as agent, modality and ‘future’ marker (Almuth Degener, p.c.).<sup>49</sup>

Cardona’s quote is the formulation of a dominant point of view that has not substantially changed since Chatterji. However, I disagree with these conclusions. Pischel notes that the Prakrit grammarians teach that MIA *-alla/illa/ulla-* are used in the sense of Sanskrit suffixes *-mat* and *-vat* ‘furnished with’. The oldest among the grammarians who dealt with these suffixes quoted by Pischel is Vararuci who may have lived between the 3<sup>rd</sup> and 5<sup>th</sup> Century CE. The difference between *-l-* and *-ll-* suffixes has been explained by Pischel as parallel to other Middle Indo-Aryan processes of single consonants

<sup>49</sup> Future tense markers with *-l-* are quite widespread in NIA languages and they may be identical with the *-ll-* past markers. But this question is somewhat beyond the issues that are of direct relevance right here. I discuss the matter in my forthcoming publication.

getting doubled instead of deleted due to following accentuated vowels (§ 194) (e.g. MIA *duritta-* < OIA *duritá-* ‘evil etc.’; more examples are provided in fn. 6 of § 595). With regard to Apabhramśa, Ganesh Vasudev Tagare notes (1987: 336) that the primary suffix *-illa* means ‘agent, doer’ but the secondary suffix *-illa* (1987: 338) expresses, like the Prakrit suffixes, ‘pertaining to, possessing, having’. But Pischel, after noting that there is an increase of *-l-* suffixes from OIA to MIA, has also observed additional functions (§ 595). With single *-l-*: *mīsālia-* is a preterital passive participle of a denominative *\*mīsāla-* (OIA *miśra-*) thus probably meaning ‘was mixed’. With double *-ll-*: a present participle function e.g. in *sāsilla-* (OIA *śvāsin-*) ‘breathing’; and *-illa* is also used in the sense ‘located there’ (‘belonging to’) as in Pali *gāmilla-* ‘farmer’ (lit. ‘one belonging to a village’); it functions also pleonastic and can have also a passive past participle function e.g. in *theṇillia-* ‘taken; timid’ (cf. OIA *stená-* ‘thief’). The *-ll-* suffixes could combine with other suffixes quite freely with regard to sequencing. This is an important point because it suggests that the *-ll-* suffixes were, in case of verbal formations, not always automatically added to the past stems of verbs as assumed by Southworth and others.

Besides the few traces of past participle functions observed by Pischel, there are more such cases in Old Marāṭhī (i.e. Marāṭhī in use before the Muslim conquest soon after 1300 CE [Master 1964: v]). Thus it seems likely that the increased use of the *-ll-* suffixes in past participle constructions led them become tense and aspect markers. Discussing the possible historical origin of the suffixes, Southworth says (2005: 133) that Chatterji “mentions possible OIA origins” but “he also notes (and rejects) the suggestion of a link between the Indo-Aryan *-l-* pasts and the pasts in *-l-* found in Slavic and elsewhere in Indo-European.” The relevant passages in Chatterji (volume iii: 943f.) read like this: “Another view about the origin of the NIA. << -l- >> is that it is an independent affix occurring in Indo-European itself, preserved in NIA., but ignored or left unnoticed in OIA... But this connection is not proper. We have seen that the MIA. form of the suffix was << -ll- >> is entirely different...” So it seems that the question regarding the origin of the *-ll-* suffix is still not known. It cannot have derived from Vedic Sanskrit because, even though also Vedic Sanskrit has quite a number of words showing the presence of a



Proto-Indo-European *-l-* suffix, this suffix did not possess the morphological and grammatical productivity inherent in the *-ll-* suffix.

Therefore I suggest that the *-ll-* suffix is indeed of Proto-Indo-European origin, however, not of Vedic Indo-Aryan ancestry. It originated from one or more Old Indo-Aryan lects that were different from Vedic. And it surfaced in the perception of the Indian grammarians around the same time when also other Outer Language features had begun to influence the Middle Indo-Aryan languages (see Part IV). What do we know about this Proto-Indo-European suffix?

The *\*l-* class of Indo-European adjectives was transformed into verbal forms in, basically, Slavic, Armenian and Tocharian.<sup>50</sup> The suffix *\*-lo-* attached to stems that could be nominal, verbal or adjectival; in some cases the suffix had a diminutive or an expressive function, cf. e.g. Old Greek *παχυλοζ* ‘thickish’ which corresponds with OIA *bahulá-* ‘thickish’. In Slavic *l-* participles were combined with the finite forms of the auxiliary verb to create the periphrastic forms of the verb. “... the change from a perfect to a past function is quite a common grammaticalization path” (Igartua 2014: 308) namely by losing the semantic feature of “current relevance of a past action” (ibid.). If the *-ll-* suffix surfaced between the 3<sup>rd</sup> and the 5<sup>th</sup> Century CE in the writings of the Indian grammarians with its various functions outlined above, then it took several more centuries until it developed a past tense function because Southworth notes (2005a: 170) a “... lack of evidence for the *-l-* past ... before about the eighth century CE at the earliest.” This time frame is comparable with the similar developments in Slavic. But now especially important for us is the fact that “... the suffix *\*-lo-* could be extended by means of different preceding vowels, thereby yielding secondary formations...” (Igartua 2014: 306f.). The following forms have been reconstructed for PIE:

*\*-ulo-*, *\*-elo-*, *\*-ilo-*, *\*-ālo-* (< *\*-ah<sub>2</sub>-lo-*), *\*-ē-lo-* (< *\*-eh<sub>1</sub>-lo-*)

Only *\*-ulo-* is found in OIA *bahulá-* (but of course there are more examples for the *-l-* ending in OIA). However, the reconstructed PIE

<sup>50</sup> Tocharian is an extinct ‘kentum’ language formerly spoken in oasis cities on the northern edge of the Tarim Basin (now part of Xinjiang in northwest China). Speakers of Tocharian and speakers of Niya Prakrit, which was spoken on the southern edge of the Tarim Basin, were probably in direct contact.

suffixes resemble strikingly the three MIA suffixes *-illa/alla/ulla-*.<sup>51</sup> In Tocharian, the suffix *\*-lo-* produced verbal adjectives (gerundives), in Tocharian A the suffix was *-l* and in Tocharian B *-lye* and *-lle*. According to Don Ringe (1996: 116), *-lle* developed from PIE *\*-lo-* plus adjective suffix *\*-yo-*. Instead of Pischel's suggestion that the doubling of the lateral is due to a following accent, the MIA forms rather seem to go back like in Tocharian B to *\*-il-ya-*, *\*-al-ya-*, *\*-ul-ya-* i.e. extended by the Sanskrit gerund *-ya-*<sup>52</sup> which makes 'grammatical sense' (regarding phonetics cf. e.g. Pa. *kulla-* 'winnowing basket' < OIA *kulya-* and Pischel § 286: "*lya* wird *lla*").<sup>53</sup> The fact that in case of the three Middle Indo-Aryan suffixes the preceding vowels *-e-* and *-o-* are missing can be explained by two unequally likely reasons: (a) the more unlikely reason is that there was only one 'archimorpheme' *-alla-* whose initial vowel sometimes changed to *-i-* or *-u-*. Such changes are well-known from IA history, but as a result one would expect either geographical or grammatical differences which seem not to exist. Moreover, under such a scenario also development of *-a-* > *-e-* and *-o-* should be expected which is also not the case. Therefore more likely is alternative (b), namely that the lack of *-e-* and *-o-* is an effect of the common Indo-Iranian merger of PIE *\*e* and *\*o* with *a*.

<sup>51</sup> There is no other Prakrit suffix with three different initial vowels which makes the suggested origin of *-illa/alla/ulla-* even more likely.

<sup>52</sup> Strings of grammatical suffixes are found in MIA, and an NIA language like Bangani has a plurality of suffixes that can be decomposed into separate suffixes. Note also that the OIA rule that *-ya-* is to be affixed to verbs with prefixes does not hold good in later OIA (see Whitney 1973: 355).

<sup>53</sup> As much as I can see, Chatterji does not offer a solution for the historical origin of MIA *-illa/alla/ulla-*, Ghatake (1948: 336) makes the unlikely suggestion with question mark for derivation < *-ra* or *-la*, Pischel's suggestion would have led to irregular results (sometimes *-l-* and sometimes *-ll-* aside from the problem that it is generally assumed that MIA had lost the OIA accent), and Southworth appears quite clueless: he mentions Romani examples with *-l-* (all of which, of course, go back to *-t-*), he quotes Geiger who had explained past forms containing *l* in Sinhala from composite verbs built with the light verb *lanu* 'put, place' (all 2005a: 133), and he even considers borrowing from Dravidian (2005a: 150) where he mentions Malayalam *-ulla* used to form attributive adjectives. All this does not work: in case of a Dravidian borrowing because of the retroflex laterals, and also the suggestion with *lanu* is wrong (at least for northern India) for the simple reason that those NIA languages which distinguish phonologically between *l* and *ʃ* (e.g. Bangāṇī) the past *-l* is always dental and thus *must* go back to an older geminate *-ll-*. Therefore I am sure that my above-presented suggestion offers the most convincing solution.

The *l*-suffix is also found with Anatolian Hittite adjectives and nouns, producing sometimes agent nouns through substantivization of adjectives with the meaning ‘pertaining to/who deals with ...’ (Igartua 2014: 307). Compare this with the before quoted example from Pali: *gāmilla-* ‘farmer’; and in some cases Hittite forms are kind of *l*-participles (typologically) closely related to those in Slavic. “This functional development in Hittite seems to anticipate the evolution of *l*-formations in the other Indo-European languages in which participles and even finite verbal forms arose out of adjectival formations ... In Lydian, another Anatolian language, there are infinitival as well as past tense forms in *-l*, some of which are strikingly parallel in their formal structure to their Slavic correlates: cf. for example Lyd.[ian] *esl* ‘was’ and OCS [Old Church Slavic] *bulъ* ‘(has) been’, coming from different roots of the verb ‘to be’” (Igartua 2014: 312). In fact, there is a further parallel to OCS *bulъ* ‘(has) been’ in the West Pahārī variety Kūthalī *bhūlā* ‘was’ (LSI ix,iv: 552), in Bihārī *bhela* ‘became’ (LSI v,ii: 93) and in Dardic Palūla perfective *bhil-u* (m.sg.) and *bhil-a* (m.pl.) (Liljegren 2008: 84, 127, 148). I may refer here to Cardona regarding the relationship of Indo-Iranian with other IE languages. He writes (2003: 20): “The most definitely established and accepted subgroup within Indo-European is Indo-Iranian, a subgroup adjacent to Slavic...” However, the dialectal Indo-Aryan development of the PIE *\*-lo-* suffix, which resembles more that in Tocharian B, is thus probably an independent development.

*Further evidence in support of the first proposition*

It has long-since been known that the Dardic language name Khovar which means ‘Kho language’, and the Nuristani language name Vasi Veri<sup>54</sup> which correspondingly means ‘the Vasi language’ contain the reflex of a Proto-Indo-European verbal root *\*uer-* ‘to speak, talk formally’ (found in English ‘word’) which is not found in Old Indo-Aryan. Thus, Turner reconstructs OIA *\*vari-* ‘speech’ (11327) and he quotes three modern reflexes from Nuristani languages.<sup>55</sup> One reason that this did not raise more curious astonishment may be due to the widespread perception that Nuristani is so much more archaic than the

<sup>54</sup> Same language as Prasun.

<sup>55</sup> He considers it possible that the lemma is etymologically related with OIA *vārṇa* ‘color’ and/or *vāṇá-* ‘voice, music’. None of the suggestions has been accepted by Manfred Mayrhofer (EWA).

rest of Indo-Aryan. Above we have seen that this is questionable. Turner probably did not quote ‘Khowar’ because he must have assumed that ‘-war’ is a borrowing from Nuristani. However, the word is also found in Dardic Kalasha *var* ‘language’ (a borrowing of a borrowing?) and in geographically quite distant Dardic Indus Kohistani as *gošvārī* ‘the Burushaski language’ (Zoller 2005) with *goš-* being a shortening of Burushaski *gušpūr* ‘prince, male member of a Rajah-family’ which itself is a borrowing from Iranian *\*wisya-puōra* ‘son of the community/clan’; *gošvārī* thus means ‘language of the Burush nobles’. The question comes up, is this an accidental isolated case or not? According to the majority view (see also above Morgenstierne’s dictum p. 83), all later reflexes of ultimately Proto-Indo-European-derived words stem from Old Indo-Aryan (with a few exceptions mentioned above). However, if my first proposition is not completely erroneous it has to be shown that indeed a convincing number of Proto-Indo-European-derived words are found in the Outer Languages without known parallels in Old Indo-Aryan. So far I have gathered much more than hundred lemmata, many of them found in more than one language. Here I can present only a small selection of such words whose derivations are more or less straight forward (in not few other cases complex argumentations are required). Here follow the examples:

- Prasun *pul-*, *āpul-* ‘say, speak’ as e.g. in *kuk’a polū* ‘speak loud!’; cf. PIE *\*(s)pel-* ‘to speak loudly, emphatically’; note also with *a-* prefix Greek *apeilēō* ‘hold out in promise or in threat’.
- Kalasha *lep* ‘flat, smashed’ as in *ónḍrak pe átav hav, lep híu* ‘if an egg falls it will become smashed’ and *lep nástan* ‘flat-nosed, squashed-nosed’ — Sindhī *lap* ‘the full of one hand open’ < PIE *\*lēp-*, *lōp-*, *ləp-* ‘be flat, flat; plane, hand, shovel’.
- Kalasha *vi-čái-k* ‘to rest from working’ as in *adhék vičáio krom kári* ‘take a little rest and then work’ with verb root ultimately < PIE *\*k<sup>u</sup>eih<sub>1</sub>-* ‘rest, quiet’. The word has the same prefix as OIA *viśramate* ‘rests’ and it has a ‘Nuristani’ shape but I am not aware of Nuristani parallels and it is missing in OIA. But comparable are Old Church Slavic *pokořǐ* ‘peace,

- quiet, rest’ and *počijō, počiti* ‘to rest’, Old Persian *šiyāti* ‘comfort’, Avestan *šyāta-*, *šāta-* ‘pleased, delighted’.
- Kalasha *drázik* ‘to load (something) onto one’s own back for carrying’ < PIE *\*dherǵh-* ‘bind fast’ (but OIA *DARH* ‘be firm’ with *h* < *\*jh*).
  - Bangānī *lepno* ‘to peel, skin; to snatch’, *lepno* ‘rag, cloth, shred’, *lepiano* ‘to peel, skin; seize s.th.’, *lephro* and Khaśdhārī *lepra* ‘(thin) bark of tree (which can be peeled)’ — perhaps here also Multānī *lāpan* ‘to cut the ears of jawār, bājhrā, and those tall crops the straw of which is not taken to the threshing-floor’ < PIE *\*lep-* ‘peel’ and ‘to sliver’; cf. e.g. Greek *lépo* ‘peel!’
  - Kati *vór* ‘any male relative 2 generations above Ego’, Waigalī *aveli* ‘parent’s mothers agnate’, Kalasha *váva* ‘grandfather (father of father or mother)’ < PIE *\*h<sub>2</sub>eu<sub>h</sub>₂os* ‘grandfather’ (cf. Latin *avus* ‘grandfather’).
  - Bangānī *śopno* or *śopkano* and Deogārī *śepno* and *śopno* all meaning ‘to slurp, swallow (loudly); to harm s.o. magically (as a witch who is swallowing [‘slurping up’] the ‘life-force’ of the victim)’ — Indus Kohistani *šapáiv* ‘to lick up (e.g., a spilled liquid, leftovers)’, Burushaski *šap -it-* ‘to slurp (up) (vulgar)’, Khowar *šruph* ‘slurp’, Kalasha *šurúp kárik* ‘to sip’, and Kashmiri *šrapun* ‘to be digested, be soaked up’ < PIE *\*srebh-* ‘slurp; gulp, ingest noisily’ (cf. e.g. Lithuanian *sriaubiu*, *sriaubti* ‘to slurp’, Hittite *s(a)rap* ‘gulp’). The different forms of the Indo-Aryan words are all regular with regard to their respective historical sound changes (e.g. with regard to stem-final devoicing).

Note: The last lemma PIE *\*srebh-* is somewhat complicated: on the one hand, there is regular Proto-Iranian *\*hrab-* ‘to sip’ with modern reflexes having also been borrowed into Burushaski *huúp -t-* ‘to slurp, devour, inhale (smoke)’ and Sh. *huúp th-* ‘to pull up s.th.’, on the other hand there are Munda Kharia *suru<sup>b</sup>* ‘to sip, suck’ and Santali *si<sup>ru</sup>p* ‘to sip, suck, to suck in audibly’ (and IA Sadani *surp-* ‘to sip’) which are possibly of Austro-Asiatic origin. Proto-Mon-Khmer has practically the same root: *\*srup*, *sruup* ‘to suck, drink’ and even a parallel to the Iranian debuccalization is found in Proto-Palaungic

\**hruup* ‘to drink’. Do we have before us a mingling of two accidentally similar roots from Indo-European and Austro-Asiatic?

Apart from the last ‘lemma’, which is somewhat complicated, I do not know how the above examples could be explained in terms of slight dialect variations in Old Indo-Aryan. Other examples, not quoted here because too complicated to be explained in a few lines, seem to go back to Indo-Iranian with common reflexes in Nuristani, Dardic and other Indo-Aryan languages. In yet other cases it seems that Outer Languages have preserved the Proto-Indo-European meaning of a lemma which already in Old Indo-Aryan is not any more clearly identifiable. Note e.g. OIA *TUL* ‘lift up, weigh’ < PIE \**telh<sub>2</sub>*- ‘raise, lift’ with modern reflexes mostly meaning ‘weigh, balance, scales’ (see CDIAL). This differs significantly e.g. from Bangānī *tùlkeṇṇ* and Deogārī *tulkōṇṇ* and *tulkaṇṇ* (with a *-k-* suffix) all meaning ‘to swell, rise as water, brim over as water’ which have a close semantic parallel in Nuristani Prasun *tol-* ‘to swell, bulge, mass together’ as in *āv tol’ogso* ‘water rose’. Note also that Bangānī *tùlkeṇṇ* with tone contrasts with e.g. *tulṇṇ* ‘to be weighed’ without tone which suggests that the latter form is a later borrowing from Hindi or a related language. So how is it possible that West Pahārī Bangānī and Deogārī can share such an archaic trait with a geographically quite distant language belonging to a different branch of Indo-Iranian?

*Incomplete satemization?*

The following two examples may be suited for a comparison with Balto-Slavic history with regard to possible incomplete satemization:

- PIE \**ghaisos* ‘dart; staff, stick’ is reflected in Bng. *gèsṇ*, *gèsḷ* ‘a stick used for driving cattle and for fighting’,<sup>56</sup> Jaun. *ghesli* ‘a stick for killing’, Deog. *gesri* ‘a stick for fighting’; Rj.mev. *ghesḷo* ‘long stick’ (but OIA *hēsās-* ‘missile, weapon’ < \**jhešas-*). In western Garhwal, the word is also used in hero songs, the so-called *hāruḷ*. Here two lines from such a song about the hero Hāku (Lakshmikant Joshi 2007: 46):

*leuṣi ri ghesli kāṭe muṭeia ri chīṭi*  
*chiṭkāre muṭaia ri lai bheḍa pīṭi*

<sup>56</sup> A Bangānī tone as here on *-è-* is frequently, but not always, the result of loss of aspiration.

‘(Haku) cuts a fighting-stick from a *Leushi* tree (and) a stick  
from a *Mutaia* tree,  
(and) he is killing the sheep with the stick from the *Mutaia* tree’

In another article (Zoller 2016) I discuss the importance of reflexes of \**ǵhaisos* in traditional Indo-European pastoral and martial societies and I refer to the Russian Indologist Yaroslav Vassilkov who says (2012: 165): “The attribute of the hero, common to all branches of the ancient Eurasian tradition, is the *shepherd’s staff* or crook” which is called e.g. in Old Greek tradition *khaĩos* (also transliterated *chaion*) ‘herdsman’s staff’.<sup>57</sup>

- a) Mayrhofer suggests (EWA) that OIA *HARS* ‘get excite, happy’ is reflex of the conjunction of PIE \**ǵ<sup>h</sup>ers-* and \**ǵ<sup>u</sup>ers-* ‘stiffen (of hair), bristle’. The second form is not found in the usual sources, for instance Mallory and Adams (2006) present only the first form. But here relevant is perhaps also PIE \**ǵ<sup>h</sup>eis-* ‘frighten’ which cf. with Bng. *gə̌isɪŋə* ‘to get frightened’, *gə̌rs(ɛ)ŋə* ‘to feel sheer terror, be terrified’ and *gə̌rsɪŋə* ‘to be bristly, spiny (e.g., as pig or porcupine)’ and *gə̌si* ‘shocked, frightened, scared; fright, scare’, Deog. *gə̌isɪŋə* and *gə̌rsɪŋə* ‘to get terribly frightened (with body hair standing upright)’. Whereas the above forms seem to have not undergone satemization, the following words, if they have the same origin, have undergone satemization, however in a ‘Nuristan’ way:

---

<sup>57</sup> The OIA reflex of \**ǵhaisos* is *hə̌sas-* ‘missile, weapon’ but it does not seem to have any associations with martial shepherds and pastoralism, traces of which are still found in the mountains of north-western South Asia. Vassilkov writes about hero stones in India (2011: 198): “The territories with the hero-stones form a kind of belt around the subcontinent. They have something in common: we often find in them cattle-breeding societies with strong vestiges of an archaic social organization and traditions of cattle-raiding. Interestingly, there are no “hero-stones” in Madhyadeśa, i.e. the northern part of Uttar Pradesh, the cradle of the Vedic (Brahminic) civilization.” He suggests further (2011: 199) that the tradition of hero stones “... could possibly represent one of the non-Vedic waves of Aryan migration to India.” Not surprisingly, hero stones and associated hero traditions are very common in many parts of the Himalayas (see Zoller 2007, 2016). Unfortunately, here is not the place to deepen further linguistic and cultural parallels characteristic of parts of the Outer Languages.

b) K. *zarzar*, *zarazar* ‘fear, apprehension’, Ind. *zhár ho-* ‘to get frightened, frighten, startle’ and *zhár karáy̐* ‘to frighten or startle s.o.’. The two Kal. forms *záraš žúni*<sup>58</sup> *hik* ‘to become frightened’ and *zran* ‘afraid’ are allomorphs of a form that must have been very similar to Av. *zarāšiiamna-* (Mallory and Adams *zaršayamna-* [2006: 347]) ‘feathers upright’. The Av. form is directly reflected e.g. in Iranian Wan. *zerāž* ‘rough, stiff, rigid’.

This phenomenon – namely the occasional reflex of palatal PIE \**k* or *ǵ* as velar *k* or *g* instead of palatal *s* or *j* – characterizes a number of words especially in Balto-Slavic, but individual cases seem also to be found in Old Indo-Aryan. Some possible examples from Old Indo-Aryan are: *gnā-* ‘wife’ ~ *jāni-* which reflect PIE \**g<sup>h</sup>neh<sub>2</sub>-* and \**g<sup>h</sup>enh<sub>2</sub>-*, *DEH* ‘smear’ ~ *digdhá-*, *HAN* ‘strike, kill’ ~ intensive *jānganti* ‘strikes, kills’ (reflected in West Pahārī [5081]). The above-quoted examples may suggest, although the evidence is perhaps less plain than in Balto-Slavic, that during the Proto-Aryan phase some lects of the linkage underwent only an incomplete satemization and thus preserved some ‘kentum’ characteristics. With regard to Balto-Slavic, Thomas Gamkrelidze writes (1997: 79): “Certain discrepancies among the individual areas of the *satem* group can be observed in the choice of positions where the opposition of velar and palatovelar phonemes was neutralized.” Gamkrelidze quotes several doublets like Lithuanian *akmuõ*, *ašmuõ* both ‘stone’ (OIA *ásman-*), and he continues “[t]hese pairs reflect, in general form, an ancient Indo-European alternation of palatovelar and velar phonemes within the paradigm under combinatory conditions that cannot be more precisely defined.” Whereas in Balto-Slavic the suggestion that the doublets reflect different dialects has been rejected with the arguments that all doublets have exactly identical meanings, this does not apply to the above examples which definitely are due to language and dialect differences.

---

<sup>58</sup> *žúni* < OIA *yóni-* ‘womb, birthplace, abode’, the phrase thus means literally ‘to be (in the) form (of) fear’.



*Differences in satemization*

Whereas labiovelar Proto-Indo-European \*g<sup>u</sup> is reflected in Old Iranian as the palatal affricate ʃ and palatovelar \*ǵ reflected as the dental sibilant z, in Old Indo-Aryan the respective reflexes are aspirate h (< \*ǵh) and palatal affricate j (see Huld 1997). And whereas Nuristani is known to follow the Iranian type, it has been generally assumed that Dardic has followed the Indo-Aryan type. Both Old Iranian and ‘Old Nuristani’ had two affricate phoneme types ʃ and z (i.e. a palatal order opposing a dental order), but Old Indo-Aryan had only j (i.e. only a palatal order). Obviously, the phoneme z is the result of a depalatalization process. Since, as I have pointed out above, depalatalization is also very common in Dardic, one may wonder about its ‘purpose’. And indeed it seems that also ‘Old Dardic’ had two phonemes j and (d)z. Depalatalization of reflexes of PIE \*ǵ did, however, not occur in Dardic (and probably also in Nuristani) when j- was followed by a [+back] and/or [-low] vowel, in case of coronal consonant harmony (CCH), and in some cases even phonologically unconditioned.<sup>59</sup> An example for non-depalatalization due to CCH is found in Indus Kohistani ʒhuiṭā ‘defiled’ < OIA juṣṭa- ‘remnants of a meal’ where the retroflex ṭ(h) prevented depalatalization.<sup>60</sup> Examples:

Proto-Indo-European	Nuristani	Dardic	Sanskrit
*sk <sup>u</sup> el-		Ind. čhál ‘deceit’	chala-
*g <sup>u</sup> eǵh <sub>3</sub> -		Ind. žīl ‘life’ <sup>61</sup>	jīvala-
*ǵenh <sub>1</sub> -tú-		Ind. zān ‘snake’	jantú-
*g <sup>uh</sup> en-	Dm. žan- ‘kill’ Pr. žon-		HAN
*ǵneh <sub>3</sub> -	Dm. zaan- ‘know’	Shum. zāni <sup>62</sup>	jānāti

<sup>59</sup> Some details are still unclear to me.

<sup>60</sup> The Indus Kohistani form goes back to MIA juṭṭha- and shows aspiration fronting which is a typical feature of Outer Languages and never found in Inner Languages. Note therefore also Panjabi jhūṭā, Dardic Phalura jhuṭá, Bengali jhuṭhā and Deśya Prakrit jhuṭṭha, but Hindi jūṭhā.

<sup>61</sup> I have found not a single case where PIE \*g<sup>u</sup> turned into a dental phoneme in Dardic.

<sup>62</sup> In Kalasha jhónik ‘to know’ there is no depalatalization because of following [+back, -low] vowel (?).

The examples prove the inaccuracy of Morgenstierne's dictum because Dardic shares with Nuristani and Iranian a feature not found in Old Indo-Aryan and its direct daughter languages. The correctness of my above deliberations is also supported by the following fact: we have seen that Old Iranian has a contrast between a palatal affricate  $\check{z}$  and a dental fricative  $z$ . This parallels the fact that in those Dardic languages (and, at least partially in Nuristani languages) which have a similar contrast (to my knowledge found practically everywhere), only the palatal affricate phonemes are phonetically realized as affricates in all word positions whereas the dental and retroflex affricates<sup>63</sup> have a more or less clear proclivity for phonetic fricative pronunciation. This leads to the question of how to explain the presence of the retroflex phoneme  $z$  in Dardic and Nuristani.

The phoneme must have belonged already to 'Old Dardic' and 'Old Nuristani' at the time of Old Indo-Aryan. First argument: Indus Kohistani  $zhár$  'gushing down' – which is only used together with  $vī$  'water' as in  $vī zhār hōt$  'water gushes down' – is connected with Old Indo-Aryan  $jhara-$  'waterfall' and goes back to Pre-Vedic  $gzharati$  'oozes, flows' (Oberlies 1999: 45). Whereas Pre-Vedic  $*z(h)$  had disappeared until the Vedic period, as is well-known, it has been preserved in Dardic.<sup>64</sup> Second argument: It can be assumed that its unvoiced counterpart  $ç$  is of comparable antiquity. This is suggested by Nuristani Prasun  $ṣə$  'cattle' which, I suggest, derives < Rigvedic  $kṣú-$  <  $*pśu-$  < Old Indo-Aryan  $paśú-$  'cattle' (see EWA) with the well-known change of  $kṣ > ç (> ṣ)$ . Also these examples prove the inaccuracy of Morgenstierne's dictum and they are yet another example for lectal differentiations between 'the' Old Indo-Aryan language and other Old Indo-Aryan lects.

*Old and new graphic conceptualizations of the hypothesis*

We have meanwhile come across several features I regard as characteristic for Outer Languages and as confirming the second proposition. They are different from the features usually quoted as

<sup>63</sup> It is obvious that the inherited affricates are phonemically affricates no matter whether they are realized as affricates or fricatives.

<sup>64</sup> I am not aware of a comparable parallel in Nuristani. Together with the more authentic preservation of the Proto-Indo-European-inherited accent in Dardic when compared with Nuristani, this is a strong argumentative package against the elevated linguistic status of Nuristani.

defining South Asia as a linguistic area (*Sprachbund*). The most prominent features are the following: retroflex consonants, echo-word formations, quotative constructions, the so-called absolutive, SOV word order, morphological causatives, a ‘second causative’ construction (see Thomason 2000: 10), and, as a later contact phenomenon, dative subjects (see Hobbs 2016). As I have pointed out above (p. 74), these features are mainly ascribed to contact-induced influence between Dravidian and Indo-Aryan. I do not discuss these features here because they do not reveal anything about the prehistoric linguistic landscape situation before the arrival of the Indo-Aryan speakers and its impact on the Outer Languages. Besides the features already mentioned above, I discuss in my forthcoming book also these: ‘spontaneous’ affricatization of sibilants and conversely also deaffricatization; *r* and *l* metatheses; two variants of epenthesis; vowel length opposition limited to *a* vs. *ā*; past tense forms built with *ta*; the auxiliary *t-* ‘is, was’;<sup>65</sup> a variant of Grassmann’s law;<sup>66</sup> SVO-like syntax,<sup>67</sup> and, of course, very much lexical evidence. Here, that is in the next section, I discuss voice and aspiration fluctuations, loss of initial consonants and the important so-called sesquisyllabic syllables. Since the second proposition proposes solely a temporal model of language immigration, the previous geographical divisions are regarded as inadequate. Compare the language map of Grierson right below with my following.

As can be seen in the map by Grierson, the first visual presentations of the model were language maps with clear geographical delimitation between what was claimed to be Outer and Inner (and Transitional) Languages.

---

<sup>65</sup> Connected with OIA *STHĀ* ‘stand’.

<sup>66</sup> Many Outer Languages allow only one aspirated consonant in a word.

<sup>67</sup> This is geographically limited to some varieties in West Pahārī and Dardic.

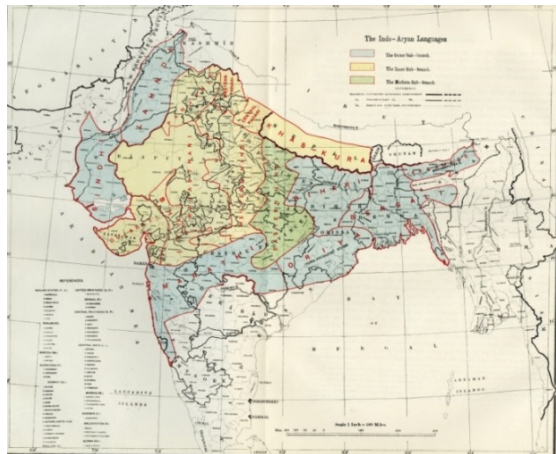


Figure 1. Grierson 1927.

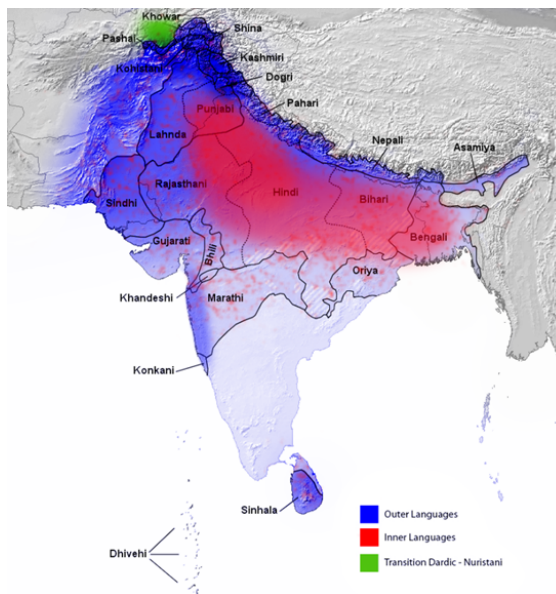
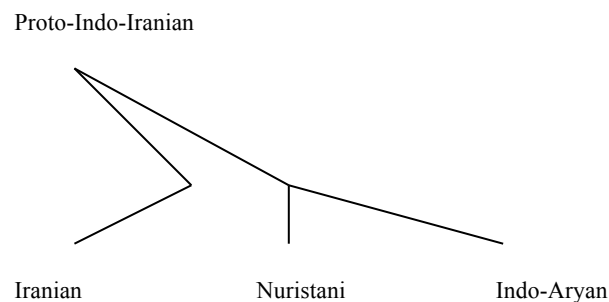


Figure 2. Zoller.

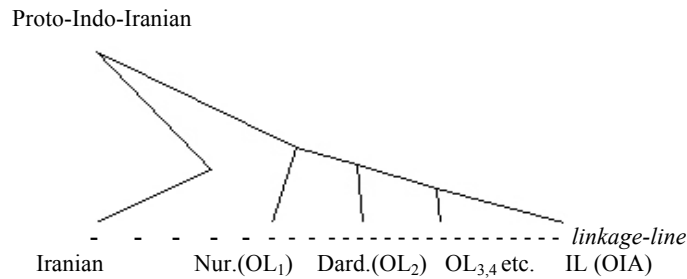
The above map in three colors is to be read: the small green area in the north-west is the homeland of Nuristani and Dardic. The other Outer

Languages do not have an own homeland because they and the Inner Languages are found in the same geographical area of the northern parts of South Asia. This map is neither a historical family tree nor a synchronic two-dimensional language map. It does not claim to display a specific moment in the history from OIA to NIA but endeavors to convey the underlying idea of the relationship between Outer and Inner Languages. Therefore it is three-dimensional. In the blue areas typical Outer Language features dominate and thus reflect the older OIA layer, whereas in the crimson areas typical Inner Language features dominate and thus reflect the younger OIA layer. Changes of shadings of the blue from darker to lighter are meant to suggest a decrease in the number of Outer Language features from (north-)west to (south-)east. One can see areas where both colors interpenetrate and there are squirts of crimson all over the bluish area intended to illustrate the impact of the Inner Language over the whole Outer Language area. Thus, the map is designed to show that there is no clear distinction between Outer and Inner Languages: an individual language is either more Outer and less Inner Language or vice versa, depending on the amount of typical Outer Language features characterizing that individual language. Thus, when the terms Outer and Inner Languages are quoted they should always be understood as abbreviations for more Outer and more Inner Languages. What the map does not show: for instance, Hindi is a typical Inner Language but Braj Bhasha, whose spread is partially identical with that of Hindi, has many Outer Language features; Standard Bengali has only few Outer Language features, but some of its dialects (like that of Chittagong) have quite many, as Assamese has, etc.

Whereas the above map is 'non-historic' the previous historical Indo-Iranian language tree model looked like this (see Degener 2002):



The revised model can be presented thus:



The linkage-line in the above diagram indicates that Nuristani, Dardic and the other Outer Languages are closer connected with Vedic Sanskrit and its descendants (- - -) than with Iranian and its descendants (- - -).

#### Part IV The second proposition

##### *Old Indo Aryan, New Indo-Aryan and Munda/Austro-Asiatic*

Southworth (2005a: 67ff.) refers to F.B.J. Kuiper's claim (on whom more below) that the oldest loanwords in Vedic are of Munda or Austro-Asiatic origin and that therefore there must have been a "... presence of the speakers of these languages in the Panjab as early as the second millennium BCE. The Rigveda alone contains more than 300 such words." However, this position has been sharply criticized by the Munda specialist Toshiki Osada (2006) who quotes his colleague Norman Zide (p. 1): "The identification of words in Indo-Aryan and Dravidian as Munda loans, even when this has been done by careful scholars, is not often convincing, in the light of newer data." He himself sums up (p. 2): "... I generally find that the role of Munda languages for the South Asian linguistic area is overemphasized ... the Munda or Austroasiatic role for convergence in South Asia including loan words is overestimated." He then goes on to review the following words: plough, banana, pepper, lemon, cotton, gourd which, as formerly being seen as the clearest evidence, were widely accepted as being of Munda/Austro-Asiatic origin. Osada arrives at the conclusion (p. 17) that "... Mayrhofer has drastically

changed his position on the Austroasiatic origin for Sanskrit words from KEWA to EWA. When I examined the indexes of KEWA and EWA I saw that the number of entries for Austric<sup>68</sup> are clearly diminishing; e.g., 278 in KEWA but 30 in EWA. That is to say, Mayrhofer apparently recognizes that both Kuiper and Przulski works are not reliable at all. I, as a Munda specialist, agree with his basic position. It is no exaggeration to say that simple calculation has been done; i.e., foreign words in Sanskrit minus Dravidian origins equal to Munda origins.” Finally, he remarks (p. 17f.) that Donegan and Stampe (2004: 27) favor a South Asian origin for Munda/Austro-Asiatic whereas he is convinced of a westward movement from Southeast Asia to South Asia. Not surprisingly, Parpola welcomes this quite authoritative verdict (2015: 165). He takes up some examples from Witzel (see Witzel 2005a: 176-180) – who follows Kuiper’s thesis – in order to question their Austro-Asiatic provenance and offer his own ideas of Proto-Dravidian being the language of the Indus Valley Civilisation (IVC) and thus the appropriate candidate as source of borrowings.<sup>69</sup>

Before the here-sketched background of scholarly contention, which seems to ascribe to Munda a rather moderate role in the linguistic history of South Asia, it is now interesting to see that D.D. Sharma (2003) has claimed that the West Himalayish Tibeto-Burman languages contain a Munda substrate. And that is not all, Roger Blench has recently claimed (2013: 1) the existence of “... an apparent Austroasiatic substrate in Lepcha (Róng) an isolated branch of Tibeto-Burman spoken principally in Sikkim” and Paul Sidwell (2002) has shown an Austro-Asiatic substrate in Chamic<sup>70</sup> which leads him to the conclusion that (2002: 120) “... it is quite likely that much of the Indo-Chinese hinterland now or recently occupied by Bahnaric and Katuic speakers was inhabited by speakers of other M[on-K[hmer] languages.” When I add below additional data from north-

<sup>68</sup> Mayrhofer also used the term ‘Austric’ though it is highly hypothetical.

<sup>69</sup> Parpola’s book, which is a kind of résumé of the decades of his research on this question, contains indeed a number of quite compelling arguments in favor of Dravidian, and in my eyes it is not unlikely that Dravidian was *also* spoken in the IVC, perhaps especially or exclusively by religious, political and economic elites.

<sup>70</sup> “The Chamic languages are a Malayo-Polynesian sub-grouping, with speakers located today in Vietnam, Cambodia, Hainan Island (China) and Sumatra (Aceh Province of Indonesia)” (Sidwell 2002: 113).

western South Asia (West Himalayish and related Zhang Zhung [spoken in Upper Tibet]), north-western Outer Languages, and Burushaski, then the strong suspicion can hardly be put away that large areas of northern India (including the high mountains) and of Southeast Asia were once dominated by a huge family of Austro-Asiatic languages. That this is to a large extent a matter of prehistory has been confirmed by Blench and others before him who point out that the present-day Austro-Asiatic languages present an image of fragmentation. According to him, a major linguistic factor for this fate was this (p. 3): “The notion that Austroasiatic has been fragmented and assimilated by the expansion of Sino-Tibetan and other language phyla is usually accepted.” Before I can present examples of data, I need to explain what all this has to do with my second proposition.

*Consonant fluctuations in Munda*

In Kuiper’s article on consonant variation in Munda (1965), he discusses the notion of ‘sporadic sound laws’, i.e. phonological changes limited to certain words only, which permeate the Munda languages. Kuiper lists the following variations:

- Between voiced and voiceless stops
- *gh/g : kh/k : h : 0*
- *b(h) : h : 0*
- *d(h) : h : 0*
- Between dental and retroflex consonants<sup>71</sup>

For these changes he usually cannot find any explanation (p. 55f. and 59); and “it is impossible to decide with certainty where the domain of variation ends and that of parallel rhyme words derived from etymologically different roots begins” (p. 56). The quoted variations are also typical for Outer Languages but untypical for Old Indo-Aryan and Inner Languages like Hindi, Bhojpuri and Bihari. See the following comparisons:<sup>72</sup>

---

<sup>71</sup> I have left outside a few items which are either not found in Outer Languages or which are not diagnostic, in my eyes, like the dental versus retroflex examples.

<sup>72</sup> I have grouped together Kuiper’s not so systematically ordered types of variation into just two rows.



Variation type	Munda	Outer languages
±voice and ±aspiration	<p><i>baḍaʔj- buḍuʔj</i> ‘bubbling’ = <i>pɔdɔ-pɔdɔ</i> Khas. <i>blup-blup</i> ‘bubbling’ = <i>phlup-phlup</i> Kh. <i>saʔdhay</i> ‘torment’ = H. <i>satānā</i> <i>bhondra</i> ‘large’ = <i>phandra</i> <i>jibir jibir</i> = <i>jhibir jhibir</i> ‘drizzle’ = IA Deog. <i>chirbir chirbir</i> ‘drizzle’</p>	<p>P. <i>bājhli</i> ‘flute’ &lt; OIA <i>vaṃsī</i> Phal. <i>jhāb</i> ‘be quiet’ = H. <i>cup</i> P. <i>ūt</i> and <i>udh</i>, Sh. <i>uś</i> all ‘otter’ &lt; OIA <i>udrā</i> Wg. <i>grop</i> ‘womb’ &lt; OIA <i>gārbha</i> Klm. <i>khum</i> ‘bottom’ &lt; OIA <i>gambhan</i><sup>73</sup></p>
Loss of initial closure or consonant <sup>74</sup>	<p>Kur. <i>kɔn</i> ‘son’ = Sant. <i>hɔn</i> Kur. <i>kākū</i> ‘fish’ = Sant. <i>hako</i> <i>duluduñ</i> ‘type of snake’ = Or. <i>hulhula</i><sup>75</sup> <i>bangam</i> ‘finish’ = <i>hangam</i></p>	<p><b>Deśya Prakrit</b> <i>kaṅkelī</i> = <i>aṅkelī</i> ‘Ashoka tree’ &lt; OIA <i>kaṅkeli-</i> ‘the tree Jonesia Asoka’ <i>caviām</i> = <i>aviām</i> ‘said’ &lt; OIA *<i>cavati</i> ‘says’ (4724) <i>jūā</i> = <i>ūā</i> ‘louse’<sup>76</sup> &lt; OIA <i>yūka-</i> ‘louse’</p> <p><b>Nuristani Prasun</b> <i>iri</i> or <i>īri</i> ‘horse’ &lt; OIA <i>ghoṭa-</i> ‘horse’ <i>vuč</i> ‘five’ &lt; OIA <i>pāñca-</i> ‘five’ <i>uḍyō</i> ‘loom’ &lt; OIA *<i>khadda-</i> ‘pit’ (3790)<sup>77</sup></p>

Regarding the ‘special’ status of *Deśya* Prakrit vis-à-vis the ‘normal’ Prakrits, Bhayani (1988: 150) writes: “Many of these words are familiar to us from Prakrit and Apabhraṃśa literatures. They form a part of the common stock of the literary vocabulary and there is

<sup>73</sup> The change *a > u*, which is very widespread in Outer Languages, has also many parallels in Munda.

<sup>74</sup> This is practically absent in Old Indo-Aryan with the exception of *kaikaya-* = *haihaya-* ‘name of a warrior-tribe’ but which does not look at all ‘Aryan’.

<sup>75</sup> Cf. also Burushaski *tol* ‘snake’ and Aslian Jahai *dadəl* ‘reed snake’.

<sup>76</sup> See 10512 for modern reflexes of the form lacking OIA *y-*.

<sup>77</sup> This refers to such traditional looms where the weaver sits in a hole. This and the preceding example again show the typical Outer Languages change *a > u*.

nothing regional or dialectal about them.” That means that even *deśya* ‘indigenous’ words had undergone a process of koinéization! One can only imagine how widespread in Outer ‘village languages’ loss of word-initial consonants and voice and aspiration fluctuations must have been in the north Indian Middle Ages.

*Sesquisyllabic words*

Sesquisyllabic words are a central characteristic of Austro-Asiatic. Their specific structure has left deep marks in all the different language families (perhaps with the exception of Tibetan) in north-western South Asia and perhaps even in Sindhī (for which see further below), and they are an unmistakable sign for the former presence of Munda/Austro-Asiatic. This crucial feature has slipped the attention of previous scholars in the field of Munda and Indo-Aryan studies. Jenny and Sidwell write (2014: 15): “A characteristic feature of many Austroasiatic languages (although less so especially in Munda, Vietnamese, and Nicobarese) are phonological words that consist of two syllables, whereby an initial unstressed syllable (often called ‘minor syllable’ or ‘presyllable’) is followed by a stressed full syllable (‘main syllable’). This word structure has also been called ‘sesquisyllable’ (‘one-and-a-half’ syllables long) since Matisoff ... [s]esquisyllabicity has also been postulated for Proto-AA”,<sup>78</sup> and (p. 19): “Minor syllables that are created by partial reduplication of the main syllable are also attested in AA.”

The fact that languages in north-west South Asia have many words with sesquisyllabic-like structures does not mean that they can be called sesquisyllabic languages (like many AA languages in Southeast Asia). It is rather a phonotactic feature borrowed from Austro-Asiatic that has influenced these languages which otherwise have also many other types of phonotactic patterns. In case of sesquisyllabic-like bisyllabic words, the first syllable is ‘phonetically subordinate’ and the second syllable is ‘phonetically superordinate’. Or in the words of Becky Ann Butler (2014: 9): “First, prosodic prominence (i.e. stress or tone) must be word-final” and (p. 10): “The second property of sesquisyllables is that non-final syllables are

---

<sup>78</sup> In some of the examples following below, the stress is on the third syllable or on the second syllable followed by a third unstressed syllable, but the mechanism seems to be basically the same as in AA.

phonologically reduced. This includes segmental properties, syllable shape and weight.” Butler’s and my definitions are vaguer than the various standard definitions (see Butler 2014 for different definitions by different authors) which refer e.g. to reduced phoneme inventory (especially a tendency for schwa [ə] as the only vowel) and monomeric V, CV or VC tendency in the subordinate syllable and ‘fully developed’ syllable structure (whatever this means) in the superordinate syllable.

Here for us important is also the following observation by Jenny and Sidwell (2014 20): “... although it has been recognized that there is a tendency for sesquisyllabic words to become monosyllabic ... sesquisyllabic words in AA appear to be diachronically very stable...” This is possibly one condition for the survival of sesquisyllabic structures in north-western South Asia. However, there is probably an additional factor which has supported their survival, namely the presence of stress and pitch accent in the languages of the area. It seems that in the discussions about sesquisyllabic structures in Austro-Asiatic languages the role of pitch accent is not of special relevance. But an important characteristic of the vast majority of the Nuristan, Dard and East Iranian languages, and Burushaski, is that they are pitch accent languages. Apart from the case of Burushaski, the pitch accent is historically frequently an inheritance of the Vedic accent<sup>79</sup> (Zoller forthcoming) but the system was generalized and includes now also all non-inherited words (i.e. also words borrowed from English as some examples below show). Even though the position of the Vedic accent was free, as is well-known, my preliminary statistical surveys e.g. of Nuristani Prasun and Dardic Indus Kohistani clearly show that in case of bisyllabic words the pitch accent is much more often on the second than on the first syllable. My guess is, although more research will be necessary, that a phonetic conspiracy took place between the borrowed Austro-Asiatic sesquisyllabic structures and the inherited Vedic accent which secured the survival of sesquisyllabic-like word patterns till today. Several examples below also show that originally non-sesquisyllabic words got transformed into sesquisyllabic-like words. I use the term

---

<sup>79</sup> In case of East Iranian, e.g. Pashto, one cannot, of course, speak of a Vedic accent. But even though the conditions regarding inheritance of the Old Iranian accent are less clear than e.g. in Dardic, there do exist some pretty clear examples.

sesquisyllabic-like because the languages in the north-west of South Asia have developed various phonetic strategies for marking the phonetically superordinate syllable in a sesquisyllabic word which are probably not current (or at least common) in Austro-Asiatic. I include here also the phenomenon of syncope of initial syllables which is an indirect way of identifying the former superordinate syllable. Here is a (perhaps incomplete) list of features for marking superordinate syllables:

- (a) Syncope of initial syllable
- (b) Pitch accent on superordinate syllable
- (c) The north-western pitch accent corresponds frequently with Munda/Austo-Asiatic glottal stop (or checked consonant) or, more rarely, with a raised vowel (e.g. *a* > *o*)
- (d) Change of *a* to *ē* (only in Baṅgānī, but similar to the Munda *a* > *ɔ*, *o* change) in superordinate syllable.

Examples:

- (a) Wg. *pšik* ‘cat’ < OIA \**puśsi-* (8298); Him. *rgāṇu* ‘to color’ (cf. H. *raṅgānā*); Mult. of Afghanistan *prešān* either ‘operation’ (← English) or ‘troubled’ (← Persian *parešān*),<sup>80</sup> compare the latter variant with Kal. *periśān* and Ind. *parēśān* both ‘worried’ where the Persian word did not undergo syncope but got an accent attached to its superordinate syllable in the borrowing process.<sup>81</sup>
- (b) Ind. *ektār* ‘an actor (on stage)’ (← Eng.) and *kažāl* ‘collyrium’,<sup>82</sup> < OIA *kajjala-* ‘lamp-black’ (2622) (but H.

<sup>80</sup> The phenomenon is not found in the homeland of Multānī in Pakistan. Multānī in Afghanistan is mainly spoken in Kabul, Kandahar and Jalalabad by Hindus and Sikhs (though most of them meanwhile have fled the country because of growing Islamic fundamentalism). The dialect has been considerably influenced by surrounding Pashto. I have heard the two quoted words and other structurally similar words in quite many talks from speakers now living in Germany.

<sup>81</sup> That this phenomenon really has nothing to do with vowel length can be shown e.g. with this example of the borrowing reflexes of Ar. *naẓar* ‘sight’: Dardic Kal. *nizēr* and Ind. *nazār*, Munda Sant. *najer*.

<sup>82</sup> Practically the same pronunciation is widespread all over the north-west and includes also Pashto. The word is possibly of Munda/Austo-Asiatic origin because of Sanskrit lexicographic *ajjhala-* ‘coal’ with loss of initial consonant; note also Ho

- kājal*); Pr. *üč'ū* 'a tear' < pre-OIA \**áčru-* (cf. OIA *ásru-* and note shift of accent).
- (c) Bur. *ḍaḍaáko* 'severe labor pains; agony, death throes' — Sant. *ḍhaṛaʔk ḍhoṛoʔk* 'jerkingly, joltingly, limpingly; jolt, limp';<sup>83</sup> Bur. *ḡaḍaḡaḍáp* 'sound of horse's hooves', Sh. *gaṛáp* 'ditto', Ind. *gáṛáp-gaṛáp<sup>h</sup>* 'sound of clatter (of horses)' — Sant. *kaṭaʔp kaṭaʔp* 'clattering (sound of hoofed animals walking; of wooden clogs)'.  
 (d) Bng. *agēḷ* 'door latch' < OIA \**argaḍa-* (629), *bamēṇ* 'Brahmin' < OIA *brāhmaṇá-* (9327), *gṛēṇ* 'solar or lunar eclipse' < OIA *gráhaṇa-* (4364), *ṭimaṭēr* 'tomato' (also Ind. *ṭamāṭár* but H. *ṭamāṭar*) ← English.

Note: With regard to Bur. *ḡaḍaḡaḍáp* 'sound of horse's hooves', a reviewer of the present paper has pointed out that the 'clattering' words are ideophones which resemble e.g. the *kantapper kantapper* ideophone found in a Grimm's fairy tale that is related with the English tale of the *Gingerbread Man* but in which a pancake is 'cantering' along its way. The objection seems to imply that ideophones are not very useful for etymological comparisons. I will deal with this question in detail in my forthcoming publication where I present arguments to show that this conjecture is widespread but not well-founded. Here just the following points: Ideophones are understood to evoke an idea in sound, e.g. movement, color, shape, sound, or action. They are understood as realizations of elementary sound gestures, whereby it remains unclear what 'elementary' means. For instance in Bengali, all following words mean 'to hum': *guñjarā*, *gunaguna*, *jhañkāra*, *ḍāka*, *bhanabhana*, *bhāmjā*, *sura*. But Palaungic Riang has *ṛṛṛ*, Katic Ngeq *parwew*, etc. And several of the Bengali ideophones have IA etymologies, which means, they are not beyond the reach of historical reconstruction. Another obvious characteristic

---

*kajr* 'lampblack' suggesting second vowel as marked ( Sant. *kajal*, *kajar* 'collyrium' is, on the other hand, apparently not sesquisyllabic) and PMK \**kcaas*, *kcas*, *kcah* 'charcoal', Katuic Ngeq *kajah* 'charcoal', Monic Nyah Kur *kəcáh* 'charcoal'.

<sup>83</sup> There does not seem to exist a PMU or PMK reconstruction for this lemma. But since the Burushaski word shows partial reduplication resulting in a subordinate syllable and since both Burushaski and Santali have stress respectively checked pronunciation with the superordinate syllable (the second in case of Burushaski) it is clear that the words are of Munda origin.

of many ideophones is reduplication. The *garáp* lemma has the additional feature of accent on the superordinate syllable. It is thus sesquisyllabic. I found association of the sound *garáp* with the gait of a horse or hooved animal only in Burushaski, Dardic and Munda. And even though one can always overlook something, I did not find convincing parallels in Mon-Khmer, Dravidian, Tibeto-Burman or Iranian. The somewhat similar ideophone M. *gaḍagaḍañēm* ‘to clatter’ seems not to be associated with horse gait. Munda languages are well-known for their enormous wealth in onomatopoeia, ideophones, etc. and it is standing to reason to assume that at the time of the arrival of Old Indo-Aryan in South Asia, the prehistoric linguistic landscape of North India was also characterized by an abundance of onomatopoeic, ideophonic etc. modes of word formation. Here follows a small example of how this linguistic ‘pre-giveness’ may have affected further developments of Indo-Aryan. Whereas *garáp* has a sesquisyllabic structure CVCVC, Marāṭhī *gaḍagaḍa-* does not have this; there is instead CVCV. Nevertheless, that sesquisyllabic word structures, which are a characteristic of Austro-Asiatic, must have had their impact on Indo-Aryan also in the field of ideophones, is demonstrated by the following example: Hindi and Panjābī *dapaṭ* mean also ‘gallop’ besides ‘rush; attack (commonly used in compounds with *daur-* ‘run’)’ but, of course, their second syllable is not superordinate. They have a morphological near-parallel in Prasun *dəp'ak*, *ḍəp'ak* ‘hustling with force and vehemence’, here with sesquisyllabic structure because of the accent. Hindi *dapaṭ* is derived by Platts < Pk. *dabaḍī*<sup>84</sup> < OIA *DRAV* ‘run, hasten’ (cf. also Pa. *davo-* ‘quick motion’, Pk. *davadavā-* ‘veg vālī gati – swift motion’ [Sheth: *A comprehensive Prakrit-Hindi dictionary*] and Old G. *davadavāe* ‘with speed’ [6623 *dravā-* ‘quick motion’], and S. *drokaṇu* ‘to gallop’ [6624 *drāvati* ‘runs’]).<sup>85</sup> The Hindi word is not found in the CDIAL because Turner must have thought that Platts’ derivation < OIA *drava-* + *-ṭa-* could not be correct because of the *-p-* instead of expected *-b-*. But we have seen above that both in Munda and in different Outer Languages voice fluctuations are quite common (cf. e.g. Kt. *tapip* ‘doctor’ ← Ar. *ṭabīb*). Therefore the Hindi and Panjābī

<sup>84</sup> But I cannot locate this form in my Prakrit dictionary.

<sup>85</sup> The dental-retroflex alternation in Pr. *dəp'ak*, *ḍəp'ak* is, I think, not a problem because there are many cases found in Outer Languages for OIA *dr* > *ḍ*. Thus also Pr. *(v)uṭ'us* ‘avalanche’ < OIA *utrāsa-* ‘fear, terror’ (see CDIAL 6013).

word must be a borrowing from an Outer Language – cf. here again Prasun *dəp'ak*, *dəp'ak* – which had used an Indo-Aryan lexeme to which an ‘ideophoneme’ *-t*<sup>86</sup> was suffixed in order to form an ideophone based on the very common Munda/ Austro-Asiatic morphological pattern of sesquisyllabicity. The morphological relationship between Bur. *gəḍəgəḍəp* and M. *gəḍəgəḍəṇēm* parallels the morphological relationship between Prasun *dəp'ak*, *dəp'ak* and Prasun *dəbədəbə* ‘rumbling’ with regard to ± sesquisyllabicity. The latter word appears in the text published by Buddruss and Degener (2016) together with a verb of motion: *dəbədəbə tī am'aso* ‘(he) comes rumbling’. Tellingly, Prasun *dəbədəbə* has also a sesquisyllabic-like allomorph *dəbədəb'i* ‘rumbling’ which also qualifies in the text a verb of motion. We thus observe that an IA verb of motion (*DRU*) has been ‘ideophonized’ by using a Munda/Austro-Asiatic pattern of sesquisyllabicity created with the help of ‘ideophoneme’ suffixes so that it then can qualify other verbs of motion. The remaining question is, whether a similar historical development can be discovered in case of Santali *kaṭa'p kaṭa'p* ‘clattering’. I provisionally suggest derivation < Proto-Kherwarian (North.Munda) \**kaṭa* ‘leg’ (see SEAlang *Munda Languages Project*) which is, so to say, a noun of motion, and which is typically used e.g. in Sant. *məɾəm kaṭa* ‘goat’s trotters’, *sukri kaṭa* ‘pig’s trotters’, both of which are hooped animals. So there can hardly be any more a doubt that the Bur., Sh. and Ind. words are direct borrowings from Munda/Austro-Asiatic. Last of all, this is seemingly further supported by Bur. *gaṭál* ‘by foot’ as in *gaṭál bulá* ‘polo by foot (played by children)’ and *kaṭál* ‘on foot’ (Willson), which also seems to be a borrowing of Munda *kaṭa* ‘leg’, and where also the second word *bulá* shows the typical tendency in Burushaski for voicing of unvoiced stops in borrowings.

#### *Additional lexical evidence*

The equations with languages in north-west South Asia can be divided into the following three types: (i) parallels with Munda without Proto-Munda reconstruction;<sup>87</sup> (ii) parallels with Munda with Proto-

<sup>86</sup> Other ‘ideophonemes’ found in the here-discussed words are *-p* and *-k*.

<sup>87</sup> Parallels of this group are only clearly of Munda origin if they display a sesquisyllabic structure. Otherwise they may be simply of unknown origin in which case I call them “North Indian.”

Munda/Proto-Austro-Asiatic reconstruction; (iii) parallels with Austro-Asiatic languages (and no parallels in Munda) with or without Proto-Mon-Khmer reconstruction. The following examples are classified according to the following principle: (a) parallels only found in Burushaski; (b) parallels only found in West Himalayish including Zhang Zhung language; (c) parallels found in other languages of west and north-west South Asia and sometimes including Burushaski; (d) Sindhi and Munda.<sup>88</sup>

(a) Parallels only found in Burushaski

- i. Bur. *hará-* ‘to pee, urinate’ — Kh. *haḍa* ‘to urinate; urin’, Sant. *aḍo*, Mu. *ḍḍḍ, aṛu* ‘to urinate’, etc., see Pinnow (1959: 153).
- ii. Bur. *ha* ‘house’ — Kh. *hoʔ* ‘house’ (also without initial aspiration *oʔ*) — the lemma seems to be of PAA origin, cf. e.g. PMK *\*[j]aa[ʔ]* ‘house’, Bahnaric Mnong *hih* ‘house’, Mon *haeʔ* ‘house’.
- iii. Bur. *čočoq -i-* ‘to pound, crush with a stone; to castrate’ — Sant. *čoʔk čoʔk* ‘the sound heard when rice is stamped in a *dhiñki*’ (husking machine) — the lemma may have further AA parallels, cf. PMK *\*[k]bok* ‘to pound; mortar’, Katuic Ngeq *co:k* ‘to pound (in small motor)’ (sic) and Kui *cuʔ* ‘to pound (vegetables, fruits) with mortar and pestle’.

(b) Parallels only found in West Himalayish and Zhang Zhung language

- i. Zz. *mang* ‘red’, West Himalayish Darma *mangnu* ‘red’ and Rp. *mənd* ‘red’ — cf. Proto-South-Bahnaric *\*bro:ŋ* and Proto-Bahnaric *\*bro:ŋ* ‘red’, Bahnaric Mnong *mbro:ŋ* ‘red’,

---

<sup>88</sup> As in the previous section, only a few of the actually found parallels are presented here. The special relationship between Sindhi and Munda consists in the fact that in a few cases a Sindhi word-medial implosive corresponds in the Munda parallel with a glottal stop. I have described the details in Zoller 2016.



Bahnaric Halang *məhaŋ* ‘red pepper’, Monic Nyah Kur *ml̩ɔŋ* ‘bright red’.<sup>89</sup>

- ii. Zz. *sam* ‘cold’ — PMK *\*ksaam* ‘(to catch) cold’, Khas. *səm sàm* ‘bitter (cold)’.
- iii. Zz. *rtsa* ‘a cubit, the distance from the elbow to the tip of the middle finger’ — cf. Palaungic Danaw *ʔs<sup>h</sup>aʔ<sup>l</sup>* ‘cubit’.

(c) Parallels found in other languages of north-west South Asia

- i. Bur. *ćamáat̩ ét-* ‘to bestir oneself, make an effort, endeavor’, Sh. *ćamáat̩* ‘ditto’, Ind. *ćamāṭ̩<sup>h</sup>* ‘effort’ — Khmer *camʔaət* ‘to try to reach (something), make an effort to stretch up in order to see better; to try to look taller.’
- ii. Ind. *khin* ‘a blanket made of rags’ (formerly worn in winter) also in *khinpōṣ* ‘the black dress of a Faqir’ with second element ← Pers. *pošāk*, K. *khüñ<sup>u</sup>* ‘a kind of warm woolen blanket’ — cf. Bahnar *khan* ‘blanket, especially the bahnar type’.
- iii. Pr. *√kud* ‘to vomit’<sup>90</sup> — Surin Khmer *kʔv:t* ‘to vomit, throw up, puke’, Khmer *kʔuət* ‘to vomit’.

(d) Sindhī, West Pahārī, West Himalayish and Munda<sup>91</sup>

- i. Ralf Turner has shown (1924) that – put in a simplified way – in Sindhī inherited voiced initial consonants changed into implosives (injectives) whereas the same happened word-medially with voiced double consonants inherited from Prakrit which themselves derived from Sanskrit consonant clusters:<sup>92</sup> *ḡambhīr<sup>u</sup>* ‘sedate’ < OIA *gambhīrā-* ‘deep’ (4031), *ubāraṇu* ‘to save’ < Pk. *uvvārēi* ‘releases’ < OIA *\*udvārayati* ‘opens’ (2082). Some Indo-Europeanists tried to show that reconstructed PIE glottalized stops are reflected in Sindhī

<sup>89</sup> In my eyes, the Zz. and Darma forms are clearly of Austro-Asiatic origin but note also the unconvincing attempt for a Tibeto-Burman etymology by James Matisoff (2001: 15); STEDT database mentions the form but without etymological suggestion.

<sup>90</sup> According to Degener (p.c.), the basic meaning seems to be ‘squirt’ which would not match completely with the AA forms.

<sup>91</sup> The following section is also found in Zoller in press.

<sup>92</sup> There are also a few exceptions.

implosives. However, this is very unlikely for good reasons named by Kümmel (2012). However, besides the many IA inherited words displaying this sound change in Sindhī, there is also a small number of non-inherited words (I have so far collected only a handful) with word-medial implosive corresponding with a Munda glottal stop. Examples: Sindhī *aḍanu* ‘to build’ < OIA \**aḍd-* ‘obstruct, stop’ (mentioned by Turner sub 188) — Kharia *aḍe* ‘to stay, stand firm; block someone’s way’, Sindhī *guḍanu* ‘to pound, thrash’ < OIA \**gudd-* ‘dig’ (3934.6) — Bōnda *gu?* ‘to dig (earth, etc.)’. It seems that the prolonged delay in the release of the double consonant facilitated the development of implosives and glottal stops. Both are also articulatory similar in that for the articulation of an implosive a glottalic ingressive airstream is required. Note also that very many Austro-Asiatic languages have glottal stop phonemes, and also implosives are a feature of many languages of Mainland Southeast Asia (Jenny and Sidwell 2014: 23) even though they have largely disappeared from Munda. Yet, I think it is worth considering that the historical development of the Sindhī implosives was perhaps influenced by Austro-Asiatic languages once existing in its vicinity which still had implosives in their phoneme inventories. Conversely, it seems also possible that word-medial double consonants changed in Proto-Sindhī first into glottal stops and only later into implosives also because glottal stops and related phonetic phenomena like checked consonants or creaky voice are quite widespread in Outer Languages of north-western South Asia (Zoller forthcoming).

- ii. West Himalayish Kinnauri *gaḍ* ‘rivulet’ (Sharma 2003: 14) has an almost exact phonetic parallel in Munda Kharia *gaḍha* ‘river’ (Pinnow 1959) respectively *gaḍha dhoḥa* ‘ravine; ravine in which a river flows; stream, brook’ (Peterson 2009) and further corresponds with the West Pahārī varieties Kōṭgarhī *gāhr* ‘brook’ and Inner Sirājī *gāhd* ‘depth’ (see Hendriksen 1976: 39). They are perhaps connected with OIA \**gaḍḍa-* ‘hole, pit’ (3981) but Pinnow (1959: 351) considers the possibility of contamination of two different lemmata ‘river’ and ‘hole’. In any case, he regards Kharia *gaḍha* to be a genuine Munda word. Glottal stop (respectively checked

consonant) and aspiration stand phonetically in a mirror image relation to each other with regard to the laryngeal features ‘space between the focal cords’ and ‘tension in the folds’: aspiration has the features [-constricted] [+spread] and glottal(ized) articulation has [+constricted] [-spread]. In the West Pahārī variety between the towns of Jubbal and Shimla one finds *gōʔro* ‘horse’ (Hendriksen 1986: 23f.) which compares with *gohā* ‘horse’ in the Chinali variety of West Pahārī. Both words derive < OIA *ghoṭaka-* ‘horse’ (4516) and both display right-shift of the initial aspiration which is a common phenomenon in a number of West Pahārī varieties. The two examples quoted here indicate an occasional oscillation of *ʔ* ↔ *h* in West Pahārī. This can be compared with the above Sindhī examples and thus again old Austro-Asiatic influence as ultimate cause seems possible (which is anyway likely because of the form of the Kinnauri word).

### Conclusions

I have presented a series of different, yet interrelated arguments that in my opinion confirm the linguistic reality of the old hypothesis of Outer and Inner Languages, however with arguments and linguistic data that differ notably from previous attempts. I have shown that at the time of Old Indo-Aryan there must have existed a linkage of lects, with Vedic just one of them. These lectal differentiations seem to suggest that the standard model of the three branches of Indo-Iranian is in need of a revision. Their existence also supports the idea of the earlier immigration of the ancestor(s) of the Outer Language which led to a strong encounter with Munda/Austro-Asiatic languages (but to a weak encounter in case of Vedic and Classical Sanskrit) which must have dominated the prehistoric linguistic area of northern India. This dominance must have extended far into prehistory because of the many parallels in the language isolate Burushaski.

### General abbreviations

< historically deriving from

> historically developing into

→ borrowed from another language

← borrowed into another language

CDIAL *A Comparative Dictionary of the Indo-Aryan Languages*  
(Turner)

IPA International Phonetic Alphabet

IVC Indus Valley Civilization

### Languages and language abbreviations<sup>93</sup>

AA – Austro-Asiatic

Ar. Arabic

Aslian – Austro-Asiatic: spoken on Malay Peninsula

Av. Avestan

Bahnar – Austro-Asiatic: spoken in southern Viet Nam

Bahnaric – Austro-Asiatic: the languages of this group are spoken in Vietnam, Cambodia, and Laos

Bahnaric Halang – Austro-Asiatic: spoken in the southern Laotian province of Attapu

Bahnaric Mnong – Austro-Asiatic: spoken in Vietnam and Cambodia

Bng. – Baṅgāṇī (West Pahārī)

Bonda – Munda

Bur. – Burushaski (isolate?)

Chinali – West Pahārī

Deog. – Deogārī (West Pahārī)

Garh. – Gaṛhwālī

Gta? – Munda

Him. – Himachali (West Pahārī)<sup>94</sup>

Ho – Munda

IA – Indo-Aryan

IL – Inner Languages

Ind. – Indus Kohistani (Dardic)

---

<sup>93</sup> Mon-Khmer languages are left without abbreviations as they are less familiar to South Asianists than languages from South Asia.

<sup>94</sup> Actually cover term for the West Pahārī varieties spoken in Himachal Pradesh.

- Inner Sirājī – West Pahārī  
 K. – Kashmiri  
 Kal. – Kalasha (Dardic)  
 Katuic – Austro-Asiatic: Katuic languages are spoken in the borderlands of Thailand, Cambodia, Laos, Vietnam  
 Katuic Ngeq – Austro-Asiatic: spoken in Laos  
 Kh. – Kharia (Munda)  
 Khas. – Khasic, Khasian Austro-Asiatic: group of languages in Meghalaya and surroundings  
 Khmer – Austro-Asiatic: spoken in Cambodia  
 Kt. – Kati (Nuristani)  
 Kṭg. – Kōṭgarhī variety of West Pahārī  
 Kur. – Kurku (Munda)  
 M. – Marāṭhī  
 Mon – Austro-Asiatic: spoken in Myanmar and Thailand  
 Monic Nyah Kur – Austro-Asiatic: spoken in north-eastern Thailand  
 MIA – Middle Indo-Aryan  
 Mu. – Mundari (Munda)  
 Mult. – Multānī  
 NIA – New Indo-Aryan  
 Nur. – Nuristani  
 OIA – Old Indo-Aryan  
 OIr. – Old Iranian  
 OL – Outer Languages  
 P. – Panjābī  
 PAA – Proto-Austro-Asiatic<sup>95</sup>  
 Palaungic – Austro-Asiatic: spoken in mountainous areas of Myanmar, southern Yunnan Province (China), and northern Thailand  
 Palaungic Danaw – Austro-Asiatic: spoken in Myanmar  
 PIE – Proto-Indo-European  
 PII – Proto-Indo-Iranian  
 PMK – Proto-Mon-Khmer  
 PMU – Proto-Munda  
 Pr. – Prasun (Nuristani)  
 Rj.mev. – the Mēvātī dialect of Rājasthānī  
 Rp. – Rañ-pō bhāsa (West Himalayish)

---

<sup>95</sup> PAA comprises PMK plus PMU but some authors make a difference between Munda and Austro-Asiatic (Mon-Khmer).

S. – Sindhī  
 Sant. – Santali (Munda)  
 Sh. – Shina (Dardic)  
 So. – Sora (Munda)  
 Surin Khmer – Austro-Asiatic: a variety of Khmer spoken in north-eastern Thailand  
 Wan. – Wanetsi (Iranian)  
 Wg. – Waigali (Nuristani)  
 Werchikwar dialect of Burushaski  
 Zz. – Zhang Zhung an extinct Tibeto-Burman language formerly spoken in Upper Tibet

### Select Literature

- Anderson, Gregory D.S. 2004. Advances in proto-Munda reconstruction. In: *Mon-Khmer Studies: A Journal of Southeast Asian Linguistics and Languages* 34: 159–184.
- (ed.). 2008. *The Munda languages*. London and New York: Routledge.
- Arsenault, Paul Edmont. 2012. *Retroflex Consonant Harmony in South Asia*. A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy, Department of Linguistics, University of Toronto. <https://tspace.library.utoronto.ca/handle/1807/33911>
- Auer, Peter. 2001. Silben- und akzentzählende Sprachen. In: Martin Haspelmath et al. (eds.), *Language typology and language universals. An international handbook*. Bd. 2; pp. 1391–1399. Berlin/New York: Mouton de Gruyter.
- Berger, Hermann. 1998. *Die Burushaski-Sprache von Hunza und Nager; Teil III: Wörterbuch Burushaski-Deutsch, Deutsch-Burushaski*. (Neuindische Studien 13). Wiesbaden: Harrassowitz Verlag.
- 2008. *Beiträge zur historischen Laut- und Formenlehre des Burushaski* (Neuindische Studien 15). Wiesbaden: Harrassowitz Verlag.

- Bhayani, H. C. 1988. *Studies in Deśya Prakrit*. Ahmedabad: Kalikāla Sarvajña Śrī Hemacandrācārya Navam Janma Śatābdī Smṛti Śikṣaṇ Saṁskār Nidhi.
- Blench, Roger. 2013. Rongic: A vanished branch of Austroasiatic. [https://www.academia.edu/5562335/Rongic\\_a\\_vanished\\_branch\\_of\\_Austroasiatic](https://www.academia.edu/5562335/Rongic_a_vanished_branch_of_Austroasiatic)
- Bodding, Paul Olaf. 1932-1936. *A Santal dictionary* (7 vols.). Oslo: Dybwad.
- Buddruss, Georg and Almuth Degener (Hrsg.) 2016. *Materialien zur Prasun-Sprache des afghanischen Hindukusch. Teil I: Texte und Glossar*. Harvard: Harvard University Department of South Asian Studies.
- Butler, Becky Ann. 2014. *Deconstructing the Southeast Asian sesquisyllable: a gestural account*. PhD dissertation: Cornell University. <https://ecommons.cornell.edu/handle/1813/36006>
- Cardona, George and Dhanesh Jain. 2003. *The Indo-Aryan languages*. London: Routledge.
- CDIAL Turner, Ralph L. 1966. *A Comparative Dictionary of the Indo-Aryan Languages*. London: School of Oriental and African Studies.
- Duden 2013. *Das Herkunftswörterbuch: Etymologie der deutschen Sprache*. Mannheim.
- EWA: Mayrhofer, Manfred. 1986-2001. *Etymologisches Wörterbuch des Altindoarischen*. 3 Bde. Heidelberg: Winter.
- Chatterji, S. K. 1926. *Origin and Development of the Bengali language* (in three volumes). Calcutta: University of Calcutta.
- Degener, Almuth. 1998. *Die Sprache von Nisheygram im afghanischen Hindukusch*. Wiesbaden: Harrassowitz Verlag.
- 2002. The Nuristani languages. In: *Proceedings of the British Academy*, 116: 103–117.
- Donegan, Patricia Jane and David Stampe. 2002. South-East Asian Features in the Munda Languages: Evidence for the Analytic-to-Synthetic Drift of Munda. In: *Proceedings of the Twenty-*

*Eighth Annual Meeting of the Berkeley Linguistics Society: Special Session on Tibeto-Burman and Southeast Asian Linguistics*, pp. 111–120.  
<http://journals.linguisticsociety.org/proceedings/index.php/BLS/article/viewFile/1041/825>

- Elfenbein, Josef. 1997. Pashto phonology. In: Kaye, Alan S. (ed.), *Phonologies of Asia and Africa*, pp. 733–760. Winona Lake, Indiana: Eisenbrauns.
- Gamkrelidze, Thomas V. 1995. A relative chronology of the shifts of the three stop series in Indo-European. In: Douglas Q. Adams (Ed.), *Festschrift for Eric P. Hamp. Vol. I. Journal of Indo-European Studies Monograph Number 23*: 67–82.
- Grierson, George Abraham (ed.) 1903. *Linguistic Survey of India, Vol. V, Part II: Indo-Aryan family: Eastern group: specimens of the Bihārī and Oriyā languages*. Calcutta: Government of India, Central Publication Branch.
- 1907. *Linguistic Survey of India, Vol. IX, Part IV: Indo-Aryan family: Central group: specimens of the Pahārī languages and Gujurī*. Calcutta: Government of India, Central Publication Branch.
- 1927. *Linguistic Survey of India, Vol. I, Part I: Introductory*. Calcutta: Government of India, Central Publication Branch.
- Hendriksen, Hans. 1976. *Himachali studies; I. Vocabulary*. Det Kongelige Danske Videnskabernes Selskab, Historisk-filosofiske Meddelelser 48,1. Kommissionær: Munksgaard: København.
- 1986. *Himachali studies; III. Grammar*. Det Kongelige Danske Videnskabernes Selskab, Historisk-filosofiske Meddelelser 48,3. Kommissionær: Munksgaard: København.
- Hobbs, David. 2016. *Dravidian influence on Indo-Aryan: The case of the Dative-Subject construction*. Undergraduate Thesis: The Ohio State University.  
[https://kb.osu.edu/dspace/bitstream/handle/1811/76784/1/Thesis\\_final\\_draft.pdf](https://kb.osu.edu/dspace/bitstream/handle/1811/76784/1/Thesis_final_draft.pdf)



- Hoernle, August Friedrich Rudolf. 1880 (1975). *A comparative grammar of the Gaudian languages: With special reference to the Eastern Hindi accompanied by a language-map and a table of alphabets*. London: Trübner.
- Hoffmann, Karl und Bernhard Forssman. 1996. *Avestische Laut- und Flexionslehre*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- Huld, Martin E. 1997. Satəm, Centum and Hokum. In: Douglas Q. Adams (Ed.), *Festschrift for Eric P. Hamp. Vol. I. Journal of Indo-European Studies Monograph Number 23*: 115–138.
- Igartua, Iván. 2014. The Indo-European adjective class with the suffix \*-lo and its development in Slavic. In: *The Journal of Indo-European Studies*, 42,3-4: 302–331.
- Jenny, Mathias and Paul Sidwell (eds.). 2014. *The handbook of the Austroasiatic languages* (two vols.). Leiden, Boston: Brill.
- Joshi, Lakshmikant. 2007. *Hārul. Jaunsār-Bāvar ke paurāṇik lokgīt*. Dehradun: Vinsar Publishing Company.
- Krishnamurti, Bhadiraju. 2003. *The Dravidian languages*. Cambridge Language Surveys. Cambridge: Cambridge University Press.
- Kuiper, F.B.J. 1948. *Proto-Munda words in Sanskrit*. Amsterdam: N.V. Noord-Hollandsche Uitgevers Maatschappij.
- 1965. Consonant variation in Munda. In: *Lingua* 14: 54–87.
- 1991. *Aryans in the Rigveda* (Leiden Studies in Indo-European 1). Amsterdam: Rodopi.
- Kümmel, Martin Joachim. 2014. Syllable- and word-related developments in earlier Indo-Iranian. In: Javier Caro Reina and Renata Szczepaniak (eds.), *Syllable and word languages*, pp. 210-224. Berlin: de Gruyter.  
<http://www.academia.edu/379001/Syllable-\and\word-related\developments\in\earlier\Indo-Iranian>
- 2012. Typology and reconstruction: The consonants and vowels of Proto-Indo-European. In: Benedicte Nielsen Whitehead, Thomas Olander, Birgit Anette Olsen, and Jens Elmegård Rasmussen (eds.), *The sound of Indo-European*:

- Phonetics, phonemics, and morphophonemics*. University of Copenhagen: Museum Tusulanum Press.  
<http://www.martinkuemmel.de/typrec.pdf>
- Lehr, Rachel. 2014. *A descriptive grammar of Pashai: The language and speech community of Darrai Nur*. A dissertation submitted to the Faculty of the Division of the Humanities in candidacy for the degree of Doctor of Philosophy. Chicago, Illinois.
- Liljegren, Henrik. 2008. *Towards a grammatical description of Palula: An Indo-Aryan language of the Hindu Kush*. Stockholm University: Stockholm.
- LSI, *Linguistic Survey of India Vol. IX, Part IV: Specimens of the Pahārī languages and Gujuri*. Calcutta: Superintendent Government Printing.
- Mallory, J. P. and D. Q. Adams. 2006. *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world*. Oxford: Oxford University Press.
- Martin, Dan. Zhangzhung dictionary. In: *Revue d'Etudes Tibétaines*, 18: 5–32.  
[http://himalaya.socanth.cam.ac.uk/collections/journals/ret/pdf/ret\\_18\\_01.pdf](http://himalaya.socanth.cam.ac.uk/collections/journals/ret/pdf/ret_18_01.pdf)
- Masica, Colin P. 1979. Aryan and non-Aryan in North Indian agriculture. In: M.M. Deshpande and Peter E. Hook (eds.), *Aryan and non-Aryan in India*, pp. 55–151. Ann Arbor: Michigan Papers on South and South-East Asia.
- 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Master, Alfred. 1964. *A grammar of Old Marathi*. Oxford: Clarendon Press.
- Matisoff, James A. 2001. The interest of Zhangzhung for comparative Tibeto-Burman. In: Yasuhiko Nagano and Randy J. LaPolla (eds.), *New Research on Zhangzhung and related Himalayan languages* (Bon Studies 3: 155–180). Senri Ethnological Reports 19. Osaka: National Museum of Ethnology.

- Morgenstierne, Georg. 1961. Dardic and Kāfir languages. In: *The Encyclopaedia of Islam*, New Edition, Vol. 2, Fasc. 25. Leiden: E. J. Brill.
- 1973. Traces of Indo-European accentuation in Pashto? In: *Nordic Journal of Linguistics*, 27: 61–65.
- Oberlies, Thomas. 1999. Middle Indo-Aryan and (the) Vedic (dialects) (Miscellanea Palica VII). In: *Historische Sprachforschung / Historical Linguistics*, 112,1, H.: 39–57.
- Osada, Toshiki. 2006. How many Proto-Munda words in Sanskrit? With special reference to agricultural vocabulary, pp. 151–74. In: Toshiki Osada (ed.), *Proceedings of the Pre-symposium of RIHN and the 7<sup>th</sup> ESCA Harvard-Kyoto Roundtable*. Kyoto: Research Institute for Humanity and Nature.
- Parpola, Asko. 2010. New etymologies for some Finnish words. In: *Studia Orientalia* 108: 305–318.
- 2015. *The roots of Hinduism: The early Aryans and the Indus Civilization*. New York: Oxford University Press.
- Peterson, John. 2009. A Kharia-English lexicon. In: *Himalayan Linguistics Archive* 5, i-xv: 1–212.
- Pinnow, Heinz-Jürgen. 1959. *Versuch einer historischen Lautlehre der Kharia-Sprachen*. Wiesbaden: Otto Harrassowitz.
- Pischel, Richard. 1900. *Grammatik der Prakrit-Sprachen (Grundriss der indo-arischen Philologie und Altertumskunde; 1,8)*. Strassburg.
- Pollock, Sheldon. 2006. *The language of the gods in the world of men*. Berkeley etc.: University of California Press.
- Ringe, Jr., Don. 1996. *On the chronology of sound changes in Tocharian. Vol. 1: From Proto-Indo-European to Proto-Tocharian*. New Haven: American Oriental Society.
- Ross, Malcolm. 1997. Social networks and kinds of speech-community events. In: Roger Blench and Matthew Spriggs (eds.), *Archaeology and language 1: Theoretical and methodological orientations*. London and New York: Routledge.

- SEAlang *Mon-Khmer Languages Project*.  
<http://sealang.net/monkhmer/index.htm>
- SEAlang *Munda Languages Project*.  
<http://sealang.net/munda/dictionary/>
- Sheth, Pandit Hargovind Das Trikamacandra. 1923-1928. *Päia-sadda-mahaṅṅavo. A comprehensive Prakrit-Hindi dictionary*.  
 Calcutta: Calcutta University.
- Sidwell, Paul. 2002. The Mon-Khmer substrate in Chamic: Chamic, Bahnaric and Katuic contact.  
<http://sealang.net/sala/archives/pdf4/sidwell2002mon-khmer.pdf>
- . 2010. The Austroasiatic central riverine hypothesis. In:  
*Journal of Language Relationship*: 117–134.  
[http://www.jolr.ru/files/\(51\)jlr2010-4\(117-134\).pdf](http://www.jolr.ru/files/(51)jlr2010-4(117-134).pdf)
- Southworth, Franklin. 2005a. *Linguistic archaeology of South Asia*.  
 London: Routledge-Curzon.
- . 2005b. Prehistoric implications of the Dravidian element in the NIA lexicon, with special reference to Marathi. In:  
*International Journal of Dravidian Linguistics* 34.1: 17–28.
- STEDT *Sino-Tibetan Etymological Dictionary and Thesaurus*.  
<http://stedt.berkeley.edu/~stedt/cgi/rootcanal.pl>
- Stuhrmann, Rainer. 2016. *Die Zehnkönigsschlacht am Ravifluß*.  
<http://crossasia-journals.ub.uni-heidelberg.de/index.php/ejvs/article/view/933>
- Tagare, Ganesh, Vasudev. 1948 (1987). *Historical grammar of Apabhraṃśa*. New Delhi: Motilal Banarsidass.
- Thomason, Sarah Grey. 2000. *Linguistic areas and language history*.  
<http://www-personal.umich.edu/~thomason/papers/areas.pdf>
- Turner, Ralf L. 1924. The Sindhi recursives or voiced stops preceded by glottal closure. In: *Bulletin of the School of Oriental Studies* 3,2: 301–315.
- Vacek, Jaroslav. 1989. The Dravidian and Mongolian linguistic comparison. In: *Dravidian Studies* 36: 22–27.

- Whitney, William Dwight. 1973 (1924). *Sanskrit grammar: Including both the classical language and the older dialects of Veda and Brāhmaṇa*. Delhi etc.: Motilal Banarsidass.
- Willson, Stephen R. 1999. *Basic Burushaski vocabulary*. Islamabad: National Institute of Pakistan Studies and Summer Institute of Linguistics. (Studies in Languages of Northern Pakistan Vol. 6).
- Witzel, Michael. 1999. Early sources for South Asian substrate languages. In: *Mother Tongue* (extra number), pp. 1–70.
- 2005. Central Asian roots and acculturation in South Asia: Linguistic and archaeological evidence from western Central Asia, the Hindukush and northwestern South Asia for early Indo-Aryan language and religion. In: Toshiki Osada (ed.), *Linguistics, Archaeology and the human past: Occasional Paper 1*. Kyoto: Indus Project, Research Institute for Humanity and Nature.
- Zoller, Claus Peter. 2005. *A grammar and dictionary of Indus Kohistani. Volume 1: Dictionary*. Berlin, New York: Mouton de Gruyter.
- 2007. Himalayan Heroes. In: Brückner, Heidrun, Hugh van Skyhawk and Claus Peter Zoller (eds.), *The Concept of Hero in Indian Culture*; pp. 237–274. New Delhi: Manohar.
- 2011. Pahari (in Russian). In: G. A. Zograph, T. I. Oranskaia, L. I. Kulikov, P. K. Pandey (eds.), *Languages of the world: The New Indo-Aryan languages (Jazyki mira: novye indoarijskie jazyki)*, pp. 219–261. Moscow: Academia.
- 2016. *Proto Indo-European – Indo-Aryan – Himalayan Aryan languages: New perspectives on some old questions*. Dehradun: Bishen Singh Mahendra Pal Singh.
- 2017. Northern India before and after the arrival of Indo-Aryan. In: B.K. Joshi and Maheshwar P. Joshi (eds.), *Unfolding central Himalaya: The cradle of culture*, pp. 1–45. Doon Library and Research Centre: Dehradun, and Almora: Almora Book Depot.

——— Forthcoming: Origins and developments of Indo-Aryan: An exploratory study.