

## 1 *Where We Stand in the Study of the Nûristânî Languages*

Good morning. Happy to see all you participants. Kudos to all you who have had a hand in making this workshop happen.

Are there any Nuristanis here today? How many of you have been to Nûristân? How many of you have heard a Nûristânî language?

I want to run through a quick list of the usual linguistic topics, highlighting where we stand and what we ought to know or see done going forward.

### 1.1 *Ethnicity*

#### 1.1.1 *What We Know*

We have a proper accounting of the ethnic groups in Nûristân (Strand *Website*: 1997–present).

The recognition of the Kom (Prs. *Kâmôzî*) as modern remnants of the ancient Kambojas connects them to at least the 6<sup>th</sup> century BC .

#### 1.1.2 *What We Would Like to Know*

What were the relations of the Kom (ancient Kambojas) vs. the Kâta and other Kâmkata-vari speakers of central Nûristân?. Do Kâmviri and Kâta-vari have a common origin, in which case they probably continue the ancient Kambojan language, or did one replace a former language of the other (*e.g.*, did the Kambojas abandon their language for an early form of Kâta-vari)?

Whence the Vâsi? Are they an Indo-Aryan client tribe that got attached to the Kâta?

Whence the Vâi? According to their traditions, the Vâi fled the Ghaznavid invasion of Kâma, following the Kunar up to Mâdeš and Samâlâm in the Šigal Valley and thence over the watershed to their main community of Vâigal. Accounts of the Gawâr people state that the Vâi expropriated the current site of Vâigal from the Gawâr, who fled to the Kunar Valley. As the Vâi expanded, they established other communities in the upper Vâigal and further east.

Whence the Kalaša? Did the Čima-Niše-alâ Kalaša abandon their original Grâmsaňâ-viri for Vâi-alâ, while giving their ethnonym to the Vâi and, further, to the Kâsio of Chitral? Are there lexical remnants of Grâmsaňâ-viri in Čima-Niše-alâ?

Note that the thrust of ethnic movements has been primarily from central Nûristân toward Chitral, not vice versa, as evidenced by the early movement of the Kalaša Nûristânîs from Laghmân through the Pech basin and into southern Chitral. Thus we would expect little impact on the Nûristânî languages from Chitral or further east, but we should not be surprised to find sprinkles of Nûristânî influence in words like Kalaša-mund *'iĉ*, Khowar *orĉ* or *ohĉ*.

### 1.2 *Phonology*

#### 1.2.1 *What We Know*

We have improved phonemic analyses of Kâta-vari (Kul'em Dialect: Grjunberg 1980, Kt'ivi dialect: Strand *Website*:2011a), Kâmv'iri (Strand *Website*:1997–2007), and Kalaša-alâ

(Samiullâh Tâza 1995, Degener 1998, Strand 1999), which supersede the phonetic transcriptions of predecessors and remove spuriously marked features such as length.

We must recognize the roles of the following phonetic processes in the development of Indo-Irânian (Strand *Website*: 2002–2011, 2010):

(1) Anterior vs. posterior voicing: the voicing of the Nûristânî and Irânian languages is exclusively ANTERIOR, vs. the mixed POSTERIOR AND ANTERIOR voicing for neighboring Indo-Aryan languages; Note the contrasts of anterior vs. posterior phonation in Khowar and Âcharêtâ'.

(2) “Prognathizing,” which produces changes such as  $\check{c} > \acute{c}$ ,  $dv > b$

### 1.2.2 What We Ought to Know

There is a tradition of mediocre phonetic observation and poor phonemic analysis in Nûristânî research. This is due primarily to the limited access to speakers, but current researchers need to pay better attention to the phonetic details and base their phonemic analyses on sound vocal-tract physiology and not on abstract phonological models.

I highlight here some problematical issues in regional phonological analysis.

#### 1.2.2.1 Vowels

##### 1.2.2.1.1 Long vowels: phonemic or not?

- Sub-phonemic length results from:
  - openness: Kâmviri, Kalaşa mun
  - accent: Kâmviri, Vâsi vari, Sañu viri[?] (how much indicated length is spurious in Vâsi-vari and Sañu-viri?); also in Chitral: in Kalaşa mun and in Khowar posterior voicing (in Khowar posterior voicing has been interpreted as length, which is systemically wrong)
  - nasalization: Sañu viri[?]
  - tenseness: Âcharêtâ' (in Âcharêtâ' note that the tense/long vowels only occur accented), Paşaî
- Phonemic length results from:
  - placement of the above processes into a contrastive position
  - glide reduction: Kâmviri
  - syllable-onset loss: Kâmviri
  - long-short systems occur in Kâmviri, Sañu-viri, and perhaps Varĵan-alâ and Vâsi-vari, if the latter are not of the next type.

		Kâmv'iri				
		labialized	fronted	default		
short	Close	<i>u</i>	<i>i</i>	<i>a</i>	Close	
		<i>ü</i>				
	Mid	<i>o</i>	<i>e</i>		Open	
	Open			<i>â</i>		

		Sañu-viri				
		labialized	fronted	default		
short	Close	<i>u</i>	<i>i</i>	<i>a</i>	Close	
	Mid	<i>o</i>	<i>e</i>		Open	
	Open			<i>â</i>		

  

		Kâmv'iri				
		labialized	fronted	default		
long	Close	<i>ua</i>	<i>ia</i>	<i>aa</i>	Close	
		<i>üa</i>				
	Mid	<i>oa</i>	<i>ea</i>		Open	
	Open			<i>âa</i>		

		Sañu-viri				
		labialized	fronted	default		
long	Close	<i>u:</i>	<i>i:</i>	<i>a:</i>	Close	
	Mid	<i>o:</i>	<i>e:</i>		Open	
	Open			<i>â:</i>		

In Kâmviri vocalic length is an allophone of the phoneme *a*. The default pronunciation of this phoneme is [i]. But [i] cannot occur after a vowel in Kâmviri, and length [:] only occurs after vowels; therefore, [i] and [:] are in complementary distribution after vowels and are both allophones of the single phoneme *a*.

Contrastingly, in Sañu-viri *a* ([i]) occurs after vowels (e.g., *ia*), and therefore cannot be a candidate for indicating vocalic length in that language. Length contrasts in Sañu-viri, with antecedents, include:

<i>bâs</i> 'vapor'	<i>b<sup>h</sup>a:s'a-</i>	<i>bâ:s</i> 'twelve'	<i>dv'â-daša</i>
<i>vâs</i> 'adze'	<i>v'âši:-</i>	<i>vâ:s</i> 'day'	<i>vâs-'a-</i>
<i>kâṇ</i> 'arrow'		<i>laukâ:ṇ</i> 'blood vessel'	
<i>pié</i> 'cream'	<i>pičcha:-</i>	<i>pi:é</i> 'pine'	<i>pauči:-</i>
<i>bi</i> 'seed'	<i>b'i:ja-</i>	<i>bi:</i> 'spade'	
<i>ko</i> 'whom'	<i>k'a-</i>	<i>ko:</i> 'hump'	<i>kaku-b<sup>h</sup>a-</i>
<i>gom</i> 'I went'	<i>ga-t'a-</i>	<i>go:m</i> 'wheat'	<i>god<sup>h</sup>u:ma-</i>
<i>šo</i> 'six'	<i>šuvaṣ*</i>	<i>šo:</i> 'information'	<i>šr'av-as-</i>
<i>no</i> 'nine'	<i>n'ava</i>	<i>no:</i> 'water trough'	<i>nâv-y'a:-</i>
<i>daru</i> 'copper'		<i>dâ:ru:</i> 'gunpowder'	
<i>âsku</i> 'wire doorkey'	<i>â-skau-ti</i>	<i>ku:</i> 'where'	<i>k'a-</i>

It is not clear what the conditioning factors were that produced such contrasts: perhaps accent or length in a following syllable?

- If the contrast between long and short is not systemic and pervasive, the occasional long vowel should be analyzed as a sequence of two identical vowels; e.g., Vâsi *l'â-* 'have' vs. *l'ââ-* 'had'; Khowar *hay`a* 'this' vs *hay`aa* 'here'. These mostly arise across morpheme boundaries.

### 1.2.2.1.2 Evolution of the Nûristânî Vowel Systems from long-short to tense-lax to close-open

The Proto-Aryan vowel system was probably best characterized as tense-lax, having evolved from a system that had both tense-lax (*â-a*) and long-short (*a:-a*) contrasts. Tense vowels were incidentally long, indistinguishable from the true long vowels that resulted in PAr. from a PIE vowel plus laryngeal. Thus length and tension merged, and either could be marked as phonemic.

Proto-Indo-European				
	labialized	fronted	back	
<b>Long</b>	<i>u:</i>	<i>i:</i>	<i>r:</i>	Close
<b>Short</b>	<i>u</i>	<i>i</i>	<i>r</i>	
	<i>o</i>	<i>e</i>	<i>a</i>	Open
<b>Long</b>	<i>o:</i>	<i>e:</i>	<i>a:</i>	
<b>Short</b>	<i>eu, ou, au</i>	<i>ei, oi, ai</i>	<i>er, or, ar</i>	Diphthongs
<b>Long</b>	<i>e:u, o:u, a:u</i>	<i>e:i, o:i, a:i</i>	<i>e:r, o:r, a:r</i>	

Proto-Aryan				
	labialized	fronted	back	
<b>Long</b>	<i>u:</i>	<i>i:</i>	<i>r:</i>	Close
<b>Short</b>	<i>u</i>	<i>i</i>	<i>r</i>	
			<i>a</i>	Open
<b>Long</b>			<i>a:</i>	
<b>Short</b>	<i>au</i>	<i>ai</i>	<i>ar</i>	Diphthongs
<b>Long</b>	<i>a:u</i>	<i>a:i</i>	<i>a:r</i>	

The linguists' tradition has been to mark them as long.

But if there are qualitative differences between long and short vowels, the long ones should be marked as tense. Only if long and short vowels do not differ in quality is length phonemic.

- Modern regional tense-lax systems include those of Fârsî, Paxt'o, Pašaî, Gawar-bati, and Âcharêtâ'.

Fârsî				
		labialized	fronted	default
<b>Close</b>	<b>Tense</b>	<i>û</i>	<i>î</i>	
<b>Mid</b>	<b>Lax</b>	<i>u</i>	<i>i</i>	<i>a</i>
<b>Open</b>	<b>Tense</b>	<i>ô</i>	<i>ê</i>	<i>â</i>

Pašaî				
		labialized	fronted	default
<b>Close</b>	<b>Tense</b>	<i>û</i>	<i>î</i>	
<b>Mid</b>	<b>Lax</b>	<i>u</i>	<i>i</i>	<i>a</i>
<b>Open</b>	<b>Tense</b>	<i>ô</i>	<i>ê</i>	<i>â</i>

Âcharêtâ'				
		labialized	fronted	default
<b>Close</b>	<b>Tense</b>	<i>û</i>	<i>î</i>	
	<b>Lax</b>	<i>u</i>	<i>i</i>	
<b>Mid</b>		<i>o</i>	<i>e</i>	<i>a</i>
<b>Open</b>	<b>Tense</b>	<i>ô</i>	<i>ê</i>	<i>â</i>

The Nûristânî languages have no contrast of tense vs. lax vowels, the ancient phonemes having coalesced.

All the languages have three phonetic degrees of vocalic opening, but only Kalaşa-alâ has three phonemic (contrasting) degrees of openness. The remaining Nûristânî languages have only two phonemic degrees of openness.

Close—Mid—Open Systems:

Kalaşa-alâ

Nişei-alâ				Varjan-alâ			
	labialized	fronted	default		labialized	fronted	default
Close	<i>u</i>	<i>i</i>		Close	<i>u</i>	<i>i</i>	<i>i</i> ("ə")
	<i>ü</i>				<i>ü</i>		
Mid	<i>o</i>	<i>e</i>	<i>a</i>	Mid	<i>o</i>	<i>e</i>	<i>a</i>
	<i>ö</i>				<i>ö</i>		
Open		<i>ä</i>	<i>â</i>	Open		<i>ä</i>	<i>â</i>

(The phonemic status of [ɐ] vs. [i̠] in Varjan-alâ is unclear. Is there a phonemic contrast, or are they positional allophones?)

Close—Open Systems:

Vâsi-vari, Kâmk'ata-vari, Sañu-viri, also Khowar and Bhaṭ'e-sa zib.

Vâsi-vari					Kâmv'iri					Sañu-viri				
	labialized	fronted	default			labialized	fronted	default			labialized	fronted	default	
Close	<i>u</i>	<i>i</i>	<i>a</i>	Close	Close	<i>u</i>	<i>i</i>	<i>a</i>	Close	Close	<i>u</i>	<i>i</i>	<i>a</i>	Close
	<i>ü</i>					<i>ü</i>					<i>ü</i>			
Mid	<i>o</i>	<i>e</i>		Open	Open	<i>o</i>	<i>e</i>		Open	Open	<i>o</i>	<i>e</i>		Open
	<i>ö</i>					<i>ö</i>					<i>ö</i>			
Open			<i>â</i>					<i>â</i>					<i>â</i>	

### 1.2.2.2 Approximants and Flaps: Phonemic Treatment of [ɹ], [ɹ̠], [ɹ̥], [r], and [r̥]

In Sañu-viri and Vâsi-vari initial \**r* was strengthened to fricative *z*, and in Kalaşa-alâ to *vr* [vɹ], while in Kâmkata-vari it was weakened to *ř* ([ɹ̠]). No Nûristânî language retains the flapped [r] (trilled [r] does not occur in Nûristânî) as the realization of the initial phoneme \**r*.

#### 1.2.2.2.1 Vâsi-vari: (no phonemic approximants)

*r* = [r] (occurs initially in loanwords only; \**r*- > *z*-)

*ɹ* = [ɹ] in C\_ (dialectally, < CVɹV).

#### 1.2.2.2.2 Kâmkata-vari: (approximants are phonemic):

*r* = [r] (occurs initially in loanwords only; \**r*- > *ř*-)

*ř* = [ɹ̠]

*ñ* = [ɹ̥] (does not occur initially)

*ŋ* = [ŋ]

*ɹ* = [ɹ] in V\_V (Kâmviri only).

#### 1.2.2.2.3 Nişei-alâ: (approximants are sub-phonemic)

*r* = [r] (occurs initially in loanwords only; \**r*- > *vr*-)

*ɹ* = [ɹ] in C\_V<sub>frn</sub> and v\_

[ɹ̥] in C\_V<sub>-frn</sub> (i.e., CV̥, a rhoticized vowel), and

[ɹ] medially and finally (does not occur initially)

**ñ** = [ɲ] in C\_V<sub>frn</sub> and v\_<sub>-</sub>  
 [˘] in C\_V<sub>-frn</sub> (i.e., C<sup>˘</sup>V<sup>˘</sup>, a rhoticized, nasalized vowel), and  
 [ɲ] medially and finally (does not occur initially)  
**ŋ** = [ɲ̃] in V\_V,  
 [ŋ] elsewhere.

#### 1.2.2.2.4 Sañu-viri: (approximants are sub-phonemic)

**r** = [r] (does not occur initially; \*r- > z-)  
**ṛ** = [ɹ] after labial and velar consonants,  
 [r] elsewhere (does not occur initially)  
**ñ** = [ɲ̃] in V\_V, finally (does not occur initially)  
**ŋ** = [ŋ] finally (does not occur initially or intervocalically).

Alternatively, if [r] does not occur after labial and velar consonants in loanwords,

**r** = [ɹ] after labial and velar consonants,  
 [r] elsewhere (does not occur initially; \*r- > z-)  
**ṛ** = [ɹ̣] (does not occur initially),

and, if final [ɲ̃] does not occur,

**ŋ** = [ɲ̃] in V\_V,  
 [ŋ] finally (does not occur initially or intervocalically).

#### 1.2.2.3 Prognathizing in Southern Nûristânî

Prognathizing promotes:

- dentalization of laminals (**č, ĵ, š** > **ć, ź, s**), seen in Proto-Aryan and again in Sañu-viri (and Paxt'ô.)
- \***dv** > **b** (sporadic Indo-Aryan loans in Southern Nûristânî: Skt **dv'âdaša** > K-a **bâš** '12'),
- **v** is an "unstable sound," as Morgenstierne remarked.
  - as a result of prognathizing, **v**
    - = [β], [v], or [w] as language-dependent default sounds before back or central vowels, but
    - = [v] before front vowels.
  - note the change of postvocalic **v** to **a** = [ɨ] in Kâmviri.

#### 1.2.2.4 Relative Timings of Articulations

- Kâmviri syllables and dissyllables: "burst" delivery;
- Vâsi-vari anticipation of labialization and fronting;
- Kalaşa-alâ anticipation of fronting.

## 1.3 Grammar

### 1.3.1 What We Know

The ubiquity of “directionality” in the Nûristânî languages has impressed all field researchers (Morgenstierne 1949: 229–231, Grjunberg 1980: 265–281, Buddruss 1977b, Strand 1997) and has inspired the spatial-visual cognitive approach to Nûristânî grammar that Strand has advocated since the early 1980’s (1985, 1991, 1997, 1999). This approach has yielded grammatical insights that are applicable to other languages of the region; notably:

- There are two ways of looking at the image of discourse: 2-dimensionally (sideways) or 3-dimensionally (in perspective). Nouns with (at least) two case forms (certain classes of nouns including pronouns) imply three dimensions; nouns uninflected for case imply two dimensions. Three-dimensional images have a foreground and a background zone; uninflected nouns stand in the foreground, while case-marked nouns stand in the background (Strand *Website:2000–2002*).
- Looking backward into time with a RETROSPECTIVE PERSPECTIVE onto foreground and background zones accounts for split ergativity and case marking more insightfully than a traditional statement like “the subject is in the oblique case, and the object is in the direct case and indicated by the verbal ending,” which confuses the concepts of verbal subject, which always stands in the foreground, and oblique-case agent, which stands in the background (Strand 1999, *Website: 1999b–present*, *Website: 2000–2002*).
- Remote agency is the basis of “causative” verbs, along a distance scale of internal vs. surface vs. external vs. remote zones (Strand 1985, *Website: 1999b–present*).
- The stative verb *âsa-* ‘is’ is a special type of motion verb that moves or projects nouns into the image of discourse, and the verb *bu-* ‘happen; become’ depicts nouns “popping up” in the discourse in different locations, directions, or guises (Strand 1985, *Website:1999b–present*).
- Certain directional suffixes have probable ancient gestural precursors (Strand *Website: 2011*).

### 1.3.2 What We Ought to Know

Go Depictive! See my online presentations.

#### 1.3.2.1 Objects: Case

See my online presentation.

#### 1.3.2.2 Changes:

Mode and Modal Participles: See my online presentation.

Projection of modal participles with “is”

“Pop-up” situation with “happen”

#### 1.3.2.3 Embedding:

See my online presentation.

## **1.4 Lexicography**

### **1.4.1 What We Know**

From Buddruss and Degener (2016), a lengthy lexicon of Vâsi-vari; from Degener (and Buddruss) (1998), a lengthy lexicon of Nišei-alâ; from Strand (*Website*: 1999a–present), online phonemically transcribed lexicons of Kâmviri (5,500 native words), plus shorter lexicons of Kâta-vari (Kt'ivi dialect), Kalaša-alâ (Nišei-alâ dialect), and Saņu-vîri, in addition to the Indo-Aryan languages Khowâr, Ačharêtâ', Degano (Eastern Pashaî), and Bhaṭesa-zib. Strand's *Nûristâni Etymological Lexicon*, initially with some 1,800 entries, appeared on his website in 2012.

### **1.4.2 What We Need to Do**

A database combining Buddruss's and my data, supplemented by Grjunberg's and Morgenstierne's data to fill in. Should be available at nuristan.org in some widely-used commercial form, not in some locally-developed university database.

## **1.5 Phylogeny**

### **1.5.1 What We Know**

Details of phylogenetic classification (Morgenstierne 1974, Strand 1973, 2001, 2010, Buddruss 1977a), with a proper accounting of languages, dialects, and ethnic groups (along with supporting origin myths: Grjunberg 1980: 36–38, Degener 1998: 237–252, Strand *Website*: 1997–99), a confirmation of the division between Northern and Southern (“Kalaša,” characterized by prognathizing) groups of Nûristânî languages (passim), and the exclusion of the Indo-Aryan language Dameļi as a possible Nûristânî language (Strand *Website*: 1998–2002, 2001: 254).

### **1.5.2 What We Need to Know**

We need a closer look at lexical comparisons to determine layers of borrowing. Especially Vâsi-vari requires a close comparison with Kâmkata-vari to determine the latter's influence on the former, and to determine the “core” vocabulary of Vâsi-vari to better determine its phylogenetic position. Also, can close lexical inspection determine whether Kâmviri and Kâta-vari have a common origin, or whether one has replaced a previous language of the other; and similarly, the relationship between Vâi-alâ and Čima-Nišei-alâ.

I should the results of the CoBL-IE cognancy study, and the strange position of the Nûristânî clade.

## **2 Prospects for Nûristânî studies**

The immediate future does not look bright for studies of the Nûristânî languages, as long as hostilities in the region remain. Nûristân now is still too dangerous for outsiders to work in. Studies with displaced consultants outside of Nûristân are a poor second best to immersed study in a fully functioning linguistic community within Nûristân.

Detailed grammatical descriptions built on solid phonemic grounds are still needed for Âškuņu-Saņu-vîri and Tregâmi, along with more lexical data. Cognitive formats of grammatical description would be more informative than traditional linguistic paradigms in elucidating the strong sense of space that underlies the grammar of the Nûristânî languages.

Transcriptions of local oral history are sorely lacking, and such knowledge is fast disappearing.