Bhāṣā

Journal of South Asian Linguistics, Philology and Grammatical Traditions

Vol. 3 – Num. 2 October 2024

e-ISSN 2785-5953

Edizioni Ca'Foscari

e-ISSN 2785-5953

Bhasha

Journal of South Asian Linguistics,
Philology and Grammatical Traditions

Edizioni Ca' Foscari - Venice University Press
Fondazione Università Ca' Foscari
Dorsoduro 3246, 30123 Venezia
URL https://edizionicafoscari.unive.it/it/edizioni/riviste/bhasha/

Bhasha

Journal of South Asian Linguistics, Philology and Grammatical Traditions

Semestral journal

Head office Ca' Foscari University of Venice | Department of Asian and North African Studies | Dorsoduro 3462, 30123 Venice, Italy | bhasha_journal@unive.it

Publisher Edizioni Ca' Foscari | Fondazione Università Ca' Foscari | Dorsoduro 3246, 30123 Venice, Italy | ecf@unive.it

Certificazione scientifica delle Opere pubblicate da Edizioni Ca' Foscari: tutti i saggi pubblicati hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima sotto la responsabilità del Comitato scientifico della rivista. La valutazione è stata condotta in aderenza ai criteri scientifici ed editoriali di Edizioni Ca' Foscari.
Scientific certification of the works published by Edizioni Ca' Foscari: all essays published in this issue have received a favourable opinion by subject-matter experts, through an anonymous peer review process under the responsibility of the Advisory Board of the journal. The evaluations were conducted in adherence to the scientific and editorial criteria established by Edizioni Ca' Foscari.

Bhasha
Vol. 3 – Num. 2 – December 2024
e-ISSN 2785-5953

Table of Contents

ARTICLES

REVIEWS

URL https://edizionicafoscari.unive.it/it/edizioni4/riviste/bhasha/2024/2/
DOI http://doi.org/10.30687/bhasha/2785-5953/2024/02

Articles

e-ISSN 2785-5953

Bhasha
Vol. 3 — Num. 2 — October 2024

Buddhist Hybrid Sanskrit: How Did It Originate?

Hans Henrich Hock
University of Illinois at Urbana-Champaign, USA

Abstract Since Edgerton 1953, Buddhist Hybrid Sanskrit (BHS) has been considered a sanskritization of Prakrit, based on strong linguistic evidence: lexical items are Sanskritic, endings Prakritic. Sanskritization has been argued as motivated by a need to compete with Sanskrit-using brahmins. The issue of how sanskritization might have been accomplished is the topic of this paper. In early AD, Sanskrit was spread by brahmins as language of technical and fine literature. The curriculum of schools imparting Sanskrit instruction started with memorization of a Sanskrit lexicon and a version of Pāṇini's grammar. The link between these was established in later years. Sanskritization of BHS can be explained in terms of early Buddhist students only completing the initial stage of instruction. This would provide them with a Sanskrit lexicon for replacing Prakrit words. However, not yet knowing how to apply the grammatical rules, students would use Prakritic endings. Support for this hypothesis comes from Kapstein's (2018) account of grammatically deficient, but lexically accurate Sanskrit compositions by medieval Tibetans, as resulting from acquiring grammar and lexicon separately, 'with almost no training in practical application'. I conclude by considering the implications of my proposal as well as the similarities and differences between BHS and 'Bilingual Mixed Languages'.

Keywords Buddhist Hybrid Sanskrit. Historical development. Sanskritization. Tibetan parallel. Bilingual Mixed Languages.

Edizioni Ca'Foscari

Peer review

Submitted 2024-06-06
Accepted 2024-11-08
Published 2024-12-18

Open access

Citation Hock, H.H. (2024). "Buddhist Hybrid Sanskrit: How Did It Originate?". *Bhasha*, 3(2), 175-192.

DOI 10.30687/bhasha/2785-5953/2024/02/001

1 Introduction

Since Edgerton (1953) Buddhist Hybrid Sanskrit (BHS), the language attributed to the Lokottaravāda Mahāsāṅghika school of Buddhism, has been considered a sanskritization of an original Prakrit, where 'Prakrit' is understood as a variety of early Middle Indo-Aryan.[1] A likely motivation for this sanskritization has been proposed in earlier literature, namely the desire to compete with Sanskrit-using brahmins in religious disputations at royal courts (e.g. Salomon 1998; Bronkhorst 2010). The manner in which such a massive lexical sanskritization could have been accomplished, however, has not been satisfactorily addressed. This paper attempts to answer that question.

Section 2 presents a brief overview of the characteristics of BHS, followed by a discussion of the similarities and differences between BHS and Epigraphic Hybrid Sanskrit (EHS) in Section 3. Chronological issues regarding BHS are dealt with in Section 4, together with the evidence of EHS. Motivations for the development of BHS are the topic of Section 5. Section 6 discusses the likely mechanism for the sanskritization of BHS. Section 7 examines two texts that show mixtures of Prakrit or Apabhraṁśa and Sanskrit similar to BHS but are sufficiently different to merit separate discussion. Finally, Section 8 presents a summary of my findings, as well as a comparison to 'Bilingual Mixed Languages'.

2 Characteristics of BHS

Buddhist Hybrid Sanskrit is characterized by the following features. The stems of nouns and adjectives almost always are Sanskrit in form. Verbs and function words, by contrast, tend to show various degrees of Prakrit features, and so do nominal and adjectival case endings. In addition, external sandhi generally follows Prakrit rules.

Detailed coverage of the features of Buddhist Hybrid Sanskrit can be found in Edgerton 1953. For present purposes, suffice it to consider examples (1) and (2) from the early text *Mahāvastu* (Marciniak 2019 edition), as illustrations of how Buddhist Hybrid Sanskrit differs from standard Sanskrit.[2] Sandhi (small cap italics), inflectional

This paper is a thoroughly revised and expanded version of a presentation at the 2023 SALA meeting in Venice. I have benefited from comments at that meeting as well as feedback from two anonymous referees. The responsibility for any errors and omissions rests with me.

1 The term, thus, does not (normally) cover the literary Prakrit discussed in detail by Ollett 2017.

2 Edgerton's grammar does not present any textual examples.

endings (bold), and entire forms (italics), especially verbs, are of Prakrit origin, while the majority of lexical items are Sanskrit (unmarked). Sanskrit equivalents are given for comparison below each line; a third line offers a virtual Pāli version that illustrates the difference of BHU from both standard Sanskrit and Prakrit.

(1)	*abhūṣi*	rājā	ikṣvāku	vārāṇasyāṁ	mahābalo
	abhūd/āsīd	rājā	ikṣvākur	vārāṇasyāṁ	mahābalaḥ
	a(b)hosi	rājā	okkāko	bārāṇasiyā	mahābalo
	be.AOR.3SG	king.NOM.SG.M	Ikṣ.NOM.SG.M	Vār.LOC.SG.F	great.strength.NOM.SG.M

'There was a strong king, Ikṣvāku, in Vārāṇasī.'

(2)	so	ca	jīrṇo	*bhavi***tvāna**
	sa	ca	jīrṇo	bhūtvā
	so	ca	jiṇṇo	bhavitvā(na)/bhavittā*
	dem.nom.sg.m	&	old.NOM.SG.M	be(come).CVB
	vepamāne**hi**	gātr**ehi**	rājādvāram	upāgato
	vepamānair	gātrai	rājadvāram	upāgataḥ
	vepamānehi	gattehi	rājabbāraṁ**	upāgato
	trembling.INST.PL.N	limb.INST.PL.N	king.gate.ACC.SG.N	go.up.PFV.PPL.NOM.SG.M

'And he (Indra) turning himself old, […] came to the royal gate with trembling limbs.'

* The latter form is non-canonical.

** This is the expected Pāli form and the general Prakrit version; the usual Pāli form, *-dvāra*, may reflect sanskritization (von Hinüber 1986, 123).

As can be seen, nominal and adjectival roots and stems are Sanskrit in form. Contrast, e.g., the name *ikṣvāku-* with Pkt. *okkāka*, *jīrṇa-* 'old' with *jiṇṇa*, *gātra-* 'limb' with *gatta-*.

However, only a few inflected forms are identical to their Sanskrit counterparts – *rājā* 'king.NOM.SG', *vārāṇasyāṁ* 'in Vārāṇasī', and *jīrṇo* 'old.NOM.SG'. Nominal stems are generally identical to their Sanskrit counterparts, such as *ikṣvāku-* in (1) or *vepamāna-* in (2), an exception being *rājādvāra-* in (2), rather than *rājadvāra-*, with long *ā* vs. Skt. short *a* in the second syllable. Nominal endings, however, are Prakrit, not only those violating Sanskrit sandhi such as *ikṣvāku* rather than *ikṣvākur* and *mahābalo* instead of *mahābalaḥ* but also forms where Middle Indo-Aryan made suffix choices different from contemporary Sanskrit, such as *vepamān**ehi*** rather than *vemapānair/ḥ*.[3]

Especially important are the verbal forms *abhūṣi* 'was' and *bhavitvāna* 'having been/become'. The latter differs from the corresponding Sanskrit form both in the form of the root (*bhavi-* instead

3 Vedic Sanskrit offered an option between *-ebhiḥ* and *-aiḥ* and their sandhi variants, but that was lost in favor of *-aiḥ* in Classical Sanskrit.

of *bhū-*) and in its suffix (*-tvāna* vs. *-tvā*). The former is, in fact, quite complex with an unusual mixture of Prakrit and Sanskrit features. Its retroflex sibilant *ṣ* follows Sanskrit phonology (Prakrit would have dental *s*), and its root shape *bhū-* looks like Skt. *bhū-* in the aorist form *abhūd*,[4] but its overall formation as an *s*-aorist is Prakrit; compare Pali *a(b)hosi* vs. Skt. *abhūd*.[5] A virtual, but unattested Sanskrit counterpart of *abhūṣi* would be *abhūṣīd**, but the final short *-i* of the BHS form is Prakritic. Moreover, the choice of the aorist formation is typical of Prakrit, which merges imperfect and aorist, with heavy dominance of aorist forms; Sanskrit would instead have the imperfect *abhavad* or, preferred in this context, *āsīd* (from a different root).

In the larger text, however, there are instances of sanskritized verb forms, such as *tiṣṭhati* 'stands' vs. Pkt. *tiṭṭhati*. These are mainly found in forms that are relatively similar to Sanskrit in their morphology; forms like *abhūṣi*, whose morphology differs considerably from Sanskrit, remain Prakritic.[6]

Other features, not found in the sample texts of (1) and (2) include those listed in (3) below. (3a) illustrates the common Prakritic extension of the *a*-stem genitive singular ending *-sya,* sanskritization of Prakrit *-ssa,* to the *i*-stems and *in*-stems on the model in (3c).[7] The retroflex sibilant in the rarer variant *riṣiṣya* is a case of hypersanskritization. (3b) provides examples of function words of Prakritic origin or shape; note that BHS *sace* (sometimes partly sanskritized as *saced*) and Skt. *ced* differ in their syntax: *sace* is clause-initial, whereas *ced* is a second-position clitic.

(3)	a.	BHS	Sanskrit		
		vārisya	vāriṇaḥ	'water.GEN.SG.N'	
		riṣisya/riṣiṣya	ṛṣeḥ	'seer.GEN.SG.M'	
	b.	kāci	kāś-cid	'some.NOM.PL.F'	
		sace	ced	'if'	
	c.	Prakrit	*pakkha*:	pakkha-ssa	'wing, side'
			vāri:	X = *vāri-ssa*	'water'

4 Sanskrit forms are cited here in the sandhi version appropriate to the context in (1).

5 However, the root form differs (*ū* vs. *o*), and as Edgerton (1953, 157) notes, there is no exact Prakrit counterpart.

6 Note also, elsewhere in the text, forms such as *āsi* 'be.AOR.3PL' vs. Skt. *āsan* 'be.IMPF.3PL'.

7 On this ending, see also Edgerton 1953, 74.

Bhasha
3, 2, 2024, 175-192
e-ISSN 2785-5953

3 BHS, Epigraphic Hybrid Sanskrit (EHS), and the Issue of Buddhist Sanskrit

Although details differ, a similar mixture of Sanskrit and Prakrit features is found in the language of inscriptions dated (roughly) from the first century BC to the first century AD, a language which has been called Epigraphic Hybrid Sanskrit (EHS) by Damsteegt (1978) and Salomon (1998, 81). For an example see (4). Here again virtual versions in Sanskrit and Prakrit are added for comparison.

(4)	bhaṭārakā *añātiyā*	ca	gatosmi	*varṣāratuṁ* [= varṣā + ratuṁ]
	bhaṭāraka+ājñātyā*,**	ca	gatosmi	varṣartuṁ [= varṣā + ṛtuṁ]
	bhaṭāraka añātiyā	ca	gato mhi	vassāratuṁ
	lord+order.INS.SG.F	&	gone+be.PRS.1SG	rain.season.ACC.SG.M
	mālay**ehi**	ru(d)dha**ṁ**	ut(t)amabhadraṁ	mocayituṁ
	mālayai	ruddham	uttamabhadraṁ	mocayituṁ
	mālayehi	ruddhaṁ	uttamabhaddaṁ	mocayituṁ
	Mālaya.INS.PL.M	besieged.ACC.SG.M	Uttamabhadra.ACC.SG.M	free.CAUS.INF
	tatosmi	gato	pokṣarā**n**i	tatra ca mayā
	tatosmi	gataḥ	puṣkarāṇi	tatra ca mayā
	tato mhi	gato	pokkharāni	tatta ca mayā
	then+be.PRS.1SG	gone.NOM.SG.M	Puṣkara.ACC.PL.N	there & I.INS.SG
	abhiseko	kṛto	tīṇi	gosahasrā**n**i
	abhiṣekaḥ	kṛtas	trīṇi	gosahasrāṇi
	abhiseko	kito	tīni	gosahassāni
	bath.NOM.SG.M	made.NOM.SG.M	3.NOM.PL.N	cow.thousand.NOM.PL.N
	da(t)tāni			
	dattāni			
	dattāni			
	given.NOM.PL.N			

(Nasik Cave Inscription No. II; Bühler 1881, 99-100)

"And by order of the supreme lord I went in the rainy season to liberate the Uttamabhadra chief who was besieged by the Mālayas […] Afterwards I went to the Puṣkaras (Pokṣaras), and I bathed there, and gave three thousand cows". (Bühler's translation)

* This is a virtual Sanskrit form; a more appropriate form would be *ājñayā*.

** The sign + indicates sandhi applying or failing to apply across linguistic forms.

Here again, many nominal and adjectival stems are Sanskrit in form, such as *varṣā* 'rain, rainy season' vs. Pkt. *vassā* or the name *u(t)tamabhadra-* vs. *uttamabhadda-*. The place name *pokṣarāṇi* for Pkt.

pokkharāni is more complex: its *o* is Prakritic, and its *kṣ* is hyper-Sanskrit, instead of the proper Sanskrit *ṣk*.[8]

Many other forms are lexically Prakritic, such as the forms *abhiseko* and *tīṇi*; the *ti*-stem form *añātiyā* which is both lexically and morphologically Prakritic (the correct Sanskrit form would be *ājñayā* from the *ā*-stem *ājñā*); and the ending of *mālayehi* which is likewise Prakritic (vs. Skt. *-ai(ḥ)*). Beyond that, there are many cases of Prakritic sandhi or lack of sandhi, but also a few Sanskritic sandhi forms, such as *tatosmi* (/tataḥ+asmi/ 'then I.NOM.SG.M'). Elsewhere in this inscription, there is an example of hyper-Sanskrit sandhi (*devatābhy**aḥ*** *brāhmaṇebhyaś ca* 'for the deities and the brahmins', with *-aḥ* before voiced *b-* where Sanskrit would have *-o)*. Relevant verbal forms are rare in EHS, but note the verbal-noun form *bhoj**āpayi**trā* 'causer.to.eat.INS.SG.M' (vs. Skt. *bhojayitrā*), with Prakritic 'double causative' marking,[9] found elsewhere in the text. Other inscriptions offer examples like *sahisya* (vs. Skt. *saheḥ*), with the same Prakritic development as in (3a).[10]

Damsteegt also documents that there were at least three different historical stages of EHS, with sanskritization becoming increasingly stronger in later stages, but with some Prakritic forms, especially the instrumental-plural suffix *-ehi*, stubbornly persisting throughout the entire period. Sanskritization, thus, was a continuing process.

While Damsteegt and Salomon see in EHS a stage of sanskritization, similar to that in BHS, Sen argues that both types of language use reflect a 'Spoken Sanskrit', an 'unstable literary or business language varying according to time and place', and he refers to Buddhist Hybrid Sanskrit as simply 'Buddhist Sanskrit'.

In the case of EHS, Sen's characterization ignores the fact that EHS is a transitional stage in the inscriptional record between earlier Prakrit and later Sanskrit, with sanskritization proceeding through several chronological phases. Salomon (1998) therefore is probably correct in considering BHS, too, to reflect a transitional phase of sanskritization.[11] Moreover, Salomon (1983) demonstrates

8 *kṣ* and *ṣk* both resulted in Ptk. *kkh*, but the correspondence *kṣ*: *kkh* is more common.

9 On this formation see Edgerton 1946.

10 For a detailed study of EHS morphology see Damsteegt 1977.

11 For a discussion of alternative theories see Salomon 1998, 81-6. Yet a different perspective is that of Ollett (2017, 44) who seems to look at the replacement of Prakrit forms like *khattapa* by Sanskritic *kṣatrapa* as involving phonological replacement of clusters like *kh* and *tt* by *kṣ* and *tr,* possibly under the influence of the extreme northwestern Gandhari Prakrit which retained such clusters. However, this proposal fails to explain a large number of other BHS (and EHS) phenomena such as the *au* of *k**au**tūhala* 'curiousness' (vs. Pkt. *k**o**tūhala*); for all the other Middle Indo-Aryan varieties, including Gandhari, changed Old Indo-Aryan *au* to *o*. Phenomena like these can only be explained in terms of lexical transfer from Sanskrit.

that Buddhist Hybrid Sanskrit is only one form of Buddhist Sanskrit – the more or less standard Sanskrit of Aśvaghoṣa clearly differs from the Hybrid Sanskrit of the Mahāsāṅghika school of Buddhism (even though it contains a few partly sanskritized Prakritic forms such as *saced* 'if', probably reflecting technical language of Buddhist argumentation).

4 Chronological Problems and the Testimony of EHS

As noted e.g. by Edgerton (1953, 4), BHS underwent increasing sanskritization 'from the very beginning of its tradition as we know it (that is, according to the mss. we have)'.

Especially instructive is the *Mahāvastu* of which we now have two editions, based on chronologically different manuscripts (Senart 1897 and Marciniak 2019). The edition by Senart, based on later manuscripts, shows a higher incidence of sanskritization than Marciniak's edition, which is based on recently found earlier medieval manuscripts; see e.g. the examples in (5). Example (5a) shows a Prakritic neuter form of the demonstrative *eta-* in Marciniak vs. a Sanskritic form in Senart; (5b) and (5c) exhibit Prakritic loss of final stop vs. Sanskritic presence of the stop (but the Prakritic *kā-* instead of *kāś* is retained even in Senart); and (5d) exemplifies Prakritic final *-o*, while Senart applies Sanskrit sandhi appropriate to the phonological environment. Evidently, later scribes introduced forms that were more in conformity with the Sanskrit norms that they were familiar with. This might also explain the fact that verb forms that are relatively similar to their Sanskrit counterparts as in (6a) are commonly fully sanskritized, whereas forms that are markedly different are not, as in (6b). In some cases, metrical concerns may also have prevented sanskritization.[12] For instance *bhavitvāna*, containing four syllables, cannot be replaced by disyllabic *bhūtvā*.

(5)		Marciniak	Senart	Sanskrit	
	a.	etaṁ	etad		'this.ACC.SG.N'
	b.	yāva	yāvat		'as long as'
	c.	kāci	kācit	kāścit	'some.NOM.SG.F'
	d.	upāgato	upāgataḥ	upāgataḥ	'(having) come up'
		tato	tataḥ		'then, after that'

12 Much of the *Mahāvastu* is in verse.

(6)		Prakritic	Sanskritized	Sanskrit	
	a.	bhāṣati	bhāṣate		‘speaks’
		vardhati	vardhate		‘grows’
	b.	abhūṣi		āsīt/abhavat/abhūt	‘was’
		bhavitvāna		bhūtvā	‘having been’
		bhavāhi		bhava	‘be IMPV’

This evidence raises the question of whether the original BHS of the early centuries AD might have been even less sanskritized – a question that for lack of attestations is difficult to answer. However, the late BC/early AD evidence of sanskritization in EHS is ‘set in stone’, not subject to the predilection of later scribes, and early BHS texts like the *Mahāvastu* are remarkably similar (making allowance for differences in genre) to the later phases of EHS. This evidence supports the view that even in its earliest, first-century AD stages, BHS would have undergone a significant degree of sanskritization.

5 Motivation and Institutionalization

In the early centuries AD, Sanskrit spread as the language of statecraft and of technical and fine literature throughout South Asia. Bronkhorst (2010) plausibly argues that this spread was propelled by Sanskrit-using brahmins who were experts in these matters. My own work (Hock 2019) suggests that an important vehicle for imparting the knowledge of Sanskrit consisted in brahmin-dominated, Sanskrit-medium schools.

As regards the use of Sanskrit by Buddhists, whose original texts were in different forms of Prakrit, it has been suggested that a switch toward Sanskrit was motivated by a desire to acquire competence in Sanskrit in order to compete with Sanskrit-speaking brahmins and to defend the Buddhist faith against brahmins in disputations at royal courts; see e.g. Salomon 1998, Bronkhorst 2010. The hybrid language of one school of Buddhism, then, might be considered to be an intermediate stage in this sanskritization process, comparable to EHS.

For reasons that are not recoverable, texts reflecting this intermediate stage became institutionalized and their Hybrid Sanskrit was adopted as the sacred language of one school of Buddhism, just as Pali was the sacred language of Theravada Buddhism, and Sanskrit was the sacred language of Brahmanism and became the language of most forms of Buddhism. In fact, Bronkhorst (1993) presents evidence that just as brahmins rationalized the difference between the Vedic language of their sacred texts and the Sanskrit of their current usage by declaring the two to actually be the same language, so members of the Mahāsāṁghika branch of Buddhism claimed that BHS and Sanskrit actually are the same language; and just as the

brahmins were able to account for Vedic pecularities in terms of rules in Pāṇini's grammar, Mahāsāṁghikas employed some of the same rules to justify peculiarities of BHS.

6 Mechanism

While sanskritization must clearly be recognized as a historical process, the term 'sanskritization' itself is merely descriptive, its mechanism is left unaccounted for.

We can safely rule out the idea that Buddhists employed brahmins competent in Sanskrit to sanskritize their texts. If they had done so, the brahmins would surely have produced grammatically proper Sanskrit texts. A more likely account for the mechanism underlying sanskritization would proceed along the following lines.

As noted earlier, an important vehicle for instruction in Sanskrit consisted in brahmin-dominated, Sanskrit-medium schools. Now, the curriculum of such schools started with the memorization of a thesaurus of Sanskrit words and of a simplified version of Pāṇini's grammar. A link between these memorized texts was established in later years through composition and the study of texts. The existence of such a form of instruction into the early twentieth century was confirmed by me in a 1980-81 research project on spoken Sanskrit, funded by the American Institute of Indian Studies.

Against this background, the mainly lexical sanskritization of BHS can be explained as the result of some of early Buddhist students only completing an elementary level of Sanskrit instruction.[13] This would leave them with a ready-made Sanskrit lexicon for replacing Prakrit words – hence the massive lexical sanskritization. As regards grammar, however, they would not yet have a full grasp of how to apply the memorized rules in practice – hence they would tend to use Prakrit sandhi and inflectional forms. Occasional hypersanskritizations further attest to the fact that Sanskrit grammar was only incompletely grasped.[14]

13 A reviewer objects that '[s]urely *some* of the Buddhist students would have gone beyond the first year and mastered 'proper' Sanskrit. It is hard to see how the weak attempts by beginners would have been codified into a literary vehicle.' Now, it is in fact true that 'some Budhhists' went beyond an elementary study of Sanskrit; but they produced a different literary vehicle – that of the bulk of Indian Buddhist tradition and also of poets and dramatists such as Aśvaghoṣa, whose language (as noted in Section 3) is more or less 'proper Sanskrit'. The problem is that Buddhist Hybrid Sanskrit cannot possibly accounted for as resulting from more than elementary exposure to Sanskrit grammar.

14 The evidence of EHS suggests a limited increase in familiarity with aspects of Sanskrit grammar during the late centuries BC and the early centuries AD.

The result, then, evidently became institutionalized as the language of the Mahāsāṁghika branch of Buddhism. Other branches, except for the Theravādins who used Pali, followed the general trend to full acquisition of Sanskrit.

Unexpected support for the likelihood of the present hypothesis comes from Kapstein's account of grammatically deficient, but lexically accurate Sanskrit compositions by Tibetans during the Middle Ages:

> Errors such as this [...] were likely due in part to the practice of teaching vyākaraṇa [grammar] and abhidhāna [practice] quite separately, [...] with almost no training in practical application. (2018, 470)

7 Some isolated Mixed Sanskrit varieties similar to BHS and EHS

Beside BHS and EHS there are several other forms of language use that exhibit a mixture of Prakrit (or Apabhraṁśa) features and Sanskrit, but these are limited to individual texts, not parts of broader attested textual traditions; and they exhibit features sufficiently different from BHS and EHS to merit separate discussion. One of the these is the so-called *Patna Dhammapada* (Cone-Ānandajyoti Bhikku 2017, Ānandajyoti Bhikku 2020), the other the mathematical "Bakhshali Manuscript" (Hoernle 1887, Hayashi 1995).

The *Patna Dhammapada* (PDh), considered to be associated with the Sāmmatīya branch of Buddhism, has been variously characterized as Buddhist Hybrid Sanskrit, Prakritic, Buddhist Prakrit, and Sanskritized Prakrit (Cone-Ānandajoti Bhikkhu 2017, 4), but Karpik (2023, 78) considers it 'an important milestone in the Sanskritisation of Pali'.

There are in fact a number of features shared with BHS, but there are also differences. Consider the sample passage in (7). (Prakrit forms are in italics, Prakrit inflectional endings in bold, Prakrit sandhi in small cap italics; unambiguously Sanskrit forms are underlined; plain roman indicates forms that could be both Prakrit and Sanskrit). Here again a virtual Sanskrit version is added for comparison, as well as a real Prakrit counterpart (from the Pali *Dhammapada*).

(7)	ākroś**i**	maṁ	avadh**i**	maṁ
	ākrośan	mām	avadhiṣur	mām
	akkocchi	maṁ	avadhi	maṁ
	abuse.PST.3PL	I.ACC.SG	hit.PST.3PL	I.ACC.SG
	ajini	maṁ	ahāsi	me
	ajayan	mām	aharan	me
	ajini	maṁ	ahāsi	me

Bhasha
3, 2, 2024, 175-192
e-ISSN 2785-5953

defeat.PST.3PL	I.ACC.SG	take.away.PST.3PL	I.OBL
ye		tānı	upanahyanti
ye		tāny	upanahyanti
ye	ca	taṁ	upanayhanti
REL.PRON.NOM.PL.M		DEM.ACC.PL.N	combine.PRS.3PL
veraṁ	tesaṁ	na	śāmyati ‖ 5 ‖
vairaṁ	teṣāṁ	na	śāmyati
veraṁ	tesaṁ	na	sammati
hatred.NOM.SG.N	DEM.GEN.PL.M	neg	quiet.down.PRS.3SG

'They abused me, they hit me, they defeated me, they robbed me; Those who combine these (actions), their hatred does not cease.'

As in BHS, some forms are sanskritized, such as *ākroś-* (vs. Pali *akkocch-*) and *śāmyati* (vs. *sammati*), and many others could be either Prakrit or Sanskrit, such as *na* NEG. However, many parts of example (7) are purely Prakritic, such as *veraṁ tesaṁ* 'their hatred' (vs. Skt. *vairaṁ teṣāṁ*); and throughout the text the ratio of Prakritic vs. Sanskritic forms is much higher than in BHS.[15]

Morphologically, past-tense forms exhibit the Prakrit generalization of the aorist, where Sanskrit would tend to use the imperfect (or the perfect), as in *ajini* 'they defeated' vs. Skt. *ajayan*; moreover, the endings of these forms are Prakritic, as in *-i* vs. Skt. *-an*. Forms that are relatively similar to Sanskrit, however, tend to be sanskritized; e.g. *upanahyanti* (vs. Pali *upanayhanti*). Elsewhere in the text we find Prakritic function words such as *sace* 'if' (vs. Skt. *ced*). External sandhi follows Prakrit norms, as in *tāni**i*** *upanahyanti* vs. *tān**y*** *upanahyanti*; similarly *man**o**śreṣṭhā* 'having mind as their leader' vs. *mana**ḥ**śreṣṭhā* elsewhere in the text. There are also some instances of hypersanskritisms, such as *ku**ś**īdaṁ* 'idle' (vs. Skt. *ku**s**īdaṁ*, Pali *ku**s**ītaṁ*) and *śīghra**śś**o* 'fast horse' (Skt. *śīghrā**śv**a*, Pali *sīgha**ss**o*).

Following Karpik (2023), these facts could be interpreted as indicating an early stage of the sanskritization found in BHS texts, in which case it might be tempting to attribute the greater degree of sanskritization in BHS to Sanskrit-proficient scribes. However, it is also possible that the *Patna Dhammapada* reflects a different development from the one underlying BHS, with a more minimal amount of sanskritization, in a different branch of Buddhism than the one associated with BHS.

15 Terminology central to Buddhism, such as *dhamma* '(cosmic) order; proper conduct' (Skt. *dharma*), *kamma* 'action' (Skt. *karman*), *bhikkhu* 'monk' (Skt. *bhikṣu*), *nibbāṇa* 'release' (Skt. *nirvāṇa*), remains unsanskritized in virtually all cases. On the other hand, as in Pali, *brāhmaṇa* 'brahmin' always appears in Sanskritic form (vs. Pkt. *bamhana*).

The possibility of different, independent sanskritization receives some support from the Bakhshali Manuscript (BM) whose base is not Prakrit (including Pali) but more likely Apabhraṁśa, a later form of Middle Indo-Aryan.

Like the *Patna Dhammapada*, the Bakshali Manuscript has been variously characterized. Hoernle labels its language as "the literary form of the ancient Northwestern Prakrit", with "a strange mixture of what we should now call Sanskrit and Prakrit forms" (1887, 10), 'an imperfect sanskritisation of the vernacular Prakrit' (1887, 14). Kay (1933, 11) describes it as "an irregular Sanskrit', whose 'peculiarities of spelling, sandhi, grammar […] are exceedingly common in the eleventh and twelfth centuries found in north-west India" (without, however, giving examples of other texts with the same characteristics). Hayashi (1995, 15) notes that the language 'has a number of peculiarities in common with the so-called Buddhist Hybrid Sanskrit', but also has 'strong affinity with Apabhraṃśa […], and, as expected from the find-spot (Bakhshālī near Mardan) of the manuscript, also with Old Kashmiri' (with reference to Grierson 1929); a final part of BM, however, is in an entirely different dialect, which is not discussed below.

An examination of the 'peculiarities' of the text suggests that Hayashi's comparison with Apabhraṁśa, the latest stage of Middle Indo-Aryan and the (near-)ancestor of the Modern Indo-Aryan languages, points in the most likely direction.[16]

As in the other mixed varieties, let us start with a short sample of BM; see (8), with similar formatting as for PDh.; for the forms with dotted underline see the discussion below. Here again, an attempt at a Sanskrit version is added for comparison.[17] The grammatical glosses refer to what may be the correct Sanskrit equivalent.[18]

16 Hayashi's comparison with the Old Kashmiri text described by Grierson is less likely to be correct. Grierson describes the language of that text as similar to learnèd texts in Modern Indo-Aryan languages, with a heavy amount of lexical borrowing from Sanskrit, but with the grammatical endings of Old Kashmiri (1929, 77).

17 I don't feel competent to add a virtual Apabhraṁśa version, especially since this is a technical, mathematical text. Where individual virtual Apabhraṁśa forms are cited, the information they are based on comes from Tagare 1948.

18 Problems with the grammatical endings of the BM sometimes make it difficult to guess what the correct Sanskrit equivalent should be.

(8)	[rā]japutr**o**	dvay**o**	*k**e**ci*	nṛpati**s**sevy**a**	
	rājaputrau	dvau	kaucin	nṛpatisevyau	
	Rajput.NOM.DU.M	2.NOM.DU.M	some.NOM.DU.M	king.master.NOM.DU.M	
	santi	vai**ḥ** |			
	sto	vai			
	be.PRS.3DU	particle			
	*mekā*syāhne	*dvaya**ṣ**ṣa**d**bhāg**ā***	dvitīyasya		
	ekasyāhne	dviṣaḍbhāgo	dvitīyasya		
	1.GEN.SG.M + day.LOC.SG.N	2.[1]/6.part.NOM.SG.M	2[nd].GEN.SG.M		
	*div*arddhika**ṁ**				
	dvyardhikasya				
	1[1]/2.GEN.SG.M				
	prathamena	dvitīyasya	daśa	dīnāra**s**	dattavān |
	prathamena	dvitīyāya	daśa	dīnārā	dattāḥ
	first.INS.SG.M	second.DAT.SG.M	10	dīnāra.NOM.PL.M	given.NOM.PL.M
	kena kālena	samatā**ṁ**	gaṇayitvā		
	kasmin kāle	samatā	gaṇayitvā		
	what time.LOC.SG.M	equality.NOM.SG.M	calculate.CVB		
	vada+āśu	me ||			
	vadāśu	me			
	tell.IMPV.2SG+quickly	I.obl			

'Two Rājputs are the servants of a king. The wages of one per day are two and one-sixth of the other one and one half. The first gives to the second ten *dīnāras*. Calculate and tell me quickly, in what time will there be equality (in their possessions)?' (Hoernle's translation)

As can be seen, almost all lexical items are in Sanskritic shape. Contrast, e.g., the Apabhraṁśa numeral *doṇṇi* or *beṇṇi* '2' with the *dvayo* of our text, Ap. *pahila-* 'first' with *prathama-*, Ap. *pu**tt**a* 'son' with *-pu**tr**o,* or Ap. *v**aahi*** 'say IMPV' with *v**ada***.

However, the case usage, such as *dvitīyasya* (GEN) 'to the second' for *dvitīyāya* (DAT) and *kena kālena* (INS) 'at what time' for *kasmin kāle* (LOC) reflects Prakrit or Apabhraṁśa grammar. Similarly, the use of plural instead of dual forms, as in *keci* (PL) 'some' vs. Skt. *kau-cin* (/kau-cid/ DU) is attributable to Prakrit or Apabhraṁśa influence.[19]

More significantly, several forms and constructions suggest that the composer's (or composers'?) grasp of Sanskrit was limited.

19 Some peculiarities of BM do not seem to be explainable in terms of any other known language use. These include the doubling of word- or stem-initial consonants in (8) in such forms as *nṛpati-**s**sevya* for *nṛpati-sevya-*, as well as the use of *m, r,* or *s* as hiatus breakers. There is also the form *meka-* 'one', which Hayashi (1995, 35) plausibly explains as reanalyzed from *ekameka-* 'one-another'.

Consider the form *dvaya-* 'two'. The correct Sanskrit formation is *dvau* (and inflectional variants) and it is used only in the dual; Skt. *dvaya-* means 'of two kinds, double' and is normally used only in the singular or the plural, not in dual reference.[20]

Further, the verb form in *prathamena* (INS) *daśa dīnāra dattavān* 'the first one […] gave' is ungrammatical in Sanskrit, which would require the *ta*-participle form *dattā(ḥ)* instead, agreeing with *dīnārā(ḥ)*.[21] The past participle in *-tava(n)t* is construed as active, requiring a nominative subject (*prathama(ḥ)* NOM *daśa dīnāra dattavān* 'the first one gave'); the past *ta*-participle is construed as ergative, with instrumental agent marking (*prathamena* INS *daśa dīnāra dattavān* 'by the first one was given' = 'the first one gave'). Apparently the composer did not understand this difference and, focusing on the fact that both the *tava(n)t-* and the *ta*-form indicate past tense, used the *tava(n)t*-participle instead of the *ta*-participle because it better fits the meter (which calls for the scansion ⏑⏑– in this position). Note that Apabhraṁśa only uses the reflex of the *ta*-participle for past-tense reference.[22]

Further evidence that the composer(s) were insufficiently familiar with Sanskrit is the hyper-Sanskritism *vai**ḥ*** for Skt. *vai*.

Hoernle (1887, 15) and Hayashi (1995, 26-55) note that there is a great amount of confusion as regards case, gender, and number. As it turns out, this 'confusion' is characteristic of Apabhraṁśa, where sound change and analogy led to the loss of distinction between masculine and neuter, the attentuation of a formal distinction of feminine gender, and the reduction of the case system (Tagare 1948, 27, 105-6). More specifically, in his listing of morphological peculiarities, Hayashi (1995, 26-55 with general reference to Tagare 1948) gives numerous examples of case endings that are found in Apabhraṁśa, but not in earlier Prakrits, as well as endings that are found in both. These include the masculine/neuter nominative/accusative singular endings ***-a**, -aṁ, -**ā**, -āḥ, -āṁ, -**u**, -o* and the corresponding plural

20 The second occurrence of *dvaya-* seems to be a dittography; something like *d(v)iṣaḍbhāga-* would better fit the meter.

21 A reviewer suggests that *daśa dīnāra dattavān* should be interpreted as singular, because *daśa* is singular; but *dīnāra* would have to be in the plural (*dīnārā(ḥ)*) and the verb should agree with that plural.

22 A reviewer comments that "[m]ixed passive/active constructions are common in OIA and MIA languages; see especially Jamison 2000". To my knowledge, such structures are not at all common in either Old Indo-Aryan or the mainstream of Middle Indo-Aryan. Jamison (2000), to be sure, notes structures that are superficially similar in the extreme northern Niya Prakrit, but the morphosyntax differs considerably from Sanskrit or the mainstream of Middle-Indo Aryan. For instance, forms of 'be' are cliticized to first- and second-person forms of the *ta*-participle, leading to the loss of ergative alignment; third-person antecedents may be marked in the absolute case or the instrumental; instrumental marking is often found in traditional formulas or seems to be employed as a kind of 'differential agent' marker.

endings ***-a***, *-aṁ, -aḥ,* ***-ā***, *-āṁ,* ***-e***, *-o* (where bold indicates exclusively Apabhraṁśa variants). Forms like *rājaputr**o***, *dīnār**a***, *samatā**ṁ***, thus, can be accounted for as having Apabhraṁśa morphology. (Note that the final short vowel of *dīnār**a*** is guaranteed by the meter).

There remain certain forms that suggest that in some cases the composer(s) used correct Sanskrit grammatical forms and thus showed at least some familiarity with Sanskrit grammar. Consider e.g. *santi* 'they are' in example (8). While the singular present forms of the verb *as-* 'to be' are attested in Apabhraṁśa, the third plural *santi* is not found in the extensive list of attested present forms in Prakash (1975, 260-7). Perhaps, then, *santi* replaced a metrically equivalent Ap. *honti* 'they are' (from the Sanskrit root *bhū*), but perhaps *santi* was still used in Apabhraṁśa but did not happen to be found in the sample texts examined by Prakash (even for the earlier Prakrits, Pischel (1900, 350) notes that *santi* is 'selten' (rare).[23]

We can thus conclude with Hayashi that the base language of the Bakhshali Manuscript is most likely to have been Apabhraṁśa and that the composer(s) of the text had access to the Sanskrit lexicon but were insufficiently informed about Sanskrit grammar.[24]

Although details differ, the language of BM, thus, exhibits a similar pattern of sanskritization to BHS (and EHS), but with a different base language – Apabhraṁśa vs. Pali/early Prakrit.

Most important, however, the example of BM clearly shows that the sanskritization of BHS (and the *Patna Dhammapada*?) is not unique, but that similar developments could arise independently, presumably because of the overarching prestige of Sanskrit for religious and scholarly discourse.

8 Conclusions

Buddhist Hybrid Sanskrit reflects incomplete sanskritization, mainly lexically and especially in noun and adjective stems. The process was most likely motivated by trying to defend Buddhism against Sanskrit-using brahmins in disputations at royal courts. A likely account for the mechanism of sanskritization is that it is the result of incomplete learning in a school system which required memorization of a Sanskrit thesaurus and a formal grammar of Sanskrit, and that the educational process did not continue to later stages at which the

23 Note however that the plural *santi* vs. the expected Sanskrit dual *staḥ/sto* reflects the Prakrit and Apabhraṁśa loss of the dual category.

24 A reviewer suggests to consider the language of the BM to be 'Vernacular Sanskrit', along the lines of Salomon 1989. However, doing so should not preclude attempts at trying to determine a vernacular, non-Sanskrit base for this form of language.

two components would be firmly linked with each other. The memorized thesaurus, then, would serve as a source for replacing Prakrit words with Sanskrit ones. Function words and inflectional suffixes, as well as verb forms that strongly differed from Sanskrit, however, to a great extent were not replaced. The result, then, is similar to general tendencies in linguistic borrowing, with function words and inflectional affixes less likely to be adopted (Hock 2021, 414). Occasional hypersanskritizations further attest to the fact that Sanskrit grammar was only incompletely acquired. The entire process is similar to what is found in the EHS of late first century BC/early centuries AD inscriptions, in the *Patna Dhammapada* and the Bhakshali Manuscript, as well as in medieval Sanskrit compositions in Tibet. Unlike these forms of language use, however, early Buddhist texts composed in BHS came to be institutionalized and their language was canonized as a defining feature of a specific school of Buddhism.

In many ways, then, BHS resembles the kinds of languages that have been referred to as 'Bilingual Mixed Languages' (BML), discussed e.g. in Bakker & Matras 2003 and the contributions to that volume. BMLs are commonly characterized as having undergone 'relexification' of one language based on the lexicon of another, as a deliberate process of creating a separate linguistic identity. The fact that sanskritization in BHS was largely confined to the lexicon could be considered an example of relexification, and the fact that it appears to have been motivated by an attempt at competing with Sanskrit-speaking brahmins could be considered a deliberate action. However, it is by no means clear that sanskritization itself was a deliberate act of creating a separate linguistic and social identity. Rather, the separate linguistic identity of BHS seems to be the result of the institutionalization of a product of incomplete acquisition, in terms of lexical borrowing from Sanskrit without comparable acquisition of the grammar. But even in the case of paradigm examples of 'Bilingual Mixed Languages', the creation of a separate linguistic and social identity may well have been an after-effect of massive lexical borrowing (Hock 2021, § 13.6).

With these reservations, then, BHS could be added to the set of hybrid languages subsumed under the label 'Bilingual Mixed Languages'.

Bibliography

Ānandajoti, B. (2020). *A Comparative Edition of the Dhammapada: Pāḷi Texts with Parallels from Sanskritized Prakrit, Edited Together with a Study of the Dhammapada Collection*. 4th rev. ed.
https://www.ancient-buddhist-texts.net/Buddhist-Texts/C3-Comparative-Dhammapada/Comparative-Dhammapada.pdf

Bakker, P.; Matras, Y. (eds) (2003). *The Mixed Language Debate: Theoretical and Empirical Advances*. Berlin; New York: Mouton de Gruyter.

Bronkhorst, J. (1993). "Buddhist Hybrid Sanskrit: Its Original Language". Mishra, K.N. (ed.), *Aspects of Buddhist Sanskrit = Proceedings of the International Symposium on the language of Sanskrit Buddhist Texts* (Sarnath, 1-5 October 1991). Sarnath; Varanasi: Central Institute of Higher Tibetan Studies, 396-412.

Bronkhorst, J. (2010). "The Spread of Sanskrit". Franco, E.; Zin, M.; Davidson, R.M. (eds), *From Turfan to Ajanta: Festschrift for Dieter Schlingloff on the Occasion of His Eightieth Birthday*, vol. 1. Bhairawaha; Rupandehi: Lumbini International Research Institute, 117-39.

Bühler, G. (1881). "Inscriptions". *Archaeological Survey of Western India*, vol 4. Bombay: Government Central Press.

Cone-Ānandajoti Bhikku (2017) = *Patna Dhammapada: Patna dhamma verses*.Transcribed by M. Cone; translated (and edited) by Ānandajoti Bhikku.
https://www.ancient-buddhist-texts.net/Texts-and-Translations/Patna-Dhammapada/index.htm

Damsteegt, Th. (1978). *Epigraphical Hybrid Sanskrit*. Leiden: Brill.

Edgerton, F. (1946). "Indic Causatives in *-āpayati* (*-āpeti, -āvei*)". *Language*, 22, 94-101.

Edgerton, F. (1953). *Buddhist Hybrid Sanskrit Grammar and Dictionary*, vol. 1. New Haven: Yale University Press.

Grierson, G.A. (1929). "The Language of the *Mahā-Naya-Prakāśa*: An Examination of Kāshmīrī as Written in the Fifteenth Century". *Memoirs of the Asiatic Society of Bengal*, 11(2), 73-120.

Hayashi, T. (1995). *The Bakhshālī Manuscript: An ancient Indian mathematical treatise*. Groningen: Egbert Forsten.

Hinüber, O. von. (1986). *Das ältere Mittelindisch im Überblick*. Wien: Österreichische Akademie der Wissenschaften.

Hock, H.H. (2019). "Foreigners, Brahmins, Poets, or What? The Sociolinguistics of the Sanskrit 'Renaissance'". *Journal of Asian Linguistic Anthropology*, 1(2), 10-7.

Hock, H.H. (2021). *Principles of Historical Linguistics*. 3rd rev. ed. Berlin; Boston: de Gruyter Mouton.

Hoernle, R. (1887). *On the Bakshali manuscript*. Vienna: Alfred Hölder.

Jamison, S.W. (2000). "Lurching Towards Ergativity: Expressions of Agency in the Niya Documents". *Bulletin of the School of Oriental and African Studies*, 63(1), 64-80.

Kapstein, M. (2018). "Other People's Philology: Uses of Sanskrit in Tibet and China, 14th-19th Centuries". *L'espace du sens: Approches de la philologie indienne/The space of meaning: Approaches to Indian philology*.
https://hal.archives-ouvertes.fr/hal-02190087

Karpik, S. (2023). "Light on Epigraphic Pali: More on the Buddha Teaching in Pali". *Journal of the Oxford Centre for Buddhist Studies*, 23, 41–89.

Kay[e], G.R. (1933). "The Bakshali Manuscript: Early Hindu Mathematics: A Study in Mediaeval Mathematics". *Indian Historical Researches*, 24(1).
https://ia601505.us.archive.org/4/items/in.ernet.dli.2015.33644/2015.33644.The-Bakhshali-Manuscript--Vol-1.pdf

Marciniak, K. (ed). (2019). *The "Mahāvastu"*: *A New Edition*, vol. 3. Tokyo: Soka University.

Ollett, A. (2017). *Language of the Snakes: Prakrit, Sanskrit, and the Language Order of Premodern India*. Oakland: University of California Press.

Pischel, R. (1900). *Grammatik der Prakrit-Sprachen*. Strassburg: Trübner.

Prakash, R. (1975). *Verb-Morphology in Middle Indo-Aryan*. Delhi: Munshiram Manoharlal.

Salomon, R. (1983). "The Buddhist Sanskrit of Aśvaghoṣa's Saundarananda". *Wiener Zeitschrift für die Kunde Südasiens*, 27, 97-112.

Salomon, R. (1989). "Linguistic variability in Sanskrit". Caillat, C. (ed.), *Dialectes dans les littératures indo-aryennes*. Paris: Collège de France, 275-94.

Salomon, R. (1998). *Indian Epigraphy: A Guide to the Study of Inscriptions in Sanskrit, Prakrit, and the Other Indo-Aryan Languages*. New York; Oxford: Oxford University Press.

Sen, S. (1977). "On Buddhist (Hybrid) Sanskrit". *Bulletin of Tibetan Studies*, 1, 5-8. https://api.repository.cam.ac.uk/server/api/core/bitstreams/acc1569f-bd84-494b-87cd-d2b0f37432f6/content

Senart, É. (ed.) (1897). *Mahāvastu avadānam: Le Mahâvastu*, vol. 3. Paris: Imprimerie Nationale.

Tagare, G.V. (1948). *Historical grammar of Apabhraṁśa*. Poona: Deccan College Postgraduate and Research Institute.

e-ISSN 2785-5953

Bhasha
Vol. 3 — Num. 2 — Ottobre 2024

Observations on Gāndhārī Orthography and Phonology: ST Clusters and Related Problems

Jakob Halfmann
University of Würzburg, Germany

Abstract This article re-evaluates the phonological development of Old Indo-Aryan consonant clusters of the shape sibilant + coronal plosive in Gāndhārī and the representation of their outcomes in the Kharoṣṭhī script. In this context it also deals with several developments involving retroflex plosives and their written representation. For the solution of orthographic and phonological problems related to these areas in Gāndhārī, a primary role is given to evidence from those modern Indo-Aryan languages of the far northwest which can be assumed to be the closest living relatives of written Gāndhārī.

Keywords Gāndhārī. Kharoṣṭhī. Kohistani. Orthography. Phonology.

Edizioni Ca'Foscari

Peer review

Submitted 2024-07-11
Accepted 2024-09-17
Published 2024-12-18

Citation Halfmann, J. (2024). "Observations on Gāndhārī Orthography and Phonology: ST Clusters and Related Problems". *Bhasha*, 3(2), 193-212.

DOI 10.30687/bhasha/2785-5953/2024/02/002

The value of the testimony of the modern Indo-Aryan languages of the far northwest for the evaluation of questions in Gāndhārī orthography and phonology has been recognized early on by Gāndhārī scholars, who have drawn on the documentation of these languages time and time again to help solve their philological problems.[1]

One of the thorniest issues in Gāndhārī orthography is the seeming graphemic overabundance in the area of coronal plosives and clusters of sibilants followed by coronal plosives and the puzzling patterns of variation and substitution that exist between these graphemes. As I will attempt to show in this paper, the evidence from the modern Indo-Aryan languages of the northwest has introduced additional confusion in this area, but, at the same time, it can also lead the way to a possible solution.

Though it still often appears in the literature, I will avoid using the word 'Dardic' for these languages, which, as Morgenstierne (1961, 139) famously and correctly pointed out, is

> simply a convenient cover term to denote a bundle of aberrant Indo-Aryan hill languages, which in their relative isolation [...] have been in a varying degree sheltered against the expanding influences of IA Midland (Madhyadesha) innovations, being left free to develop on their own.

As the discussion will show, this term is not just irrelevant for linguistic subclassification, it has even actively hindered an understanding of the actual relation between Gāndhārī and the linguistic landscape of the Indo-Aryan northwest, because its use as a category tends to encourage a uniform treatment of such widely differing languages as Pashai and Kashmiri, and to lead to the assumption that these two and all languages in between are equally capable of shedding light on literary Gāndhārī.[2]

The Kharoṣṭhī script has a number of graphemes that appear in words where the Old Indo-Aryan (OIA) etymology suggests the original presence of sibilants, coronal plosives or both. Three of these were certainly graphemes for voiceless sibilants – alveolar 𐨯 <s>, palatal 𐨰 <ś> and retroflex 𐨱 <ṣ>. Two other graphemes certainly stood for voiceless coronal plosives – dental unaspirated 𐨤 <t> and dental aspirated 𐨣 <th>.[3] In addition to this, there were five

1 I would like to thank Robert Tegethoff and the two anonymous reviewers for helpful comments on the first draft

2 Kashmiri in particular does not have much in common with the languages of the far northwest beyond having been lumped together with them by Grierson (1919), but quotations of Kashmiri forms can still sometimes be found in analyses of Gāndhārī phonology.

3 These graphemes stood exclusively for voiceless sounds at least at the time of the initial conception of the script and they continued to do so in word-initial position.

more graphemes with less certain phonetic interpretation that are also in some way related to the domain of sibilants and coronal plosives – 𐨯, 𐨰, 𐨱, 𐨲 and 𐨳 – respectively transcribed with the conventional symbols <st>, <ṭh>, <ḍh>, <ṭ> and <ḍ>. It is the interpretation of these five graphemes that this article is concerned with.

1 The Grapheme 𐨯 <st>

For 𐨯 <st>, Franke (1906, 511) suggested an interpretation as /st^{h}/, based on the shape of the *akṣara*, which resembles a 𐨣 <th> with a modifying stroke. Brough (1962, 75) cautiously accepted this idea, but found that the transcription /st/ would be more justified, as it generally stands for etymological unaspirated /st/. This interpretation is maintained by Baums (2009, 164). In later sanskritized texts modified graphemes are introduced, which seem to reflect an attempt to distinguish Sanskrit /sth/ from /st/ in Kharoṣṭhī writing (Strauch 2012, 153), indicating that this distinction was originally absent from the script.

Based on the data of Baums and Glass (2002), it seems that words with etymological /st/ are very consistently spelled with <st>, the only notable exception being <thuba ~ thuva> 'stupa' ~ OIA *stūpa-*, which is attested almost exclusively with <th> (once as <tubha> with <t>). Since words with etymological /st/ are otherwise only very rarely spelled with <th> (or <ṭh> etc.) and in this way are clearly distinguished from those with etymological /sth/, we may assume a general preservation of the cluster /st/ (and also of /str/) in spoken Gāndhārī.

The few attested spellings with <th> (and twice <dh> ~ OIA /ṃst/, a development specific to the Khotan Dharmapada), which are summarized in Table 1, can be assumed to be either loanwords from Pali, which is very likely in the case of amply attested <thuba ~ thuva> (cf. P. thupa), or – in the case of manuscripts – interferences from prototypes originally written in Pali or other central Prakrits, where the development /st/ > /(t)th/ is regular **[tab. 1]**.[4]

Lenition processes in intervocalic position later offset this one-to-one relation of graphemes to speech sounds.

4 The spelling of the prepositions corresponding to OIA *purastāt* and *adhastāt* with <ṭh> may be due to a early reanalysis of these forms as containing the suffix *-stha-*. If this is the case they would then have gone through the developments suggested for *sth* below.

Table 1 Attested cases of Gāndhārī spelling <th> for etymological /st/ (except <thuba ~ thuva> 'stupa'). Data, etymologies and text abbreviations from Baums, Glass 2002

Attested Spellings	Cognates
Mān XI 12 <dhramasaṃthave> 'familiarity with the dharma'Dhp^{K} 250 <sadhavu>; Arthp 844 <sathavaṇi>'familiarity'	Skt. dharmasaṃstava-, P. dhammasaṃthavaSkt. saṃstava-, P. saṃthava
Dhp^{K} 60 <paḍisadharaguti> 'with guarding of goodwill'	Skt. pratisaṃstāragupti-, P paṭisaṃthāragutti
Dhp^{K} 154 <avathaṇi> 'cast off' (N.Pl.Dir)	Skt. apāsta-, P. apattha
<śatha>, <śathu> 'teacher'3x in CKM 415, once in Av^{L1}, beside many more attestations with <st>	Skt. śāstar-, P satthar
<thiṇa> 'woman' (Gen. Pl.) Once (Dhp^{K} 174), beside many more attestations with <str>	Skt. strī-, P. itthī, itthi, itthikā

2 The Grapheme 𐨞 <ṭh>

The situation is more complicated with regard to <ṭh>. Here, the OIA etymological correspondences are both *ṣṭ* and *ṣṭh*, seemingly without regard for aspiration. Despite the conventional transcription as <ṭh>, the phonetic value of this grapheme is now generally held to be a cluster /ṣṭ/ [ʂʈ] (von Hinüber 2001, 182; Baums 2009, 164). This is based on Brough's (1962, 77) argument that the corresponding OIA clusters are preserved in some modern northwestern Indo-Aryan languages. However, this interpretation causes a number of phonological and orthographic issues. First, it necessitates the assumption that the aspiration contrast was lost in the cluster /ṣṭ/. While some orthographic vacillation with regard to aspiration is well-known in Gandhari, the identification of <ṭh> as [ʂʈ] would necessarily mean that the expression of aspiration on this cluster was never even possible in Kharoṣṭhī. This parallels the situation with <st> to some extent, where no corresponding <sth> existed originally, but there the reflexes of OIA /st/ and /sth/ do not merge. Giving <ṭh> the value [ʂʈ] also creates an odd gap in the phonological system of Gandhari: the velar, dental and labial consonants all contrast a voiceless unaspirated variant (/k/, /t/, /p/) with a voiceless aspirated (/k^{ʰ}/, /t^{ʰ}/, /p^{ʰ}/), voiced unaspirated (/g/, /d/, /b/) and voiced aspirated (/g^{ɦ}/, /d^{ɦ}/ /b^{ɦ}/)

variant. The retroflex series stands alone in lacking a voiceless aspirate under the analysis given by Baums (2010).[5]

On the other hand, the neutralization of aspiration would be easy to understand as a result of a debuccalization of /ṣ/, during which the original sibilant becomes an aspirating element on the remaining, now geminated plosive, as is well attested in the other Middle Indo-Aryan languages: *st* and *sth* > *(t)th*, *ṣṭ* and *ṣṭh* > *(ṭ)ṭh* (von Hinüber 2001, 181). Later conjunct spellings of <ṣ> and <ṭh>,[6] which would be unexpected if <ṭh> on its own already stood for [ʂʈ], are also perfectly understandable as a (sanskritizing) spelling of [ʂʈʰ] if <ṭh> stood for [ʈʰ]. Additionally, in words like <nigaṭho> 'Jain' ~ Skt. *nirgrantha-*, Pali *nigaṇṭha* (Baums, Glass 2002) a sibilant would be completely out of place, but a plosive would not be surprising.[7]

While structural factors would thus favor an interpretation of <ṭh> as [ʈʰ], Brough's (1962) interpretation as a cluster is accepted by Baums (2009, 164) primarily "[o]n the evidence of the modern Dardic languages". This evidence, as presented by Brough (1962, 77), consists of a number of words in Pashai and Khowar: Pashai *aṣṭ*, Khowar *ošṭ* 'eight' ~ OIA *aṣṭá-*; Pashai *ǰeṣṭa:* ~ OIA *jyeṣṭha-ka-* 'elder' *kaṇiṣṭa:* 'younger' ~ OIA *kaniṣṭha-ka-*.[8] These languages indeed preserve the cluster, as do Khowar's closest relative IA Kalasha (*aṣṭ* 'eight'), the languages of the Gawar-Bati group (Gawar-Bati *aṣṭ* 'eight') and the languages of the Shina group on the northeastern mountain periphery (Gilgit Shina *ãṣ* 'eight', with later loss of final *ṭ*).[9] The same preservation is evident in the Indo-Aryan loanword layer in the Nuristani languages (e.g. northeastern Katë *uṣṭ* 'eight', *ǰiṣṭ* 'elder').[10]

Of the two languages cited by Brough (1962), Khowar in particular is hardly a reliable witness to the ancient language of Gandhāra, since, until a few centuries ago, it was spoken hundreds of kilometres away near Mulkhow and Torkhow directly south of Wakhan, where it was in contact with earlier forms of Wakhi, Yidgha and Burushaski

5 The situation is different for the dentals since a separate <th> exists.

6 Attested in EĀ^{Bm} v3 <kidriṣṭhiyo>, EĀ^{Bm} 9.3r3 <dreṣṭhavya>, CKD 511 o6 <tiṣṭhatu> (Baums, Glass 2002). See also Glass (2000, 133).

7 An anonymous reviewer points out that this word is also attested in a spelling with <ṭh> and suggests that this would point to <ṭh> in this word being an earlier, underspecified spelling for later <ṭh> (see section 1 below on the interpretation of <ṭh>). However, given the etymology, this would not be any less odd, especially since the same reviewer would like to interpret <ṭh> as a sibilant-plosive cluster as well (cf. fn. 27).

8 Forms cited here from Darra-yi Nūr Pashai as given by Morgenstierne (1956, 20, 85, 94); Khowar from Bashir (2023, 94).

9 Sources: IA Kalasha – Trail and Cooper (1999, 17); Gawar-Bati – FLI (2016, 10); Shina – Degener (2008, 26).

10 Author's own data. There are some possible exceptions with the reflex *ṭ*, e.g. Kt. NE *piṭ* 'mountain spur', *pṭi* 'back' ~ OIA *pṛṣṭha-*, *pṛṣṭha-ka-*. These may have been borrowed from the literary variety as opposed to the local *lingua franca*.

and lay far outside the orbit of lowland civilization (see Bashir 2022, 2-3, 31; Morgenstierne 1936, 661-2).[11] It lacks such clearly attested Gandhari innovations as /śr/ > /ṣ/ (cf. Khowar *ašrú* 'tears' (Bashir 2023, 6) ~ OIA *aśru-ka-*). The many varieties of Pashai, though they likely descend from the original language of the western Kabul valley, i.e. the Indo-Aryan language spoken in Lampāka and Nagarahāra (modern Laghmān and Nangarhār) (Morgenstierne 1967, 11),[12] cannot automatically be equated with the language of Gandhāra proper either. Morgenstierne (1934, 172) in fact concludes from his examination of isoglosses between Pashai and languages spoken further east "that the ancient dialect of the Peshawar District, the country between Tirah and Swat, must have belonged to the Tirahi-Kohistani type, and that the westernmost Dardic language, Pashai, which probably had its ancient centre in Laghman, has enjoyed a comparatively

11 Language shift to Khowar in what is today the southern half of Chitral could only have begun after the conquest of this area by the state of Chitral, which can be dated to the end of the seventeenth century (Cacopardo, Cacopardo 2001, 50). Shift from IA Kalasha to Khowar is culturally associated with conversion to Islam in Chitral, which began in the areas further to the south in the 19th century (Cacopardo, Cacopardo 2001, 53-4). In many areas that are today Muslim, conversion was only completed in the 20th century and the completion of language shift lagged behind by some decades (Cacopardo, Cacopardo 2001, 75-6).

12 Morgenstierne's connection of Pashai with the language of Lampāka and Nagarahāra has been rejected from an anthropological perspective by Keiser (1974) and Ovesen (1983, 325-7, 329; 1984, 397-400). Ovesen (1983) is oddly dismissive of historical linguistics in general and presents the entire field as something like a curiosity of the past, but some of the more cautious arguments presented in Keiser (1974) and Ovesen (1984) are probably correct in some regards: it is certainly unrealistic to imagine the displacement of the Pashai language in terms of invading Pashtuns literally chasing the Pashai up the mountains. The language more likely lost ground via language shift than via population displacements. This probably began earliest in the more well-connected areas of the Kabul valley whereas people in more remote mountain valleys held onto their original language longer. It is also right to question the idea that today's Pashai speakers in some way 'hold the inheritance' of the civilization of Lampāka and Nagarahāra, whereas the speakers of Nuristani languages (or the biological ancestors of the Pashtuns living in the region today) had no relation with it. Clearly the lowland civilization must have had an influence on the cultures of both linguistic groups and it is well known that the Nuristani languages received a large amount of loanwords from Indo-Aryan languages, including, e.g., religious vocabulary (see Halfmann 2023). It is also likely that the culture of the remote mountain regions differed in some regards from that of the main Kabul valley already in antiquity. Still, this in no way lessens the point that the speakers of earlier forms of Pashai must have had a closer linguistic connection to the former language of the surrounding lowlands than the speakers of earlier forms of Nuristani. Overall there can be little doubt that the Pashai varieties descend from sections of the same Indo-Aryan dialect continuum that also covered the area of the main Kabul valley in antiquity and in this way they can surely be considered the closest thing to modern continuations of the language of Lampāka and Nagarahāra. The same process of language shift to the prestigious language of the plains proceeding slowly upwards into the more remote valleys likely happened at least twice, first from Nuristani varieties to Indo-Aryan, and in more recent times from Indo-Aryan varieties to Pashto, with some regions remaining unaffected in both cases.

independent position since early times". Since the Peshawar district is precisely the area where we would expect a variety most closely corresponding to literary Gāndhārī to have been spoken, the best sources for Gāndhārī phonology would seem to be precisely these languages of the "Tirahi-Kohistani type".

Among these, Tirahi, the original language of the Tirah valley in the Spīn Ghar south of the Khyber Pass,[13] has a cluster *xt* as the reflex of OIA /ṣṭ/ (*axt* 'eight'), with *ṣ* shifted to *x* probably under the influence of northeastern Pashto which underwent the same sound change (Morgenstierne 1934, 166). In two attested words *ṣṭ* was preserved (*guṣṭa* 'house', *čuṣṭīē* 'hip'), possibly conditioned by the preceding *u*. We can therefore assume an earlier preservation of /ṣṭ/ also for Tirahi, but Tirah is geographically still closer to Nagarahāra than it is to central Gandhara and we must expect that linguistic innovations continuously spread from the southeast (the plains of central India) to the northwest, reaching the central Peshawar basin earlier than the more remote mountain regions. Furthermore, Tirahi is an outlier in this regard among the languages of the 'Kohistani type' referred to by Morgenstierne. Torwali, spoken in the lower part of upper Swat, shows a development of *ṣṭ(h)* > *ṭh* (*āṭh* 'eight', *mīṭh* 'fist' < *muṣṭi-*, *pīṭh* 'back' < *pṛṣṭha-*, *aŋūṭ* 'thumb' < *aṅguṣṭha-*),[14] and so does Indus Kohistani, on the northeastern edge of Gandhāra (*ȁṭʰ* 'eight', *muíṭʰⁱ* 'a handful').[15] The language of Woṭapūr and Kaṭārqalā, today all but extinct,[16] also has the outcome *ṭ* (*aṭ* 'eight', *pīṭ* 'flour' < *piṣṭa-*).[17] This language, though spoken in the Pech valley in Afghanistan near the end of its life, may have descended from the original language of Bajauṛ, based on what can be deduced from isoglosses, loanwords, and the oral history of the speech community (Buddruss 1960, 71-4;

13 This language was (fragmentarily) documented at a time when its speakers were settled in Nangarhār, having apparently been driven out of Tirah by Pashtuns. Their origin from Tirah is affirmed by Leech (1838, 782-3), who connects the displacement of the Tirahis with a campaign against the Roshaniya sect. This information was presumably gained from the Tirahi speakers he interacted with, but this is not made explicit in his report. Later authors (Stein 1925; Morgenstierne 1934) provide no independent confirmation of this story, but their informants seemingly also did not contradict it. The language is in all likelihood extinct today.

14 Data from Torwali (2020). It seems that aspiration is lost at the end of disyllabic words, as is suggested by *aŋūṭ* and also (though from the same root) *æŋīṭ* 'ring' < *aṅguṣṭhya-*.

15 Data from the Jijālī dialect as recorded by Zoller (2005, 74, 338).

16 On a recent survey in Afghanistan, Sviatoslav Kaverin still encountered some old men who remembered a few isolated words in the language.

17 Data from Buddruss (1960, 90, 121). Buddruss (1960, 17) notes his difficulties with the aspiration contrast in the language of Woṭapūr and Kaṭārqalā, which may have been either unstable or already lost under the influence of Pashto. In any case, the *ṭ* recorded by Buddruss may reflect an earlier **ṭh*.

Morgenstierne 1952, 125-6). This would place it on the northwestern edge of ancient Gandhāra. It seems, then, that the most reliable witnesses to the spoken language of Gandhāra rather point to <ṭh> standing for [ʈʰ], thus vindicating the conventional transcription.

It must be remarked that the languages of the 'Kohistani type' (to the exception of Tirahi) also show the development *st* > *th*, something that we can hardly presuppose for literary Gāndhārī, as has been shown above. At first glance, this would appear to weaken the argument that the phonological development of these languages can be adduced for the interpretation of Gāndhārī orthography. This issue, however, will become less problematic once we turn to a discussion of <ṭh>, which provides suggestive, though unfortunately not fully conclusive, evidence for the assumption that the development *st* > *th* postdates the assimilation of the other two ST clusters not just in literary Gāndhārī, but also in these modern languages.

The only remaining argument for the interpretation of <ṭh> as [ʂʈ] is the representation of loanwords with /ṣṭ/ as <ṭh> in Gāndhārī spelling. Only two examples of this are mentioned by Bailey (1949, 123-6) and taken up by Brough (1962, 76) and Baums (2009, 164). The first is the administrative title *ṣoṭhaṃga* (a kind of clerk), which may be connected with the Agnean (Tocharian A) word *ṣoṣtaṅk-* of similar meaning (attested once). This title is likely also attested in Bactrian, where it has the form *σωταγγο* without a second sibilant (Sims-Williams 2007, 266). The etymology of the word is quite unclear. Bailey's (1949, 123-6) Iranian derivation is entirely ad-hoc both phonologically (**fr* > *š*; in any case incompatible with Bactr. σ) and morphologically (an otherwise unknown agent noun suffix *-*tana*-). It is rejected by Carling and Pinault (2023, 497) and ignored by Sims-Williams (2007, 266). Carling and Pinault (2023, 497) in fact try to explain Agnean *ṣoṣtaṅk-* as a re-sanskritization of the Gāndhārī word (seemingly taking the conventional transcription <ṭh> literally), which they in turn derive (quite implausibly) from a compound of a Chinese loanword (either 寫 MChin. *sjaeX* 'to depict, to write' or 書 MChin. *syo* 'to write, written document', cf. Carling, Pinault 2023, 495) and a reflex of OIA *sthānika*-.[18] The probative value of *ṣoṣtaṅk-* can only be very limited as long as its etymology and the direction of borrowing are uncertain. The Bactrian form also calls the necessity of deriving the Gāndhārī form from a source word with a sibilant-plosive cluster into question.[19]

18 *ṭhaṃga* would not be an expected outcome of *sthānika-* in Gāndhārī, cf. *dhaniga-* 'wealthy' < *dhanika-*.

19 If the Bactrian form was borrowed from Gāndhārī, this would prove the reality of the assimilation to a plosive. If the Gāndhārī form is a borrowing, it is equally possible to assume that it is borrowed from Bactrian, so that <ṭh> need not necessarily represent a cluster.

The second example is the name of the Western Satrap *Cā̆ṣṭana-*, thus spelled in Brahmi inscriptions and possibly connected with Khotanese *caṣṭeṃ* 'master, as a proper name' (Bailey 1979, 100) and Pashto *ċəx̌tán* 'master'.[20] This name is attested on coins with the Kharoṣṭhī spelling <Caṭhana> (Bailey 1949, 125). As an example this is more convincing than the first, but if we proceed from the idea that the cluster /ṣṭ/ was assimilated and that this assimilation happened before the conception of the script, it is not surprising that writers of Gāndhārī would have encountered difficulty in representing such clusters in writing (and perhaps also in producing them in speech). The grapheme for [ʈʰ] may simply have been the closest possible approximation of a foreign [ʂʈ] that the writer could think of. One may also consider the possibility that writers with Sanskrit education were aware of the regular correspondence between Sanskrit [ʂʈ(ʰ)] and Gāndhārī [ʈʰ:], which could then be applied to the name of *Cā̆ṣṭana-*.[21] In any case, truly convincing evidence for the preservation of the cluster could only be provided by loanwords from Gāndhārī into other languages which would show the cluster, but, as far as I am aware, such evidence does not exist.[22]

20 Bailey's (1949, 125) etymology has no basis, as both the form of the root and the suffix he proposes have no parallels elsewhere.

21 Another point worth considering is that some areas in the sphere of Gāndhārī literacy, in particular Lampāka and Nagarahāra, had retained the cluster /ṣṭ/ in their spoken dialects. The writing system seems not to have been conceived with their speech in mind, but we cannot exclude the possibility that speakers of these dialects were not just aware of a diachronic/register-based correspondence between Sanskrit and Gāndhārī but also of a spoken dialect correspondence [ʂʈ(ʰ)] ~ [ʈʰ:] and therefore adopted the grapheme <ṭh> also for the representation of [ʂʈ] (and perhaps even pronounced it that way in reading).
An anonymous reviewer is not convinced by the idea that Gāndhārī speakers may have substituted [ʈ(:)] for [ʂʈ], given the availability of the cluster [st] in their language and the representation of the same name in Greek as <Τιαστανης> (*Tiastanēs*) etc. However, the phonological system of Greek is structurally rather different, having only one sibilant and no distinction between retroflex and dental consonants at all, so that different patterns of substitution are not surprising. We may also note that <st> is never used as a sanskritizing spelling of /ṣṭ/ in later texts.

22 An anonymous reviewer points out that the name of *Cā̆ṣṭana-* is also attested in Kuchean (Tocharian B) Brāhmī spelling as <caṣṭane> (nom. sg.) (`https://cetom.univie.ac.at/?F_B_caṣṭane`) and suggests that this would be evidence of the kind that I demand, since – in their view – the name is likely to have reached Kuchean via Gāndhārī. However, this need not be the case. The Brāhmī spelling with <ṣṭ> was certainly widely known, since *Cā̆ṣṭana-* ruled in Brāhmī-writing central India and since all coins which attest the Kharoṣṭhī spelling with <ṭh> also bear a parallel inscription in Brāhmī with <ṣṭ>. The transmission of personal names may in any case follow different trajectories than that of ordinary loanwords.

3 The Grapheme 𐨤 <ṭ́h>

We can now turn to a discussion of the Kharoṣṭhī grapheme <ṭ́h>. Brough (1962, 76-7), based on its appearance in place of etymological *sth*, suggested the reading /sth/, but already noted the issue of its unexpected use also for etymological *rth* in words like <aṭha> 'meaning, profit, sake' < OIA *artha-* and <caüṭha> 'fourth' < OIA *caturtha-*. In general, the use of <ṭh> for etymological *sth* is much less regular than the correlation of <st> with etymological *st*, since other spellings also occur quite frequently in the same places. A simple count of all attested spellings appearing in place of etymological *sth* in Baums and Glass (2002) gives the distribution summarized **[tab. 2]**.

Table 2 Spellings appearing in place of etymological sth attested in Baums, Glass 2002 [23]

Grapheme	Occurences
<sth>	9
<st>	24
<th>	83
<ṭ́h>	80
<ṭ>	14
<ṭh>	7*
<rth>	1
<h>	1

* 37 with the inclusion of the word <puraṭhita>, excluded here as an outlier.

The two most common spellings are <th> and <ṭh>, which shows that the sounds represented by these two graphemes must have been quite similar. On the other hand, retroflex spellings also appear with more than chance frequency. The frequent use of <th>, but also the appearance of <ṭ>,[24] renders Brough's (1962) assumption of a sibilant-plosive cluster /sth/ unlikely, even if one were to accept the idea that <ṭh> stood for such a cluster. Instead, a voiceless aspirated plosive seems quite probable. The relatively less common variants <st>

23 Only those forms are counted where the etymology and the relevant *akṣara* are listed as unambiguous by Baums, Glass 2002.

24 See the discussion below on the identity of this grapheme. An anonymous reviewer wonders why <ṭ> would appear at all in this context, referring especially to cases where it appears to represent the /ṭh/ of borrowed central MIA words, and why <ṭh>, if it really had the value [ʈʰ], was not used in all such cases instead. As there are several cases in which the spelling with <ṭ> is attested beside a spelling with <ṭh>, e.g. <kaṭha> beside <kaṭa> < OIA *kāṣṭha-* 'wood' or <ṭh(*ido)> (and <ṭhidaga> = *sthita-ka-*) beside <ṭido> < OIA *sthita-* 'stood', I would tentatively assume that these are instances of the general vacillation in the representation of aspiration in Gāndhārī.

and <sth> can easily be discarded as attempts at historical spelling. <st> would have been the only remaining sibilant-plosive cluster and therefore the only real option at first to represent Skt. /sth/. The conjunct spelling <sth> is a late, even more sanskritizing variant. On the other hand, the assumption of a voiceless aspirated plosive raises the question of the place of articulation. Dental and retroflex voiceless aspirated plosives are already represented by <th> and <ṭh>, so why would another grapheme <ṭh> be needed? And why would this grapheme also appear in the place of etymological /rth/, a cluster that has no sibilant element at all?

A more recent attempt to reconcile these facts was made by Baums (2009, 164-7), who diverged slightly from Brough's (1962) interpretation and suggested, based on the parallel of the Pali development *sth* > (*ṭ*)*ṭh*-, that <ṭh> may have been a newly invented grapheme for a new kind of sound, an alveolar that resulted from retraction of dental /th/ in the clusters /sth/ and /rth/, but without fully merging into the retroflex series. This explanation is very likely to be correct, and in fact it seems to be the only way to explain the existence of <ṭh> as well as its interchangeability with other graphemes. In the details, however, the idea runs into several issues, at least in the version of it that Baums proposes. His suggestion is the following:

> Within Gāndhārī, the earliest representation of this alveolar articulation would then be ṭh (primarily used for [ʂʈ]) in the Aśokan edicts and in the British Library avadānas, apparently doing duty both for an alveolar cluster [s̪t̪] (< OIA [st^{h}], parallel to G [st] and [ʂʈ]) and for an alveolar aspirated plosive [t̪^{h}ː] (< OIA [ɾt^{h}], parallel to G [t^{h}ː]). The modified sign ṭ̪h was then introduced to distinguish these two alveolar articulations clearly from the retroflex one. (Baums 2009, 165-6)

The idea that <ṭh> supposedly did 'double duty' for a sibilant-plosive cluster and an aspirated plosive is clearly a weak point in the theory. In fact, this interpretation discards the explanatory power of the alveolar hypothesis by assuming different phonetic results for /rth/ and /sth/ after all. Baums (2009, 164) considers the alternative assumption of an assimilation of OIA /sth/ into a voiceless aspirated plosive "unlikely on systematic phonetic grounds since none of the three OIA clusters [st], [ʂʈ] and [ʂʈ^{h}] underwent such assimilation in Gāndhārī". As we have seen, however, there is good reason to assume that the retroflex cluster did in fact also undergo assimilation. Significantly, the cluster /sth/ is not preserved in a single northwestern Indo-Aryan language, even in the most conservative ones that do preserve /st/ and /ṣṭ/. This can easily be illustrated with descendants of the OIA root √*sthā*, most of which reflect the stem *sthiya*- 'to be stood', which evolved into a copular verb in many languages of the region, or the

participle *sthita-* ‘stood’: Pash. *th-* ~ *t-* ‘to be’; GB. *th-* ‘to be’; Khow. *tʰiík* ‘to stay firmly/securely in one place’; Tir. *thī-* ‘to be’; Torw. *thu*, pl. *thi* ‘is, are’; IKoh. *thṵ̀* (m.), *thī̀* (f.) ‘is’, Shin. *th-* ‘to do’.[25] The same development is evident in the Indo-Aryan loanword layer in Nuristani (northeastern Katë *ti-* ‘to stand’, *tul* ‘field’ ~ OIA *sthala-*). We therefore have good grounds to assume that <ṭ́h> was in fact invented to represent a voiceless alveolar aspirated plosive [t̪ʰ], and never stood for a sibilant-plosive cluster.[26]

For Baums (2009), the retroflex spellings in the same places where <ṭ́h> is also used are reflections of an earlier convention, used to represent a more retracted articulation at a time when a grapheme had not yet been invented to distinguish alveolar from retroflex pronunciation. He furthermore assumes that the unstable intermediate position of the newly arisen alveolars led to “mergers with the retroflex and dental series in Pali and, ultimately, with the dental series in Gāndhārī and the Dardic languages” (Baums 2009, 165). I would argue, however, that the mergers in Gāndhārī and the surrounding varieties involved both the dental and the retroflex series. The key to the problem is again provided by the modern languages of the region.

As the examples < *√sthā* quoted above demonstrate, the alveolars indeed merged with the dentals in initial position. The consonant cluster /sth/ is not very common in Old Indo-Aryan, outside of derivatives of the root *√sthā*. However, one good example that is unrelated to this root can be found in the word *ásthi-* ‘bone’.

This word appears in Gāndhārī in the spellings <aṭ́hi>, <aṭhi> and <aṭi>. Excluding the latter spelling from the discussion for now, we thus see spellings with the graphemes for the alveolar and for the retroflex aspirated plosive under the interpretation suggested in this article. All modern Indo-Aryan languages of the region that have retained this word show reflexes of a development of the consonant cluster to *ṭṭh* (with secondary loss of aspiration and/or development of a nasal cluster from the geminate in some cases): Pash. *á:ṭṭhiː ~ aṇṭíː*;

25 Data sources: Pashai (Laurowan and Darra-yi Nūr) – Morgenstierne (1956, 179), Gawar-Bati – Morgenstierne (1950, 53), Khowar – Bashir (2023, 136), Tirahi – Morgenstierne (1934, 188), Torwali – Torwali (2020), Indus Kohistani (Jijālī) – Zoller (2005, 241), Shina (Gilgit) – Degener (2008, 308). Morgenstierne’s (1934, 169) assertion that in Tirahi “Postvocalic *st* remains (*ast* ‘hand’, *nast* ‘nose’), but initial *st-* results in *t-* (*thān* ‘house’, *thī* ‘he is’)” is not quite right. The relevant factor is not position inside the word but original aspiration, as his examples demonstrate (cf. OIA *hasta-* ‘hand’, *sthāna-* ‘place’).

26 An anonymous reviewer wonders why the result of the merger of /rth/ and /sth/ could not have been a sibilant-plosive cluster. While this is theoretically possible, it is probably less expected in terms of sound change typology (at least Kümmel 2007, 162, 231 lists more examples of retraction of plosives after [ɾ] than of fricativization of [ɾ] before plosives), and it certainly has less precedent within the region (/rth/ > /(ṭ)ṭh/ and /sth/ > /(ṭ)ṭh/, on the other hand, are both attested MIA sound changes). Additionally, we would expect at least one modern language to show traces of such a cluster as the outcome of /sth/ or /rth/, but not a single one can be found.

IA Kal. *aṭhí*; Gr. *aṇṭî*; Sh. *ãṭi* (all ~ suffixed *asthi-ka-*).[27] The same development is evident in its borrowed reflexes in the Nuristani languages (e.g. northeastern Kt. *aṭí* 'bone').

Previous etymologists, who assumed that *ṭh* could only derive from *ṣṭ(h)*, were puzzled by the fact that this development showed up also – even primarily – in those languages that otherwise regularly preserve *ṣṭ(h)*. Their way out of this problem was to assume a derivation from *aṣṭi-* (attested in the *Bhāgavatapurāṇa*) or *aṣṭhi-* (attested lexicographically), both f. and meaning 'seed, kernel, stone (of fruit)',[28] and to attribute its appearance in areas with preserved *ṣṭ(h)* to borrowing from "a language that has disappeared today" ("*une langue aujourd'hui disparue*") (Fussman 1972, 263) or from 'Ind.' (presumably meaning more central Indo-Aryan languages) (Turner 1962, T. 958).[29] This explanation has two weak points: first, the normal word for 'bone' in OIA is *asthi-* and it would be simplest to assume that this is also the lexeme that is continued by the normal words for 'bone' in modern Indo-Aryan languages. Forms only or primarily attested by Sanskrit lexicographers, on the other hand, have frequently turned out to be unreliable witnesses to Old Indo-Aryan as it was spoken, since they also include, e.g., forms that are re-sanskritized in an unetymological way. Secondly, borrowings from central India diffusing this far to the northwest are very rare. Accordingly, it would be very surprising if this happened to such an extent with a word that, to all appearances, lies outside of the realm of cultural vocabulary.

Taken together with the orthographic information from written Gāndhārī, which suggests an articulatory retraction of original *sth*, it seems more likely that *ṭh* is simply the regular intervocalic outcome of *-sth-* in the languages of the northwest. This also accords well with the orthographic merger of original *sth* and *rth* in written

27 Data sources: Pashai (Laurowan and Darra-yi Nūr) – Morgenstierne (1956, 22), IA Kalasha – Trail and Cooper (1999, 18), Gṛangali – Buddruss (1979, 32); Shina (Gilgit) – Degener (2008, 246). See also Fussman (1972, 262-4). Khowar *astī* 'bone' cited by Turner (1962, T. 982) appears to be a ghost word. Like Fussman (1972, 264), I was unable to confirm its existence in any Khowar lexical resource and the usual Khowar word for 'bone' is *koół* ~ *kʰół* (Bashir 2023, 67). IA Kal. *aṣ* 'shoulder' (only this form, not *aṣṭ*, is recorded by Trail, Cooper 1999, 17) would appear to be unrelated on semantic grounds.

28 The meaning 'bone' given with an asterisk in Turner (1962, T. 958) is nowhere attested as such, but is reconstructed in order to fit the data. The meaning 'bone of elbow or knee', which is attributed to *aṣṭhī-* by lexicographers, is probably extracted from the genuine forms listed in the bracket behind it by Turner (1962, T. 958), on which see Mayrhofer 1992, 143-4.

29 Turner (1962, T. 958) attributes also Nepalese *ãṭh* 'the ribs' and Sinhalese *aṭaya* 'bone' to *aṣṭhi-*, but for these the development *sth* > *ṭṭh* can already easily be assumed based on the more central MIA dialects that they must descend from, so that they, too can be derived from *asthi-*. The preservation of the old word for 'bone' only in these two marginal languages outside the far northwest is a typical pattern resulting from the outward spread of central innovations.

Gāndhārī, since the modern languages show reflexes of *ṭṭh* also as the outcome of *rth* (which could only occur intervocalically in OIA), e.g. Torwali *čoṭʰəm* 'fourth' (Torwali 2020) < OIA *caturtha-* (with a secondarily added ordinal suffix *-əm* extracted from forms such as OIA *pañcamá-* 'fifth' (Turner 1962, T. 4600) or from the Pashto ordinal suffix). Further evidence for this development may perhaps be found in the Indo-Aryan loanword <gāṭhaa> 'householder' (~ OIA *gṛhastha-[ka-]*) in Khotanese, which Loukota (2023, 24) derives from a Gāndhārī source. Here, the Brāhmī spelling with <ṭh> unambiguously expresses an aspirated voiceless retroflex plosive.[30]

With the assumption of such a conditioned sound change, we can also explain modern forms like GB. *ṭhān* 'place' (FLI 2016, 47), where the development of *sth* contrasts with GB. *thun* 'pillar' (FLI 2016, 44), as extractions from compounds with *-sthāna-* as a final member (amply attested in written Gāndhārī). The conditioned sound change could also give us a basis for the assumption that the Kohistani languages underwent the sound change *st* > *th* later than *sth* > *th-* ~ *-ṭh-* and are in this way closely comparable to written Gāndhārī, since intervocalic representatives of both clusters would have to have contrasting outcomes. Unfortunately, the central witness for the development of intervocalic *-sth-* has been replaced in all of them by the most common New Indo-Aryan word for 'bone' (~ Hindi *hāṛ* etc.)[31] and I have so far been unable to find another surviving lexeme with original intervocalic *-sth-* in the limited lexical resources that are available for these languages. The adoption of the innovative lexeme for 'bone' does again illustrate their status as 'early adopters' of linguistic innovations from central India, but makes it difficult to contrast words like Torwali *hāth* 'hand' < OIA *hasta-* directly with a retroflex reflex of *-sth-*. We can in this case only appeal to the likelihood that these languages earlier had a reflex of this word with a retroflex plosive, since that is what is attested in written Gāndhārī. The fact that *st* was also eventually assimilated may be related to the later spread of Panjabi varieties into the region, with which at least the

30 An anonymous reviewer disagrees with Loukota's (2023, 24) judgment that this loanword came to Khotanese from Gāndhārī, arguing instead that that it must be a loanword from central MIA (cf. Pali *gahaṭṭha*) based on its vocalism. The reviewer likely has in mind the usual Gāndhārī development of *r̥* > *i* as against the usual Pali development *r̥* > *a*. They add that the attested Gāndhārī forms <grahatha-> and <ghahaṭha-> of the same word would have to be considered "semi-naturalized loanwords". If this is correct, Khot. <gāṭhaa> would have less probative value.

31 Turner (1962, T. 13952), Fussman (1972, 262-4) and Mayrhofer (2001, 531) are right to keep this lexeme apart from OIA *asthi-*. The form *haḍḍa-*, attested late in Sanskrit and mostly by lexicographers, is likely not a genuine OIA form, but an introduction from MIA, cf. Pkt. *haḍḍa*. It can be plausibly derived from OIA *hārda-* 'located in the heart' (with metaphorical extension of 'heart' > 'center, inside [of the flesh]') with *rd* > *ḍḍ* and shortening of the vowel via the Two-Mora Rule.

language of Woṭapūr and Kaṭārqalā must have been in contact (Buddruss 1960, 74).

Returning to Gāndhārī spelling issues, it seems therefore that retroflex spellings in place of etymological /sth/ may have more than one explanation. In earlier texts we may indeed see an underspecified orthography in which alveolar sounds are represented as retroflex.[32] This is made plausible also by the secondary nature of the alveolar grapheme. But we would also not be surprised by retroflex spellings appearing again in later documents, at a time when the intervocalic alveolar plosive had already merged with the retroflex. Given that retroflex <ṭh> was likely the general outcome in intervocalic position, we would expect that late retroflex spellings appear more commonly in intervocalic position.

4 The Graphemes 𐨟 <ṭ> and 𐨠 <ṭ́>

The last problem that remains to be discussed is the more recently identified grapheme 𐨠 and how it relates to the other ones discussed above. This grapheme first came to the attention of researchers when a Kharoṣṭhī fragment was discovered that contained an acrostic poem arranged according to the so-called Arapacana sequence, the order in which the Kharoṣṭhī script was likely taught and memorized. The status of 𐨠 as a separate grapheme had not emerged clearly from other attestations of the Arapacana, which were either indirect (transmitted in Brāhmī) or incomplete.

Graphically 𐨠 appears to be an unmodified form of the grapheme 𐨟. This is surprising, since 𐨟 is conventionally transcribed as <ṭ>, but in the aspirated series the more basic grapheme is 𐨡 <ṭh> and the modified one is 𐨢 <ṭh>, which accords well with the hypothesized secondary nature of <ṭh>.

Structurally we would expect there to be a grapheme for a retroflex voiceless unaspirated plosive, devised for the expression of the reflex of OIA *ṭ* in clusters like *ṇṭ* or *ṭṭ*, as well as a grapheme for an alveolar voiceless unaspirated plosive, devised for the expression of the reflex of OIA *rt*, which would have become a geminated alveolar plosive in parallel with aspirated *rth*. That *rth* and *rt* developed in parallel is also suggested by the evidence of the modern languages of the

32 In contrast to Baums (2009, 165-6) whose arguments are based on the idea that the Aśokan inscriptions contain only <ṭh> and never <ṭh>, Melzer (2020, 34) argues that the Aśokan inscriptions contain only <ṭh>. In any case it is clear that only one type of character is used. Based on its shape alone, it would be equally justified to see in this a paleographic variant of <ṭh> or of <ṭh>, but, given the status of <ṭh> as a modification of the more basic shape of <ṭh>, which is clearly visible in later texts, it seems more likely that Baums (2009) is right to classify this earlier type as <ṭh>.

region, which almost universally show a retroflex outcome of *rt* and *ṛt* (e.g. Woṭ. *muṛó* 'now' < OIA *muhūrta-* + X; *muṛ* 'died (m.)' < OIA *mṛta-*, cf. Buddruss 1960, 114).[33] The only exceptions to this are Khowar and IA Kalasha, which likely form a relatively isolated subgroup that was rather distant from Gāndhārī on the dialect continuum.

While Strauch proposed to transliterate 𐨯 as <ṭ> (Salomon 2004, 46) to fill the gap in the conventional transcription, Salomon (2004, 46-7) suggested that it may be more apt to give the value <ṭ> to 𐨯 and to change the transcription of 𐨰 to <ṭ́>, though he put off this matter until after the publication of the acrostic poem, which was accomplished in 2020 (Melzer 2020). He also calls attention to the fact that what had been taken for an earlier paleographic variant of 𐨰 in the Aśokan inscriptions actually accords better with the newly identified grapheme (Salomon 2004, 46).

It seems therefore that in this case we are dealing not just with a case of unclear phonetic interpretation of a grapheme, but with unclear assignment of tokens to graphemes. It is important to note, however, that both forms are attested in the Aśokan inscriptions (Glass 2000, 69; Melzer 2020, 31-5). The simpler form without the modifying stroke appears primarily with the vowel diacritic for *i* or the combining stroke for pre-onset *-r-*, whereas the version with the additional stroke is most common for the basic *a* syllable (Melzer 2020, 33). Some words are attested with both variants. Glass (2000, 69 fn. 19), based on the observation that *akṣara*s with vowel diacritics tend to retain more paleographically conservative forms, tentatively assumed that 𐨯 is the earlier form, but both forms are in fact in use from the beginning of attestation.

Based on the distribution of the two forms in the Aśokan inscriptions, it is conceivable that the two shapes were at this stage merely graphic variants that were used based on considerations of leaving space for diacritic strokes or distinction from other graphemes in unmodified form. We could then suspect that the development was similar to that of <ṭh> and <ṭ́h>, where the phonological distinction between the alveolar and the retroflex was not expressed in writing at the time of Aśoka. The meaningless graphic variants could later have been appropriated to express the alveolar-retroflex phonological contrast, which would explain their inclusion as separate members of the Arapacana. Later the alveolar sound merged with the retroflex and the distinction became again meaningless. The variant without the additional stroke seems to barely occur at all in later texts, at least if the graphemic identifications that are currently

33 This is contrary to Baums's (2009, 162-3) idea that *rt* was generally assimilated to dental *tt*. Spellings as <t> could instead be seen as comparable to the variation of <th> with <ṭ́h> in intervocalic contexts and perhaps in part as influences from central MIA.

available can be relied upon.[34] This could perhaps indicate that making the distinction it was meant to express did not catch on in general usage (cf. Baums 2009, 166). In this regard it would be unlike <ṭh>, which did become established in general usage, though with varying consistency.

This still does not clarify the matter of which of the two graphemes stood for the alveolar and which stood for the retroflex at the time when the two were presumably distinguished in speech and could potentially be distinguished in writing. Since both variants are present in the Aśokan inscriptions, the graphic relations between the two characters are not unimpeachable evidence.

In the acrostic poem the example word chosen for 𐨗 is <a-𐨗>. Strauch, cited from personal communication by Salomon (2004, 46), had earlier suggested that this might be a descendant of either OIA *aṣṭa-* 'eight', or *artha-* 'purpose'. Baums (2009, 166), relying on the version of the Arapacana preserved in Sanskrit and Chinese transmission, leaned towards the latter option and accordingly saw in the introduction of 𐨗 an attempt to distinguish an alveolar plosive in writing from an alveolar ST cluster (according to him represented by <ṭh>). However, Melzer's edition of the poem now shows that the intended word is in fact a descendant of OIA *ārta-* 'tormented' (Melzer 2020, 94-5). If the author of the acrostic poem still had a phonological contrast between alveolars and retroflexes in their speech – which is not at all certain given the use of all of <th>, <ṭh> and <ṭ̣h> for word-initial etymological *sth* (Melzer 2020, 88, 90, 98) – this would lead us to conclude that 𐨗 stood for the alveolar arising from *rt* and should indeed be transcribed as <ṭ>, even though it has the more basic character shape of the two. Neither the retroflex nor the alveolar would be very suited to an acrostic illustration, since neither of the two would be expected to occur at the beginning of words. Nevertheless the author of the poem only resorted to using a word-internal example in the case of 𐨗. The example word for 𐨯 is the unclear word <𐨯 𐨯>, which Melzer (2020, 173) tentatively associates with OIA **tartar-* 'crosser' or *trātar-* 'savior'. The former is otherwise unattested, the latter is only attested in Gāndhārī in a thematized trisyllabic form <tratarasa> (gen.sg.) (Baums, Glass 2002). The context is too decayed to provide clarification.

Since 𐨗 is so rare and 𐨯 also appears in words derived from Skt. forms with *ṭṭ* and *ṇṭ*, which likely never went through an alveolar stage (e.g. <pa-𐨯> 'silk' ~ Skt. *paṭṭa-*, <(graḥma)-ka-𐨯-ka> '(village) bamboo' ~ Skt. *kaṇṭaka-* 'bamboo', cf. Baums, Glass 2002) as well as in cases of aspiration slips in words with /ṭh/ < *ṣṭ(h)*, which equally

34 A new paleographic study in light of the evidence from the acrostic poem would be quite helpful.

never had an alveolar sound (e.g. <ka-𐨵> 'stick, firewood' ~ Skt. *kāṣṭha-*, cf. Baums, Glass 2002), it seems best to retain the transcriptions 𐨷 = <ṭ́> and 𐨵 = <ṭ> for now, until further clarifying evidence emerges.

5 Conclusions

If the arguments made in this article are sound, the "difficult problem" (Salomon 2004, 47) of the development of ST clusters and retroflexes in Gāndhārī and their representations in Kharoṣṭhī can be resolved into a consistent, symmetrical system **[tab. 3]**. This system is supported by the evidence of those modern languages that can be assumed to have been closest on the dialect continuum to literary Gāndhārī, it accounts for all patterns of orthographic substitutability observed in the written sources and the phonology implied by it is typologically realistic.

Table 3 Results of the analysis for written and spoken Gāndhārī

OIA sources	Grapheme	Transcription	Phonology	Later mergers
st	𐨶	<st>	/st/ [st̪]	(> /th- ~ -tth-/ in post-Gāndhārī)
(ṭ)ṭh, ṣṭ, ṣṭh	𐨶	<ṭh>	/ṭh/ [ʈʰ]/ʈʈh/ [ʈʰː]	
sth, rth	𐨹	<ṭ́h>	/ṭh/ [ʈʰ]/ʈʈh/ [ʈʰː]	> /th-/> /-ṭṭh-/
(ṭ)ṭ	𐨵	<ṭ>	/ṭ/ [ʈ]/ʈʈ/ [ʈː]	
rt	𐨷	<ṭ́>	(/ṭ/ [ʈ])*/ʈʈ/ [ʈː]	> /ṭṭ/

* In accordance with the general principles of the writing system, the grapheme could potentially have been used for non-geminates as well, but this sound likely never existed in the language in ungeminated form.

Bibliography

Bailey, H.W. (1949). "Irano-Indica II". *Bulletin of the School of Oriental and African Studies, University of London*, 13(1), 121-39.

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

Bashir, E. (2022). "Khowar and Kalasha: Similarities and Differences in their Micro-areal Contexts". Degener, A; Hill, E. (eds), *Proceedings of the International Workshop on the Languages of Nuristan: The Nuristani Languages in Synchrony and Diachrony* (University of Cologne, 1-2 November 2019). *International Journal of Diachronic Linguistics and Linguistic Reconstruction*, 19, 1-59.

Bashir, E. (2023). *A Khowar-English Lexicon – with Cultural and Etymological Notes.* Himalayan Linguistics Archive 12.
https://doi.org/10.5070/H90057297.

Baums, S. (2009). *A Gāndhārī Commentary on Early Buddhist Verses: British Library Kharoṣṭhī Fragments 7, 9, 13 and 18* [PhD Dissertation]. Seattle: University of Washington.

Baums, S. (2010). *Outline of Gāndhārī Grammar*. https://gandhari.org/grammar

Baums, S.; Glass, A. (2002). *A Dictionary of Gāndhārī*. https://gandhari.org/n_dictionary.php.

Brough, J. (1962). *The Gāndhārī Dharmapada*. London: Oxford University Press. London Oriental Series 7.

Buddruss, G. (1960). *Die Sprache von Woṭapūr und Kaṭārqalā: Linguistische Studien im afghanischen Hindukusch*. Bonn: Selbstverlag des Orientalischen Seminars der Universität Bonn.

Buddruss, G. (1979). "Gṛaṅgali. Ein Nachtrag zum Atlas der Dardsprachen". *Münchener Studien zur Sprachwissenschaft*, 38, 21-39.

Cacopardo, A.M.; Cacopardo, A.S. (2001). *Gates of Peristan: History, Religion and Society in the Hindu Kush*. Rome: Istituto Italiano per l'Africa e l'Oriente. IsIAO Reports and Memoirs, Series Minor 5.

Carling, G.; Pinault, G-J. (2023). *Dictionary and Thesaurus of Tocharian A*. Wiesbaden: Harrassowitz.

Degener, A. (2008). *Shina-Texte aus Gilgit (Nord-Pakistan)*. Wiesbaden: Harrassowitz. Beiträge zur Indologie 41.

FLI (Forum for Language Initiatives). (2016). *Gawar Bātī: Maǧmūʿa-yi Alfāż*. Islamabad: FLI (Forum for Language Initiatives).

Franke, R.O. (1906). "Zum Manuskript Dutreuil de Rhins". *Zeitschrift der Deutschen Morgenländischen Gesellschaft*, 60, 477-511.

Fussman, G. (1972). *Atlas Linguistique des Parlers Dardes et Kafirs.* Vol. 2, *Commentaire*. Paris: École Française d'Extrême-Orient.

Glass, A. (2000). *A Preliminary Study of Kharoṣṭhī Manuscript Paleography* [M.A. thesis]. Seattle: University of Washington.

Grierson, G. (1919). *Specimens of the Dardic or Piśācha Languages (Including Kāshmīrī)*. Calcutta: Superindendent Government Printing, India. Linguistic Survey of India 8.2.

Halfmann, J. (2023). "Nuristani Theonyms in Light of Historical Phonology". Cacopardo, A.M.; Cacopardo, A.S. (eds), *Roots of Peristan: The Pre-Islamic Cultures of the Hindukush/Karakorum = Proceedings of the International Interdisciplinary Conference* (ISMEO, Rome, Palazzo Baleani, 5-7 October, 2022), vol. 1. Rome: ISMEO; Scienze e Lettere, 317-58. Serie Orientale Roma n. s. 37.

Hinüber, O. von. (2001). *Das ältere Mittelindisch im Überblick*. 2e Aufl. Vienna: Verlag der Österreichischen Akademie der Wissenschaften. Philosophisch-Historische Klasse, Sitzungsberichte 467; Veröffentlichungen der Kommission für Sprachen und Kulturen Südasiens 20.

Keiser, R.L. (1974). "Social Structure in the Southeastern Hindu-kush: Some Implications for Pashai Ethno-history". *Anthropos*, 69(3/4), 445-56.

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

Leech, R. (1838). "Epitome of the Grammars of the Brahuiky, the Balochky and the Panjábi Languages, with Vocabularies of the Baraky, the Pashi, the Laghmani, the Cashgari, the Teerhai, and the Deer Dialects". *The Journal of the Asiatic Society of Bengal*, 7(2), 780-7.

Loukota, D. (2023). "Ne Hāḍe Vajrropamä Vaśärä: Indic Loanwords in the Khotanese Book of Zambasta and the Chronology of the Spread of Buddhism to Khotan". *BuddhistRoad Papers*, 1.7.
https://doi.org/10.46586/rub.br.274.246.

Mayrhofer, M. (1992). *Etymologisches Wörterbuch des Altindoarischen*, Bd. 1. Heidelberg: Universitätsverlag Winter.

Mayrhofer, M. (2001). *Etymologisches Wörterbuch des Altindoarischen*, Bd. 3. Heidelberg: Universitätsverlag Winter.

Melzer, G. (2020). *An Arapacana Acrostic Poem in Gandhari. Bajaur Collection Kharoṣṭhī Fragment 5*. Munich.
http://130.223.29.184/editions/melzer_bc5_202007.pdf.

Morgenstierne, G. (1934). "Notes on Tirahi". *Acta Orientalia*, 12, 161-89.

Morgenstierne, G. (1936). "Iranian Elements in Khowar". *Bulletin of the School of Oriental and African Studies, University of London*, 8(2-3), 657-71.

Morgenstierne, G. (1950). *Notes on Gawar-Bati*. Oslo: Jacob Dybwad.

Morgenstierne, G. (1952). "Linguistic Gleanings from Nuristan". *Norsk Tidsskrift for Sprogvidenskap*, 16, 117-35.

Morgenstierne, G. (1956). *The Pashai Language*. Vol. 3, *Vocabulary*. Oslo: Instituttet for Sammenlignende Kulturforskning, Universitetsforlaget. Indo-Iranian Frontier Languages III.3.

Morgenstierne, G. (1961). "Dardic and Kāfir Languages". *Encyclopaedia of Islam*. Leiden: Brill.
http://dx.doi.org/10.1163/1573-3912_islam_COM_0155.

Morgenstierne, G. (1967). *The Pashai Language*. Vol. 1, *Grammar*. Oslo: Instituttet for Sammenlignende Kulturforskning, Universitetsforlaget. Indo-Iranian Frontier Languages III.1.

Ovesen, J. (1983). "The Construction of Ethnic Identities: The Nuristani and the Pashai of Eastern Afghanistan". Jacobson-Widding, A. (ed.), *Identity: Personal and Socio-Cultural. A Symposium*. Uppsala: Almqvist & Wiksell International, 321-33.

Ovesen, J. (1984). "On the Cultural Heritage of the Pashai". *Anthropos*, 79(4-6), 397-407.

Salomon, R. (2004). "An Arapacana abecedary from Kara Tepe (Termez, Uzbekistan)". *Bulletin of the Asia Institute, New Series*, 18, 43-51.

Sims-Williams, N. (2007). *Bactrian Documents from Northern Afghanistan*. Vol. 2, *Letters and Buddhist Texts*. London: The Nour Foundation & Azimuth Editions.

Stein, A. (1925). "Notes on Tirahi: The speakers of Tirahi". *Journal of the Royal Asiatic Society*, 57(3), 399-404.

Strauch, I. (2012). "The Character of the Indian Kharoṣṭhī Script and the 'Sanskrit Revolution': A Writing System Between Identity and Assimilation". de Voogt, A.; Quack, J.F. (eds), *The Idea of Writing*. Vol. 2, *Writing Across Borders*. Amsterdam: Brill, 131-68.

Torwali, Z. (2020). *Torwali-English Student Dictionary*.
https://www.webonary.org/torwali/en/.

Trail, R.; Cooper, G. (1999). *Kalasha Dictionary - with English and Urdu*. Islamabad: National Institute of Pakistani Studies, Quaid-i-Azam University.

Turner, R. (1962-66). *A Comparative Dictionary of the Indo-Aryan Languages*. London: Oxford University Press.

Zoller, C.P. (2005). *A Grammar and Dictionary of Indus Kohistani*. Vol. 1, *Dictionary*. Berlin: De Gruyter Mouton. Trends in Linguistics. Documentation [TiLDOC] 21-1.

e-ISSN 2785-5953

Bhasha
Vol. 3 – Num. 2 – October 2024

Umlaut in Jerusalem Domari

Moe Kitamura
The University of Tokyo, Japan

Abstract This study conducts a phonetic and phonological analysis of the umlaut phenomenon within the gender system of Jerusalem Domari. Using descriptive research and acoustic analysis of recordings, the study establishes several key findings: A contrast is observed in the integration of pre-Arabic loanwords (Persian, Kurdish, Turkish) with Indo-Aryan native words, which follow the umlaut rules, whereas the loanwords from Arabic, the most recent contact language, do not. A clear phonemic distinction is identified between the two open vowels, [a(:)] and [ɑ(:)], in pre-Arabic words, while these vowels exhibit allophonic values in the Arabic loanwords.

Keywords Domari. Umlaut. Descriptive linguistics. Language contact. Phoneme identification.

This work was supported by Grant-in-Aid for JSPS Grant Number JP23KJ0762 and JSPS Overseas Challenge Program for Young Researchers. I would like to extend a special word of appreciation to Prof. Eva Schultze-Berndt for her invaluable guidance and insightful suggestions during my visiting research project from September 2023 to February 2024. Also, I would like to extend my sincere gratitude to Prof. Yaron Matras for providing data that has been indispensable to my research. My sincere thanks also go to the members of the Doctoral Academy and the Faculty of Humanities at the University of Manchester for their constructive feedback and support.

Edizioni Ca'Foscari

Peer review

Submitted 2024-06-06
Accepted 2024-07-17
Published 2024-12-18

Citation Kitamura, M. (2024). "Umlaut in Jerusalem Domari". *Bhasha*, 3(2), 213-236.

DOI 10.30687/bhasha/2785-5953/2024/02/003

1 Introduction

This study investigates the phenomenon of umlaut observed in the nominal or adjectival gender system in Jerusalem Domari. Conducting descriptive research and acoustic analyses of fieldwork data from one Domari speaker recorded in Jerusalem during 2019-20, as well as acoustic analyses of sample audio data from two additional speakers in Matras (2012), it seeks to establish the following arguments:

1. Jerusalem Domari exhibits umlaut on nouns or adjectives with open vowels in their final syllables. Consonant-ending nouns and adjectives, which have lost the final vowels that would have triggered umlaut, also exhibit umlaut.
2. Pre-Arabic loanwords, including Persian, Kurdish, and Turkish, integrate with Indo-Aryan native words and follow the umlaut rules. In contrast, loanwords from Arabic, the latest contact language, do not follow these rules.
3. The two open vowels, front [(ɑː)] and back [(aː)], are recognized as distinct phonemes in pre-Arabic words of Jerusalem Domari. It should be noted, however, that the distinction between these open vowels has existed independently of the umlaut phenomenon.
4. In Arabic loanwords, the open vowels [ɑ(ː)] and [a(ː)] show allophonic values with complementary distribution. This indicates the existence of Parallel System Borrowing in phonology in the bilingual setting of Domari.

Regarding the third argument on distinct phonemes in Jerusalem Domari, Matras (2012, 51) highlights the challenges in identifying phonemes, particularly around the sounds [a], [ʌ], [ɔ], and [o]. This study aims to address these challenges and clarify phoneme identification for these specific sounds.

All Domari examples in this paper, unless otherwise noted, are based on the fieldwork research which I conducted in Jerusalem in 2019-20 with a male Jerusalem Domari speaker, born and raised in Jerusalem.

2 Introduction to Domari

2.1 Domari Language

Domari is an Indo-Aryan language spoken by the Dom people, an ethnic diaspora group in the Middle East. Its relationship with Romani is often mentioned due to shared ethnic and linguistic characteristics. It is indicated in previous research (Matras 2012; Herin 2012; 2014) that there are dialectal variations, specifically Northern

Domari in Syria and Lebanon, and Southern Domari in Palestine and Jordan, although there is still a need for more fieldwork data to establish Domari dialectology. The focus of this study is on Jerusalem Domari, a variant of the Southern dialect. Domari is severely endangered, with most speakers being bilingual in Arabic. Most Domari speakers also have proficient knowledge of Arabic. Due to the intense and prolonged bilingual situation, Domari has been significantly influenced by Arabic. In particular, Jerusalem Domari is in a severely endangered situation, with almost all Dom people speaking Arabic as their first language and having little knowledge of Domari. Currently, only two fluent Domari speakers are found in Jerusalem among the elderly generation.

Through the analysis of phonological innovations, Turner (1926) demonstrated that it belongs to the Central group of the Indo-Aryan branch. It is estimated that the Dom people resided in the Midland of the Indian subcontinent from the Old Indo-Aryan to the Middle Indo-Aryan periods, subsequently relocating to the Northwest from the Middle Indo-Aryan to the New Indo-Aryan periods.

In figure 1, the estimated migration path of the Dom people is illustrated **[fig. 1]**. It depicts their movement from the midland of the Indian subcontinent towards the Northwest, their traversal of the Iranian-speaking area, and their eventual arrival in Arabic-speaking areas. Consequently, the term 'Pre-Arabic words' is used in this study to refer to the vocabulary in Domari that existed prior to contact with Arabic. This category includes native Indo-Aryan words, along with loanwords from Iranian languages such as Persian and Kurdish, as well as from Turkic languages.

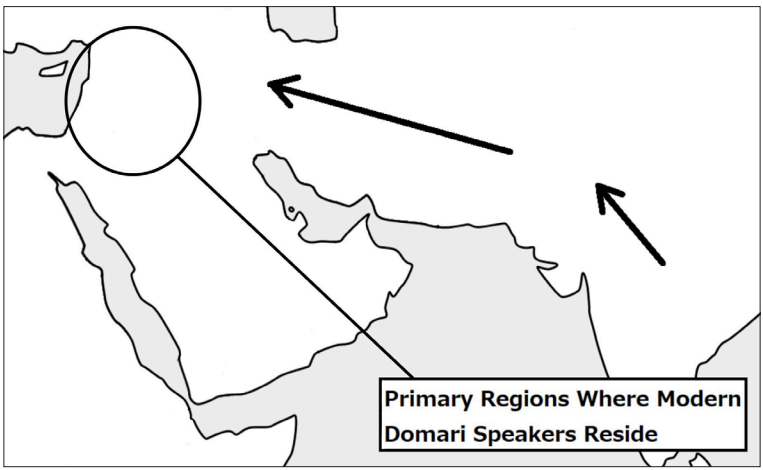


Figure 1 Estimated Migration Path of Domari

2.2 General Grammatical Background

The basic word order in Domari is VSO, and subject nouns are often omitted when they are clear from the context, a pattern that aligns with Jerusalem Arabic.

(1) a. *fe:-r-a* *za:r-a* *o:r-as.*
hit-PRF-3.SG.M boy-M.SG.NOM that-OBL.M
'The boy hit that man.'

b. *lah-am-r-i* *kull* *di:s.*
see-1.SG.A-2.SG.P-PRS every day
'I see you every day.'

Domari has predication markers, which construct the subject complement and agree in number and gender with the subject, as illustrated in example (2).

(2) a. *ame* *gif* *do:m-e:ni.*
1.PL.NOM all Dom-PRED.PL
'We are all Doms.'

b. *ama* *till-e:k.*
1.SG.NOM old-PRED.M.SG
'I am old.'

c. *ato* *ʃtət-ik.*
2.SG.NOM young-PRED.F.SG
'You are young.'

d. *kari:m* *do:m-i.*
Kareem Dom-PRED.SG
'Kareem is a Dom man.'

2.3 Domari Vowel System

Figure 2 displays the vowel system based on the one presented in Matras's (2012) *A Grammar of Domari*. Additionally, it should be noted that Domari distinguishes between short and long pairs of each vowel phoneme. Figure 2 only displays short vowels; however, each vowel also has a long counterpart, which is recognized as a distinct phoneme **[fig. 2]**.

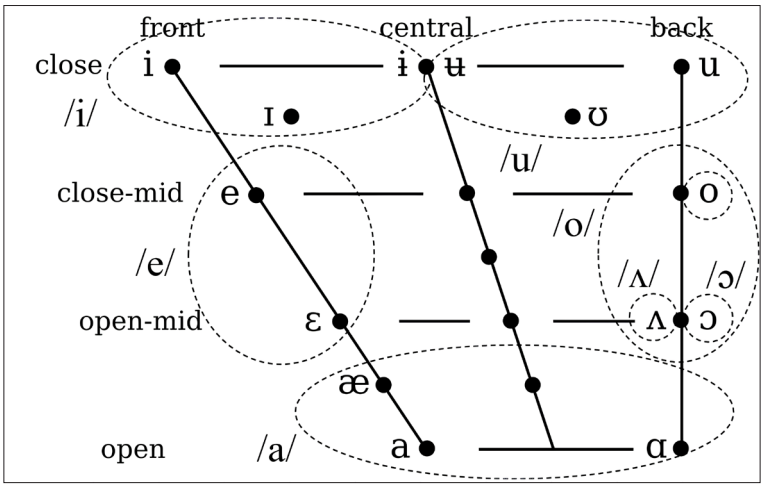


Figure 2 Vowel System based on Matras 2012, 36

In discussing the Jerusalem Domari vowel system, Matras (2012) mentioned the challenges associated with identifying the vowel sounds [a], [ʌ], [ɔ], and [o].

In Matras (2012, 51), it is stated that the contrast between [a] and [ʌ] is distinctive despite the fact that there is only one lexeme including the phoneme [ʌ], as illustrated by this near minimal pair:

[ˈpandʒi]	*pandži*	'he/she'	(Matras 2012, 51)
[ˈpʌndʒes]	*pandžes*	'five'	(Matras 2012, 51)

The phonemic contrast between /ɔ/ and /o/ is described as ambiguous, suggesting the possibility of their status as free variants. /ɔ/ predominantly precedes semi-vowels, and has minimal pairs contrasting to /a/, as demonstrated by the following examples:

[rɔˈwari]	*rɔwari*	'he/she cries'	(Matras 2012, 51)
[raˈwari]	*rawari*	'he/she travels'	(Matras 2012, 51)
[bɔˈjom]	*bɔyom*	'my father'	(Matras 2012, 51)
[baˈjom]	*bayom*	'my wife'	(Matras 2012, 51)
[dɔˈwari]	*dɔwari*	'he/she washes'	(Matras 2012, 51)
[daˈwari]	*dawari*	'he/she dances'	(Matras 2012, 51)

Additionally, it is implied that [ɔ] and [o] do not show complementary distribution, and evidence suggests that native speakers regard these two vowels as distinct phonemes. This study aims to address the challenges associated with identifying vowel phonemes in Jerusalem Domari.

2.4 Domari Nouns and Gender System

Domari shows a two-way gender system: masculine and feminine. Typically, masculine nouns in Domari are marked with the suffix *-a* (*-ɑ* in later transcription), and feminine nouns with the suffix *-i* in the nominative singular. Furthermore, Domari has what are known as consonant-ending nouns. These nouns exhibit gender distinction but lack suffixes in the nominative singular form. Consequently, their gender cannot be identified based on suffixes alone.

Table 1 presents Domari nouns in the nominative singular form, based on Matras 2012. For this table, I have adopted the phonological system outlined by Matras 2012. The nouns in the leftmost column are marked with masculine suffixes *-a*. Those in the middle column are marked with feminine suffixes *-i*, while the nouns in the rightmost column are consonant-ending nouns, lacking gender-specific suffixes **[tab. 1]**.

Table 1 Domari nouns in Nominative Singular based on Matras 2012

Masculine suffix	Feminine suffix	Consonant-ending
qrar-a 'Bedouin man'	*qrar-i* 'Bedouin woman'	
šōn-a 'non-Dom boy'	*šōn-i* 'non-Dom girl'	
zar-a 'Dom boy'	*lāč-i* 'girl'	*ūyar* 'market, Jerusalem' (F)
ṣnoṭ-a 'dog'	*gor-i* 'horse'	*qar* 'donkey' (M)

In Domari, adjectives follow the same declension system as nouns. While most Domari adjectives have suffixes, there is also an example of the consonant-ending adjective, *qar-* 'stupid; donkey'. This is a result of pattern replication or a calque, modelled on the Arabic *ħimaːr* 'donkey; stupid'.

According to Masica (1991, 217-23), nominal gender is common across many Indo-Aryan languages. Sanskrit originally had three genders, a system preserved in Pali and Prakrit, although there was some confusion between the masculine and neuter genders. In New Indo-Aryan languages, the most prevalent gender system is a two-gender system, resulting from the merger of the old masculine and neuter genders.

Thus, it is evident that the Domari gender system aligns with the common pattern in modern Indo-Aryan languages. In languages where the original final vowels have been lost, 'unmarked' nouns ending in consonants are often found. The gender assignment of these nouns generally depends on the vowels that have been lost. This appears to be the origin of the consonant-ending nouns in Domari.

Considering these facts, the Domari gender system is inherited from Indo-Aryan languages and appears to be shared in many New Indo-Aryan languages.

3 Umlaut in Other Languages

Umlaut is a type of sound change originally named after the phenomenon observed in Germanic languages. Hock (1991, 66) defines umlaut as "the assimilation of a class of vowels to a set of [+vocalic] segments in an immediately neighbouring syllable".

This term indicates not only the diachronic sound change itself but also frequently denotes the synchronic phenomenon resulting from this historical linguistic development. Table 2 provides examples of synchronic umlaut phenomena in various languages **[tab. 2]**.[1]

Table 2 Umlaut in other languages

Languages	Examples	
Modern Standard German	*Huhn* /huːn/ 'hen.SG'	*Hühn-er* /hyːnɐ/ 'hen-PL'
	Vogel /foːgl/ 'bird.SG'	*Vögel* /føːgl/ 'bird.PL'
	Hund /hʊnt/ 'dog.M'	*Hünd-in* /hʏndɪn/ 'dog-F'
Kashimiri	*šur* 'child.M'	*šuɨr* 'child.F'
	koṭ 'boy'	*kəṭ* 'girl'
Palula	*ǰaanu* 'masculine person'	*ǰeeni* 'feminine person'
	moomu 'mother's father'	*meemi* 'mother's mother'
	praaču 'guest'	*preeči* 'female guest'
Nsong (Bantu B80)	*-lam-* 'cook'	*-lɛmɛn* 'cook for'
Mpur (Bantu B80)	*-lám-* 'cook'	*-lɛ́ɛ́m* 'cook for'

He also described some additional observations on the conditions of umlaut.

> Note however that crosslinguistically, umlaut most frequently is conditioned by final syllables. The reason seems to be that word-final position is a highly conducive environment for the loss of segments and syllables, including vowels, the most common conditioning environments for vowel assimilations. (Hock 1991, 68)

Consequently, umlaut vowels may remain in certain morphological contexts, even in the absence of the original morpheme that provided the phonological context necessary for the assimilation. For instance, in the German examples in Table 2, *Hühner* 'hen-PL' maintains the suffix *-er* which was originally *-ir* in Old High German and triggers the umlaut of the vowel in the word stem, while *Vögel* 'birds.PL' has lost the suffix that triggered the umlaut but still preserves the vowel

1 Modern Standard German: Wiese 1996; Kashmiri: Koul 2003; Palula: Liljegren 2019; Nsong and Mpur: Bostoen, Koni Muluwa 2014.

quality conditioned by the umlaut. Regarding the High German Umlaut, various scholars have studied the process whereby the sounds /u(:)/ or /o(:)/ change to /y(:)/ (<ü>) or /ø:/ (<ö>), conditioned by /i(:)/ or /j/ in subsequent syllables. This change led to these vowels obtaining distinct phonemic statuses, evolving from Old High German to Middle High German (Janda 2003).

In Indo-Aryan languages, umlaut phenomena in the gender system, similar to the umlaut in Domari that will be reported in this study, have been observed in the so-called Dardic languages of the Hindu-Kush area, located in the northwest region of the Indian subcontinent.

In Kashmiri, five types of vowel alternation are observed: 1. Lowering of /ə/, /ə̄/, and /ū/ of monosyllabic stems to /a/, /ā/, and /ō/ respectively before a plural suffix *-i* or *-i̯*. 2. Raising of /a/ and /ā/ in CVC stems to /ə/ and /ə̄/ respectively before a suffix with *-i*. 3. Centralization of /u/, /ū/, /o/, and /ō/ to /ɨ/, /ɨ̄/, /ə/, and /ə̄/ respectively before a suffix with *-i* or *-y*. 4. Centralization of the second vowel /u/ of disyllabic words with the structure CVCVC to the central vowel /a/ before a plural suffix *-ø* (Koul 2003, 904). The second and third alternations are assimilation processes, which match the definition of umlaut. These umlaut phenomena are widely observed in Kashmiri, including in the gender system, similar to the Domari case discussed in section 4. In the gender formation process from masculine to feminine in Kashmiri, the following two umlaut processes regarding vowels are observed:

a. /u, ū, o, ō/ in masculine nouns with the structure CVC are diphthongized or replaced by the central vowels at the same height: masculine *šur* 'child', *gūr* 'milkman', *gob* 'heavy', *koṭ* 'boy': feminine *šuɨr* 'child', *gūər* 'milkwoman', *goəb* 'heavy', *kəṭ* 'girl'.
b. Penultimate /u/ of masculine nouns with the structure CVCVC is replaced by /ɨ/: *kōtur* 'pigeon', *kɔkur* 'cock': feminine *kōətɨr*, *kɔkɨr* 'hen' (Koul 2003, 905-6).

The condition which causes the first case of umlaut is similar to the case of umlaut in the Domari gender system, as described in section 4.

Similar umlaut phenomena in the gender system have been reported in some other Dardic languages. For instance, as described by Liljegren (2019), in Palula, one of the Dardic languages in the Hindu-Kush area, 'masculine person' is expressed as *ǰaanu*, while 'feminine person' is *ǰeeni* (Liljegren 2019, 302).

As noted in section 2.1, Domari belongs to the Central group of Indo-Aryan languages, so the similarity regarding umlaut between Domari and Dardic languages does not seem to be due to the preservation of an inherited genetic feature. Although language contact between Domari and Dardic languages was suggested in Turner's

(1926) analysis of phonological changes, there is no evidence so far to suggest that the similarity of umlaut in the gender systems of Domari and Dardic languages, such as Kashmiri, is due to contact-induced change.

In addition, Bostoen and Koni Muluwa (2014) reported umlaut phenomena in some Bantu B80 languages. They identify four types of umlaut:

> (a) the raising of the low central vowel a to mid front vowel (y)ɜ or æ; (b) the fronting of the open-mid back vowel ɔ to (w)ɛ or œ; (c) the fronting of the open-mid back vowel o/ʊ to wi/we or ø; (d) the fronting of the closed back vowel u to wi or ü (=[y])'. (Bostoen and Koni Muluwa 2014, 228)

Among these languages, in Yans, Mpur, Ding, and Ntsambaan, the umlaut leads to a phonemic split. In contrast, in Nsong, Mpiin, Mbuun, and Ngong, it does not lead to such a split because, in these languages, the outcome vowel of the umlaut was already present in the inherited Proto-Bantu vowels.

In both Germanic and Hindu-Kush Indo-Aryan languages, but not in some of the Bantu languages such as Nsong, Mpiin, Mbuun, and Ngong, umlaut phonemes emerged as new phonemes, distinct from certain existing phonemes. This occurred through the process of umlaut producing allophones of existing phonemes, and these allophones being phonemicized due to the loss of the conditioning environment. However, in Domari, the phonemic distinction between the open vowels /a/ and /ɑ/, which primarily interacts with umlaut rules in modern Jerusalem Domari, seems to have existed prior to the occurrence of umlaut. This is evidenced by the fact that the front vowel /a/ frequently occurs in verbal stems unrelated to the umlaut rule. Similarly, the back vowel /ɑ/ is also present in a few verbal stems.

ex.	*ga-r-a* 'go-PRF-3.SG.M'	*bag-id-om* 'break-PRF-1.SG'
	raw-r-i 'cry-PRF-3.SG.F'	*waz-r-a* 'flee-PRF-3.SG.M

Investigating the diachronic processes of these vowels could be a topic for future research.

4 Umlaut Rules in Domari

The analysis in this section is based on the fieldwork research which I conducted in 2019-20 with a 67-year-old male Jerusalem Domari speaker. Five rules have been identified that govern umlaut in the speaker.

1. Fronting of Open Vowel:

The back open vowels /ɑ(:)/ in masculine nouns become front vowels /a(:)/ in feminine nouns.

ex. *qrɑːr-ɑ* 'Bedouin-M' *qraːr-i* 'Bedouin-F'
ɑːn-ɑ 'egg-M' *kaːb-i* 'door-F'

2. Centralization of Close-Mid Vowel:

The back close-mid vowels /o(:)/ in masculine nouns become central vowels /ɵ(:)/ in feminine nouns.

ex. *ʃoːn-ɑ* 'non-Dom.boy-M' *ʃɵːn-i* 'non-Dom.girl-F'
ʃtot-ɑ 'small-M' *ʃtɵt-i* 'small-F'

3. Application to a Consonant-Ending Feminine Noun:

Although most consonant-ending nouns are masculine (ex. *wɑt* 'stone', *sɑːl* 'rice'), only one documented example of a consonant-ending feminine noun *wijar* 'market, Jerusalem' also follows these rules. It exhibits the front vowel in the last syllable, even though it lacks the suffix that would have triggered umlaut.

4. Application to Declined Feminine Nouns with the Palatal Glide *-j-*:
5. The umlaut is also observed in declined feminine nouns accompanying the feminine suffix with the palatal glide *-j-* instead of *-i*.

ex. *ʃɵːn-j-ɑ* 'girl-F-OBL.F' *kaːb-j-ɑ* 'door-F-OBL.F'

6. Non-Application to Masculine Nouns:

These rules are not applicable to masculine nouns, even when they share the same phonological environments as feminine nouns.

ex. *wɑt-i* 'stone-PRED' *doːm-i* 'Dom-PRED'

Table 3 displays Domari nouns and adjectives that have a shared word stem and variable gender, while Table 4 presents Domari nouns with invariable gender **[tabs 3-4]**.

Table 3 Domari Nouns/Adjectives with Variable Gender

Masculine	Feminine	Consonant-ending
qraːr-a 'Bedouin man'	*qraːr-i* 'Bedouin woman'	
ʃoːn-a 'non-Dom boy'	*ʃɵːn-i* 'non-Dom girl'	
	qar-i 'stupid (F)'	*qar* 'stupid (M)'
ʃtot-a 'small (M)'	*ʃtɵt-i* 'small (F)'	

Table 4 Domari Nouns with Invariable Gender

Masculine	Feminine	Consonant-ending
zaːr-a 'Dom boy'	*laːʃ-i* 'Dom girl'	*wat* 'stone' (M)
aːn-a 'egg'	*kaːb-i* 'door'	*saːl* 'rice' (M)
aːt-a 'tahini'	*ʃmaːl-i* 'chicken'	*ʃaːl* 'well' (M)
man-a 'bread'	*baːn-i* 'water'	*wijar* 'market, Jerusalem' (F)
snoːt-a 'dog'	*kɵr-i* 'house'	

Concerning rules 1-3 (Fronting of Open Vowel; Centralization of Close-Mid Vowel; Application to a Consonant-Ending Feminine Noun), it can be observed that the nouns with masculine suffixes in the left-most column in Table 3 and 4 contain back vowels [ɑ(ː)] or [o(ː)] in the word stem. In contrast, the nouns with feminine suffixes in the middle column have front or central vowels [a(ː)] or [ɵ(ː)] in their word stems. Regarding the consonant-ending nouns in the right-most column, masculine nouns typically contain back vowels, while only one feminine noun exhibits the front vowel [a].

Examples (3) and (4) demonstrate Rule 4: Application to Declined Feminine Noun with the Palatal Glide *-j-*. In example (3), the feminine noun *ʃɵːn-i*, which means 'non-Dom girl', shows the front vowel [ɵː] in the oblique case *ʃɵːn-j-a*. Similarly, in example (4), the feminine noun *kaːb-i*, meaning 'door', exhibits the front vowel [aː] in the oblique case *kaːb-j-a*.

(3) *f-ar-i* *ʃɵːn-j-a.*
hit-3.SG-PRES girl-F-OBL.F
'He/She hits the girl.' < *ʃɵːn-i* 'non-Dom girl'

(4) *qoːl-am-i* *kaːb-j-a.*
open-1.SG-PRES door-F-OBL.F
'I open the door' < *kaːb-i* 'door'

This rule is typically observed in kinship terms, as Domari kinship terms generally appear with possessive pronouns. Table 5 displays Domari kinship terms accompanied by the first-person singular pronominal suffix *-oːm*, meaning 'my'. Notably, the front vowel [aː] is consistently retained in the feminine kinship terms **[tab. 5]**.

Table 5 Umlaut in Kinship Terms

Masculine	Feminine
baːj-oːm 'my father'	*daː-j-oːm* 'my mother'
baːd-oːm 'my grandfather'	*daːd-j-oːm* 'my grandmother'
maːm-oːm 'my paternal uncle'	*maːm-j-oːm* 'my paternal aunt'
xaːl-oːm 'my maternal uncle'	*xaːl-j-oːm* 'my maternal aunt'
baːr-oːm 'my brother'	*baː-j-oːm* 'my wife'

Examples (5), (6), and (7) demonstrate Rule 5: Non-Application to Masculine Nouns. In example (5), the masculine noun *wat*, meaning 'stone', is followed by the predication marker *-i*. This makes a phonological environment the same as that of nouns with the feminine suffix *-i*. However, it retains the back vowel [ɑ] in the word stem. A similar pattern is observed in example (6) with nouns containing the close-mid vowel [oː]. Although the masculine noun *doːm*, meaning 'Dom man', precedes the predication marker *-i*, it maintains the back vowel in the word stem. Likewise, in example (7), the masculine noun *bɑːr,* meaning 'brother', is followed by the pronominal suffix *-im* but still retains the back vowel [ɑː] in the word stem.

(5) a. *aha* *wat-i.*
this stone-PRED
'This is a stone.' < *wat* 'stone' (Masculine)
b. **aha* *wat-i.*

(6) a. *ama* *doːm-i.*
1.sg.nom Dom-PRED
'I am *a* Dom man.' < *doːm* 'Dom man'
b. **ama* *dɐːm-i.*

(7) a. *aha* *botr-oːs* *baːr-im-ki.*
This son-3.SG.POSS brother-1.SG.POSS-ABL
'This is my brother's son.' < *baːr* 'brother'
b. **aha* *botr-oːs* *baːr-im-ki*

Based on Rules 3-5 (Application to a Consonant-Ending Feminine Noun; Application to Declined Feminine Noun with the Palatal Glide *-j-*; Non-Application to Masculine Noun), it can be concluded that this is not solely a phonological phenomenon; rather, the umlaut has morphologised.

In addition, the application of umlaut rules is observed in pre-Arabic loanwords, including those from Persian, Kurdish, and Turkish origins, but these rules are not applied to Arabic loanwords.

ex.	*zɑːr-ɑ* 'boy'	Kurdish or Persian
	zard 'gold'	Kurdish or Persian
	kaːb-i 'door'	Turkish or Kurdish

There are the following exceptions to these umlaut rules.

1. Feminine nouns marked with the Arabic suffixes *-iːja*

ex. *doːm-iːja* 'Dom woman'

2. The masculine noun *kmɑːl-i*, meaning 'policeman', which is exceptionally marked with the feminine suffix *-i*
3. Three masculine consonant-ending nouns: *kaːn* 'ear', *ag* 'fire', *lɵːn* 'salt'

While the exceptional cases typically involve Arabic loanwords or pre-Arabic words preceding the Arabic-origin feminine suffix *-iːja*, there are also a few exceptions among pre-Arabic nouns with the Indo-Aryan native gender system, as illustrated in exceptions 2 and 3. Specifically, from the observations made in this research, there are three exceptions among 60 pre-Arabic nouns or adjectives that feature front vowels [a(ː)] or back vowels [ɑ(ː)] in the last syllable. Additionally, there is one exception among 22 pre-Arabic nouns or adjectives that include back vowels [o(ː)] or central vowels [ɵ(ː)] in the last syllable.

Regarding exception 3, the reason why these three words are exceptional is not clear. However, it might be due to the fact that they are all consonant-ending nouns, which could somehow neutralize the umlaut rules.

5 Acoustic Analysis

5.1 Acoustic Analysis Based on Fieldwork Data in 2020

This section presents the results of an acoustic analysis of the umlaut. The audio data analysed consisted of around 28 minutes of elicitation recorded in 2020, with the aim of capturing the pronunciation of nouns and adjectives. The speaker was a 67-year-old male Jerusalem Domari speaker, the same speaker with whom I conducted the descriptive research introduced in section 4. During this session, I asked him to repeat carrier sentences in Domari such as "*ɑhɑ* __" or "*ihi* __" (This is __) so that the gender of the words is indicated by the demonstrative *ɑhɑ* 'This.M' or *ihi* 'This.F'. For example, "*ɑhɑ ɑːn-eːk*" (This.M egg-PRED.SG.M) 'This is an egg' or "*ihi baːj-oːm-i*" (This.F mother-1.SG-PRED.SG) 'This is my mother'. After I said the sentence, the speaker repeated it with accurate Domari pronunciation. I recorded

some additional words on another day in the same manner and setting, in the same room. I selected nouns and adjectives from the recorded file and manually extracted vowel areas from them, measuring the mean values of the first formant and second formant of the extracted vowel areas using Wavesufer. There were 50 tokens of 31 words, including multiple repetitions of the same words with long open vowels [ɑː] or [aː]; 31 tokens of 22 words with short open vowels [ɑ] or [a]; 17 tokens of 7 words with long close-mid vowels [oː] or [ɵː]; and 9 tokens of 5 words with short close-mid vowels [o] or [ɵ]. As evident from the number of tokens, the number of nouns or adjectives including close-mid vowels is considerably less than those with open vowels.

In the following charts, I have plotted the first formant (F1) and the second formant (F2) of all tokens of those vowels in the last syllable of nouns and adjectives. A higher F1 value indicates a more open vowel, while a higher F2 value means a more frontal position of the vowel. 'F' represents feminine words, and 'M' indicates masculine words **[charts 1-2]**.

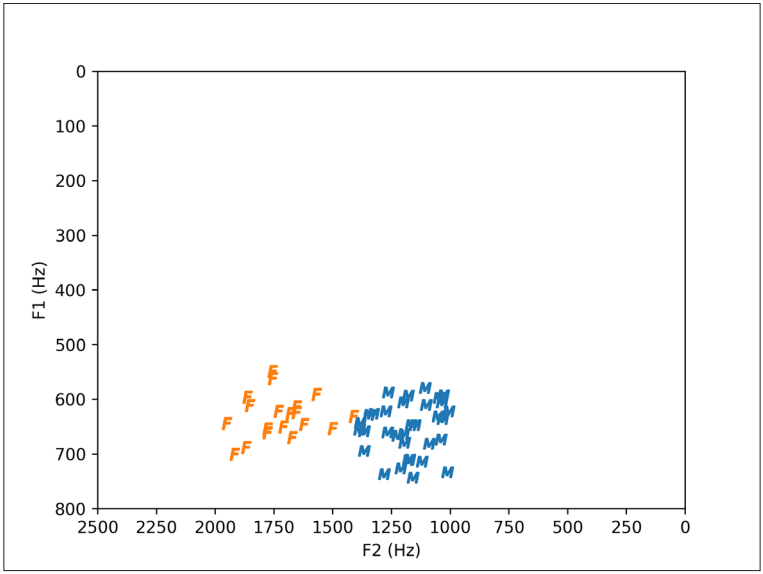


Chart 1 Umlaut in Long Open Vowels [ɑː]/[aː]

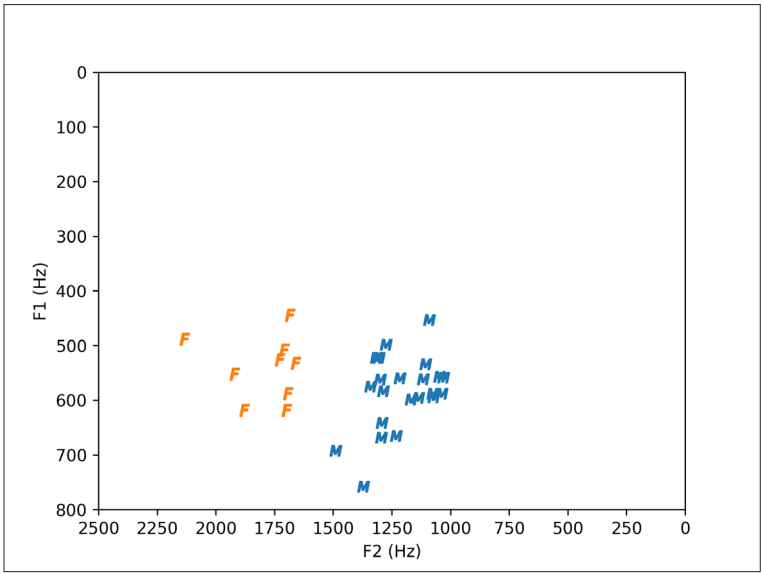


Chart 2 Umlaut in Short Open Vowels [ɑ]/[a]

Chart 1 illustrates the umlaut phenomenon in long open vowels. It shows that the open vowels in feminine words exhibit a higher F2 than those in masculine words, indicating that the open vowels in feminine words are more fronted.

Chart 2 displays the umlaut phenomenon in short open vowels. It shows that the open vowels in feminine words exhibit a higher F2 compared to those in masculine words, suggesting that the open vowels in feminine words are more fronted. Furthermore, the border between front and back vowels in short open vowels appears to be more distinct than in long open vowels.

Chart 3 illustrates the umlaut phenomenon in long close-mid vowels. The close-mid vowels in feminine words exhibit a higher F2 compared to those in masculine words, suggesting that the close-mid vowels in feminine words are more centralized. Additionally, the close-mid vowels in feminine words exhibit a slightly lower F1 than those in masculine words, indicating that they are more closed.

Chart 4 demonstrates the umlaut phenomenon in short close-mid vowels. Observation shows that the close-mid vowels in feminine words have a higher F2 than those in masculine words, indicating centralization in the former **[charts 3-4]**.

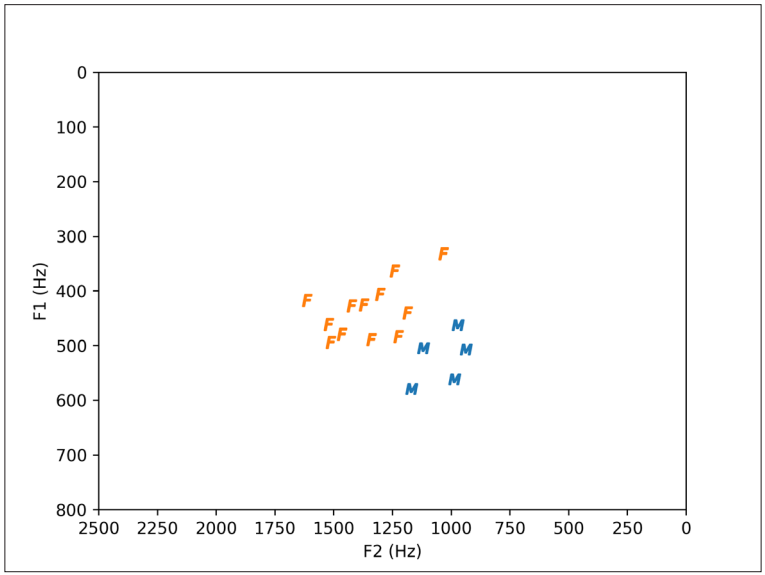


Chart 3 Umlaut in Long Close-mid Vowels [oː]/[eː]

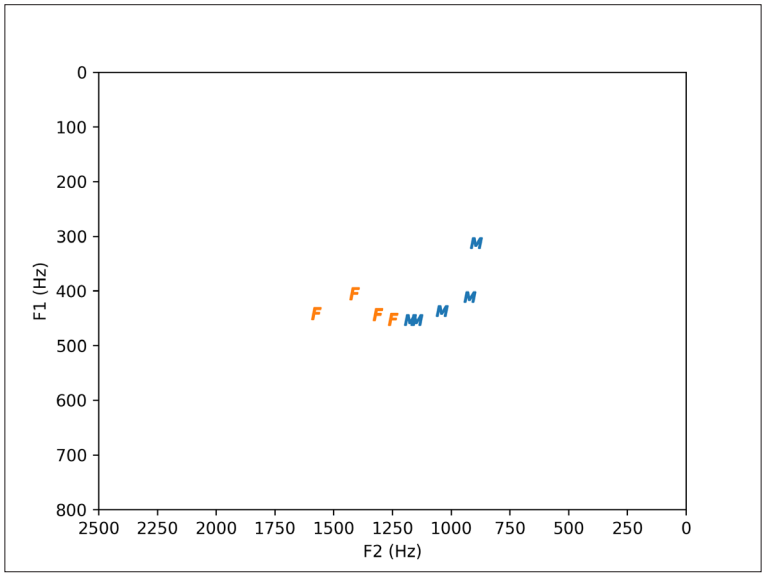


Chart 4 Umlaut in Short Close-mid Vowels [o]/[e]

Chart 3 and 4 have differences in the F1 values. The F1 values of the short close-mid vowels in Chart 4 do not show lower values in feminine nouns. This observation suggests that the defining aspect of the umlaut phenomenon would relate to the F2 values, which reflect the front and back positions of vowels, and variations in F1 seem to be influenced by other phonological factors. However, a more detailed analysis is needed for an extensive discussion of this topic.

5.2 Acoustic Analysis in Audio Samples from Matras (2012)

Recently, I had the opportunity to read the work of Schubert (2007), compiled under the supervision and direction of Professor Matras, which offers a detailed phonetic analysis of audio data of a female Domari speaker recorded by Professor Matras in 2000. This analysis helped inform the phonetics and phonology sections in Matras's 2012 publication, *A Grammar of Domari*. In Schubert (2007), there is a report of an umlaut phenomenon similar to the one described in this study, a detail that was prompted by Professor Matras, but not further discussed in Matras 2012.

> A phoneme /æ/ was suggested to occur in particular lexical items (*næmos, gæmos*), and as an umlaut vowel in feminines (e.g. *māmyom*), with a fronted raised quality compared to the vowel in their masculine counterparts (*māmom*). The feminine umlaut is probably better analysed as fronting from /ɑ/ to /a/. (Schubert 2007, 2)

This led me to apply a methodology similar to that used with my audio data to analyse the audio data of narratives that were transcribed in Matras (2012, 391-425), which were kindly made available to me by Professor Matras (personal communication). Due to the limited number of tokens in the sample audio, the analysis was restricted to long vowels.

There were 27 tokens of 7 words, including multiple repetitions of the same words with long open vowels [ɑː]/[aː] in the final syllable of their stems (*bɑːj-* 'father', *bɑːd-* 'grandfather', *xɑːl-* 'maternal.uncle', *daːd-j-* 'grandmother-F', *maːm-j-* 'paternal.aunt-F', *baː-j-* 'mother-F', *maːs-i* 'meat-F'), from Sample 2, 'Life after retirement', a 3-minute, 46-second narrative, and 12 tokens of two words with long close-mid vowels [oː]/[eː] in the final syllable of their stems (*ʃoːn-ɑ* 'non-Dom.boy-M' and *ʃoːn-i* 'non-Dom.girl-F'), from Sample 3, 'A love tale', an 8-minute, 56-second narrative. The two audio files were recorded from distinct male Jerusalem Domari speakers, each of whom was in his sixties during the period from 1997 to 2000, when the fieldwork was conducted.

Applying the same methodology as in the preceding subsection, I charted the first and second formants of both the long open vowels and long close-mid vowels located in the final syllables of nouns or adjectives.

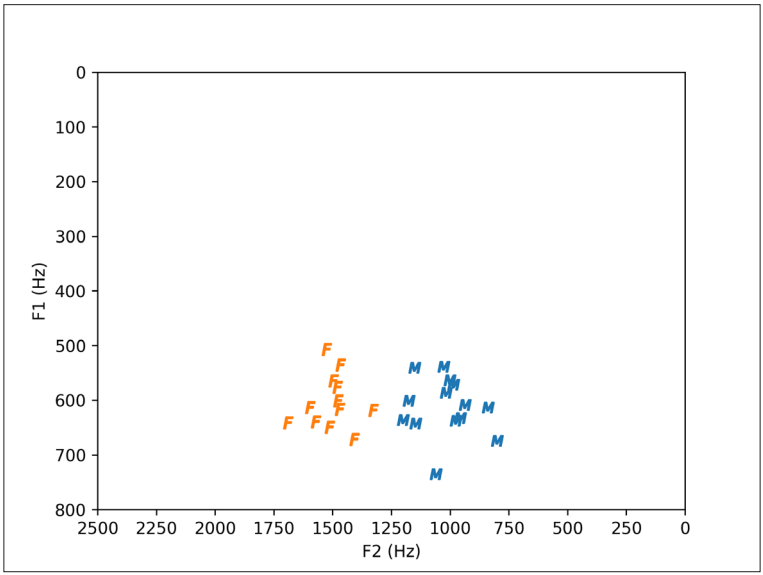


Chart 5 Umlaut in Long Open Vowels [ɑː]/[aː]

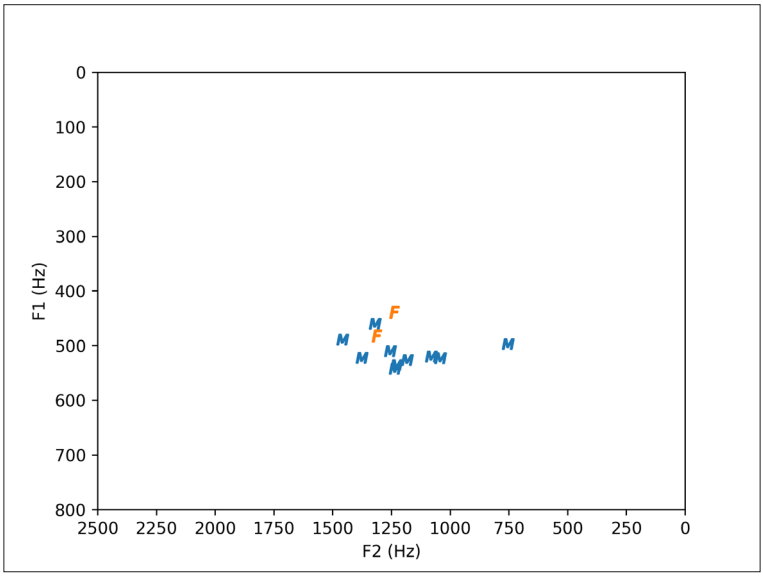


Chart 6 Umlaut in Long Close-mid Vowels [oː]/[ɵː]

Chart 5 shows that the open vowels in feminine words possess higher F2 values compared to those in masculine words, indicating a more fronted articulation of open vowels in feminine words.

However, in Chart 6, the presence of umlaut in long close-mid vowels was not observed: the F2 value in long close-mid vowels does not vary significantly with gender. However, the number of tokens is too low to draw definitive conclusions. Tentatively, I conclude that while umlaut in open vowels is a prevalent feature in Jerusalem Domari, it is not observed prevalently in the close-mid vowels [o(:)] and [ɵ(:)], or at least not for all speakers **[charts 5-6]**.

6 Discussion: A New Perspective on the Vowel System of Jerusalem Domari

In the preceding section's acoustic analysis, a phonetic distinction between open vowels in masculine nouns/adjectives and those in feminine nouns/adjectives was observed in both speakers. This distinction is evidenced by lower F2 values in masculine words, indicating a more back tongue position, transcribed as [ɑ(:)]. In contrast, feminine words exhibit higher F2 values, suggesting a more front tongue position, represented as [a(:)]. As for the close-mid vowels, one speaker displayed a pattern of lower F2 values in masculine words and higher F2 values in feminine words, while this pattern was not observed in the other speaker.

Considering these results, it seems reasonable to suggest a revised phonemic system for Jerusalem Domari.

- There is a contrast between [+BACK] and [-BACK] in open vowels, resulting in two distinct open vowel phonemes: /ɑ(:)/ and /a(:)/. The following examples illustrate some minimal pairs. As discussed in Section 3, these two vowel phonemes also occur independently of the umlaut context.

ex. *bɑ:j-om* 'father-1.SG' vs *ba:-j-om* 'wife-F-1.SG'
qar-i 'stupid.M-PRD' vs. *qar-i* 'stupid-F.NOM'
mɑ:s-i 'month-PRD' vs. *ma:s-i* 'meat-F.NOM'

This poses a typological difference between the vowel system of Romani, which has one open vowel phoneme /a/ in most of its dialects, and that of Domari, which has two open vowel phonemes: /ɑ(:)/ and /a(:)/.

- Variation between two speakers is observed regarding the umlaut in close-mid vowels [o(:)] and [ɵ(:)], suggesting a phonological contrast between the two vowels for one speaker but not the other. Given the low number of tokens for close-mid vowels, especially for the second speaker, this conclusion is tentative.

ex. *ʃoːn-a* 'non-Dom.boy-M.NOM'
vs. *ʃəːn-i* 'non-Dom.girl-F.NOM' in one speaker
ʃoːn-a 'non-Dom.boy-M.NOM'
vs. *ʃoːn-i* 'non-Dom.girl-F.NOM' in the other speaker

This proposal addresses the challenges associated with identifying vowel phonemes in Jerusalem Domari, as explained in the previous section 2.3.

Figure 2 in § 2.3 illustrates the vowel system as presented by Matras (2012), while Figure 3 shows the newly proposed vowel system in this study **[fig. 3]**.

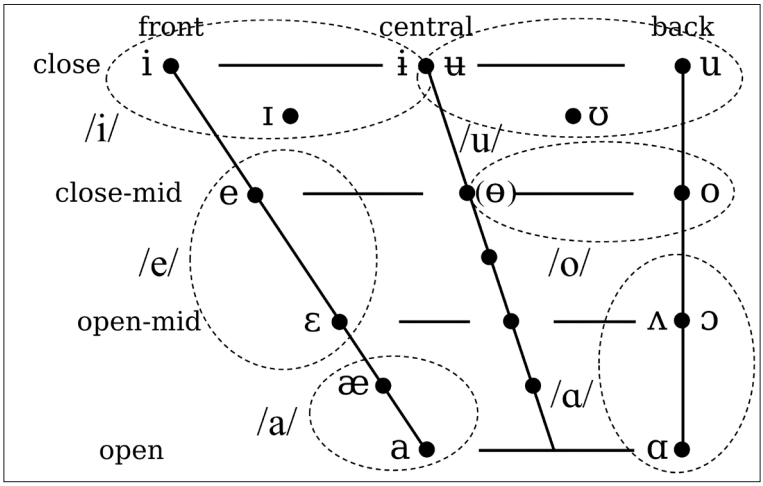


Figure 3 Proposed Vowel System

The primary differences involve the open vowel, as well as the back close-mid and open-mid vowels. As discussed in 2.3, the phonemic statuses of [ʌ] and [ɔ] have been regarded as ambiguous. I propose to consider the open-mid vowels [ʌ] and [ɔ] as allophones of the back open vowel phoneme /ɑ/. Establishing /ɑ/ as a phoneme contributes to solving the problem of the unclear phonemic status of [ʌ] and [ɔ] mentioned in Matras (2012).

Furthermore, the representation of the central close-mid vowel [ɵ] is enclosed in brackets. Considering the inter-speaker variation in the umlaut of close-mid vowels, it is challenging to decisively determine whether the central close-mid vowel [ɵ] functions as an allophone of the back close-mid vowel /o/ or is an independent phoneme.

In addition, the phenomenon of particular interest is the functioning of the open vowels [a] and [ɑ] as allophones in Arabic loanwords. Matras mentioned this point:

> The realisation of /a/ as [ɑ] is in fact consistent or obligatory in the immediate environment of pharyngeals, and so one might speak of a pharyngealizing effect on the vowel, similar to that found in Arabic. (Matras 2012, 39)

ex. [ˈzˤaːbɪt] Arabic *ẓābiṭ* 'officer' (Matras 2012, 47)
[tˤɑwˈlɛ] Arabic *ṭawle* 'table' (Matras 2012, 44)

This effect is widely common in Arabic. Cowell (1964), in his descriptive research on "Syrian Arabic", the colloquial Arabic of what is called "Greater Syria", including Syria, Lebanon, Jordan, and Palestine, explains as follows:

> Velarization is usually not limited to a single sound in a word, but commonly affects whole syllables and often whole words: *ḍaḷḷ, maḅsũṭ, ẓāḅeṭ.* (Cowell 1964, 7)

In addition, Cowell (1964) mentions the velarized glottal stop sound [ʔˤ], showing an example of a minimal pair: *ʔáššaṛ* 'he signalled' vs. *ʔ́áššaṛ* 'he peeled' (Cowell 1964, 7). There is a view that attributes the pharyngeal effect to the open vowel in these words rather than to the glottal stop and establishes the pharyngeal vowel phoneme /aˤ/ as a distinctive vowel from /a/, as in Obégi (1971, 25-8), the analysis of phonemic systems of Lebanese Arabic. This phenomenon is, however, observed as a marginal case limited to some areas, as Cowell (1964, 8) points out:

> In a large part of the central area, including Damascus, and most of Lebanon, the distinction between ʔ and ʔ́ is likewise obliterated, and is likewise subject to much vacillation elsewhere.

Thus, the view in which the back open vowel /aˤ/ is treated as a distinct phoneme from /a/ in Arabic seems unrelated to the Domari open vowels discussed in this paper.

Consequently, the phenomenon of open vowels in Domari is summarized as follows: These open vowels /a/ and /ɑ/ behave as distinct phonemes in pre-Arabic Domari lexicon, yet they function as allophones with complementary distribution in Arabic loanwords. This indicates that Domari possesses complex phonological layers, exhibiting dual phonological systems within its structure.

Cross-linguistically, this kind of accurate phonological borrowing without phonological adaptation, as seen in Arabic loanwords in Domari, occurs under an intense bilingual setting (Matras 2009, 342). In that situation, speakers have full knowledge of the donor language and make an effort to replicate the original phonology. In most cases, this type of phonological borrowing causes enrichment of the

phonemic system in the recipient language by adding new phonemes to its inventory, without causing conflict between the phonological environments of native words and loanwords. In the case of Domari discussed in this study, however, the phonological rule for the open vowel [ɑ] as an allophone of /a/ in Arabic loanwords, modelled on Arabic, adds another phonological layer to the existing phonemic system of pre-Arabic words, where the back open vowel /ɑ/ is a distinct phoneme from the front open vowel /a/.

In the work of Kossmann (2010, 459), the phenomenon called Parallel System Borrowing is explored. This refers to the type of borrowing which leads to "a coexistence of borrowed and native paradigms in one and the same language".

Kossmann describes two key typological or sociolinguistic factors that influence Parallel System Borrowing. The first is defined as

> contact situations with a high-prestige language, which is used for purposes related to religious and scientific learning, and which is formally taught. (480)

The second factor is described as a bilingualism setting where

> relatively small language communities in a setting with a foreign language, which is the dominant language in most communicative domains that extend outside the community. (481)

Although Kossmann's study focuses on morphology, the latter sociolinguistic setting is notably similar to that of Domari, and the results of this study imply Parallel System Borrowing in the field of phonology, particularly in the phonemic system.

7 Conclusion

In conclusion, through descriptive research and acoustic analyses of fieldwork data from one Domari speaker recorded in Jerusalem during 2019–2020, along with acoustic analyses of sample audio data from two additional speakers as presented in Matras (2012), this study can be summarized by the following four points:

1. In Jerusalem Domari, umlaut in nouns and adjectives with open vowels in their final syllables is observed. Consonant-ending nouns and adjectives, which have lost the final vowels that would have triggered umlaut, also exhibit umlaut.
2. Pre-Arabic loanwords, including those from Persian, Kurdish, and Turkish, are assimilated into the same morpho-phonological rules of umlaut as Indo-Aryan native words, whereas loanwords from Arabic, the latest contact language, do not follow these rules.
3. The two open vowels /ɑ(:)/ and /a(:)/ are recognized as distinct phonemes in pre-Arabic words. It should be noted, however, that the distinction between these open vowels has existed independently of the umlaut phenomenon.
4. In Arabic loanwords, the open vowels [ɑ(:)] and [a(:)] function as allophones of /a(:)/ with complementary distribution. This implies the possibility of Parallel System Borrowing in a phonemic system in the bilingual setting of Domari.

There remains a challenge to be addressed in understanding this phenomenon. The umlaut on the close-mid vowels [o(:)] and [ɵ(:)] exhibits apparent inter-speaker variation.

In addition, it is essential to address the historical explanation of this phenomenon: is the umlaut in question a recent innovation or an archaic retention? The umlaut phenomenon has not been documented in Northern Domari or Romani, despite the observed phonemic distinction between the back open vowels /ɑ(:)/ and /a(:)/ in the former (Herin 2012; 2014). However, there has been no research specifically concentrating on the phonetics or phonology of Northern Domari. It means that there is still a possibility of related phenomena in this dialect.

To investigate the inter-speaker variation regarding close-mid vowels [o(:)] and [ɵ(:)] in detail and to determine whether this umlaut constitutes the retention of an archaic feature or represents a recent innovation, it is essential to collect additional fieldwork data from other dialects.

Bibliography

Boeston, K.; Koni Muluwa, J. (2014). "Umlaut in the Bantu B80 Languages of the Kwilu (DRC)". *Transactions of the Philological Society*, 112(2), 209-30.
https://doi.org/10.1111/1467-968x.12033

Cowell, Mark W. (1964). *A Reference Grammar of Syrian Arabic.* Washington, D.C.: George Town University Press.

Herin, B. (2012). "The Domari Language of Aleppo (Syria)". *Linguistic Discovery*, 10(2), 1-52.
https://doi.org/10.1349/ps1.1537-0852.a.412

Herin, B. (2014). "The Northern Dialects of Domari". *Zeitschrift der Deutschen Morgenländischen Gesellschaft*, 164(2), 407-50.
https://www.jstor.org/stable/10.13173/zeitdeutmorggese.164.2.0407

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

Janda, R.D. (2003). "'Phonologization' as the Start of Dephoneticization – or, on Sound Change and Its Aftermath: Of Extension, Generalization, Lexicalization, and Morphologization". Joseph, B.D.; Janda, R.D. (eds), *The Handbook of Historical Linguistics.* Oxford: Blackwell, 401-22.

Kossmann, M. (2010). "Parallel System Borrowing: Parallel Morphological Systems Due to the Borrowing of Paradigms". *Diachronica,* 27(3), 459-87.
https://doi.org/10.1075/dia.27.3.03kos

Koul, O.N. (2003). "Kashmiri". Cardona, G.; Jain, Dh. (eds), *The Indo-Aryan Languages*. London: Routledge, 891-948.
https://doi.org/10.4324/9780203945315

Liljegren, H. (2019). "Gender Typology and Gender (In)Stability in Hindu Kush Indo-Aryan Languages". Di Garbo, F.; Olsson, B.; Wälchli B. (eds), *Grammatical Gender and Linguistic Complexity.* Vol. 1, *General Issues and Specific Studies*. Berlin: Language Science Press, 279-328.
https://doi.org/10.5281/zenodo.3462772

Matras, Y. (2009). *Language Contact.* Cambridge: Cambridge University press.

Matras, Y. (2012). *A Grammar of Domari.* Berlin: De Gruyter Mouton.

Masica, C.P. (1991). *The Indo-Aryan Languages.* Cambridge: Cambridge University Press.

Obégi, M. (1971). *The Phonemic System of a Lebanese Arabic Dialect* [Master's thesis]. Burnaby: Simon Fraser University.
https://summit.sfu.ca/item/2958

Schubert, C. (2007). *Domari Results of Phonetic Analysis* [Unpublished].
https://domari.org/wp-content/uploads/2024/08/domari_results_of_phonetic_analysis.pdf

Turner, R. (1926). "The Position of Romani in Indo-Aryan". *Journal of the Gypsy Lore Society*, 3(5), 145-89.

Wiese, R. (1996). "Phonological Versus Morphological Rules: on German Umlaut and Ablaut". *Journal of Linguistics*, 32(1), 113-35.
https://doi.org/10.1017/s0022226700000785

e-ISSN 2785-5953

Bhasha
Vol. 3 – Num. 2 – October 2024

Mortal Combat and the Hereafter: *saṃparāya* and *sāṃparāya* in Sanskrit Literature

Dominik A. Haas
Leiden University, Netherlands

Abstract *Sāṃparāya* is a comparatively rare word in Sanskrit literature that has been translated in various ways, among them 'battle', 'the next world', and 'transit to the next world'. It is a nominalized adjective derived from *saṃparāya*, which is in turn derived from the rarely used verb *sam+parā+i* (to pass away, to decease). The aim of the present paper is to establish the basic meanings of *sāṃparāya* as well as of its base, *saṃparāya*, and to investigate how they are used in Sanskrit literature. To this end, text passages from a range of sources are discussed. It is shown that in pre-medieval literature, *saṃparāya* and *sāṃparāya* are generally used as variants of one and the same noun, and that this noun has two basic meanings: 'mortal combat, battle' and 'the postmortal, the hereafter'. The plethora of meanings recorded in modern dictionaries were mostly derived from later, highly context-specific commentaries.

Keywords sam+parā+i. Death. Battle. Afterlife. Prakrit.

Edizioni Ca'Foscari

Peer review

Submitted 2024-05-08
Accepted 2024-07-12
Published 2024-12-18

Open access

Citation Haas, D.A. (2024). "Mortal Combat and the Hereafter: *saṃparāya* and *sāṃparāya* in Sanskrit Literature". *Bhasha*, 3(2), 237-262.

DOI 10.30687/bhasha/2785-5953/2024/02/004

1 Introduction

Sāṃparāya is a rare word in Sanskrit literature that has been translated in various ways, among them 'battle', 'the next world', and 'transit to the next world'.[1] The aim of the present paper is to establish the basic meanings of *sāṃparāya* as well as of its base, *saṃparāya*, and to investigate how they are used in Sanskrit literature, with a focus on pre-medieval literature. To this end, a range of passages from texts pre-dating the eighth century CE are discussed. To cover different genres and historical periods, passages from the following sources have been selected:

- *Taittirīya-Brāhmaṇa* (TB)
- *Mahāvastu* (Mvu)
- *Mahābhārata* (MBh)
- *Rāmāyaṇa*
- three *Dharmasūtra*s: *Āpastamba*-° (ĀpDhS), *Baudhāyana*-° (BaudhDhS), and *Vasiṣṭha*-° (VasDhS)
- two *Gṛhyasūtra*s: *Hiraṇyakeśi*-° and *Baudhāyana*-°
- *Daśakumāracarita* (Daś)
- *Kaṭha-Upaniṣad* (KU)

I am aware of several other Sanskrit texts in which the two words occur.[2] However, a cursory examination has not revealed any pre-eighth-century instances in which the two words are used in meanings other than those identified using the small corpus defined above.

As will be shown, two basic meanings for the words *saṃparāya* and *sāṃparāya* were well established in the pre-medieval period: 'mortal combat, battle' and 'the postmortal, the hereafter'. There is

I would like to thank Andrey Klebanov, Georgi Krastev, Vitus Angermeier, and Channa Li for their valuable suggestions and Dragomir Dimitrov for his numerous comments on a previous version of this paper. I also thank Vincent Tournier, Valters Negribs, Christophe Vielle, and Marcus Schmücker for their help with various passages. The paper is partly based on an appendix in my Master's thesis, which was supervised by Marion Rastelli; see Haas 2018, 138-43. Its completion was made possible in part by the J. Gonda Foundation of the Royal Netherlands Academy of Arts and Sciences and the International Institute for Asian Studies.

Unless otherwise indicated, translations are by the Author. Square brackets are used to mark explanatory additions to or modifications of quotations and translations; if they are part of original quotations, this is indicated. Hyphens after Sanskrit words indicate that the word is a sandhi form. Regarding the transcription of the Anusvāra sign with *ṃ*, I have opted for the spelling used in the editions I have consulted.

1 Translations of the word are mentioned in this paper when discussing the relevant text passages; see also the dictionary entries cited below.

2 E.g. the *Amarakośa*, the *Bhāgavata-Purāṇa*, the *Bodhisattvabhūmi*, the *Divyāvadāna*, the *Kāśikā-Vṛtti*, the *Kāvyaprakāśa*, the *Kumārasambhava*, the *Mahāparinirvāṇa-Sūtra*, the *Mahāsudarśana-Avadāna*, the *Mahāyānasūtra-Alaṅkāra*, the *Pañcaviṃśatisāhasrikā Prajñāpāramitā*, the *Rāṣṭrapāla-Paripṛcchā*, and the *Śrāvakabhūmi*.

nothing to suggest that these two meanings have ever become obsolete over the course of the last two millennia. However, medieval commentators also introduced alternative interpretations that were often guided by their own exegetical agenda. In this paper, only a few text passages dealing with such reinterpretations and explanations are discussed: Śaṅkara's commentary on the *Kaṭha-Upaniṣad* as well as his *Brahmasūtra-Bhāṣya*, Raṅgarāmānuja's commentary on the *Kaṭha-Upaniṣad* (Prakāśikā), Śāṇḍilya's *Bhaktisūtra*s with Svapneśvara's commentary, and Nīlakaṇṭha's commentary on the *Mahābhārata* (the *Bhāratabhāvadīpā*).

2 *sam+parā+i* and *saṃparāya*

The verb *parā+i* means 'to go away, to depart' and is already used in the Ṛgveda in the sense of 'to pass away' or 'to decease' (see e.g. *Ṛgveda* X 14.1-2). It seems that the verb *sam+parā+i* essentially means the same as *parā+i* (the prefix *sam* may serve the purpose of intensification: 'to go away completely'). In contrast to *parā+i*, the verb *sam+parā+i* is not well attested in Vedic and Sanskrit literature: the pw (I, 200) only knows the participle *saṃpareta* in *Aitareya-Āraṇyaka*,[3] the PW only *saṃpareta* in the much later *Bhāgavata-Purāṇa*.[4] A search (June 4, 2023) in the *Digital Corpus of Sanskrit*, the corpus of the *Göttingen Register of Electronic Texts in Indian Language*, and the *Thesaurus Indogermanischer Text- und Sprachmaterialien* yielded only a few more results, but essentially confirmed that this verb was used comparatively rarely.

The verb *sam+parā+i* can theoretically be nominalized as an action noun in at least four ways: *saṃparāya* (m.), **saṃparāyaṇa* (n.), **saṃpareta* (n.), and **saṃpareti* (f.); however, only *saṃparāya* is

3 According to *Aitareya-Āraṇyaka* III 2.4 (cf. *Śāṅkhāyana-Āraṇyaka* VIII 7), a man whose self (*ātman*) and the sun 'gape' (*vi+hā*; in reality, these two are one and the same; cf. Norelius 2023, 310) will not live much longer. The text explains that 'his self has gone away' (*saṃpareto 'syātmā*), which most likely means that the most important vital power has left the body.

4 PW V 1137: "Bhāg. P. 5, 2, 22. [= *Bhāgavata-Purāṇa* V 2.23 in the edition used for the present study] 10, 44, 38 [in both editions]".

attested.[5] The PW, pw,[6] MW, and Apte[7] attribute a variety of meanings to the noun *saṃparāya*. MW (whose dictionary is largely based on the PW),[8] for instance, lists a total of six meanings:

1. decease, death;
2. existence from eternity;
3. conflict, war, battle;
4. calamity, adversity;
5. futurity, future time;
6. a son.

Most of these meanings seem to have been simply derived from various context-specific Sanskrit commentaries and glosses composed centuries after the word came into use – that alone makes it seem worthwhile to review them. In the following, I will argue that the word has only two basic meanings.

First, *saṃparāya* denotes the 'passing away' of a person as a consequence of the complete cessation of all vital functions of the body, and may in that sense be translated as 'death'. Second, there is also a metonymic[9] understanding of the word, which generally seems to be restricted to – and most likely originated in – Buddhist usage: *saṃparāya* does not only denote the process or event of passing away, but also the destination or target of this 'movement'.[10] Like

5 Curiously, *sāṃparāyaṇa* is attested as a name of Death (*mṛtyu*) in the *Mantrārṣa-Adhyāya* of the *Kaṭha-Saṃhitā* (Weber 1855, 459). In *Yama-Smṛti* V 16cd, the word is used (obviously in another meaning) to qualify sons (*aputrasya ca putrāḥ syuḥ kartāraḥ sāṃparāyaṇāḥ*), possibly because they are responsible for the fate of their ancestors in the hereafter; cf. Contexts 6-9 below.

6 pw VII, 737: 1. "Tod", 2. "das von-Ewigkeit-her-Sein", 3. "Kampf", 4. "Ungemach, Unglücksfall", 5. "Zukunft"; PW has the same meanings with slight variations.

7 1. "Conflict, encounter, war, battle", 2. "A calamity, misfortune", 3. "Future state, futurity", 4. "A son".

8 Cf. fn. 6; see generally Steiner 2020.

9 According to Macdonell (1927, 159, §182), nouns derived from roots with a primary suffix (such as a in *saṃparāya*) "may be divided into the two classes of abstract action nouns (cognate in sense to infinitives) and concrete agent nouns (cognate in sense to participles) used as adjectives or substantives […] Other meanings are only modifications of these two". If one considers these two classes primary, using *saṃparāya* to denote both the action itself and its target or result involves a 'metonymic modification' of the meaning of the word. However, while primary suffixes are most often used to derive action and agent nouns, it remains open if these two constitute primary classes: one could argue that they may just as well be used to form 'result nouns' (e.g. in the case of *bhāga* [from *bhaj* 'to share'], which denotes the result of sharing: 'a share') or 'object nouns' (e.g. in the case of *veda* [from *vid* 'to know'], which denotes the object of knowing: 'knowledge'), without any 'modification' being involved.

10 At least theoretically, speaking about the process of dying as a departure presupposes the existence of something that does not simply vanish after the moment of death, but leaves the body, and probably also moves to another place. However, verbs expressing a movement away from something can also easily be used metaphorically (as also in

sāṃparāya (which will be discussed in the next Section), it refers in a general way to a place or state characterized by being reached or attained after death, and thus can be translated as 'the postmortal' or 'the hereafter', leaving open whether it refers to a state ('death') or a world and the life one leads therein ('the next world, the afterlife').

As a search in the UVC shows, the earliest text to mention the word *saṃparāya* in the sense of 'passing away' or 'death' is the TB:

> *nā́vedavin manute táṃ bṛhántam, sarvānubhū́m ātmā́naṁ samparāyé.*
>
> One who does not know the Veda, does not think of the great, all-perceiving Self at death.[11]

The passage probably suggests that it is important to think of the 'cosmic' Self upon passing away, that is, at the moment of death. In doing so, it expresses a common notion in South Asian religions, namely that the thoughts one has when (or possibly also before) dying have a strong influence on one's destiny after death (e.g. in a next reincarnation).

In post-Vedic literature, the word frequently appears in Buddhist texts, where it is generally used in the meaning of 'the postmortal, the hereafter'. It seems to be used in the same way as *abhisamparāya* and similarly to *gati*, words that are used to denote the various 'destinations' that beings can reach after death, that is, their 'destinies' in the cycle of rebirth. It is also found in Pāli texts, where it is apparently only used in its second original sense.[12]

The *Mahāgovinda-Sūtra*, a text contained in the *Mahāvastu* that was composed in Buddhist Hybrid Sanskrit and also has a (partial) parallel in the Pāli canon, illustrates the Buddhist usage of *saṃparāya*. In one passage, we read about a conversation between a Brahmin called Govinda and the god Brahmā. After waiting on the god, who is his guest, the Brahmin says the following:

---

the case of the English verb 'to pass away'), even without elaborate notions of a moving soul or of a hereafter. While Buddhism embraces the idea of rebirth, it generally rejects the existence of an individual essence (such as an *ātman* or 'self') that is reborn (or at least it discourages clinging to the idea that such an essence exists; see Wynne 2011).

11 TB III 12.9.7. Cf. the somewhat inaccurate translation by Dumont (1951, 674): "He who does not know the Veda, does not perceive, after death, the all-perceiving great Ātman". For remarks on the often-problematic translation of locatives with the word 'after', see Hopkins 1903, 4-6.

12 PTSD, see under *samparāya*; cf. also the entry on *abhisamparāya*: "future lot, fate, state after death, future condition of rebirth; usually in foll. phrases: *kā gati ko abhisamparāyo* (as hendiadys) 'what fate in the world-to-come'".

"dṛṣṭe dharme hitārthaṃ saṃparāyasukhāni vā
kṛtāvakāśo pṛccheyaṃ yaṃ me manasi prārthitaṃ"

evam ukte bhavanto mahābrahmā mahāgovindaṃ brāhmaṇaṃ gāthāya pratyabhāṣati

"dṛṣṭadharme hitārthaṃ vā saṃparāyasukhāni vā
kṛtāvakāśaḥ pṛcchāhi yaṃ tava syābhiprārthitaṃ"

"Whether it be for the sake of welfare in the visible condition or happiness in the *saṃparāya*,
having the opportunity, I would like to ask what's on my mind!"

Thus addressed, O honorable ones, great Brahmā replied to great Govinda, the Brahmin, in verse:

"Whether it be for the sake of welfare in the visible reality or for happiness in the *saṃparāya*,
having the opportunity, ask what may be on your mind!"[13]

Here, happiness in 'the postmortal' or 'hereafter' (*saṃparāya*) is clearly contrasted with happiness in the 'visible reality' (*dṛṣṭadharma*), an expression referring to the present world or state. In the following, the text also uses derived adjectives to distinguish between goals profitable in the visible reality (*dṛṣṭadhārmika*) and goals profitable in the hereafter (*sāṃparāyika*):

atha khalu puna bhavanto mahāgovindasya brāhmaṇasya etad abhūṣi "pravāritaṃ me khalu mahābrahmaṇā praśnavyākaraṇena. kiṃ dān(') imaṃ mahābrahmāṇaṃ praśnaṃ pṛccheyaṃ dṛṣṭadhārmikaṃ artham ārabhya utāho sāṃparāyikaṃ?". atha khalu bhavanto mahāgovindasya brāhmaṇasyaitad abhūṣi "asti tāvad ayaṃ dṛṣṭadhārmiko artho yam idaṃ pañca kāmaguṇārabhya. yaṃ nūnāhaṃ mahābrahmāṇaṃ sāṃparāyike arthe praśnaṃ pṛccheyaṃ".

But then, mind you, O honorable ones, the following occurred to great Govinda, the Brahmin: "I have been granted, mind you, an explanation of a question by the great Brahmā. Now what question should I ask that great Brahmā about? About benefit that's profitable in the visible reality, or in the *saṃparāya*?" Then, mind you, O honorable ones, the following occurred to great Govinda, the

13 Mvu III 211 (cf. the translation by Jones 1956, 207). Parallel text of this stanza in *Dīgha-Nikāya* II 240: *diṭṭha-dhamma-hitatthāya samparāya-sukhāya ca, katāvakāso +pucchassu* [ed.: *puccha ssu*] *yaṃ kiñci abhipatthitan' ti*.

> Brahmin: "It's just this much, this benefit pertaining to the visible reality, namely this: it's about the five qualities of desire here. I should rather ask great Brahmā a question about benefit pertaining to the *saṃparāya*".[14]

Since the Mahāvastu is a Buddhist text, great Govinda is subsequently advised by Brahmā to go forth into homelessness – which he does.

As we shall see in the following Section, a very similar conception is found in non-Buddhist texts too; however, it is regularly referred to with the derived form *sāṃparāya*.

3 *saṃparāya* and *sāṃparāya* in Pre-Medieval Literature

The word *sāṃparāya* is a vṛddhi derivation from *saṃparāya*, and as such could theoretically denote anything 'relating to passing way' or 'coming after death'. It is not always clear whether *sāṃparāya* is a masculine or neuter noun, or an adjective; the main reason for this is that in the vast majority of cases, the word is used in the locative (*sāṃparāye*) or as the first part in a compound.

As in the case of *saṃparāya*, the dictionaries offer an entire range of meanings and translations for *sāṃparāya*;[15] MW, for instance, gives the following:

- adjective:
  - required by necessity or calamity;
  - relating to war or battle, warlike;
  - relating to the other world or to the future.
- masculine noun:
  - the passage from this world into another;
  - need, distress, calamity;
  - a helper or friend in need;
  - contention, conflict;

14 Mvu III 212 (cf. the translation by Jones 1956, 207). Partially parallel text in *Dīgha-Nikāya* II 240-1: *atha kho bho mahā-govindassa brāhmaṇassa etad ahosi: 'katāvakāso kho 'mhi brahmunā sanaṃkumārena. kin nu kho ahaṃ brahmānaṃ sanaṃkumāraṃ puccheyyaṃ diṭṭha-dhammikaṃ vā atthaṃ samparāyikaṃ vā ti?' atha kho bho mahā-govindassa brāhmaṇassa etad ahosi: 'kusalo kho ahaṃ diṭṭha-dhammikānaṃ atthānaṃ. aññe pi maṃ diṭṭha-dhammikaṃ atthaṃ pucchanti. +yan nūnāhaṃ* [ed.: *yannūnāhaṃ*] *brahmānaṃ sanaṃkumāraṃ samparāyikaṃ yeva atthaṃ puccheyyan' ti.*

15 PW (masculine noun): 1. "der Uebergang aus dieser Welt in die jenseitige", 2. "Noth, Bedrängniss", 3. "Kampf", 4. "etwa so v. a. ein Retter in der Noth"; pw has the same meanings, but also adds an adjective: "durch die Noth geboten"; Apte (adjective): 1. "Relating to war, warlike", 2. "Relating to the other world, future"; (masculine or neuter noun): 1. "Conflict, contention", 2. "Future life, the future", 3. "The means of attaining the future world", 4. "Inquiry into the future", 5. "Inquiry, investigation", 6. "Uncertainty".

- the future, a future life;
- inquiry into the future;
- investigation (in general);
- uncertainty.

As I shall argue in the following, it is again possible to identify two basic meanings: first, *sāṃparāya* denotes an activity that might result in death, a 'death-activity' – specifically, a deadly fight, mortal combat, or battle. This meaning is connected to the meaning of *saṃparāya* (passing away). Second, *sāṃparāya* denotes 'the postmortal' or 'the hereafter' in the sense of what comes after death. This is in fact equivalent to the Buddhist usage of *saṃparāya*, which also denotes the hereafter.

In the following, I shall discuss passages from various contexts using the words *saṃparāya* and *sāṃparāya* in order to establish what meaning they have in each case.

3.1 Mortal Combat, Battle

3.1.1 Context 1: *Sabhāparvan* (*sāṃparāya*)

In one passage of the *Mahābhārata* (II 69.14cd-15ab), Vidura speaks the following words to Yudhiṣṭhira, who is about to go into exile with his brothers:

> *mā hārṣīḥ sāṃparāye tvaṃ buddhiṃ tām ṛṣipūjitām //*
> *purūravasam ailaṃ tvaṃ buddhyā jayasi pāṇḍava /*
>
> In *sāṃparāya*, do not abandon your resolve/intellect/wisdom that is venerated by the Seers!
> You surpass Purūravas Aila in resolve/intellect/wisdom,[16] O Pāṇḍava.[17]

Yudhiṣṭhira's *buddhi* – his 'resolve', 'intellect', or 'wisdom' – is what will enable him to deal with the difficulties to come, and to take action when it is time to reclaim his kingdom. Most likely, Vidura here

16 It is difficult to see why Yudhiṣṭhira's *buddhi* is compared to Purūravas's. Purūravas Aila is best known for his love for the *apsaras* Urvaśī, which, at least according to some accounts, even drove him mad; at the same time, he is also said to have brought the three sacrificial fires to earth (MBh I 70.21). Possibly, the comparison refers to the fact that Purūravas is the son of Budha, the planet mercury, whose name literally means 'the wise one'. I would like to acknowledge the help of Valters Negribs and Christophe Vielle in making sense of this passage.

17 Cf. van Buitenen's translation: "lest you lose in the world-to-come [*sāṃparāya*] this resolve that the seers honor! Pāṇḍava, with this resolve you surpass Purūravas Aila".

thinks of the battles (and other potentially fatal situations) that still await the Pāṇḍavas; *sāṃparāye* (no manuscript reads *saṃparāye*) can thus be very well understood as 'mortal combat' or 'battle'.

3.1.2 Context 2: *Karṇaparvan* (*saṃparāya*/*sāṃparāya*)

In *Mahābhārata* VIII 27.92-93, Karṇa says the following to Śalya:

> *eṣa mukhyatamo dharmaḥ kṣatriyasyeti naḥ śrutam /*
> *yad ājau nihataḥ śete sadbhiḥ samabhipūjitaḥ //*
> *āyudhānāṃ saṃparāye yan mucyeyam ahaṃ tataḥ /*
> *na me sa prathamaḥ kalpo nidhane svargam icchataḥ //*
>
> We have heard that this is the most important Dharma of a Kṣatriya:
> that he lies slain in battle, honored by good people.
> That I should escape/be released, then, in an armed *saṃparāya*
> is not the first duty for me who seeks heaven in death.

This could mean that Karṇa does not want to be liberated (*muc*) from the cycle of rebirths in a 'death' through arms, but rather seeks salvation in heaven. It is much more likely, however, that he thinks of a deadly activity involving arms, an armed combat or battle of arms (B_5 even reads *saṃprahāra* 'battle'), and would not want to 'escape' (*muc*) when being engaged in such a combat (if the author of this stanza had wanted him to say that he does not want to escape death, an ablative would be much more likely).[18] According to the critical apparatus, 'Some MSS' read *sāṃparāye* (with a long *ā*).

Later in the conversation, Karṇa again makes use of the same word (MBh VIII 29.30cd). Referring to Arjuna and Kṛṣṇa, he says:

> *adya yuddhaṃ hi tābhyāṃ me saṃparāye bhaviṣyati //*
>
> For today, I will have a fight with those two in/about *saṃparāya*!

Here, the situation is similar to that in the previous passage. If one interprets *saṃparāya* as 'death', the passage could refer to a fight about – that is, until – death. The meaning 'mortal combat', however, is more likely. Again, a number of manuscripts also read *sāṃparāye* (V_1; B; D [D_2 omits the stanza, D_{3m} reads *saṃ*°]; T_1), which never means 'death'.

18 Several MSS also read *mamaiṣa prathamaḥ kalpo-* "this is my first duty" in 93bc, in which case it is necessary to understand ab as "That I should be released/find salvation in an armed battle".

3.1.3 Context 3: *Śāntiparvan* (*saṃparāya*/*sāṃparāya*)

In a passage of the *Śāntiparvan* (MBh XII 102.15), Bhīṣma describes the ideal characteristics of combatants. About a type of particularly muscular fighters, he says that they

> *praviśanty ativegena saṃparāye 'bhyupasthite /*
> *vāraṇā iva saṃmattās te bhavanti durāsadāḥ //*
>
> enter a *saṃparāya* with great speed when it has come.
> Like infuriated elephants, they become hard to deal with.

Here, the meaning 'mortal combat' or 'battle' is the most natural one. Again, several manuscripts read *sāṃparāye* ($K_{2, 4, 5}$; B [except B_1]; Da, Dn, $D_{2, 3, 4, 5, 7}$; S), and in this case the editors even decided to put a wavy line under *saṃ* in *saṃparāya* to mark the uncertain reading.

3.1.4 Context 4: *Rāmāyaṇa* (*sāṃparāya*/*saṃparāya*)

In the *Rāmāyaṇa*, the word *sāṃparāya* is used only once. When Hanumat visits Sītā, he offers to carry her away on his back. Sītā tries to dissuade him, arguing, for example, that the *rākṣasa*s might overpower him (V 35.53):

> *atha rakṣāṃsi bhīmāni mahānti balavanti ca /*
> *kathaṃcit sāṃparāye tvāṃ jayeyuḥ kapisattama //*
>
> Then again, the *rākṣasas* – terrible, mighty, and strong –
> might somehow defeat you in *sāṃparāya*, O best of monkeys.

Here again, we must assume that 'mortal combat' or 'battle' is intended. The manuscripts $D_{1-3, 8, 10-11}$ read *saṃparāya*, a word that is otherwise not to be found in the text of the critical edition.

3.1.5 Context 5: *Daśakumāracarita* (*sāṃparāya*)

In the *Daśakūmaracaritra*, which was composed by Daṇḍin in the seventh or eighth century, *sāṃparāya* is mentioned in three passages, each time clearly meaning 'battle'. To illustrate this, I here quote Onian's translation (who has divided Daṇḍin's often long sentences into smaller units):

> Daś I 21: *mānī mānasāraḥ svasainikāyuṣmattāntarāye saṃparāye bhavataḥ parājayam anubhūya vailakṣyalakṣyahṛdayo vītadayo.*

Mānasāra is proud by name and proud by nature. After his defeat at your hands in the battle that finished his soldier's hope for a long life, shame pierced his heart and he became heartless. (Onians 2005, 49)[19]

Daś VI 18: *jagṛhe ca mahati saṃparāye kṣīṇasakalasainyamaṇḍalaḥ pracaṇḍapraharaṇaśatabhinnamarmā siṃhavarmā kariṇaḥ kariṇam avaplutyātimānuṣaprāṇabalena caṇḍavarmaṇā*.

In the great battle which was then joined, Siṃhavarman lost the entire army of his soldiers. Hundreds of violent blows smashed his armor. With superhuman innate strength Caṇḍavarman leaped from his own elephant to Siṃhavarman's. (177)

Daś VII 263: *mahati saṃparāye bhinnavarmā siṃhavarmā balād agṛhyata*.

In the great battle with his far more numerous enemy Siṃhavarman's armor was smashed, and he was captured. (275)

3.2 The Postmortal (State/World), the Hereafter

3.2.1 Context 6: *Hiraṇyakeśi-Gṛhyasūtra* (*sāṃparāya*)

The Hiraṇyakeśi-Gṛhyasūtra contains two mantras mentioning *sāṃparāya*, to be recited in the context of an Aṣṭakā, a ritual whose purpose is to feed the deceased ancestors. The first mantra reads as follows (II 14.3):

imam apūpaṃ catuḥśarāvaṃ nirvapāmi +kleśāpahaṃ[20] *pitṝṇāṃ sāṃparāye devena savitrā prasūtaḥ*.

Impelled by the god Impeller, I offer this cake (prepared) from four cups, which removes the suffering in the *sāṃparāya* of the Fathers.

The second mantra reads as follows (II 15.2):

imāṃ pitṛbhyo gām upākaromi tāṃ me sametāḥ pitaro juṣantām / medasvatīṃ ghṛtavatīṃ svadhāvatīṃ sā me pitṝn sāṃparāye dhinotu / svadhā namaḥ.

19 In this and the following quotations, the original translation reads "Mana·sara", "Chanda·varman", and "Simha·varman". I have adapted these spellings to the convention used in this article.

20 Ed. *kleśāvahaṃ*; the emendation is suggested by Kirste in his edition.

> I bring this cow to the Fathers. May my assembled Fathers be pleased with it – it has fat, it has ghee, it has *svadhā*.[21] May it satiate my Fathers in the *sāṃparāya*. *svadhā*, obeisance!

As we know from other sources, the deceased Fathers (or forefathers) abide in the *pitṛloka* (the world of the Fathers). It is possible that *sāṃparāya* in the two mantras denotes this world;[22] however, it could just as well refer to their postmortal state as forefathers.

3.2.2 Context 7: *Dharmasūtra*s (*sāṃparāya*/*saṃparāya*)

Three *Dharmasūtra*s contain a passage (with several minor variants) on the importance of keeping one's line of male descendants pure. I here quote and translate the version of the ĀpDhS (II 13.6 ≈ BaudhDhS II 3.2; the last stanza is also found VasDhS XVII 9):

> *retodhāḥ putraṃ nayati paretya yamasādane /*
> *tasmād bhāryāṃ rakṣanti bibhyantaḥ pararetasaḥ //*
> *apramattā rakṣatha tantum etaṃ, mā vaḥ kṣetre parabījāni vāpsuḥ /*
> *janayituḥ putro bhavati sāṃparāye, moghaṃ vettā kurute tantum etam iti //*
>
> Having passed away, the impregnator guides his son in Yama's abode.
> This is why one guards one's wife, fearful of the seed of others.
> Diligently guard this line (of descendants), lest the seeds of others be sown in your field!
> In the *sāṃparāya*, a son belongs to the begetter; (otherwise) a husband makes the line worthless.

Here, we are confronted with the traditional Brahminical worldview, according to which the salvation of a man depends on his male offspring. The passage makes clear that after death, only biological sons 'count'. The use of the word *yamasādana* 'Yama's abode' in the first stanza points to the idea that one reaches a certain place after death; however, whether *sāṃparāye* in the second stanza denotes this place, the place in which Yama's abode is located, or a state remains open.[23]

---

21 *svadhā* is a technical term denoting an oblation to the ancestors (according to MW, it consists "of clarified butter &c. and often only a remainder of the Havis" and is "also applied to other oblations or libations"). At the same time, it is also an exclamation accompanying such an oblation.

22 Oldenberg (1892, 232, 234) translates *sāmparāya* in both mantras as "the other world".

23 Olivelle (2000, 93, 255, 417) translates *sāmparāya* as "transit to the next world".

The ĀpDhS and the BaudhDhS both read *sāṃparāya*, the VasDhS *saṃparāya* (none of the manuscripts Olivelle used for the critical editions evince significant variants in each case). Judging from the age of the texts, the version with *sāṃparāya* is the oldest, whereas *saṃparāya* in the VasDhS (the youngest *Dharmasūtra*) is a simplification.

sāṃparāya is mentioned three more times (with basically the same meaning) in the *Dharmasūtra*s. ĀpDhS II 24.3 again thematizes that sons are responsible for the postmortal fate of their ancestors:

> *te śiṣṭeṣu karmasu vartamānāḥ pūrveṣāṃ sāṃparāyeṇa kīrtiṃ svargaṃ ca vardhayanti*.
>
> By conducting the prescribed rituals, they (i.e. the sons) increase the fame and heaven of their predecessors in the course of the *sāṃparāya*.

The (tentative) translation of this passage is somewhat difficult due to the instrumental *sāṃparāyeṇa* – as in so many other passages, one would expect a locative *sāṃparāye* here. The interpretation underlying the present translation presupposes that *sāṃparāya* denotes the postmortal state in the sense of an 'afterlife', that is, something that has a duration (including a beginning and an end).

Another passage of the ĀpDhS (II 29.8-9) deals with witnesses and their obligation to answer truthfully, and locates hell in the *sāṃparāya*:

> *anṛte rājā daṇḍaṃ praṇayet / narakaś cātrādhikaḥ sāṃparāye*.
>
> If (his answer) is an untruth, the king should impose a penalty, and in addition, hell (awaits him) in the *sāṃparāya*.

What is meant by this is that hell awaits the liar 'after death'; however, it remains open whether *sāṃparāya* denotes a postmortal state or a location.

Lastly, BaudhDhS II 11.31 quotes the passage TB III 12.9.7 mentioned above, with *saṃparāya* instead of *sāṃparāya* (which is given in the edition).

3.2.3 Context 8: Ādiparvan (*sāṃparāya*)

In the *Ādiparvan* of the Mahābhārata, the story of the Jaratkāru pair is told (both partners are called Jaratkāru). In the story, the female Jaratkāru complains to her brother, King Vāsuki, about her namesake husband Jaratkāru, who allegedly got her pregnant. Before his

claim can be proven, however, he left her. His wife considers whether his statement could be true (MBh I 44.11):

> *svaireṣv api na tenāhaṃ smarāmi vitathaṃ kvacit /*
> *uktapūrvaṃ kuto rājan sāṃparāye sa vakṣyati //*
>
> I don't remember him speaking an untruth even in jest in the past. Why, O king, should he do so concerning the *sāṃparāya*?

The background of these deliberations is that the male Jaratkāru needs offspring, because the fate of his deceased ancestors (as well as his own) depends on it. This is clearly a matter concerning the hereafter,[24] an issue one does not make jokes about. The critical edition does not report any manuscript variants for *sāṃparāye*.

3.2.4 Context 9: *Baudhāyana-Gṛhyasūtra* (*sāṃparāya*)

The *Baudhāyana-Gṛhyasūtra* contains a hymn to the Lord (*īśāna*), one stanza of which (III 7.20) reads as follows:

> *ekaḥ purastād ya idaṃ babhūva, yato babhūvur bhuvanasya gopāḥ /*
> *yam apyeti bhuvanaṃ sāṃparāye, sa no havir ghṛtam ihāyuṣe 'ttu devaḥ svāhā //*
>
> He who here came into existence alone in the beginning, from whom the world's protectors came into existence,
> he who is that world to which one goes in the *sāṃparāya* – may he eat our oblation here, ghee, for long life, (this) god! *svāhā*!

This stanza thus implies that after death, one goes to a certain 'world' (*bhuvana*) in 'the postmortal' or 'hereafter'; in this case, this world is the Lord himself.

24 Van Buitenen (1973, 108) translated *sāṃparāya* in this stanza as "a matter of life and death".

3.2.5 Context 10: *Mokṣadharmaparvan* (*sāṃparāya*/*saṃparāya*)

In a philosophical passage in the *Mokṣadharmaparvan* (MBh XII 212.1), we read the following:

> *janako janadevas tu jñāpitaḥ paramarṣiṇā /*
> *punar evānupapraccha sāṃparāye bhavābhava*u //
>
> Janaka, however, a god among people, being instructed by the supreme Seer,
> asked once again about existence and non-existence in the *sāṃparāya*.

The most natural way is to interpret *sāṃparāya* as a denotation of the hereafter. The passage suggests that it was not necessarily taken for granted that a person 'exists' (in whatever form) in the *sāṃparāya*. The word does not denote a specific state or world, but functions as a 'placeholder' for whatever comes after death (this passage also shows that translating *sāṃparāya* as 'postmortal existence' can be problematic). Only one manuscript reads *saṃparāya* ($Ś_1$), which makes it possible to read *saṃparāyabhavābhavau* as a compound: 'existence and non-existence after death'.

3.2.6 Context 11: *Kaṭha-Upaniṣad* (*sāṃparāya*)

In the KU, *sāṃparāya* is used twice, once by the young Brahmin Naciketas and once by Yama, the god of death, who is being questioned by the former about the nature of human beings after death.[25] In KU 1.20, Naciketas first formulates his question thus:

> *yeyaṃ prete vicikitsā manuṣye, 'stīty eke nāyam astīti caike /*
> *etad vidyām anuśiṣṭas tvayāhaṃ, varāṇām eṣa varas tṛtīyaḥ //*
>
> Concerning this doubt people have about a deceased one – some say he exists and some say he doesn't exist –
> I would like to know about that, instructed by you. This is the third wish of the (three) wishes.

After Yama's unsuccessful attempt to dissuade Naciketas from his wish, the latter asks again (KU 1.29):

25 See generally Haas 2024.

yasminn idaṃ vicikitsanti mṛtyo, <yat> sāṃparāye mahati brūhi nas tat /
yo 'yaṃ varo gūḍham anupraviṣṭo, nānyaṃ tasmān naciketā vṛṇīte //

That about which they have doubts here – <which is in>[26] the big *sāṃparāya* – tell us about that, O Death![27]
This wish that penetrates into the mystery – Naciketas chooses none other than that.

As already noted by Alsdorf (1950, 63), *yat* in *pāda* b (in <>) is most likely an insertion made by a later editor of the passage, as it violates the meter. If removed, *yasminn-* and *sāṃparāye* belong to the same clause: "The big *sāṃparāya* about which they have doubts here, tell me that, O Death!" In this case, the *sāṃparāya* is also the object of *brūhi*. Thus, Naciketas either wants to hear about the *sāṃparāya* itself, or – with *yat* – about that which is in the *sāṃparāya*. In either case, it is plausible that *sāṃparāya* denotes that which comes after death, which Naciketas probably calls 'big' because it is a highly controversial subject among mortals and because Yama himself treats it as a big secret.[28]

In his reply (KU 2.6-9), Yama states the following:

na sāṃparāyaḥ pratibhāti bālaṃ, pramādyantaṃ vittamohena mūḍham /
ayaṃ loko nāsti para iti mānī, punaḥ punar vaśam āpadyate me //
śravaṇāyāpi bahubhir yo na labhyaḥ, śṛṇvanto 'pi bahavo yaṃ na vidyuḥ /
āścaryo vaktā kuśalo 'sya labdhā, āścaryo jñātā kuśalānuśiṣṭaḥ[29] *//*
na nareṇāvareṇa prokta eṣa, suvijñeyo bahudhā cintyamānaḥ /
ananyaprokte gatir atra nāsty, aṇīyān hy atarkyam aṇupramāṇāt //
naiṣā tarkeṇa matir āpaneyā, proktānyenaiva sujñānāya preṣṭha /
yāṃ tvam āpaḥ satyadhṛtir batāsi, tvādṛṅ no bhūyān naciketaḥ praṣṭā //

The *sāṃparāya* is not apparent to the fool who is careless and deluded by the delusion of possessions.[30]

26 <> mark an insertion.

27 Cf. Olivelle's (1998, 381) translation: "The point on which they have great doubts – what happens at that great transit – tell me that, O Death!".

28 Cf. fn. 31.

29 Several scholars have proposed the emendation *kuśalo 'nuśiṣṭaḥ* (see Olivelle 1998, 605); the translation of this would be "blessed is he who has been taught it".

30 Cf. Olivelle's (1998, 383) translation: "This transit lies hidden from a careless fool, who is deluded by the delusion of wealth".

> Thinking "this is the world, there is no other", he falls into my power again and again.
> That which many never get to hear about, that which many, even if they hear about it, don't understand –
> rare is he who teaches it, blessed is he who obtains it, rare is he who knows it, having been taught by a blessed one.
> If it is taught by an inferior man, it is not easy to grasp, even though one may think a great deal.
> If it is not taught by someone else, one cannot gain access to it, because it is inconceivably finer than the size of the finest particle.
> This insight cannot be gained by reasoning, only when taught by someone else is it easy to grasp, my dear,
> that which you have gained. You are determined to (learn) the truth! May I have a questioner like you, Naciketas!

At first, one might think the 'next' (*para*) world mentioned in 6c might be synonymous with *sāṃparāya* in 6a: it is the world beyond that one needs to learn about in order to find salvation. The relative pronouns in 7ab (*yo-*, *yaṃ-*) could refer to either. In 9a, however, the object of knowledge is referred to with 'this insight' (*eṣā … matir-*). This suggests that *mati* and *sāṃparāya* might be coreferential: both of them refer to what comes after death as an important object of knowledge. Thus, *sāṃparāya* (what comes after death) not only denotes the next world, but also encompasses the immortal self that remains after death and is the focus of the rest of the text.

4 Reinterpretations and Mistakes

4.1 A Means to Attain the Hereafter

4.1.1 Context 12: Śaṅkara's Commentary on the *Kaṭha-Upaniṣad* (*sāṃparāya*)

For the first passage of the KU mentioning the word, Śaṅkara (c. eighth century CE) provides the following glosses (*Bhāṣya* on KU 1.29; original text in bold):

> *he mṛtyo,* ***sāṃparāye*** *paralokaviṣaye* ***mahati*** *mahatprayojananimitta ātmano nirṇayavijñānaṃ* ***yat*** *tad* ***brūhi*** *kathaya* ***no*** *'smabhyam.*
>
> O Death, **tell**: explain **us**: to us – the definite knowledge about the

> *ātman* – **that which is in the great**:[31] **which** is the cause for the great goal ***sāṃparāya***: whose object is the next world!

This passage shows that for Śaṅkara, *sāṃparāya* does not denote the *paraloka*, the 'next world', but something 'that has the next world as its object' (*paralokaviṣaya*). What he means by that becomes clearer in his commentary on the second KU passage mentioning *sāṃparāya* (2.6):

> *saṃpareyata iti* ***saṃparāyaḥ*** *paralokas tatprāptiprayojanaḥ sādhanaviśeṣaḥ śāstrīyaḥ sāṃparāyaḥ / sa ca* ***bālam*** *avivekinaṃ prati na* ***pratibhāti*** *na prakāśate nopatiṣṭhata ity etat*.
>
> Because one passes away to it (*sam*+*parā*+*i*), it is called ***saṃparāya***, the next world. The *sāṃparāya* taught in the treatises is a peculiar means whose goal is the attainment of that (i.e. of the next world). And this (*sāṃparāya*) **is not apparent**: not evident **to the fool**: to an indiscriminating one; this means it is not at his disposal.

Here, Śaṅkara first defines *saṃparāya* as something 'one passes away to' (note the passive voice of *saṃpareyate*), that is, he understands it in a meaning well known to us from pre-medieval literature: *saṃparāya* as 'the hereafter'. He also equates the hereafter with the next world (*paraloka*), which indicates that he understood it as a domain or realm rather than a state or condition. In the next sentence, he defines *sāṃparāya* as a *sādhana*, a 'means' to accomplish (*sādh*) *tat*- 'that', which refers to the next world just mentioned before. This definition is in line with his earlier characterization of *sāṃparāya* as something whose object is the next world. While this reading of the KU probably does not reflect the intention of its original author (who probably took the word to mean 'what comes after death'; see Context 11), interpreting *sāṃparāya* as the denotation of a means to pass away – or over – to the next world is not grammatically impossible.

31 In the KU, Naciketas's characterization of the *sāṃparāya* is not intended to 'glorify' it, moreover, his tone seems rather informal. This justifies the translation 'big'. In Śaṅkara's case, it is probably the other way around, which is why 'great' is more appropriate.

4.2 Transit

4.2.1 Context 13: Śaṅkara's Commentary on the *Brahmasūtra* (*sāṃparāya*)

As mentioned in the Introduction, *sāṃparāya* is frequently translated as 'transit', a translation that is in conflict with the derivation of the word. None of the contexts discussed above provide any justification for it. Most likely, this reinterpretation of the word goes back to Śaṅkara, who in his commentary on *Brahmasūtra* III 3.27 strongly deviates from his explanation in the *Kaṭha-Upaniṣad-Bhāṣya*. The *Sūtra* deals with the doctrine that a liberated one 'loses' good and evil deeds after death, and specifies that this loss takes place in the *sāṃparāya*, "because according to others, in the *sāṃparāya*, there is nothing to be crossed" (*sāmparāye tartavyābhāvāt tathā hy anye*).[32] How exactly the *Sūtra* is to be understood is a question in itself and shall not be discussed further here. According to *Śaṅkara*, in any case, it refers to the opinion that a liberated person discards their good and evil deeds *on the way* to the hereafter – in his eyes an erroneous opinion that is backed up by a passage from the *Kauṣītaki-Upaniṣad*. For him, the *Sūtra* expresses that other authoritative texts clarify that this already happens at the time of death:

> *sāmparāye gamana eva dehād apasarpaṇa idaṃ vidyāsāmarthyāt sukṛtaduṣkṛtahānaṃ bhavatīti pratijānīte.*
>
> [The author of the *Brahmasūtra-Sūtra*] affirms that this loss of good deeds and bad deeds occurs because of the power of knowledge in the *sāṃparāya*, that is, exactly at the departure, when moving out from the body.[33]

Perhaps it was this passage that inspired the authors of the dictionaries and modern scholars to translate *sāṃparāya* as 'transit'. In view of Śaṅkara's other explanation, however, one should not hastily accuse him of having understood the *word* in this meaning; probably he only used it as a 'hook' to present his interpretation of the entire *text* of the *Sūtra*.

32 Cf. Gambhirananda's (2009, 695) translation: "(A man of knowledge gets rid of virtue and vice) at the time of death, since nothing remains to be attained. For thus it is that others (i.e. the followers of the other branches) state". See generally Sharma 1978, 342-8.

33 Cf. Gambhirananda's (2009, 696) translation: "[T]he aphorist asserts that at the very time of death, at the time of moving away from the body, occurs this discarding of virtue and vice as a result of the power of knowledge".

4.3 Otherworldly

4.3.1 Context 14: Hymn in the *Ādiparvan* (*sāṃparāya*)

A curious case is given in the *Ādiparvan* of the *Mahābhārata*, where Upamanyu summons the Aśvins by reciting a hymn in order to regain his eyesight. This hymn is said to consist of *ṛcs*; however, it is not to be found in the *Ṛgveda*. Its second stanza (MBh I 3.61) reads thus:

> *hiraṇmayau śakunī sāṃparāyau, nāsatyadasrau sunasau vaijayantau /*
> *śukraṃ vayantau tarasā suvemāv, abhivyayantāv asitaṃ vivasvat //*
>
> Two golden birds belonging to the *sāṃparāya* (?), Nāsatya and Dasra, fine-beaked, belonging to the Victorious One (Indra?),
> swiftly weaving in the bright (sun?) on fine looms, weaving out the dark Vivasvat (= sun?).[34]

This stanza is very difficult to interpret. As the editors note in the edition, "[t]he text of the Aśvin hymn (st. 60-70) is still highly uncertain, in part even quite unintelligible, the MSS being at this point all very corrupt" (n. on stanza 60). Moreover, it is difficult to avoid the impression that the author deliberately wanted to make the hymn enigmatic. Possibly, *sāṃparāya* is here used as an adjective in the dual and means 'belonging to' or 'located in the postmortal'; the commentator Nīlakaṇṭha explains it to mean *sarvasya layādhiṣṭhānabhūtau*, "being in the place of the dissolution of everything"[35] (curiously, this explanation is also referred to in the PW to support the translation "ein Retter in der Noth"). Given the uncertainty of the text, it is impossible to draw firm conclusions.

4.3.2 Context 15: Raṅgarāmānuja's Commentary on the Kaṭha-Upaniṣad (*sāṃparāya*)

Raṅgarāmānuja (c. sixteenth century CE) only briefly comments on the KU passages mentioning *sāṃparāya*. He paraphrases the first passage as following (*Prakāśikā* on KU 1.29; "That about which they have doubts here"):

---

34 Cf. van Buitenen's (1973, 47) translation:
"Birds golden, fine-beaked psychopomps,
Munificent Nāsatyas, surely triumphant,
Who on fine looms swiftly weave the light in,
And swiftly weave out that darker sun".

35 *Bhāratabhāvadīpā*, n MBh I 3.61".

> *mahati pāralaukike yasmin muktātmasvarūpe saṃśerate tad eva me brūhi*.
>
> The great otherworldly proper nature of the liberated *ātman* about which they are undecided, tell me about that!

Here, Raṅgarāmānuja seems to interpret *sāṃparāya* as an adjective with the same meaning as *pāralaukika* ('belonging to the next world' or 'otherworldly'), and relates it to the nature of the Self. This interpretation could hardly be applied to the second KU passage (2.6; "The *sāṃparāya* is not apparent to the fool"), and indeed, Raṅgarāmānuja paraphrases this passage (echoing Śaṅkara's words) simply as *paraloko 'vivekinaṃ prati na prakāśate*, "the next world is not evident to an indiscriminating one".

4.4 Beginninglessness

4.4.1 Context 16: Śāṇḍilya's Bhaktisūtras (*saṃparāya*)

According to the *Bhaktisūtra*s attributed to Śāṇḍilya (eighth century CE),[36] the universe is divided into two entities: *cit*, the 'perceiver' or 'perception', and *cetyā*, the 'perceived' (*Sūtra* 40). *Sūtra* 41 states that

> *yuktau ca samparāyāt*.
>
> and they are joined because of *saṃparāya*.

The commentator Svapneśvara explains that *samparāyāt* means *anāditvāt*, 'because of beginninglessness'. While critical editions of these texts are still lacking, it seems to me that a conjecture is called for, namely replacing *samparāya* with *samavāya*, which as a philosophical term denotes "perpetual co-inherence, inner or intimate relation, constant and intimate union, inseparable concomitance" (MV, see s.v. *samavāya*).

36 Johnson 2019, see s.v. "Bhakti Sūtra".

5 Conclusion

Based on the survey conducted in this article, the following conclusions can be drawn:

1. *Saṃparāya* and *sāṃparāya* are frequently confused or used interchangeably. This can be explained in two ways: first, the two words only differ in the quantity of a single vowel. In Prakrit, long vowels in closed syllables (such as *sāmp*) are usually shortened (Pischel 1900, 72-3). While this is not true for Sanskrit, it is plausible that Sanskrit authors, too, may have (occasionally) been influenced by Prakrit pronunciation. Moreover, from a metrical perspective, the syllable containing this vowel is heavy regardless of the length of the vowel. This means that using one 'variant' or the other does not affect the meter. Second, in many South Asian scripts, the difference between *saṃparāya* and *sāṃparāya* only lies in the presence or absence of a single line; in Devanāgarī, for example, *saṃparāya* is संपराय and *sāṃparāya* is सांपराय. Regardless of what the original reading was, a copyist may easily confuse one word with the other: either by simply overlooking the *ā* in the process of copying, by adapting the word *sāṃparāya* to a pronunciation the scribe was more accustomed to, or by 'correcting' *saṃparāya* to *sāṃparāya*.
2. *Sāṃparāya* is rarely used as an adjective. In the passages discussed in this article, only the enigmatic hymn to the Aśvins in the *Ādiparvan* (Context 14) seems to use it as such (though even in this case it is possible to read it as a nominalized adjective). Notwithstanding occasional reinterpretations (or 're-adjectivizations'), *sāṃparāya* is generally used as a noun.
3. Both *saṃparāya* and *sāṃparāya* are most often used in the locative (°*parāye*). Together with the fact that they are frequently used interchangeably, this is evidence that they are generally – if not always – considered as variants of one and the same noun. In the few cases where this noun is used in the nominative, it has the masculine gender.
4. The reason that *saṃparāya*/*sāṃparāya* is almost always used in the locative is probably because it was perceived as a fixed expression with (two) fixed meanings. *Saṃparāya*/*sāṃparāya* could theoretically denote many death-related things, but in the locative always means either 'in mortal combat/battle' or 'in the postmortal/hereafter'. (One may compare it to the word 'charge' in the phrase 'in charge': while the noun 'charge' has a variety of meanings, 'being in charge' only means 'being responsible').
5. The second sense is comparatively unspecific: as shown in Section 3.2, in most contexts dealing with postmortal (rather

than mortal or lethal) matters, it remains open whether the word denotes a concrete thing and if so, what: on the one hand, *sāṃparāya* is nowhere explicitly characterized as the 'state' of an individual (i.e. the state of having passed away, or being dead). On the other hand, it is also not necessarily a 'place' (i.e. the 'hereafter' or 'next world') – even though the verb *sam+parā+i* would suggest a destination in a spatial sense (even if only metaphorically). Rather, it denotes 'what comes after passing away' – 'the postmortal' or 'hereafter' in the broadest conceivable sense.

6. In only one of the passages discussed above, *saṃparāya* has the primary meaning of 'passing away' (TB III 12.9.7). In all other pre-medieval texts, it is possible to interpret *saṃparāya* as well as *sāṃparāya* as meaning either 'mortal combat, battle' or 'the postmortal'. Both meanings are derivative, which explains why they established themselves for the nominalized adjective *sāṃparāya*. However, even an author who was fully aware of the existence of the word *sāṃparāya* might have had a reason to use *saṃparāya* in the same meaning: first, the meaning 'the postmortal' for *saṃparāya* was well established in Buddhist Sanskrit as well as in Pāli (which does not even use the form *sāṃparāya*), and possibly even in other Sanskrit traditions (as was probably the case in the VasDhS; see Context 7). As for the meaning 'decease, death', an author might have understood it as a metonymical expression for 'mortal combat, battle'.
7. The interpretations by commentators such as Śaṅkara do not reflect the general usage of the word *saṃparāya*/*sāṃparāya*, but are guided by their exegetical agenda. While they might have been aware of the general meaning of this word (cf. Context 12), they took advantage of the fact that a form such as *sāṃparāya* can easily be reinterpreted.

Primary Sources

Aitareya-Āraṇyaka = Keith, A.B. (ed. and transl.) (1909). *The Aitareya Āraṇyaka*. Oxford: Clarendon Press.

ĀpDhS = *Āpastamba-Dharmasūtra*. Olivelle 2000, 20-115.

Baudhāyana-Gṛhyasūtra = Sharma Sastri, R. (ed.) (1920). *Bodhâyana Grihyasutra*. Mysore: Government Branch Press.

BaudhDhS = *Baudhāyana-Dharmasūtra*. Cf. Olivelle 2000, 191-345.

Bhāgavata-Purāṇa = Lal Shastri, J. (ed.) (1999). *Bhāgavata Purāṇa of Kṛṣṇa Dvaipāyana Vyāsa. With Sanskrit Commentary Bhāvārthabodhinī of Śrīdhara Svāmin*. Dehli: Motilal Banarsidass.

Bhaktisūtras of Śāṇḍilya = Harshananda, S. (ed. and transl.) (1984). *Śāṇḍilya Bhakti Sūtras. With Svapneśvara Bhāṣya*. 2nd ed. Mysore: University of Mysore.

Bhāratabhāvadīpā = Mahabharatam with the Commentary of Nīlakantha. 1 Adiparva [*Bhāratabhāvadīpā Ṭīkā*] (1929). Poona: Chitrashah Press.

Brahmasūtra-Bhāṣya = Śaṅkara (1916). *The Works of Sri Sankaracharya*, vol. 1. Srirangam: Sri Vani Vilas Press.

Daś = *Daśakumāracaritra*. Onians 2005.

Dīgha-Nikāya = Rhys Davids, T.W.; Carpenter, J.E. (1903). *The Dīgha Nikāya*, vol. 2. London: Oxford University Press.

Hiraṇyakeśi-Gṛhyasūtra = Kirste, J. (ed.) (1889). *The Gṛihyasūtra of Hiraṇyakeśin with Extracts from the Commentary of Mātṛidatta*. Vienna: Alfred Hölder.

Kaṭha-Upaniṣad-Bhāṣya of Śaṅkara. Included in the ed. of the KU.

KU = *Kaṭha-Upaniṣad*. Āpṭe, V.G. (ed.) (1935). *Saṭīkādvayaśāṅkarabhāṣyopetā kāṭhakopaniṣad*. [*ānandāśramasaṃskṛtagranthāvaliḥ* 7]. Pune: Anandashrama Press.

MBh = *Mahābhārata*. Sukthankar, V.S. (1933-59). *The Mahābhārata*. 19 vols. Pune: Bhandarkar Oriental Institute.

Mvu = Marciniak, K. (2019). *The Mahāvastu. A New Edition*, vol. 3. Tokyo: The International Research Institute for Advanced Buddhology, Soka University.

Prakāśikā = Rāṅgarāmānuja's commentary on the KU. Included in the ed. of the KU.

Rāmāyaṇa = Vīra, R. (ed.) (1938). *Rāmāyaṇa of Vālmīki, India's national Epic. For the First Time Critically Edited from Original Manuscripts Derived from All Parts of India*. Lahore: International Academy of Indian Culture.

Ṛgveda = Aufrecht, Th. (ed.) (1877). *Die Hymnen des Ṛigveda*. 2 vols. 2nd ed. Bonn: Adolph Marcus.

Śāṅkhāyana-Āraṇyaka = Dev, Bh. (ed.) (1980). *Śāṅkhāyanāraṇyakam*. Hoshiarpur: Vishveshvaranand Vedic Research Institute.

TB = *Taittirīya-Brāhmaṇa*. Dumont 1951.

VasDhS = *Vasiṣṭha-Dharmasūtra*. Cf. Olivelle 2000, 346-463.

Yama-Smṛti = Āpṭe, V.G. (ed.) (1929). *Smṛtīnāṃ samuccayaḥ*. Pune: Ānandāśramasamudraṇālaya, 99-107. *Ānandāśramasaṃskṛtagranthāvaliḥ* 48.

Secondary Literature

Alsdorf, L. (1950). "Contributions to the Textual Criticism of the Kaṭhopaniṣad". *Zeitschrift der Deutschen Morgenländischen Gesellschaft*, 100, 621-37.

Apte, V.S. (1890). *The Practical Sanskrit-English Dictionary. Containing Appendices on Sanskrit Prosody and Important Literary & Geographical Names in the Ancient History of India, for the Use of Schools and Colleges*. Poona: Shiralkar.

Dumont, P.E. (1951). "The Special Kinds of Agnicayana (or Special Methods of Building the Fire-Altar) According to the Kaṭhas in the Taittirīya-Brāhmaṇa. The Tenth, Eleventh, and Twelfth Prapāṭhakas of the Third Kāṇḍa of the Taittirīya-Brāhmaṇa with Translation". *Proceedings of the American Philosophical Society*, 95(6), 628-75.

Gambhirananda, S. (transl.) [1965] (2009). *Brahma-Sūtra-Bhāṣya of Śri Śaṅkarācārya*. Kolkata: Advaita Ashrama.

Haas, D. (2018). *Vom Feueraltar zum Yoga. Kohärenz und Konzept der Kaṭha-Upaniṣad* [Master's thesis]. Vienna: University of Vienna. https://doi.org/10.25365/thesis.52786

Haas, D. (2019). "Ritual, Self and Yoga: On the Ways and Goals of Salvation in the Kaṭha Upaniṣad". *Journal of Indian Philosophy*, 47(5), 1019-52. https://doi.org/10.1007/s10781-019-09408-y

Haas, D. (2024). *Vom Feueraltar zum Yoga. Kommentierte Übersetzung und Kohärenzanalyse der Kaṭha-Upaniṣad*. Heidelberg: Heidelberg Asian Studies Publishing. https://doi.org/10.11588/hasp.1329

Hopkins, E.W. (1903). "Limitation of Time by Means of Cases in Epic Sanskrit". *The American Journal of Philology,* 24(1), 1-24.

Johnson, W.J. (2019). *A Dictionary of Hinduism*. Oxford: Oxford University Press.

Jones, J.J. (transl.) (1956). *The Mahāvastu*, vol. 3. London: Luzac & Company.

Macdonell, A.A. (1927). *A Sanskrit Grammar for Students*. London: Oxford University Press.

MW = Monier-Williams, M. [1899] (1956). *A Sanskrit-English Dictionary. Etymologically and Philologically Arranged with Special Reference to Cognate Indo-European Languages*. Oxford: Clarendon Press.

Norelius, J.P. (2023). *Soul and Self in Vedic India.* Leiden: Brill.

Oldenberg, H.; Müller, F.M. (transl.) (1892). *The Grihya-Sûtras. Rules of Vedic Domestic Ceremonies. Sacred Books of the East* XXX: [Oldenberg:] *Part II: Gobhila, Hiranyakesin, Âpastamba*. [Müller:] *Âpastamba, Yagña-Paribhâshâ-Sûtras*. Oxford: Clarendon Press.

Olivelle, P. (transl.) (1998). *The Early Upaniṣads. Annotated Text and Translation*. New York: Oxford University Press.

Olivelle, P. (ed. and transl.) (2000). *Dharmasūtras. The Law Codes of Āpastamba, Gautama, Baudhāyana, and Vasiṣṭha. Annotated Text and Translation*. Delhi: Motilal Banarsidass, 20-115.

Onians, I. (ed. and transl.) (2005). *What Ten Young Men Did. By Daṇḍin*. New York: New York University Press; JJC Foundation.

Pischel, R. (1900). *Grammatik der Prakrit-Sprachen*. Strassburg: Verlag von Karl J. Trübner.

PTSD = Rhys Davids, T.W.; Stede, W. (eds) (1925). *The Pali Text Society's Pali-English Dictionary*. Surrey: The Pali Text Society.

PW = Böhtlingk, O.; Roth R. [1855-75] (1990). *Sanskrit-Wörterbuch*, Theil I–VII. St. Petersburg: Kaiserliche Akademie der Wissenschaften.

pw = Böhtlingk, O. (ed.) [1883-86] (2009). *Sanskrit-Wörterbuch in kürzerer Fassung*. St. Petersburg: Kaiserliche Akademie der Wissenschaften.

Roland Steiner, R. (2020). "Woher hat er das? Zum Charakter des Sanskrit-English Dictionary von Monier-Williams". *Zeitschrift der Deutschen Morgenländischen Gesellschaft*, 170(1), 107-17.

Sharma, B.N.K. (1978). *The Brahmasūtras and Their Principal Commentaries (A Critical Exposition)*, vol. 3. Bombay: Bharatiya Vida Bhavan.

UVC = Franceschini, M. (2007). *An Updated Vedic Concordance. [Maurice Bloomfield's A Vedic Concordance enhanced with new material taken from seven Vedic texts]*. Part I, *a - na – Part II: pa - ha*. Cambridge, Mass.: The Department of Sanskrit and Indian Studies, Harvard University / Milan: Mimesis Edizioni.

Van Buitenen, J.A.B. (transl.) (1973-78). *The Mahābhārata.* 3 vols. Chicago: The University of Chicago Press.

Weber, A. (1855). "Einiges über das Kâthakam". Weber A. et al. (eds), *Indische Studien. Beiträge für die Kunde des indischen AlterthuMS Dritter Band*. Berlin: Ferd. Dümmler's Verlagsbuchhandlung, 451-79.

Wynne, A. (2011). "The *ātman* and Its Negation. A Conceptual and Chronological Analysis of Early Buddhist Thought". *Journal of the International Association of Buddhist Studies*, 33(1-2), 103-71.

e-ISSN 2785-5953

Bhasha
Vol. 3 — Num. 2 — Ottobre 2024

A Brief Introduction to the Turi Language of Eastern India

John Peterson
Christian-Albrechts-Universität zu Kiel, Deutschland

Abhay Sagar Minz
Dr. Shyama Prasad Mukherjee University, Ranchi, India

Prabhat Linda
Dr. Shyama Prasad Mukherjee University, Ranchi, India

Ariba Khan
Christian-Albrechts-Universität zu Kiel, Deutschland

Francis Xavier Kachhap
Dr. Shyama Prasad Mukherjee University, Ranchi, India

Manish Gari
Dr. Shyama Prasad Mukherjee University, Ranchi, India

Abstract This article presents a brief introduction to the North Munda (Austro-Asiatic) language Turi, spoken by some 1,500 speakers throughout the Indian states of Chhattisgarh, Jharkhand, Odisha, Bihar, West Bengal and Assam. After a brief introduction to the ethnic Turi group, we present a skeleton grammar of the Turi language as spoken in northwestern Odisha state, where it is still being learned by children as their home language. We then discuss the position of Turi within the Kherwarian (North Munda) group by comparing our lexical data for Turi with that for twelve other Kherwarian varieties as given in Kobayashi et al. (2003), using the software COG from the Summer Institute of Linguistics. Our results suggest that Turi is a sister language to all of the dialects of Santali and that it together with these forms the Santali-Turi branch of Kherwarian. We end with a discussion of the possible consequences of these results for the linguistic and ethnic prehistory of Eastern Central India.

Keywords Turi. Kherwarian. North Munda. Austro-Asiatic. Historical linguistics.

Edizioni Ca'Foscari

Peer review

Submitted 2023-12-13
Accepted 2024-09-15
Published 2024-12-18

Citation Peterson, J.; Minz, A.S.; Prabhat, L.; Khan, A.; Kachhap, F.X.; Gari, M. (2024). "A Brief Introduction to the Turi Language of Eastern India". *Bhasha*, 3(2), 263-304.

DOI 10.30687/bhasha/2785-5953/2024/02/005

1 Introduction – the Turi and their Language

Turi (Glottocode: 1246; ISO code: ISO 639-3:trd), spoken in six states in eastern central and northeastern India, is a member of the North Munda branch of Austro-Asiatic. Ethnic Turi are found in larger numbers in the eastern central Indian states of Jharkhand, West Bengal, Odisha, Bihar and Chhattisgarh, as well as the northeastern state of Assam. These states and the surrounding regions are shown in Map 1. At present we are not aware of any ethnic Turi groups in Bangladesh **[map 1]**.

Map 1 The states in which larger numbers of Turi live and the larger South Asian region[1]

We would like to express our gratitude to the five Odisha Turi speakers who came to Ranchi to work with us on their native language over the course of five days: Mr. Prashant Duan, Mr. Laxman Majhi, Mr. Adhikari Bhue, Ms. Bishaka Mallik and Ms. Kishori Mallik. Many thanks also to Ms. Khatkuri Suren for her help with some last-minute questions and to Lee Pratchett for suggestions on improving some formulations. We also thank two anonymous reviewers for their comments, which forced us to reconsider a few points. Needless to say, any and all remaining errors and inconsistencies are entirely our own.

We would also like to thank the German Research Council (DFG) for funding the project *Towards a Linguistic Prehistory of Eastern-Central South Asia (and Beyond)* (project no. 326697274) and the Cluster of Excellence ROOTS – Social, Environmental, and Cultural Connectivity in Past Societies, which made the work of the first and fourth authors of this paper possible.

1 Many thanks to Simon Argus for producing this map for us.

Unlike most other ethnic groups which traditionally speak a Munda language, the Turi are officially a scheduled caste, not a scheduled tribe. This is likely due to the fact that the Turi are artisans who traditionally weave various products out of bamboo, such as baskets, winnowing fans, fishing equipment, umbrellas, and other items. As the caste system is closely tied to traditional occupations, and as the vast majority of Turi no longer speak the Turi language, the one obvious trait that most Turi share is their occupation.[2] Despite this official designation as a scheduled caste and not a scheduled tribe, the Turi consider themselves – and are considered by their neighbours – to be a tribal group. They have also been greatly influenced by neighbouring ethnic groups, both Hindus and practitioners of Sarna, the collective term for the religions of all tribal groups in eastern central India, and celebrate the same annual festivals as neighbouring tribal groups, such as Nava Khani, Sarhul, Karm and Sohrai.

Turi is a severely endangered language, with less than 1% of all ethnic Turis still able to speak it, and is classified as "moribund" by the *Ethnologue* (Eberhard et al. 2024), meaning that "[t]he only remaining active users of the language are members of the grandparent generation and older".[3] Fortunately, we can state that there are still young children who speak Turi on a daily basis, albeit in very few and mostly remote small villages, and most Turi nowadays have no active or passive knowledge of their traditional language.

Turi has only been rudimentarily documented until now, both with respect to its lexicon and grammar. In this study we therefore present a basic skeleton grammar of the Odisha dialect of this language as well as a list of over 200 basic vocabulary items and two short texts. Furthermore, as the position of Turi within the North Munda group of languages is still unclear, we also present the results of a comparison of native Turi lexemes with cognate forms in 12 other North Munda varieties as these are documented in Kobayashi et al. (2003), analysing these with the program COG from the Summer School of Linguistics.[4]

Our team conducted fieldwork during several trips to Turi villages in the blocks of Raidih, Bishunpar, Lohardaga and Chainpur in Southern Jharkhand, and Bundu Block in Ranchi District, as well as a few Turi families in the city of Ranchi itself. Shorter visits were also

2 The status of groups as "Scheduled Castes" or "Scheduled Tribes" is regulated by the Constitution of India in Articles 366 (24)-(25), and Articles 341 and 342. These are legal terms and do not result automatically from a group being socially recognized as a tribal group: Both terms can be applied to tribal groups, so that a tribal group may not necessarily be classified as a "Scheduled Tribe" but can also be a "Scheduled Caste", as is the case with the Turi.

3 https://www.ethnologue.com/methodology/#Status2.

4 https://software.sil.org/cog/.

made to Jashpur and Surguja Districts in Chhattisgarh. During this work, we also came to know of Turi-speaking groups in northwestern West Bengal and the tea garden estates of Assam, with whom we have conducted telephone interviews.

Turi from all age groups were included in our fieldwork, as were both men and women, and an effort was also made to ascertain through informal conversations the attitude of the Turi community itself towards its traditional language. Finally, a group of five adult Turi speakers from Odisha – three men and two women – were invited to a five-day workshop at the International Documentation Centre for Endangered Indigenous Languages and Cultures at the Dr. Shyama Prasad Mukherjee University, Ranchi, to document the Turi lexicon as well as possible and to record and analyse Turi songs and short stories.[5] As this work has now come to an end, at least for the foreseeable future, we have decided to publish what we have learned through our work so far in the present form, so that the data are available to scholars of Munda and all those interested in the Turi and more generally in the cultural and linguistic landscape of eastern central India.

According to the *Ethnologue* (Eberhard et al. 2024), Turi was spoken in 2007 by a total of ca. 2,000 people in all of India, out of a total ethnic population of 354,000.[6] Thus, according to these figures, Turi was spoken by about 0.56% of the total ethnic Turi population in 2007. Due to the vast area in which the Turi now live, spread throughout at least six states in northeastern and eastern central India, it is exceedingly difficult to determine the number of Turi who still have an active or passive command of their traditional language. Our own estimate is that the language is only spoken by a maximum of 1,000-1,500 people in all of India, although even this figure may be somewhat inflated, based on our own experience and on second-hand accounts, and we have direct knowledge of no more than ca. 120 speakers in Jharkhand, Chhattisgarh and Odisha. But whatever the exact number of speakers found throughout this immense area may be, Turi is clearly a severely endangered language.

A large number of languages from the Indo-Aryan (Indo-European), Munda (Austro-Asiatic) and Dravidian families are spoken in

5 This workshop also resulted in the publication of Peterson, Minz 2021, a primer for Turi-speaking children to learn to write their native language.

Due to the pandemic situation, which was still acute at that time, it was deemed better to invite a small number of speakers to Ranchi to work with our team than to travel to the communities themselves and live and work with the speakers there. Although this did impede our progress considerably, it also produced an atmosphere which was conducive to this work, as all Turi speakers and all of the authors of this study were vaccinated, which was not possible in any of the Turi-speaking villages.

6 https://www.ethnologue.com/language/trd/.

these six states, so that the Turi have traditionally had to use a lingua franca in their daily lives, one which varies from one region to another. Thus the Turi are generally multilingual: Even those who do not speak their traditional language but a regional language such as Sadri/Nagpuri in Jharkhand and Chhattisgarh will still have some degree of fluency in Hindi, which is the official language of Bihar, Chhattisgarh and Jharkhand, or Odiya in Odisha, Bengali in West Bengal and Assamese in Assam. Those who still speak Turi in their daily lives are therefore generally at least trilingual. Finally, many Turi are also fluent in one or more neighbouring tribal languages, e.g., Kurux (Dravidian) or Mundari (Munda), depending on the region they live in and whether other ethnic groups also live in their village.

This high degree of multilingualism is certainly one of the main causes for abandoning the traditional Turi language in favour of a local regional or official language. Another possible cause is mentioned by Turi elders, who compare the state of their own traditional language with that of the Asur. The Asur, who speak a North Munda language of the same name, are traditionally iron smelters who live more isolated than the Turi in hill areas, closer to the sources of iron ore. Turi elders claim that the Turi also once lived higher up in the hills but chose to move to lowland forested areas, still near these mountains and hills, where they had better access to bamboo for their trade, but also better access to markets to sell their wares. These areas were also more conducive to agriculture, allowing more Turi to practice that as well.

According to these Turi elders, it is the continued relative isolation of the Asur that is responsible for the better rate of retention of the Asur language by the ethnic Asur than with the Turi language among the Turi: Living in the lowlands, the Turi are in constant contact with other ethnic groups, which is increasingly resulting in intermarriage between the Turi and these groups, as a result of which Turi is usually not passed on to the next generation, which then speaks a regional language at home. In contrast, the more isolated Asur have been better able to maintain their language due to less contact with other groups.[7]

The remainder of this article is structured as follows. Section 2 presents a skeleton grammar of Turi while Section 3 discusses the position of Turi within North Munda, based on our comparative study. The results of this study and their possible significance are then discussed in Section 4. Section 5 provides a summary of this study,

7 The *Ethnologue* lists the status of Asur (referred to there as "Asuri") as "vigorous", meaning that it is used by all generations and that the situation is sustainable (Eberhard et al. 2024), compared with Turi's status there as moribund, mentioned in the preceding pages.

which concludes with two appendices: Appendix 1 presents a list of basic vocabulary items in Odisha Turi, ordered according to the list given in Kobayashi et al. (2003), while Appendix 2 presents two short Turi texts.

2 A Brief Overview of Turi Grammar

This section presents a skeleton grammar of Turi. Only the most basic categories can be touched upon here, such as basic forms of pronouns, number and case, the most common verb categories, lower numerals, etc. Turi, like other Munda languages, is a predominantly verb-final language. It generally differs little in its basic grammatical structures from what we find in other North Munda languages, such as Santali, Mundari or Ho, although it does show some striking differences as well. In the following, all forms are from the Odisha dialect of Turi, unless otherwise explicitly noted.

2.1 Phonology

Table 1 presents the consonant phonemes in Turi which we have been able to identify. All four affricates, all consonants with breathy voice, /ɳ/ and /ʋ/ are restricted in our data to loan words from Indo-Aryan. For ease of presentation, in all Turi example sentences aspiration and breathy voice are represented by <h>. In Table 1 and in the word list in Appendix 1, however, these are given in standard IPA, with superscripted aspiration **[tab. 1]**.

Table 1 The consonant phoneme inventory of Odiya Turi

	Bilabial	Dental	Alveolar	Retroflex	Palatal	Velar	Glottal
Stops	p pʰ b bʱ	t tʰ d dʱ		ʈ ɖʈʰ	c ɟ	k kʰ g gʱ	ʔ
Affricates			ʧ ʤʧʰ ʤʱ				
Nasals	m	n		ɳ	ɲ	ŋ	
Laterals			l				
Flaps			ɾ	ɽ			
Fricatives			s				h
Approximants	ʋ				j		

Munda languages are known for their pre-glottalized, voiced stops, often with a nasal release, i.e. [ˀb] or [ˀbᵐ]. This class is also found in syllable-final position in Turi, although there is considerable variation which requires further study. Cf. e.g. from the word-list in Kobayashi et al. (2003, 353, 359) words for 'hair' (#2) from Mundari such

as [u:bˉ]/[u:b] or Santali [u:bˉ]/[u:pˉ], with Turi [uʔm], or 'eye' (#5): Ho [meɖˉ], Mundari [medˉ]/[med]/[meːdʒ] and Santali [meːtˉ]/[metˉ]/[me·tˉ] (Kobayashi et al. 2003, 347, 353, 359) with Turi [mɛʔn]. As can be seen, Turi tends to lose the stop entirely and retain the nasal – and often also glottalization – although alternative forms are also found in the texts, e.g., [dub], [dubʔm] and [duʔm] 'sit down'.

Table 2 presents the monophthong phoneme inventory of Turi as found in our database. The status of /ə/ as a phoneme is unclear. At present our data suggest that it is best viewed as an allophone of /ɑ/. All vowels can be nasalized; it is not yet clear if nasalization can be phonemic. Vowel length, however, is not phonemic, and all vowels other than /ə/ can be realized as long in certain environments, such as word-final positions **[tab. 2]**.

Table 2 Monophthong vowels in Turi

	Front	**Central**	**Back**
Close	i		u
Mid		(ə)	ɔ
Mid-low Low	ɛ		ɑ

The following diphthongs are also found in our corpus in native words: *ə̯ɛ, i̯a*.

2.2 Nouns

There is no grammatical gender in Turi. However, object indexing on the verb is sensitive to the animate/inanimate opposition as nouns with animate reference trigger verbal agreement in the 3rd persons, while nouns with inanimate reference do not.[8] Nouns may appear in the singular, which is morphologically unmarked, in the dual, marked by *=kin*, or in the plural, marked by *=kun*.

(1) a. *bandra* 'monkey'
b. *bandra=kin* 'two monkeys'
c. *bandra=kun* '(three or more) monkeys'

The morphologically unmarked noun – especially if it has inanimate reference – can also have plural reference if this is clear from context or is indicated elsewhere in the text, e.g. *sɔbu ʈɔpi* 'all the hats' in (22) or *ʈɔpi* 'the hats' in (23).

8 One exception to this rule is the locative copula, discussed further below, which exceptionally marks subjects – including inanimate subjects – at the object position. Cf. examples (26)-(27). Otherwise, non-animate subjects are not indexed on the verb.

2.3 Pronouns

Independent pronouns are found in the singular, dual and plural. Dual forms can be reinforced by adding the form *baran=kin* 'both=DU', e.g., *aliŋ baraŋ=kin* 'we both (EXCL)', and plural pronominal forms of the 1st and 2nd persons can be pleonastically marked for *=ku* 'PL', e.g. *alɛ=ku* 'we (PL.EXCL)'. Table 3 presents the independent pronouns **[tab 3]**.

Table 3 Independent pronouns in Turi

	Singular	Dual	Plural
1st (EXCL)	*iŋ*	*aliŋ*	*alɛ*
1st (INCL)		*alaŋ*	*abu*
2nd person	*am*	*abin*	*apɛ*
3rd person	*uni*	*unkin*	*unku*

2.4 Numerals

In Odisha Turi, only the numerals one-three are of Munda origin; beginning with 'four', Indo-Aryan loan words are used. Grierson (1906, 128) also cites the form *pūniā* 'four', which is not in use in Odisha Turi. These are given in Table 4 **[tab. 4]**.

Table 4 Lower numerals in Turi

miaʔn (Jharkhandi Turi: *meja*; Grierson. *mit'*, *miat'*)	'one'
barɛa (Jharkhandi Turi: *baria*)	'two'
pɛa	'three'
ʧar	'four' (IA)
panʧ	'five' (IA)

miaʔn 'one' is also commonly used as an indefinite article; cf. (2).

(2) *miaʔn* *duba* *pɛndarɛ* *duʔm-ɛn=ə=ɛ.*
one tree under sit.down-MID.PST=FIN=3SG.ANIM
'He sat down under a tree.'

The morph *=gɔɽ*, which is homophonous with a classifier in many neighbouring Indo-Aryan and Munda languages of the region, is also found in our texts, together with the Indo-Aryan numerals *ʧar* 'four' and *panʧ* 'five', but not with the lower numerals of Turi origin. However, it does not appear to be a classifier in Turi, as it is used in counting in our corpus (cf. (3)), but not with a following noun; cf., e.g., *miaʔn duba* 'a tree' in (2) or *miaʔn phɛriʋala* 'a hawker' in example (40). We therefore tentatively gloss it as "COUNT", due to its counting function.

(3) *miaʔn barɛa pɛa ʧar=gɔɽ panʧ=gɔɽ lɛkəɛ=a=bu.*
one two three four=COUNT five=COUNT count=FIN=1PL.INCL
'Let's count 1, 2, 3, 4, 5.'

2.5 Case

Odiya Turi has five cases:

- Unmarked – The unmarked (or zero-marked) case marks subjects (e.g., *maʤur* 'the peacock' in (9)) and inanimate objects (e.g., *sɔbu ʈɔpi* 'all the hats' in (22)).
- Objective – This case is used with primary objects with verbs of speech, cf. (4), and animate secondary objects (cf. *uni kɔɽa=kɛ* 'that boy' in example (13)).

(4) *bandra=kun=kɛ katha-laʔ=ə=ɛ …*
monkey=PL=OBJ say-PST.PERF.ACT=FIN=3SG.ANIM
'He said to the monkeys …'

=kɛ can also mark definite inanimate secondary objects (or P, in typological terms), as in (5), although this is not obligatory, as *sɔbu ʈɔpi* 'all the hats' in (22), mentioned above, shows.

(5) … *dan rɛ bandra=kun ʈɔpi=kɛ haʈ sɛn=ɛŋ.*
give.IMP VOC monkey=PL hat=OBJ market go=1SG
'… Give the hats [to me], oh monkeys! I will go to market.'

- Genitive – The genitive is used to incorporate a nominal phrase into a larger nominal phrase. There are two genitive markers: *=aʔ/=aʔa*, used with pronouns (6), and *=rɛn*, used elsewhere; cf. (7).

(6) *iɲ=aʔ ɲumu prʌbhat=naŋ.*
1SG=GEN name Prabhat=IDENT.COP
'My name is Prabhat.'

(7) *phɛriʋala=rɛn durum bhaŋa-ɛn=a*
hawker=GEN sleep(n.) open(itr.)-PST.MID=FIN
'The hawker woke up (= the hawker's sleep opened).'

With adnominal inalienable possession involving kinship terms, a different construction is used. Here, the noun is followed by the possessive marker *-ta* which is followed by an enclitic form of the independent pronoun, as with the two forms of *ba* 'father' in (8).

(8) a. *ba-t=ɛŋ*
father-POSS=1SG
'my father'

b. *ba-ta=m*
father-POSS=2SG
'your father'

- Locative – The locative marks the location of an entity. It has the form *=rɛ* (cf. (9)).

(9) *buru=rɛ* *maʤur* *susun-tan=ə=ɛ.*
forest=LOC peacock dance-PROG=FIN=3SG.ANIM
'In the forest the peacock is dancing.'

- Instrumental – The instrumental case marks the instrument with which an action is carried out or the cause of an action. It has the form *=tɛ*, as shown in (10).

(10) *uni* *khis=tɛ* *ɔɽaʔa* *sɛn-ɔʔ-ɛn=ə=ɛ.*
3SG.ANIM anger=INST house go-MID-PST.MID=FIN=3SG.ANIM
'He went home out of (= through) anger.'

Case markers follow number markers in Turi; cf. the form *bandra=kun=kɛ* [monkey=PL=OBL] 'to the monkeys' in (4) above or *ʧhava=kun=rɛn* [child=PL=GEN] 'of the children' below in (23).

2.6 General Introduction – Lexical Verbs

The Turi verb system shows many similarities to those of other North Munda languages, although with some differences, as we show in the following. Figure 1 gives a somewhat simplified structural overview of the affirmative finite verb in Turi and Figure 2 that of the negative verb **[figs 1-2]**.

LEXICAL BASE-TAM.ACT/MID=OBJ=FIN=SUBJ

Figure 1 The basic schema of the affirmative finite verb in Turi

NEG=SUBJ LEXICAL BASE-TAM.ACT/MID=OBJ=FIN

Figure 2 The basic schema of the negative finite verb in Turi

The first element of the predicate is the lexical base, which usually consists of a single lexical morpheme. This is followed by portmanteau TAM/basic voice markers (= active or middle), object indexing for objects with animate reference, the finite marker /a/, and subject indexing for subjects with animate reference.

Table 5 gives an overview of the enclitic subject/object markers on the verb. With the exception of the 3rd person singular, these are all highly similar to the full forms of the pronouns given in Table 3. The two forms of the 1st person singular are speaker-specific free variants, while in the 3rd person singular in Odisha Turi *=i* indexes an object and *=ɛ* indexes a subject at the respective positions shown in Figures 1 and 2. In contrast, in Jharkhandi Turi *=i* indexes both subject and object. Otherwise, all indices in Table 5 can index both subjects and objects, at their respective positions within the predicate **[tab. 5]**.

Table 5 Turi enclitic argument-indices on the verb

	Singular	**Dual**	**Plural**
1st (EXCL)	*=ɛŋ/=iŋ*	*=liŋ*	*=lɛ*
1st (INCL)		*=laŋ*	*=bu*
2nd person	*=m*	*=bin*	*=pɛ*
3rd person	*=i* (object)/*=ɛ* (subject)	*=kin*	*=ku*

The finite marker /ɑ/ is realized as [ɑ] before subject markers beginning with a consonant but is elided before the 1st person singular *=ɛŋ/=iŋ*. Before the 3rd person singular marker *=ɛ*, the finite marker is raised and realized as /ə/. These two vowels are pronounced together as the diphthong [ə̯ɛ] Cf. the respective forms in (11).

(11) *a. gitiʔ-ɛn=a=lɛ*
sleep-PST.MID=FIN=1PL.EXCL
'we slept'

b. *ɔl-ɛtan=iŋ*
write-PRES.ACT=1SG
'I am writing'

c. *susun-tan=ə=ɛ*
dance-PRES.MID=FIN=3SG.ANIM
's/he is dancing'

The P-argument, roughly corresponding to the "direct object" with mono-transitive verbs, is indexed on the predicate before the finite marker when the reference of this argument is animate. This is shown in (12), from Jharkhandi Turi, where *=i* precedes the final finite marker *=a*.

(Jharkhandi Turi)

(12) *hɔn* *ĩja* *didi* *ka=i* *dal-ɛd=i=a.*
child$_j$ REFL elder.sister$_k$ NEG=3SG.ANIM.SUB$_j$ hit-PST.ACT=3SG.ANIM.OBJ$_k$=FIN
'The child did not hit his/her own sister.'

When the finite marker is not overtly realized because the subject is the 1st person singular, the object-index directly precedes the subject index, as in (13). Note that these two indices are pronounced as two separate syllables in (13), from Odisha Turi.

(13) *iŋ* *uni* *kɔɽa=kɛ* *ɲɛl-tad=i=iŋ* (< **ɲɛl-tad=i=a=iŋ*)
1SG that boy=OBJ see-PST.ACT=3SG.OBJ=1SG.SUBJ (< *see-PST.ACT=3SG.OBJ=FIN=1SG.SUBJ)
'I saw that boy.'

Table 6 presents the TAM/basic voice categories that we have identified in Odisha Turi. As the Odisha Turi data presently do not contain examples involving the active perfect, the active present-perfect marker in Table 6 is from Grierson's (1906, 130) data for the Ranchi dialect of Turi **[tab. 6]**.

Table 6 An overview of the TAM-markers in Turi

	Middle	**Active**
Present	*-tan*	*-ɛtan*
Past	*-ɛn/-kɛn/-ɔn/-ɔʔ-ɛn*	*-ɛkɛn* – Past imperfective *-tad* – Simple past
Future	*-ɔʔ/-ɔ/-ʔ/-*	-
Present perfect	*-akan*	(*-akad*)
Past perfect	*-lɛn*	*-lɛʔ/-laʔ*

The alternate forms of the past perfect marker *-laʔ/-lɛʔ* and the middle past marker *-kɛn/-ɛn* appear to be speaker-specific (cf. (14)-(15)).

(14) a. *alɛ=kun* *landa-laʔ=a=lɛ*
1PL.EXCL=PL laugh-PST.PERF.ACT=FIN=1PL.EXCL
'we (had) laughed'

b. *iŋ* *ɟɔm-lɛʔ=iŋ*
1SG eat-PST.PERF.ACT=1SG
'I (had) eaten'

(15) a. *gitiʔ-ɛn=a=lɛ*
sleep-PST.MID=FIN=1PL.EXCL
'we slept'

b. *sɛn-kɛn=ə=ɛ*
see-PST.MID=FIN=3PL
'he went'

In a few isolated forms (cf. e.g. (16)), the middle marker *-ɔʔ,* or its shortened form *-ɔ,* appears before the middle-voice past marker *-ɛn*[9] with no apparent semantic distinction to the use of *-ɛn* alone. After the short form /ɔ/, the /ɛ/ of *-ɛn* is elided and the form is realized as [ɔn]. The decline of the use of the middle markers here is perhaps connected to the fact that these forms are still recognizable as middle-voice past forms, distinct from the active forms, even without the marker *-ɔ(ʔ).*

9 Also before *-kɛn* in Grierson's (1906, 131) data for the Jashpur dialect; cf. *jõ-y-õk'-ken=ã* [fruit-*y*-MID-PST.MID=FIN] 'fruitful-was' (Authors' gloss).

(16) *sarag* *nilija* *ɲɛl-ɔʔ-ɛn=ə=ɛ*
sky blue see-PST-MID=FIN=3SG.SUBJ
'The sky (ANIM) looks (lit. is seen [as]) blue.'

With respect to the active past forms: In our data, *-ɛkɛn* is a general past imperfective used for prolonged actions such as progressives, states and iterative actions. In contrast, *-tad* appears to be restricted to single actions; we analyse it as a pure past tense, not as a perfective past tense.

This distinction in the active past seems to hold not only for Odisha Turi but also for the data from Grierson for Ranchi, Jashpur and Sarangarth Turi, at least tendentially. Compare the use of - *ɛkɛn* in (17) from Odisha Turi for a prolonged action with that for *-tad* below in (22), also from Odisha Turi, or in (18), from Jashpur Turi, from Grierson.

(17) *duba* *tɛŋrɛ* *bandra=kun* *ɲɛl-ɛkɛn=a=ku*
tree on.top.of monkey=PL see=PST.IPFV.ACT=FIN=3PL
'On top of the tree the monkeys watched/were watching [him do this].'

(Jashpur Turi, Grierson 1906, 131)

(18) *mackam=ke* *kara* *kuca-tad=a=e.*
machkam.flower=OBJ hail smash-PST.ACT=FIN=3SG.SUBJ
The *machkam* was smashed by hail (lit.: hail smashed the *machkam* flower).'

Standard negation in Turi is marked by *ka*, which appears directly before the predicate. The subject index attaches directly to this marker instead of appearing as the final enclitic on the predicate, where it is found in non-negated clauses (cf. again Figures 1 and 2). When followed by the 3rd person singular animate marker =*ɛ*, underlying /kɑ/ is realized as [kə].

(19) a. *sɛn-tan=ɛŋ* *ka=iŋ* *sɛn-tan=a*
go-MID.PRS=1SG neg=1SG go-MID.PRS=FIN
'I am going' 'I am not going'

b. *sɛn-tan=ə=ɛ* *kə=ɛ* *sɛn-tan=a*
go-MID.PRS=FIN=3SG.ANIM neg=3SG.ANIM go-MID.PRS=FIN
's/he is going' 's/he is going'

Only one object can be marked on the verb in Turi, and only if this object is considered animate. Object indexing in Odisha Turi follows a secundative alignment pattern, i.e., a distinction is made between the indexing of primary objects (= Patients and Goals) and secondary

objects (= Themes), as opposed to indirective alignment, which distinguishes between direct and indirect objects.

A note is necessary here with respect to terminology. With (mono-)transitive verbs, the primary object is P (for 'Patient'), corresponding roughly to the direct object of English. Bitransitive predicates, on the other hand, have G (for the 'Goal' of the action) and T (for 'Theme'), i.e., that entity which 'moves' to the goal with verbs such as *give*. When P and T receive the same marking, e.g., the accusative case in Latin or Sanskrit, they together form the category of direct objects. G is then marked differently from P and T, e.g., the dative in Latin or Sanskrit, and is the indirect object.

In Odisha Turi, P is indexed on the predicate when it references an animate entity, as in (12)-(13) above, repeated here for convenience as (20)-(21).

(20) *hɔn* *ĩja* *didi* *ka=i* *dal-ɛd=i=a.*
child$_j$ REFL elder.sister$_k$ NEG=3SG.ANIM.SUB$_j$ hit-PST.ACT=3SG.ANIM.OBJ$_k$=FIN
'The child did not hit his/her own sister.'

(21) *iŋ* *uni* *kɔʈa=kɛ* *ɲɛl-tad=i=iŋ* (< *ɲɛl-tad=i=a=iŋ)*
1SG that boy=OBJ see-PST.ACT=3SG.OBJ=1SG.SUBJ (< *see-PST.ACT=3SG.OBJ=FIN=1SG.SUBJ)
'I saw that boy.'

However, an animate G is also indexed on the predicate, as (22) shows, without any applicative to raise it to object status. In (22) the *=i* of the form *ɛm-tad=i=a=ku* refers to the man to whom the monkeys gave the hats (not explicitly mentioned in (22)). As P and G are similarly indexed at the same slot on the Turi verb but not T, Turi has a primary/secondary object distinction. The secondary object is then T, the NP *sɔbu ʈɔpi* 'all [the] hats', which is not indexed on the verb.

(22) *sɔbu* *katha* *ajum=kɛtɛ* *bandra=kun* *sɔbu* *ʈɔpi* *ɛm-tad=i=a=ku*
all story hear=CVB monkey=PL all hat give-PST.ACT=3ANIM.OBJ=FIN=3PL
'After hearing his whole story, the monkeys gave him all the hats.'

The gentive-marked possessor of G can also be indexed as the primary object, as with *ʧʰavɑ=kun* 'the children' in (23), who are the possessors of *lahiʔɲ=kɛ* 'bellies.'

(23) *ʈɔpi akriŋ=kɛtɛ ʧhaʊa=kun=rɛn lahiʔɲ=kɛ ɖana ɛm=ku=ɛŋ.*
hat sell=CVB child=PL=GEN belly=OBJ food give=3PL.OBJ=1SG
'I will sell the hats and give the children food (lit.: Having sold the hats, I will give them$_i$ food, the children's$_i$ bellies).'

As examples (24)-(25) from Grierson's data for Ranchi Turi and Sarangarh Turi show, an applicative construction is used in other dialects of Turi when G is indexed on the verb, i.e., G is raised here to the status of an object. In contrast, in Odisha Turi this is the default indexing pattern, and no applicative is found.

(Ranchi Turi, Grierson 1906, 130)

(24) *oɽo ac khurji haʈiŋ-ad=kin=a=i*
and ANAPH property distribute-APPL.PST.ACT=3DU.OBJ=FIN=3SG.SUBJ
'And his father gave them both (= the two sons) their property.'

(Sarangarh Turi, Grierson 1906, 133)

(25) *aba … hukum yem-ad=i-y=a=e…*[*]
father order give-APPL.PST=3.OBJ-*y*-FIN=3SG.SUBJ
'The father … gave them the order …

* Grierson (1906, 128-9) notes that in Sarangarh Turi the distinction between singular and plural is "often confounded", hence the *=i* in *yem-ad=i-y=a=e* refers to the sons (plural), despite the apparently singular form.

2.6.1 Copulas

Odiya Turi has two copulas, *hɛn-*, a locative copula, also used with temporary states, and the enclitic identity copula *=naŋ* '=IDENT.COP'. *hɛn-* appears to derive from the form *tahɛn* 'stay, remain', related forms of which are also found as copulas in other Kherwarian languages, cf. Mundari *tae* but also Jharkhandi Turi *tai* and the Ranchi dialect of Turi form *tahi* in Grierson's data, where it also still functions as a full verb with the meaning 'stay, remain'. As a copula it is only found in Ranchi Turi in the past tenses with the form *tahi-*, while *hen-* is used in the present tense.

In the Odiya and Jharkhandi Turi data in our corpus, *hɛn-* marks for the person/number of the subject at the position which, with transitive verbs, is used to index objects, and is followed by the finite marker *=a*. We therefore gloss this index as OBJ to call attention to this.

2.6.2 Temporary/Locative Copula

(Jharkhandi Turi)

(26) *ʧɛja* *lɛkha* *hɛn=ku=a* *ʧhuʊa puta?*
what like LOC.COP=3PL=FIN children
'How (lit.: like what) are the children?'

(Jharkhandi Turi)

(27) *ʧɛja* *lɛkha* *hɛn=mɛ=a?*
what like LOC.COP=2SG.OBJ=FIN
'How are you?'

Exceptionally, with *hɛn-* inanimate subjects are indexed on the verb by *=aʔ*, whereas with other verbs inanimate subjects and objects are not indexed on the verb.

(Odiya Turi)

(28) *bahrirɛ* *lim* *duba* *hɛn=aʔ=a.*
outside Neem.tree tree LOC.COP=3SG.INAN.OBJ=FIN
'There is a Neem tree outside.'

In contrast, in Grierson's data for Turi in the Ranchi area the subject is marked in the usual subject position, verb-finally; cf. example (29).

(Ranchi Turi – Grierson 1906, 130)

(29) *miat'* *[h]oɽ=ke** *baria* *chaua* *tahi=ken=a=kin.*
one man=OBJ two child cop=PST.MID=FIN=DU
'A man had two sons (lit.: 'to one man two children were').'

* The form is given as <noɽ> in Grierson (1906, 130), although <hoɽ> was surely intended. This may also be an artefact of the reprint which we consulted.

The locative copula in Odisha Turi has the suppletive negative form *kanɔʔɔ*.

(30) *ɲɛl-laʔ=ə=ɛ* *ʤɛ* *miaʔn* *au* *ʈɔpi* *kanɔʔɔ.*
see-PST.PERF.ACT=FIN=3SG.ANIM then one and hat NEG.PRS.COP
'He saw then that there was (= is) not one single hat.'

The temporary state/locative copula marked for an inanimate 3rd person singular subject (= default subject marking) can also be used with a verbal stem and a 'subject' marked by the objective case marker *=kɛ* to denote obligation.

(31) *iŋ=kɛ* *ɔɽaʔa* *sɛn* *hɛn=aʔ=a.*
1SG=OBJ house go LOC.COP=3SG.INAN.OBJ=FIN
'I have to go home.'

2.6.2.1 Identity Copula

In Odiya Turi the identity copula is marked by *=naŋ*, the etymology of which is unclear.

(32) *iŋ=aʔ* *ɲumu* *prʌbhat=naŋ.*
1SG=GEN name Prabhat=IDENT.COP
'My name is Prabhat.'

(33) *am* *oka* *ɖihi=re=naŋ?*
2SG which village=LOC=IDENT.COP
'What village are you from?' (lit.: 'You are one in which village?')

In contrast, there is no present-tense identity copula in our Jharkhandi Turi data; instead, the two NPs are simply juxtapositioned, as in (34).

(Jharkhandi Turi)

(34) *ʌm=a* *tʃeja* *numu?*
2SG=GEN what name
'What is your name?'

2.6.3 Imperatives and Hortatives

There are a number of imperatives in our corpus, all marked for the second person singular and in the middle voice. The verb is marked by the middle voice marker *-ɔ* and the marker of the second person, singular, *-m*, resulting in *-ɔm* or the slightly irregular *=ɔʔb*. In all forms, the finite marker is lacking. See the examples in (35).

(35) a. *hiɟ-ɔm* / *hiɟ-ɔʔb* b. *sɛn-ɔm* / *sɛn-ɔʔb*
come-MID.2SG come-MID.2SG go-MID.2SG go-MID.2SG
'Come!' 'Go!'

A few exceptional forms are found in which neither voice nor person is marked, as these stems do not derive from Turi verbal roots but either from particles which have lexicalized in this function or they have been borrowed from Indo-Aryan, such as those in (36).

(36) *dan!* *ɛla!* *ɟu!*
'give!' (< Indo-Aryan) 'come!' 'go!'

Imperatives are negated by the prohibitive marker *alu*, to which the enclitic subject marker attaches, as shown in (37). Note that the imperative form lacks all TAM marking.

(37) *am* *alu=m* *sɛn!*
2sg proh=2sg go
'Don't go!'

Hortatives are formed the same way as future-tense verbs, i.e., in the active with zero marking for the future and in the middle voice with the middle-voice marker *-ɔ* attached directly to the verb stem. This is then followed by the finite marker *=a* and inclusive 1st person indexing; cf. example (38), and further forms from the texts in Appendix 2, given in (39).

(38) *ɛla* *rɛ* *tʃhaʊa=kun* *sɛn=a=bu* *iskul*
come.IMP VOC child=PL go=FIN=1PL.INCL school
'Come along, children! Let's go to school!'

(39) a. *paɽh=a=bu!*
learn=FIN=1PL.INCL
'Let's learn!'

b. *lɛkəɛ=a=bu!*
count=FIN=1PL.INCL
'Let's count!'

c. *itu-ɔ=a=bu!*
learn-MID=FIN=1PL.INCL
'Let's learn!'

2.6.4 Converbs

There are two common non-finite forms contained in our data. These are:

- the sequential converb, marked by *=kɛtɛ*, which directly follows the verb stem, as in (40).

(40) *miaʔn* *phɛriʊala* *ʈɔpi* *idi=kɛtɛ* *haʈ* *sɛn-kɛn=ə=ɛ.*
one hawker hat take=SEQ market go-MID.PST=FIN=3SG.ANIM
'A hawker took (= having taken) a hat [and] went to market.'

- the imperfective converb, usually marked by a repetition of the verb stem followed by the marker *-tɛ* .

(41) *phɛriʊala* *landa* *landa-tɛ* *haʈ* *sɛn-lɛn=ə=ɛ.*
hawker laugh laugh-SIM market go-MID.PST.PERF=FIN=3SG.ANIM
'The hawker went to market, laughing all the way.'

Some monosyllabic roots partially reduplicate internally before *-tɛ* (42), although not all (43).

(42) *sɛ-sɛn* *sɛ-sɛn-tɛ* *thəka-ɛn=ə=ɛ.*
RDP-go RDP-go-SIM become.tired-MID.PST=FIN=3SG.ANIM
'Walking along he became tired.'

(43) *dub* *dub-tɛ* *gitiʔ-ɛn=ə=ɛ.*
sit sit-SIM sleep-MID.PST=FIN=3SG.ANIM
'While sitting there he fell asleep.'

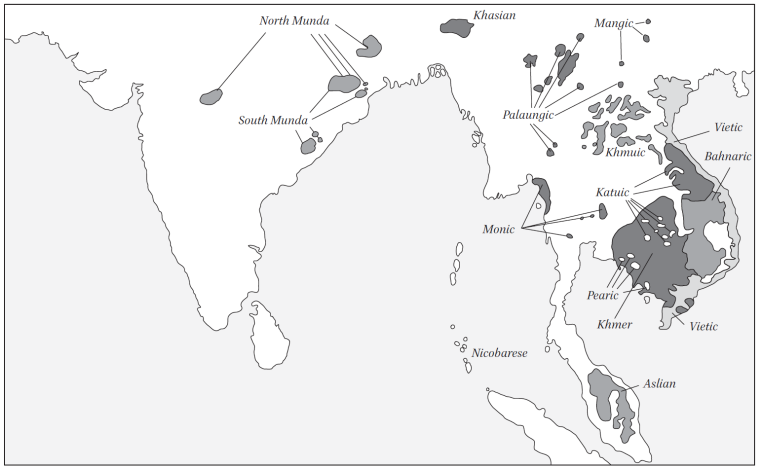
This concludes our grammar sketch. In addition to this sketch, Appendix 1 contains a list of vocabulary items for Odisha Turi according to the list used in Kobayashi et al. (2003). This is followed in Appendix 2 by two short segmented, glossed and translated Turi texts.

3 Turi and Its Relation to Other Kherwarian (North Munda) Languages

In this section we deal with the position of Turi within the Kherwarian branch of North Munda. Due to the very limited data which until now has been available, most of it stemming from the short discussion in Grierson's *Linguistic Survey of India* from 1906, Turi's position within North Munda is still unclear. In the present section we therefore present the results of an automated comparison of Turi with other varieties of the Kherwarian group. To this end, we collected data for Odisha Turi for as many of the 274 lexemes as possible discussed in Kobayashi et al. (2003) for 12 other North Munda varieties and analysed these with the help of the program COG from the Summer Institute of Linguistics in order to determine Turi's position within North Munda.

3.1 The Position of Turi Within the Munda Family – Previous Discussions

The Munda languages, to which Turi belongs, form the western-most branch of the Austro-Asiatic language family, which stretches from Central India in the west to Vietnam in the east. Map 2, from Sidwell (2015, 144), provides an overview of the spread of this family and its main branches **[map 2]**.



Map 2 The branches of Austro-Asiatic (Sidwell 2015, 144)

Figure 3 presents the traditional internal classification of Munda, from Zide (1969, 412) **[fig. 3]**. In this classification, the Munda group bifurcates into North Munda and South Munda branches. The northern branch then bifurcates into Korku, spoken in central India, and Kherwarian, spoken in eastern India. Only the Kherwarian branch will be discussed in the following.[10]

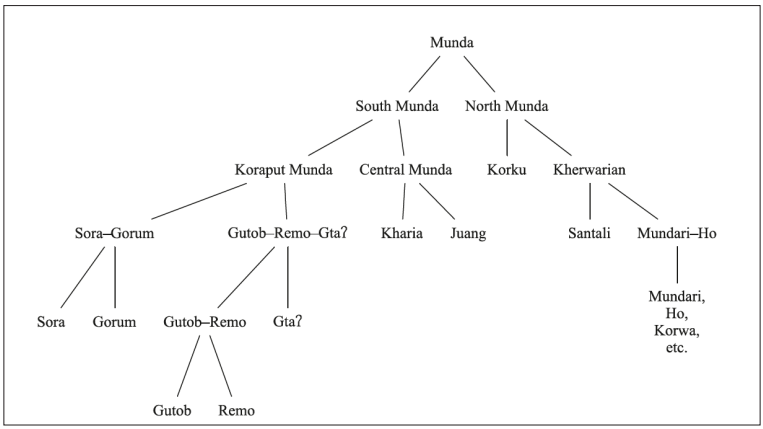


Figure 3 The Munda languages according to Zide (1969, 412)

10 For an overview of revisions of the classification of Munda languages, see Anderson 2015.

Turi belongs to the Kherwarian group and is largely mutually intelligible with the languages of both the Santali and the Mundari-Ho branches, but what has remained unclear until now is to which of these two branches it belongs, or whether it possibly constitutes a third, independent branch.

Grierson (1906, 128) writes the following of Turi: "The Birhâṛ dialect is closely related to Muṇḍārī, and the speech of the Tūrīs also agrees with that language in most essential points. In a few characteristics, however, it follows Santālī, as against Muṇḍārī". Further on the same page, he writes that "[i]n Sambalpur the Tūrī dialect is almost pure Muṇḍārī" but then goes on to note similarities with Santali, most notably with respect to phonology, which is central to the comparative method: "Forms such as *pēā*, three; *pūniā*, four, in Tūrī agree with Santālī, as does the phonology of the dialect in most points".

Munda (1968, 46-7) on the other hand notes a number of similarities between Turi and both Mundari and Santali.[11] This indeterminate status of Turi within Kherwarian persists until today, e.g. with the *Glottolog* (Hammarström et al. 2022) classifying Turi as a Mundaric language while the *Ethnologue* (Eberhard et al. 2023) groups it together with Santali.

3.2 The Data

In order to determine the position of Turi within the Kherwarian languages, the data for all 274 lexical items contained in Kobayashi et al. (2003, 347-67) for three dialects of Ho (Chaibasa, Goilkera and Ghatshila), four dialects of Mundari (Darigatu, Bandugara, Salgadih and Kera) as well as five dialects of Santali (Kadma, Heben, Tikahara, Hatsara and Simoldohi) were entered by Luna Hemmerling, a student aide at Kiel University, into a spreadsheet. Our own data on Odisha Turi were then added to this spreadsheet.[12]

11 Both Osada (1991, 175) and Anderson et al. (2008, 198) cite the following quote from Munda (1968, i-ii), which is even more explicit with respect to Munda's views on the place of Turi within Kherwarian: "The place of Turi was left undefined in Grierson's LSI but we feel that it – along with Asuri, Birhor and Korwa – is now more like Mundari than Santali. In certain respects (e. g., in sharing the same vowels in a few items and in dropping morpheme final vowels in certain forms), however, they look more like Santali than Mundari but they can be derived for the most part as simply from Pre-Mundari". Unfortunately, the first author of this study's own photocopy of Munda 1968, which does not appear to have been published, does not contain any pages with Roman numeration, so that we cannot confirm this quote, although it is largely in line with the discussion of Turi in that work on pp. 46-7.

12 Despite the wealth of data contained in Osada's (1991) publication of Father Ponette's field notes, these data were not entered into the above-mentioned list, primarily due to uncertainties with respect to the granularity of the transcriptions.

Our original aim was to enter detailed data on several dialects of Turi into this file for comparison with these other Kherwarian varieties. Unfortunately, this turned out to be more difficult than anticipated. First, we have not been able to locate any speakers in the state of Jharkhand under the age of 60, and many of the speakers that we have located are either semi-speakers or have not used the language in many years. The situation with COVID-19 further complicated our work, making it impossible to reach many villages during this time, and a number of these older speakers who we had contacted before the pandemic tragically succumbed to this illness during the second wave in 2021. We have however located a relatively large number of Turi speakers from northwestern Odisha, from communities in which Turi is still being learned by the younger generation. As Covid until very recently continued to make travel to these villages difficult, a group of five Turi adults, three male and two female, were invited to Ranchi to cooperate with the authors of this study during the course of five days to elicit stories and songs and to complete as much of this vocabulary list as possible.

As the members of this group came from different villages, there is unfortunately no one single local variety which we could record but five slightly varying varieties, although all speakers were from a relatively small geographical area and their data do not differ substantially. As a result, however, our data are not as fine-grained as that for the 12 Ho, Mundari and Santali dialects documented in Kobayashi et al. (2003). Our data for Odisha Turi were therefore entered as one form of Turi, although we have included all variant forms from these respondents in the spreadsheet. Altogether, data for 227 of the 274 items in the above-mentioned list were obtained in this manner. Data from the considerably more different Jharkhandi Turi were not included, as we still have no data for many lexical entries.[13]

The data were then cleaned to ensure maximal comparability. First, as it is imperative that only cognate forms are being compared, we removed all loan words from the 13 varieties in the data which we were able to determine, including very old loan words from Indo-Aryan such as *daru* ‘tree’ (no. 118) but of course also more recent loans, such as various words for ‘lake’ or ‘sea’ (nos. 140-1), cf. e.g. *sagar*, *samud* (from Sanskrit), *dɔrɛja* (cf. Persian *dærja*), etc. We also removed compounds from the list as the monomorphemic words for the relevant lexemes were already contained elsewhere in the list.

13 Unfortunately, no morphological data such as TAM markers, case markers, different forms of PNG markers, etc., were included in this list. Although this type of data is essential for a true comparison of such closely related language varieties, as two reviewers noted, this is presently not an option as these forms are not noted for the other Kherawarian variants in Kobayashi et al. (2003). Hence, there are no cognate forms in that list to compare with the Turi grammatical forms.

E.g., transparent compounds such as e.g. 'tear' (no. 6), consisting of morphemes for 'eye' and 'water', i.e., 'eye-water', were removed from the data, since these two lexemes are already contained in the list (no. 5 'eye' and no. 143 'water').

Complex forms where one part was cognate with forms in other dialects in the list were handled differently; here we retained the cognate lexeme from the respective complex form and removed the other element, which was either opaque, a grammatical morpheme (e.g. a nominalizer) or a lexical morpheme found elsewhere in the list. To give an example, no. 39 'sweat': In the Santali data we find the forms *ud'gɛr* (Heben) and *ud'gər* (Kadma, Tikahara and Simoldohi), but *ud'gər da·kˇ* for the Hatsara form. Here, *da·kˇ* is clearly the morpheme for 'water' and was thus removed, as it is found elsewhere in the list (no. 143), while the first element, *ud'gər*, was retained, as it is cognate with the other four entries.

Our goal was primarily to compare phonological developments in these languages in order to determine their genealogical relationships with one another, not lexical similarity per se. For this reason, we only consider phonetic similarity in the following. We also deleted a number of problematic entries, such as deictic units which consisted primarily of grammatical morphemes where it was not always clear whether the forms were cognate (e.g., nos. 257-63). These will have to be studied in more detail in a future study.

Although perhaps the most characteristic phonological trait of Munda languages, the preglottalization and non-audible release of syllable-final voiced plosives (e.g., [ʔbˇᵐ]) were removed from the data, as these forms are often realized in Turi and many other Munda languages both as preglottalized and as non-glottalized variants even by the same speaker in natural speech, and we felt that it could skew the data if e.g. in some languages both forms occurred but only one of the two happened to be documented. For this reason, only the non-preglottalized forms were used. Finally, as no studies have yet been carried out on lexical accent in Turi or how this is to be defined, all primary and secondary accents were removed from the data for all varieties before comparison. We then deleted all lexical entries for which less than five forms in total were present from the 13 different linguistic varieties.

Finally, all entries were removed which did not have a corresponding non-Indo-Aryan entry for Turi. Altogether, these measures combined to reduce the number of lexical entries considerably, from the original 274 items – of which we have 227 for Turi – to 95. It is this smaller group that forms the basis for our comparison.

We stress here that this is meant only as a preliminary attempt to determine Turi's position within Kherwarian, based on our current knowledge. As finer-grained data for Turi and other Kherwarian languages are added to this lexical database, our understanding of Turi

and the other languages in this group will continue to improve. But as we now have enough data for a first analysis of Turi, we believe that an introduction to the language such as the present one should at least offer a preliminary discussion of Turi's place in Kherwarian, however tentative that may be.

3.3 Results

The data were then analysed with the software COG by the Summer Institute of Linguistics. COG was chosen for a number of reasons: First, the software is openly available and easy to use, even when the respective researcher does not have extensive computational skills. As such, our results can be reduplicated by other researchers regardless of their computational training. Second, COG allows researchers to view the results in various formats, such as dendograms, (non-rooted) trees, networks, and also provides a similarity matrix for the different varieties. Third, COG allows the researcher to actively participate in the analysis, e.g., by correcting false analyses where the algorithm mistakenly views different forms as cognate which the researcher does not consider cognate, or conversely by marking these forms as cognate when the algorithm has not analysed them as such.[14]

Finally, COG provides analyses with respect to lexical and phonetic similarity in both a UPGMA and a Neighbour-Joining (NJ) analysis.[15] Unfortunately, these are the only two algorithms offered in COG, although in view of the preliminary nature of our investigation, this will be sufficient for an initial perspective. The results of this comparison are presented in Figures 4-5.

14 Due to the preliminary status of our study, this option was not used in the present analysis as this would involve pair-by-pair viewings of all language varieties with one another, based on detailed knowledge of historical Munda developments, which is beyond the expertise of the members of our research group.

15 For reasons of space, we will not deal here further with the different assumptions made in the UPGMA vs. Neighbour-Joining analyses. Further discussion of some aspects of these and similar approaches can be found e.g. in Nichols, Warnow 2008. For our purposes it will be enough to simply consider these to be two competing approaches to analysing the data. For information on how COG determines phonetic similarity between two different languages, see Kondrak 2000.

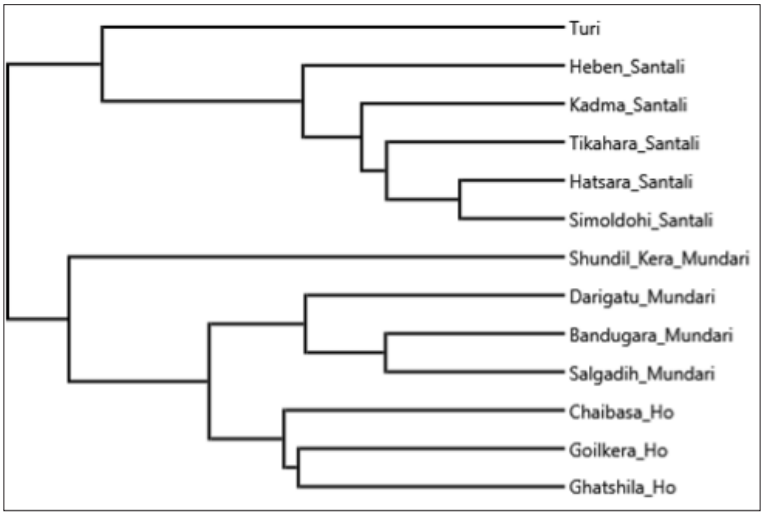


Figure 4 The genetic relationship of various Kherwarian languages and dialects – UPGMA, phonetic similarities

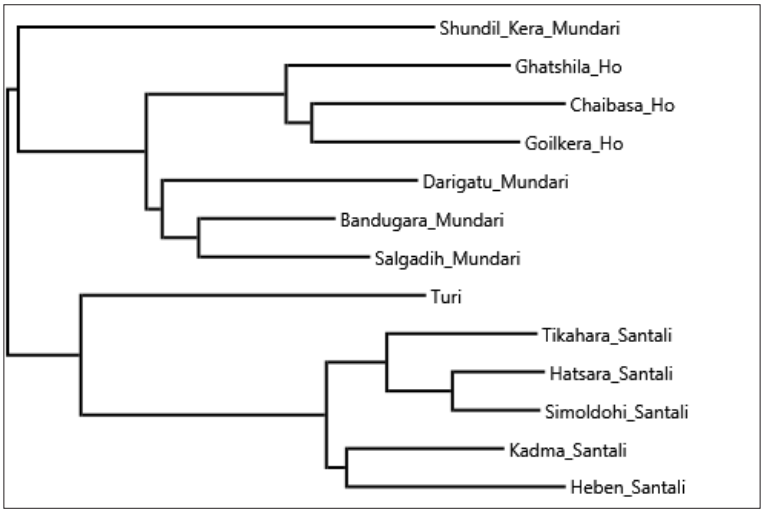


Figure 5 The genetic relationship of various Kherwarian languages and dialects – NJ, phonetic similarities

In both Figures 4 and 5, Santali dialects all group together, all Mundari dialects other than Kera Mundari (to which we will return) group together, and all Ho dialects group together. Furthermore, the Mundari and Ho branches – together with Kera Mundari – form a branch distinct from Santali **[figs 4-5]**. While this classification is not surprising, it does confirm that both analyses come to very similar

conclusions. In fact, they are identical with the exception of Kadma Santali and Heben Santali: in Figure 4, Heben forms the outermost branch of Santali, followed by Kadma and then the remaining three varieties (Tikahara, Hatsara and Simoldohi), while in Figure 5 Kadma and Heben together form a sister branch to the remaining three Santali dialects, which have the same internal relationships. Otherwise the two figures are identical with respect to their internal branching.

With respect to Turi, note that in both Figures 4 and 5 Turi is a sister language to the entire Santali branch, i.e., in neither analysis is Turi more closely related to Mundari-Ho than to Santali. This is more in line with the classification in the *Ethnologue* (Eberhard et al. 2023) than with the *Glottolog* (Hammarström et al. 2022), which classifies Turi as a member of the Mundaric branch.

Figures 4 and 5 also incidentally show that both algorithms classify Kera Mundari as a sister language to the entire Mundari-Ho branch, and not within the Mundari group. While this position of Kera Mundari may at first glance seem somewhat unexpected, it becomes more understandable when the history of this dialect is taken into consideration: Kera Mundari is known to be the result of a language shift by speakers of Kurux (Dravidian) to Mundari, as the result of which it "has unique characteristics and constitutes a distinct regional as well as ethnic dialect" (Kobayashi, Murmu 2008, 165). This shift provides a likely explanation for its relatively distant relation to the other members of this group, although it is considered a dialect of Mundari by the speakers themselves.

4 Discussion of the Results of the Comparison

With respect to the position of Turi within the Kherwarian branch, the results of both algorithms, i.e., UPGMA and NJ, come to the same conclusion, namely that Turi is a sister language to Santali, and that all of the Santali dialects cluster together as a sister branch to Turi, which joins the Santali group at a higher level. Otherwise the genealogical relationships are virtually the same in the two representations.

The question naturally arises why Turi, if it is indeed more closely related to Santali than Mundari-Ho, is spoken so far away from the Santali 'heartland', which today is considerably further to the east in eastern Jharkhand and beyond into West Bengal. We offer here the following tentative suggestion to account for this geographical separation.

Several Kherwarian-speaking ethnic groups such as the Santali, the Mundari and the Ho speak of the migration of their ancestors into their present homelands from the west, and it has recently been suggested that the Santal speakers now residing in eastern Jharkhand (and further still to the north and east) migrated there from western and central Jharkhand, perhaps from the fourteenth century onwards

(Das 2020, 1224-5). This could indicate that the break-up of this earlier, relatively homogeneous dialect group into the present distinct branches only goes back ca. 600 years, perhaps somewhat longer, which fits in well with the high degree of mutual intelligibility among the Kherwarian languages.

In this analysis, Turi and perhaps other smaller Kherwarian languages still spoken in western Jharkhand, eastern Chhattisgarh and northwestern Odisha may be remnants from the time before this eastward migration and before the differentiation of Kherwarian into distinct linguistic sub-groups. Assuming that the results of this admittedly preliminary study stand the test of time, we suggest the following scenario: Before the migration further east into eastern Jharkhand had begun, Kherwarian likely consisted of only very weakly differentiated dialects along a continuum, with at least two still very similar poles, one of which would go on to become the Santali-Turi group, the other the Mundari-Ho group. In this analysis, the ancestors of the present-day Turi would then have remained in western Jharkhand and Chhattisgarh or migrated southward into Odisha, unlike most other Kherwarian groups who migrated eastward.[16]

A word of caution is in order here, however: first, as noted in fn. 20, the data we analysed do not contain any grammatical marking such as TAM or verbal indexing, case or number, as the original list in Kobayashi et al. (2003) does not contain this information. Thus, we did not have such information from the 12 varieties in that list to compare with the Odisha Turi forms. Second, and equally important, is the small size of our database, so that even small changes in the data can lead to a slightly different analysis in either the NJ or the UPGMA analysis, or in both, with respect to the positions of Turi and Kera Mundari. When data are changed in the database for independent reasons, these two languages can either appear in their present positions in Figures 4 and 5, or they can appear together, as a sister branch to all other groups.

For example, when at a relatively late stage in our work a small number of Indo-Aryan loan words were discovered which had not previously been removed from the Santali, Mundari and Ho data from Kobayashi et al. (2003), this changed the position of Kera Mundari and Turi from where they are in Figures 4 and 5 above to clustering together as a separate branch of Kherwarian. Later, however, when a final few Indo-Aryan words were again discovered and removed, the present classification re-emerged.

This indicates to us not that there is a special relationship between Turi and Kera Mundari – there clearly is not. Rather, these

16 Presumably only much later did speakers of Turi, similar to other Munda groups, migrate to Assam (see Rau, Sidwell 2019, 36-7).

two languages do not fit well into either the Santali or the Mundari-Ho group. Kera Mundari is somewhat closer to Mundari-Ho than to Santali, and Turi is somewhat closer to Santali than to Mundari-Ho. However, the proximity to either group is apparently not strong enough to rule out another possibility: Whereas Kera Mundari's status as the result of language shift explains its special status quite well, Turi's special status may derive from it having broken off independently from the remaining Kherwarian groups at an earlier date, so that it may represent a third Kherwarian group, instead of descending from an earlier Santali-Turi group. As comparable data on different dialects of Turi and other Kherwarian varieties emerge, this question can hopefully be answered more clearly.

Based on the results of both algorithms, we can however state that Turi is not most closely related to Mundari-Ho. Whether it is a sister to the Santali branch or perhaps an independent branch of Kherwarian awaits further study.

5 Summary

In the present study we give a preliminary introduction to the Turi tribe and their traditional language. The Turi are an officially recognized Scheduled Caste residing in Jharkhand, Chhattisgarh, Odisha, Bihar, West Bengal and Assam. We also provide a skeleton grammar of Odisha Turi, including basic aspects of phonology, the nominal phrase and the verb system. Finally, the two appendices at the end of this study provide a basic vocabulary list for Turi, based on that given in Kobayashi et al. (2003), and two short texts in Turi. We also compare Turi with the 12 other North Munda varieties discussed in Kobayashi et al. (2003) to determine the position of Turi within Kherwarian using the program COG from the Summer Institute of Linguistics. The results suggest that Turi is closer to Santali than to the Mundari-Ho branch.

Our results also indicate that Kera Mundari, despite its name, is best considered not a Mundari dialect but rather a sister language to the Mundari-Ho branch from a phonological perspective, even if it is considered a Mundari dialect from a sociolinguistic perspective. This special status is no doubt due to the fact that the speakers of this language descend from earlier Kurux (Dravidian) speakers who switched to Mundari, leaving their own distinct imprint on the language in the process.

The fact that both algorithms used, i.e., UPGMA and Neighbour-Joining, assign Turi and Kera Mundari to the same respective positions vis-à-vis all other 12 Kherwarian varieties supports the results of our analysis. Nevertheless, further data are required before we can be sure of the internal structure of Kherwarian. We hope that

the present study will provide a foundation on which further studies can build.

Turi is now spoken by only very few people – almost certainly less than 2,000 and perhaps only a few hundred – but its speakers belong to an ethnic population of some 354,000. As our study suggests that Turi is a sister to Santali, which is primarily spoken considerably further to the east, this could mean that Turi emerged as a separate language when one group of speakers of Proto-Santali-Turi remained in eastern Chhattisgarh, western Jharkhand and northwestern Odisha during the general Kherwarian eastward migration. The language of those who remained in the west went on to become modern Turi, while that of the rest of this group, who continued eastwards, went on to become Santali.

There are still many open questions in this proposed development, questions that can only be answered through further fieldwork in the region. The past years have seen a number of important advances with respect to our understanding of the linguistic and ethnic history of this region, but it is clear that there is still much to be done.

While the Chotanagpur Plateau is often considered an accretion zone, following Nichols' (1992; 1997) conceptual categories,[17] it is becoming increasingly clear that Jharkhand is also a 'mini-spread zone'. That is, while the whole of the Chotanagpur Plateau can be considered an accretion zone with respect to the surrounding areas, especially the Gangetic Plain to the north, languages such as Santali, Mundari and Ho and Indo-Aryan languages such as Sadri/Nagpuri, Khortha and Kurmali have also spread throughout this zone at the expense of earlier indigenous languages. Traces of at least one of these earlier languages which are no longer spoken in the region, and which do not appear to belong to any known language family, have also recently been reported.[18] As new data from fieldwork emerge, much of what has long been considered conventional wisdom will likely give rise to new insights – and also to new questions.

17 E.g., Ivani et al. 2021; Peterson 2017; 2022; forthcoming.

18 Cf. e.g. the data on Kurmali in Paudyal, Peterson 2021, 296.

Abbreviations

1, 2, 3	1st, 2nd and 3rd persons
ACT	active voice
ANAPH	anaphoric pronoun
ANIM	animate
APPL	applicative
COP	copula
COUNT	counting morpheme
CVB	sequential converb
EXCL	exclusive
FIN	finite marker
GEN	genitive
IMP	imperative
IDENT	identity (copula)
INAN	inanimate
INCL	inclusive
IPFV	imperfective
LOC	locative
MID	middle voice
NEG	negative
OBJ	objective case
PROH	prohibitive
PST	past
PL	plural
RDP	reduplication
REFL	reflexive
SG	singular
SIM	simultaneous converb
VOC	vocative
-y-	hiatus-breaking element

Appendix 1: Odisha Turi basic vocabulary

The following list contains all Turi words elicited during our workshop together with speakers of Odisha Turi and follows the order used in Kobayashi et al. (2003). Words marked as '(IA)' for 'Indo-Aryan' were removed before entering the data into COG, as well as all compounds such as *mɛdaʔa* 'tear' (6), which likely consists of the lexemes *mɛʔn* 'eye' (5) and *daʔa* 'water' (143).

The use of '(IA)' does not necessarily imply that the respective Turi lexeme ultimately derives from Indo-Aryan but rather simply that the respective lexeme has the same or a highly similar form in an Indo-Aryan language of the region, from which it was likely borrowed. For example, *ɖihi* 'village' (247) has a very similar form in various Magadhan languages but may ultimately not be of Indo-Aryan origin, or *haphta* 'week' (165), ultimately from Persian but which has entered Turi through neighbouring Indo-Aryan languages. All loanwords which we could identify were removed from the list before comparison – both from Turi as well as from the other North Munda languages. '-' in the following list means that we were not yet able to elicit the respective form corresponding to the morpheme in the list in Kobayashi et al. 2003.

1.	head	*bɔhɔʔɔ*
2.	hair	*uʔm*
3.	forehead	*malaŋ*
4.	eyebrow	*bʱama (IA)*
5.	eye	*mɛʔn*
6.	tear	*mɛdaʔa*
7.	ear	*lutur*
8.	nose	*mũ*
9.	mouth	-
10.	lip	*limtir*
11.	tongue	*alaŋ*
12.	spit	-
13.	tooth	*ɖaʈa (IA)*
14.	chin	*ʈʰuɽʱi (IA)*
15.	cheek	*dʒɔha*
16.	moustache	*mɛtʃʰa (IA)*
17,	face	*mɛʔn muhaɽ*
18.	neck	*hɔʈɔʔɔ*
19.	throat	*saŋk*
20.	shoulder	*kʰand (IA)*
21.	back	*dɛja*

22.	waist	*majaŋ*
23.	buttock	*ʧʊtʰal (IA)*
24.	chest	*kɔɾɔm*
25.	breast	-
26.	belly	*lahiʔɲ*
27.	navel	*buʈi*
28.	arm	*tiʔi*
29.	elbow	*kuhuni (IA)*
30.	hand	*tiʔi*
31.	finger	*kaʈuʔ*
32.	nail	*ram*
33.	leg	*dʒaŋga (IA)*
34.	knee	*ʈʰɛuna (IA)*
35.	liver	*kaldʒa (IA)*
36.	heart	*dʒiu (IA)*
37.	guts	*pɔʈa*
38.	skin	*hartaʔa*
39.	sweat	*balbaldaʔa*
40.	filth	*pʰuhuɾi*
41.	pus	*sɔ̃dɒɔ*
42.	hair	*uʔm*
43.	fat	*ʧarbi (IA)*
44.	blood	*majɔm*
45.	bone	*dʒaŋ (IA)*
46.	flesh	*dʒil*
47.	body	*hɔɾɔ*
48.	disease	*dʒar (IA)*
49.	wound	*gʱaɔ (IA)*
50.	medicine	*ran*
51.	rice	*ʧaʊli (IA)*
52.	powder	*gunɖa (IA)*
52.1	flour	*aʈa (IA)*
53.	salt	*buluŋ*
54.	oil	*sunum*
55.	liquor	*arkʰi (IA)*
56.	tobacco	*tambaku (IA)*
57.	taste	*sibil*
58.	flavour	*sɔ̃ʊ̃*
59.	food	*dʒɔmɛ*
60.	meat	*dʒil*
61.	egg	*bili*
62.	chicken	*sim*
63.	bird	*ɔɾɛ*

64.	wing	*paɛŋk (IA)*
65.	feather	*pudga (IA)*
66.	nest	*kʰɔta (IA)*
67.	beak	*ʈʰɔŋ (IA)*
68.	horn	*siŋ (IA)*
69.	cow	*uriʔ*
70.	knife	*kuntʃi (IA)*
71.	sword	*kʰaɳɖa*
72.	blade	*uhula*
73.	pole	*kuʈa (IA)*
74.	bow	*dʰanu (IA)*
75.	arrow	*ʧɛl (IA)*
76.	lance	-
77.	thread	*sutam (IA)*
78.	needle	*sudʒi (IA)*
79.	clothe	*lidʒaʔa*
80.	paper	*kagadʒ (IA)*
81.	thing	*dʒinis (IA)*
82.	snake	*biŋ*
83.	worm	*tidʒu*
84.	fly	*rɔ̃*
85.	mosquito	*sikiɽi*
86.	flea	-
87.	louse	*siku*
88.	ant	*muʔn*
89.	fish	*haku*
90.	shellfish	-
91.	animal	*dʒatu (IA)*
92.	hunting	*sikar (IA)*
93.	net	*dʒal (IA)*
94.	dog	*sɛta*
95.	rope	*dãɔ̃ra*
96.	string	*baɛr*
97.	sheep	*gaɻa*
98.	horse	*gʰɔɽa (IA)*
99.	pig	*sukiri*
100.	tail	*pɔʧʰ (IA)*
101.	animal hair	*uʔm*
102.	fur	-
103.	sack	*basta (IA)*
104.	pan	*tavɒa (IA)*
105.	kettle	*gaɲɟ, gã dʒ (IA)*
106.	jar_1	-

107.	jar$_2$	*gʰamala (IA)*
108.	roof	*tʃat (IA)*
109.	wall	*katʰi*
110.	window	*dʒʰarka (IA)*
111.	door	*siniŋ, siŋduar (duar is IA)*
112.	house	*ɔɽaʔa*
113.	vehicle	*dʒanbahan (IA)*
114.	vessel	*huɖiŋ gaɲɟ*
115.	well	*bauli (IA)*
116.	job	*kami (IA)*
117.	money	*ʈaka (IA), paisa (IA), kɛtʃa*
118.	tree	*duba*
119.	stem	-
120.	branch	*ɖahuɽa*
121.	grass	*gʰãs (IA)*
122.	stalk	*hapa*
123.	root	*dʒɛri (IA)*
124.	leaf	*sɛkam*
125.	flower	*baha*
126.	fruit	*pʰal (IA)*
127.	seed	*bihɔn*
128.	bark	*tʃʰali (IA)*
129.	rice-field	*bahal dɔɛn (dɔɛn IA)*
130.	groove	-
131.	forest	*buru*
132.	road	*hɔra*
133.	hole	*lata (IA)*
134.	bridge	*pulia (IA)*
135.	river	*nai (IA)*
136.	mountain	*buru*
137.	plain$_1$	*paɽia*
138.	plain$_2$	*saman (IA)*
139.	pond	*ban (IA)*
140.	lake	*sagar (IA)*
141.	sea	-
142.	island	*ʈapu (IA)*
143.	water	*daʔa*
144.	ice	*barap, barapʰ (IA)*
145.	stone	*diri*
146.	earth	*dʰarti (IA), ɔt*
147.	sand	*dʰuri (IA)*
148.	dust	*lukum dʰuri (dʰuri IA)*
149.	smoke	*sukul*

150.	ash	*tɔrɛʔ*
151.	fire	*sɛŋgɛl*
152.	wind	*dʰuka (IA)*
153.	cloud	*rimil*
154.	fog	*kuhuri (IA)*
155.	rain	*barsa (IA)*
156.	snow	-
157.	sky	*sarag (IA)*
158.	rainbow	*in, indrɔdʰanu (IA)*
159.	sun	*siŋgi*
160.	moon	*tʃaɳɖuʔu (IA)*
161.	shadow	*umbul*
162.	star	*ipil*
163.	day	*siŋgi*
164.	daily	*sɔbu (IA), hilaŋ*
165.	week	*hapʰta (IA)*
166.	month	*mɔhina (IA)*
167.	year	*batʃʰar, baras (both IA)*
168.	morning	*sɛtaʔa*
169.	noon	*tikin*
170.	evening	*ajuʔb*
171.	night	*ninda (IA)*
172.	yesterday	*hɔla*
173.	tomorrow	*gapa*
174.	today	*tisiŋ, tihiŋ*
175.	now	*nahaʔã*
176.	when	*ɔka hilaŋ*
177.	time	*bɛra (IA)*
178.	hour	-
179.	one	*miaʔn*
180.	two	*barɛa*
181.	three	*pɛa, pɛnɛŋ*
182.	four	*tʃar (IA)*
183.	five	*pantʃ (IA)*
184.	six	-
185.	seven	-
186.	eight	-
187.	nine	-
188.	ten	-
189.	twenty	-
190.	hundred	-
191.	how much	*cimin*
192.	how many	*cimin*

193.	half	*adʱa (IA)*
194.	altogether	*sɔbu, səb, dʒɛtɛ (all three IA)*
195.	some	*midʒaŋ*
196.	number	-
197.	age	*batʃʰa (IA)*
198.	first time	*pahila kɛtɛ (pahila IA)*
199.	husband	*hɛrɛl*
200.	wife	*laŋgi, lãʄi, lani*
201.	marriage	*ʋiha (IA)*
202.	father	*aba (IA)*
203.	mother	*ajɔ (IA)*
204.	grandfather	*ɔdʒa (IA)*
205.	grandmother	*ai (IA)*
206.	son	*bap (likely IA)*
207.	daughter	*mai (IA)*
208.	child	*hɔpɔn*
209.	young	*dʒaʊan (IA)*
210.	grandchild	*natija (IA)*
211.	elder brother	*dada (IA)*
212.	elder sister	*didi, bai (both IA)*
213.	younger brother	*bʱai (IA)*
214.	younger sister	*bahin (IA)*
215.	sibling	-
216.	sister	-
217.	family	*kuɭum (IA)*
218.	friend	*gati*
219.	quarrel	*ʤʱagaɭa (IA)*
220.	force	*daɭʱi*
221.	dumb	*kuhula*
222.	deaf	*bʱaɛ̃ra (IA)*
223.	blind	*andʱa (IA)*
224.	man	*hɔɭ*
225.	woman	*laŋgi, lãʄi, lani*
226.	person	*hɔɭ*
227.	I	*iŋ, in*
228.	you	*am*
229.	he	*uni*
230.	she	*uni*
231.	we	*alɛ*
232.	you (pl.)	*apɛ*
233.	they (m.)	*unku*
234.	they (f.)	*unku*
235.	self	*ĩja*

236.	other	-
237.	who	-
238.	first name	*ɲumu*
239.	family name	-
240.	letter	*ʧíɻʈʰi (IA)*
241.	voice	-
242.	sound	-
243.	language	-
244.	mind	*mɔn (IA)*
245.	god	-
246.	festival	-
247.	village	*ɖihi (IA, Magadhan)*
248.	town	-
249.	this	-
250.	it	-
251.	that	*hanɛ, han*
252.	which	*ɔka*
253.	what	*cɛnaʔa, cɛkan, cɛa*
234.	why	-
255.	this	-
256.	how	-
257.	here	-
258.	there	-
259.	that place	-
260.	where	-
261.	this way	-
262.	that way	-
263.	away	-
264.	which way	-
265.	place	*ʈʰə (IA)*
266.	left	-
267.	right	-
268.	front	-
269.	back	*dɛja*
270.	inside	*bʱitri (IA)*
271.	out	*bahrirɛ (IA)*
272.	space	-
273.	up	-
274.	down	-

Appendix 2: Two Short Turi Texts

The following two texts were both composed by Ms. Bishakha Mallik and were translated and analysed during our five-day workshop in Ranchi (see Section 1). We are grateful to her for her permission to publish these here.

Text A: A short Turi school song for young children:

(44) *ɛla* *rɛ* *ʧhaʊa=kun,* *sɛn=a=bu* *iskul.* (repeat once)
come.IMP voc child=PL go=FIN=1PL.INCL school
'Come along, children! Let's go to school!'

(45) *iskul=rɛ* *abu* *gɔŋitɔ* *paɽh=a=bu.* (repeat once)
school=LOC 1PL.INCL math learn=FIN=1PL.INCL
'Let's learn math at school!'

(46) *alɛ* *itu-ɔ=a=bu* *lɛkəɛ=a=bu.*
1PL.EXCL learn-MID=FIN=1PL.INCL count=FIN=1PL.INCL
'Let's learn, let's count!'

(47) *miaʔn* *barɛa* *pɛa* *ʧar=gɔɽ* *panʧ=gɔɽ* *lɛkəɛ=a=bu.*
one two three four=COUNT five=COUNT count=FIN=1PL.INCL
'Let's count 1 – 2 – 3 – 4 – 5!'

Text B: The monkeys' hats - translation from the popular story in Hindi used in schools throughout India.

(48) *bandra* *ʈɔpi*
monkey hat
'The monkeys' hats'

(49) *miaʔn* *phɛriʊala* *ʈɔpi* *idi=kɛtɛ* *haɽ* *sɛn-kɛn=ə=ɛ.*
one hawker hat take=CVB market go-PST.MID=FIN=3SG.ANIM
'A hawker took some hats to market (= having taken hats, went to market).'

(50) *sɛ-sɛn* *sɛ-sɛn=tɛ* *thəka-ɛn=ə=ɛ.*
RDP-go RDP-go=SIM become.tired-PST.MID=FIN=3SG.ANIM
'Walking along he became tired.'

(51) *miaʔn* *duba* *pɛndarɛ* *dubʔm-ɛn=ə=ɛ.*
one tree under sit.down-PST.MID=FIN=3SG.ANIM
'He sat down under a tree.'

(52) *dub* *dub=tɛ* *gitiʔ-ɛn=ə=ɛ.*
sit.down sit.down=SIM sleep-PST.MID=FIN=3SG.ANIM
'While sitting there he fell asleep.'

(53) *duba* *tɛŋrɛ* *bandra=kun* *ɲɛl-ɛkɛn=a=ku.*
tree on.top.of monkey=PL see-PST.IPFV.ACT=FIN=3PL
'At the top of the tree the monkeys watched/were watching.'

(54) *phɛriʋala* *gitiʔ-ɛn=ə=ɛ* *mɛnthã* *hiʔc=kɛtɛ* *ʤɛtɛ* *ʈɔpi=kɛ*
hawker sleep-PST.MID=3SG.ANIM then come=CVB all hat=OBJ
pindh=kɛtɛ dubɑ=rɛ *rakaʔm-ɛn=a=ku.*
put.on=CVB tree=LOC climb-PST.MID=FIN=3PL
'The hawker fell asleep then, having come, they [= the monkeys] put on all the hats and climbed up (= in) the tree.'

(55) *phɛriʋala=rɛn* *durum* *bhaŋa-ɛn=a*
hawker=GEN sleep(n.) open(itr.)-PST.MID=FIN
'The hawker woke up (= the hawker's sleep opened).'

(56) *ɲɛl-laʔ=ə=ɛ* *ʤɛ* *miaʔn* *au* *ʈɔpi* *kanɔʔɔ.*
see-PST.PERF.ACT=FIN=3SG.ANIM then one and hat NEG.PRS.COP
'He saw then that there was (= is) not one single hat.'

(57) *mɔn* *dukh=kɛtɛ* *duba* *tɛŋkɛ* *saŋgil-laʔ=ə=ɛ.*
mind(n.) be.sad=CVB tree on.top.of look.upwards-PST.PERF.ACT=FIN=3SG.ANIM
'Feeling sad he looked up at the tree.'

(58) *ɲɛl-laʔ=ə=ɛ* *ʤɛ* *ʤɛtɛ* *ʈɔpi=kɛ* *bandra=kun*
see-PST.PERF.ACT=FIN=3SG.ANIM then all hat=OBJ monkey=PL
pindh-ɑkɑn=ɑ=ku.
put.on-PERF.MID=FIN=3PL
'Then he saw that all the monkeys were wearing (= have put on) all the hats.'

(59) *bandra=kun=kɛ katha-laʔ=ə=ɛ "dan rɛ bandra=kun*
monkey=PL=OBJ say-PST.PERF.ACT=FIN=3SG.ANIM give.IMP(IA) voc monkey=PL
ʈɔpi=kɛ haʈ sɛn=ɛŋ.
hat=OBJ market go=1SG
'He said to the monkeys "Give the hats [to me], oh monkeys! I will go to market.'

(60) *ʈɔpi akriŋ=kɛtɛ ʧhaʊa=kun=rɛn lahiʔɲ=kɛ dana ɛm=ku=ɛŋ".*
hat sell=CVB child=PL=GEN belly=OBJ food give=3PL.OBJ=1SG
'I will sell (= having sold) the hats [and] give the children food".'

(61) *sɔbu katha ajum=kɛtɛ bandra=kun sɔbu ʈɔpi ɛm-tad=i=a=ku*
all story hear=CVB monkey=PL all hat give-PST.ACT=3SG.ANIM.OBJ=FIN=3PL
'After hearing his whole story, the monkeys gave him all the hats.'

(62) *phɛriʊala landa landa=tɛ haʈ sɛn-lɛn=ə=ɛ.*
hawker laugh laugh=SIM market go-PST.PERF.MID=FIN=3SG.ANIM
'The hawker went off happily (= laughing, went) to market.'

Bibliography

Anderson, D.S. (ed.) (2008). *The Munda Languages*. London; New York: Routledge. https://doi.org/10.4324/9781315822433

Anderson, G.D.S. (2015). "Overview of the Munda languages". Jenny, M.; Sidwell, P. (eds), *The Handbook of Austroasiatic Languages* [Grammars and Language Sketches of the World's Languages, Mainland and Insular Southeast Asia]. Leiden; Boston: Brill, 364-414.

Anderson, G.D.S.; Osada, T.; Harrison, K.D. (2008). "Ho and the other Kherwarian Languages". Anderson 2008, 195-255.

Das, N.J. (2020). "History of Origin of the Santals of India". *Journal of Xi'an University of Architecture & Technology*, 12(5), 1222-6. https://xajzkjdx.cn/volume-xii-issue-5-may-2020/

Eberhard, D.M.; Simons, G.F.; Fennig, Ch.D. (eds) (2024). *Ethnologue: Languages of the World. Twenty-Seventh edition*. Dallas (TX): SIL International. http://www.ethnologue.com.

Grierson, G.A. (1906). *Linguistic Survey of India. Vol. IV. Muṇḍā and Dravidian languages* [Reprint]. Delhi; Varanasi; Patna: Motilal Banarsidass.

Hammarström, H.; Forkel, R.; Haspelmath, M.; Bank, S. (eds) (2022). *Glottolog 4.6.* https://glottolog.org/

Ivani, J.K.; Paudyal, N.; Peterson, J. (2021). "A House Divided? Evidence for the East-west Indo-aryan Divide and its Significance for the Study of Northern South Asia". Ivani, J.K.; Peterson, J. (eds), Special issue, *Journal of South Asian Languages and Linguistics*, 7(2), 235-74. https://doi.org/10.1515/jsall-2021-2029

Kobayashi, M.; Murmu, G. (2008). "Kera? Mundari". Anderson 2008, 165-94.

Kobayashi, M.; Murmu, G.; Osada, T. (2003). "Report on a Preliminary Survey of the Dialects of Kherwarian Languages". *Journal of Asian and African Studies*, 66, 331-64.

Kondrak, G. (2000). "A New Algorithm for the Alignment of Phonetic Sequences". *Proceedings of the First Meeting of the North American Chapter of the Association for Computational Linguistics* (ANLP-NAACL 2000) (Seattle, April 2000). Stroudsburg (PA): Association for Computational Linguistics, 288-95. https://aclanthology.org/A00-2038

Munda, R.D. (1968). *Proto-Kherwarian Phonemic System.* Chicago: University of Chicago.

Nichols, J. (1992). *Linguistic Diversity in Space and Time*. Chicago: Chicago University Press.

Nichols, J. (1997). "Modelling Ancient Population Structures and Movement in Linguistics".*Annual Review of Anthropology* 26, 359-84. https://www.jstor.org/stable/2952527

Nichols, J.; Warnow, T. (2008). "Tutorial on Computational Linguistic Phylogeny". *Language and Linguistics Compass*, 2(5), 760-820. https://doi.org/10.1111/j.1749-818X.2008.00082.x

Osada, T. (1991). "Father Ponette's field note on Turi". *Journal of Asian and African Studies,* 42, 175-89.

Paudyal, N.P.; Peterson, J. (2021). "How One Language Became Four: The Impact of Different Contact-Scenarios Between 'Sadani' and the Tribal Languages of Jharkhand". Ivani, J.K.; Peterson, J. (eds), Special issue, *Journal of South Asian Languages and Linguistics*, 7(2), 275-306. https://doi.org/10.1515/jsall-2021-2028

Peterson, J. (2017). "Fitting the Pieces Together. Towards a Linguistic Prehistory of Eastern-Central South Asia (and Beyond)". *Journal of South Asian Languages and Linguistics*, 4(2), 211-57.
https://doi.org/10.1515/jsall-2017-0008

Peterson, J. (2022). "A Sociolinguistic-Typological Approach to the Linguistic Prehistory of South Asia – Two Case Studies". Language Dynamics and Change, 12(2), 224-73.
https://doi.org/10.1163/22105832-bja10018

Peterson, J. (forthcoming). "The Spread of Munda in Prehistoric South Asia – the View from Areal Typology". *Bulletin of Deccan College Postgraduate and Research Institute*, vol. 81.

Peterson, J.; Minz, A.S. (2021). मेरी पहली तुरी चित्रमाला (My First Turi Picture Book). Ranchi: Dr. Shyama Prasad Mukherjee University.

Rau, F.; Sidwel, P. (2019). "The Munda Maritime Hypothesis". *Journal of the Southeast Asian Linguistics Society*, 12(2), 35-57.
http://hdl.handle.net/10524/52454

Sidwell, P. (2015). "Austroasiatic Classification". Jenny, M.; Sidwell, P. (eds), *The Handbook of Austroasiatic Languages* (Grammars and Language Sketches of the World's Languages, Mainland and Insular Southeast Asia). Leiden; Boston: Brill, 144-220.
https://doi.org/10.1163/9789004283572_004

Zide, N.H. (1969). "Munda and Non-Munda Austroasiatic Languages". Sebeok, Th. (ed.), *Current Trends in Linguistics*, vol. 5. The Hague: Mouton, 411-30.

Reviews

e-ISSN 2785-5953

Bhasha
Vol. 3 — Num. 2 — Ottobre 2024

Pritipuspa Mishra, *Language and the Making of Modern India*

Fabiola Nicodemo
Università Ca' Foscari Venezia, Italia

Review of Mishra, P. (2020). *Language and the Making of Modern India. Nationalism and the Vernacular in Colonial Odisha, 1803-1956*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108591263.

The monograph *Language and the Making of Modern India* by Pritipuspa Mishra, divided into six chapters and a postscript, traces a history of the creation of Odisha as a discrete linguistic province between 1866 and 1936.[1] The volume proposes an analysis of the role of the so-called vernacular languages in regional and national politics in both colonial and postcolonial India. This work represents an effort to overturn essentialist notions of weakness and powerlessness around the concept of vernacular by highlighting its political

1 Here I refer specifically to this period of time (1866-1936), as it represents the main focus of analysis of the author. In 1866 the weekly newspaper *Utkal Dipika* (The Lamp of Odisha) was established in response to the mismanagement of a severe famine that had affected the region. It aimed to inform the government about the needs of the people in the Odisha division and also worked towards the development of the Odia language. Furthermore, the year 1936 refers to the formation of Odisha as a linguistically discrete territory. However, in the subtitle of the volume the author mentions the following period of time 1803-1956. 1803 indicates the year of occupation of Odisha by the British East India Company. Lastly, the year 1956 refers to the linguistic reorganisation of Indian provinces, which had already begun in 1936 with the formation of Odisha, and that continued until the 1970s.

Edizioni Ca'Foscari

Submitted 2024-11-13
Published 2024-12-18

Citation Nicodemo, F. (2024). Review of Language and the Making of Modern India by P. Mishra. *Bhasha*, 3(2), 307-314.

DOI 10.30687/bhasha/2785-5953/2024/02/006

use on the part of regional elites. More specifically, the author is able to reveal the hegemonic power of vernacular languages in creating discrete monolingual territories through the inclusion, or rather effacement, of minority groups such as the *ādivāsī* (Indigenous) population. In this framework, Mishra refers to the case of the first province formed on linguistic basis in colonial India: Odisha. Created in 1936, the example of Odisha is particularly functional to the author's argument because of its conspicuous Indigenous presence – almost one-fourth of the population of the proposed province – and the latter's subsequent role in crafting what would be referred to as "Natural Odisha". For instance, for the Odia elite, the '*ādivāsī* element' represented both a fundamental characteristic in differentiating the Odia language and population from the Bengali neighbours and an issue of anxiety as it would qualify them as 'primitive', 'tribal' and 'uncivilised'. If, at first, this process was enabled by the "sublimation" – as Mishra borrows from Freud – of language into an imaginary geographical territory, the later use of myths associated with the Jagannath cult and the Puri pilgrimage allowed the representation of a fundamentally religious, tolerant and inclusivist, yet hierarchical, Odisha. Being sublimation a reversible process, Mishra then illustrates how the discourses around 'Odianess' are constantly challenged and manipulated to suit the needs of the Odia majority and balance regionalist and nationalist efforts through the idea of an Indian citizen 'united in diversity'.

Between Geographical Boundaries and Literary Canons

The first Chapter starts with the description of the nineteenth-century growing imperative by the colonial government to form discrete regions which could be more easily administrated by using only one Indian language. Both the concomitant British debates and colonial policies on juridical and political language, as well as education, considered mother tongues and popular common speech the most effective and ethical choices under the idea of liberal governance. Moreover, in this section, the author elaborates on how the term 'vernacular' in India has been infused with European notions that have conferred upon it a status of powerlessness. Hence, as opposed to the idea of cosmopolitan or translocal languages, in colonial India, the term vernacular came to identify major Indian languages – local and underdeveloped mother tongues, colonial vernaculars which needed to be refined and modernised as part of the broader 'civilisational mission'. One of the author's main arguments is that in the politics of colonial vernacularisation, it was this very process of viewing Indian vernaculars as completely powerless and local that created the conditions for these languages to claim their hegemonic status

as the language of state. In this framework, local debates on boundaries between geographical domains of Indian languages, such as the 1860s-70s debate between Odia and Bengali, had a considerable influence on the development of literature, literary criticism, and education in Odisha. The debate was sparked by a proposal in 1864-65 to replace Odia with Bengali as the language of instruction in schools of the Odisha division because of the lack of appropriate Odia school textbooks and qualified Odia teachers. Bengali intellectuals like Rajendralal Mitra (1822-91) supported the proposal, arguing that Odia was very similar to Bengali and that using Bengali would be more financially practical and beneficial for Odias. The debate highlighted the Odia intelligentsia's concerns about the 'backwardness' – mostly associated with the linguistic '*ādivāsī* element' and what was considered obscene pre-colonial literature –[2] of Odia in relation to Bengali. This sentiment led to efforts – sometimes paradoxically based on the example of the Bengali language and literary tradition – on the part of the Odia elite to produce new Odia textbooks and develop a distinct Odia literary canon more aligned with the image of Odisha they wanted to represent.

In Chapter 2 Mishra continues examining the subsequent literary debates on the formation and politicisation of this new Odia canon and public, which will soon prove to be a crucial step in the establishment of the province in 1936. In this framework, Mishra (2020, 28) claims that "the vernacularity of Odia was established through radical exclusion of the non-elite". In demonstrating this process, the author does not focus on the dilemma between tradition and modernity but rather emphasises its "inaugural nature" in order to present the issue of "timeliness" of literature (Mishra 2020, 78). In this framework, Mishra highlights the role of the literary and political *a* ('spirit of the time') within Odia literature and the anxieties regarding the inadequacy of the pre-colonial canon.

The Politics of Language-Based Odia Social Identity

According to the author, it was during the very first decade of the twentieth century that the politicisation of a social identity based on the Odia language started emerging. For instance, in the third Chapter of the book Mishra delves into the role of the *Utkal Sammillani*

2 The 1890s saw the publication, in all the major newspapers of the Odia-speaking tracts, of a serialised critique of the popular pre-colonial Odia poet Upendra Bhanja (1670-1740). The critics argued that his works consisted of mostly obscene materials, unsuitable to the contemporary *zeitgeist* and needs, such as Odia school textbooks.

(Odisha Conference or Union)[3] within this process. By discussing politics, citizenship, and the relationship of the Odia division with the colonial state, the *Utkal Sammillani* oriented the discourse of the inclusion of all Odia-speaking tracts under a single territory towards arguments for the *political* representation of the Odia people as a unified constituency. Although at first the organisation tried to avoid addressing political concerns directly, the impossibility of discussing the demand of a separate province without engaging with politics, ensured the shift of the organisation to a more openly political approach. Besides the *Utkal Sammillani* and the implementation of the colonial franchise between 1918-19, aiming to increase popular participation in governance, also the Indian National Congress radically changed its attitude towards regional politics. Hence, from eluding regionalist issues as they would supposedly undermine the unity of the Country, the Congress, as Mishra argues, started viewing regional issues as a fundamental part of the politicisation of the Indian masses through the creation of the liberal Indian citizen "united in diversity".

Histories of a "Natural Odisha"

As examined in Chapter 4, by the late 1910s, Odisha as a proposed province and newly imagined territorial entity (Anderson 1983) started to be referred to as "Natural Odisha". In order to present the latter as a historical reality, the Odisha advocates sustained this new ontology through the writing of histories of an 'ancient', i.e., pre-colonial, Odisha. Within the framework of nationalist pluralistic rhetoric, this process tended to associate each Indian province with peculiar underlying qualities. In that context, Odisha was conceptualised as an intrinsically religious and tolerant land, embodiment of a supposedly inclusivistic and pluralistic attitude found in the Jagannath cult and the Puri pilgrimage. Odisha was thus perceived as an entity able to embrace lower-caste people, tribal groups, and even Muslims, yet maintaining a hierarchical distance between minority groups and the Odia-speaking upper-caste. Therefore, this assumed quality of Odisha allowed it to be represented both as a local and cosmopolitan space and to assume great interest for regional and national politics.

Chapter 5 examines the consequences caused by the "Natural Odisha" paradigm, among which the conceptualisation of its history through the appropriation of the *ādivāsī* pasts and its idealisation

3 Organization established in 1903 and aimed at the inclusion of all Odia-speaking areas under a single province. The *Sammillani* rapidly became the primary pan-Odia platform for presenting Odia concerns to the colonial authorities.

as a fundamentally homogenous and ancient homeland. The establishment of the Orissa Boundary Commission in 1931 – aimed at the delineation of the boundaries of the new province – resulted in the emergence of the intrinsic contradictions regarding Odisha as a territorial and imagined entity. One of the main aspects of anxiety concerned the presence of the non-Odia-speaking *ādivāsī* communities, which represented almost one-fourth of the population of the proposed province. By examining several memoranda,[4] the author is able to outline the major justifications for the incorporation of *ādivāsī* communities into Odisha. Unsurprisingly, the rhetoric of these documents was embedded with claims on the nature of Odisha as an inclusivist, tolerant and pluralistic entity. Not to mention the employment of a fundamentally paternalistic narrative of Odia-speaking people as benign civilizers of 'tribal groups'. Despite inherent contradictions regarding the rhetoric of the memoranda, this approach proved effective, as it allowed for the incorporation of diverse regional populations into the emerging Odia political identity and territory – albeit without granting them equal social status.

Approaches to the Issue of Multilingualism in Modern India

Borrowing from Bhimrao Ramji Ambedkar's claim, the last Chapter is titled "Genius of India is to Divide" and is dedicated to the analysis of linguistic difference in the making of modern India. By analysing the approaches of Mohandas Karamchand Gandhi (1822-85), Jawaharlal Nehru (1889-1964), and Bhimrao Ramji Ambedkar (1891-1956) towards the issue of multilingualism in India, the Chapter reveals the ideologies and anxieties in the imagination of a united, yet multilingual modern India. Besides Gandhi's perspective – a non-exclusive affect-based argument towards the mother tongue – more interestingly Nehru's and Ambedkar's viewpoints uncover the inherent contradictions and tensions between regionalist and nationalist efforts in the making of modern India. Although based on Gandhi's promotion of multilingualism, Nehru and the wider Indian National Congress, attempted to present a compromise between the centrality of linguistic identity in liberal governance and an extremely limited conceptualisation of linguistic difference in India – with the acknowledgement of only fourteen major Indian languages. On the other side, Ambedkar, as a representative of the non-elite lower-caste population, feared that the division of India into extensive language-based regions would excessively increase the institutional power in the

4 Documents submitted to the Odisha Boundary Commission in 1931 by leading advocates for the formation of a separate province of Odisha.

hands of the regional elites. Nevertheless, although aware of this complex issue, Ambedkar was unable to provide a sustained critique on the formation of linguistic provinces from the point of view of the *ādivāsī* communities – that in the proposed Odisha territory did not represent, at least quantitatively, a minority. As regional languages became central to defining representation and identity in India, the process simultaneously marginalised Indigenous peoples and limited other political alternatives. According to Mishra, this exclusion became an integral part of how the modern Indian nation was conceptualised. Chapter 6 ends with the mention and analysis of speeches by Jaipal Singh (1903-70), leader of the *Ādivāsī Mahasabha* (Indigenous Great Assembly) and the movement for the formation of the *ādivāsī* majority province of Jharkhand. In the postscript, the author briefly discusses contemporary *ādivāsī* activism in order to show how Indigenous communities are still struggling within this framework of linguistic and political representation.

Methodology and Theoretical Framework

Mishra's work on the creation of Odisha as the first linguistically organised province in India is supported mainly through the textual analysis of a wide range of sources – Odia newspapers and periodicals (among which the *Utkal Dipika*), pre-colonial Odia literature, Odia literary critique, books on the history of Odisha, memoranda submitted to the Odisha Boundary Commission, speeches of prominent nationalist leaders (Gandhi, Nehru and Ambedkar) and *ādivāsī* activists (Jaipal Singh). However, regarding the methodology employed by the author, Mishra proposes a fundamentally multidisciplinary approach. Based on literary criticism (study of the creation and politicisation of an imaginary but functional Odia literary canon and public through the exclusion of the non-elite), the "institutional life of language" (language politics and rhetoric), and the spatial category of territory, Mishra's work employs concepts such as "sublimation" and "heterotopia", respectively borrowed from Freudian psychoanalysis and Michael Foucault's social theory. Considering the first term, sublimation denotes "the process of turning socially unacceptable hidden desires into more visible socially productive actions" (Mishra 2020, 10). The latter, often being a source of discontent is, thus, potentially reversible. In Mishra's use, sublimation represents the shift from defining a community on an exclusive linguistic basis to a territorial one, allowing the inclusion of non-Odia speaking people (such as the *ādivāsī*) within the imagined Odia community. Importantly, Mishra questions the givenness of language as a category of analysis in defining territorial domain by underlining that, when political circumstances change, the process of sublimation can potentially be

reversed in order to better align with the *zeitgeist* and needs of the elite. On the other side, the term heterotopia is mentioned by the author to define Odisha as a place where exception becomes the norm. Within the framework of the "Natural Odisha" paradigm, and the subsequent representation of the province as a fundamentally religious and inclusivist land, Mishra applies Foucault's concept of heterotopia to describe how Puri, and by extension Odisha, functions as a space that is both exceptional and representative. For instance, especially during the annual *Ratha Yātrā* (Chariot journey or pilgrimage) festival, Puri represents a site of exception to caste and religious exclusion. Through the concept of heterotopia, this exceptional event becomes the norm, the representation of all modern Odisha as a place where religious inclusivity is normalised, even as it maintains social and cultural distinctions.

Final Notes

Language and the Making of Modern India by Pritipuspa Mishra represents a valuable contribution to the field of linguistic politics within the regional and national histories of modern India. The main strength of the book lies in the interdisciplinary approach employed by the author in analysing a wide range of textual sources and presenting her main argument – the very process of viewing Indian vernaculars as completely powerless and local created the conditions for these languages to claim their hegemonic status as the language of state. Hence, the relevance of this volume lies especially in its ability to overturn essentialist notions of weakness and powerlessness around the concept of vernacular by highlighting its political use on the part of regional elites. In this context, Mishra's work is able to offer an understanding of the often contradictory yet fundamental relationship between regionalist and nationalist ideologies in the making of modern India, but also to reveal how the institutionalisation of language-based states and *ādivāsī* incorporation represents a contemporary issue that Indigenous communities are still dealing with throughout the Country. However, Mishra's analysis focuses on a rather limited period of time (1866-1936), potentially missing more recent developments in linguistic politics – especially the aftermaths of Indian independence (1947) which she does not cover extensively. Besides that, while the study addresses the *ādivāsī* issue, it acknowledges the challenges in providing a sustained critique from the Indigenous perspective – as she writes on Ambedkar in Chapter 6 – a point that perhaps not even Mishra's work is able to avoid completely.

Semestral journal
Department of Asian
and North African Studies

Università
Ca'Foscari
Venezia