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), 191-210.

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 𐨟 <t> and dental aspirated 𐨠 <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 𑀣 <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 *niganṭha* (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', *ǰïṣṭ* '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.

e-ISSN 2785-5953

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 <ṭ>, 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 <ṭh>, 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 <ṭh> 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 𐨷 is <a-𐨷>. 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 𐨷 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 <ṭh> for word-initial etymological *sth* (Melzer 2020, 88, 90, 98) – this would lead us to conclude that 𐨷 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 𐨷. The example word for 𐨶 is the unclear word <𐨶 𐨶>, 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 𐨷 is so rare and 𐨶 also appears in words derived from Skt. forms with *ṭṭ* and *ṇṭ*, which likely never went through an alveolar stage (e.g. <pa-𐨶> 'silk' ~ Skt. *paṭṭa-*, <(graḥma)-ka-𐨶-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-ƶ> 'stick, firewood' ~ Skt. *kāṣṭha-*, cf. Baums, Glass 2002), it seems best to retain the transcriptions 𐨷 = <ṭ́> and 𐨶 = <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>	/ṯh/ [t̪ʰ]/t̪t̪h/ [t̪ʰː]	> /th-/> /-ṭṭh-/
(ṭ)ṭ	𐨶	<ṭ>	/ʈ/ [ʈ]/ʈʈ/ [ʈː]	
rt	𐨺	<ṭ́>	(/ṯ/ [t̪])*/t̪t̪/ [t̪ː]	> /ʈʈ/

* 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.