Bulletin of the School of Oriental and African Studies (2026), 1–27
doi:10.1017/S0041977X26101025

ARTICLE

The development of Sanskrit velars into Gāndhārī

Julien Baley[1] and Alan Avdagic[2]

[1]Department of East Asian Languages and Cultures, SOAS University of London, London, UK and
[2]Chair of Comparative Philology, Julius-Maximilians-Universität Würzburg, Würzburg, Germany
Corresponding author: Julien Baley; Email: julien.baley@gmail.com

Abstract

As a language of religious and administrative importance in the early centuries of the common era, Gāndhārī came to be a donor into its neighbouring languages, such as Tocharian and Chinese. Consequently, advances in Gāndhārī historical phonology can help us discover new loanwords, refine our understanding of the historical phonology of its neighbouring languages, and eventually improve our understanding of the relationship between the communities that spoke those languages. One unresolved problem in the study of Gāndhārī phonology is the development of Sanskrit unaspirated velar stops: the relative paucity of data and variation in spelling have left previous researchers hesitant regarding the developments of those stops and their phonetic realization. In the present article, we take a bird's eye view and analyse the development of these velars across the whole edited corpus; our main contribution is the discovery of the phonetic environment conditioning the development of /k/ and /g/, thereby fully explaining the seemingly chaotic spelling observed in previous publications.

Keywords: Gandhari phonology; Sanskrit; Middle Indo-Aryan languages

1. Introduction

Gāndhārī, a Middle Indo-Aryan language spoken between the third century BCE and the fifth century CE and written in the Kharoṣṭhī script, has come to us in the form of inscriptions on stone or metal (Salomon 1998: 74–9), on coins (Salomon 1998: 209–18), as well as a range of manuscripts written on wood (Boyer, Rapson, and Senart 1920; Burrow 1937, 1940), birch bark, or palm leaf (Brough 1962).

Spoken in an area at the crossroads of several other cultures, Gāndhārī came to play an important role in the region as an administrative and religious language, and neighbouring languages came to borrow Gāndhārī words, such as Tocharian B *anās*/Tocharian A *ānās* "miserable" (Gdh. *anasa* < Skt. *anātha* "without protector", showing the regular development th > s/V _ V) (Carling and Pinault 2023) or Chinese 曇摩 /dəm.mɑ/ (Gdh. *dhaṃma*) (Coblin 1983: 248).

Beyond its administrative role, Gāndhārī is more importantly the language of the oldest extant Buddhist texts in any language (Salomon 1999: 9), and the language is also thought to be the language from which were translated the first Chinese Buddhist texts in the second half of the second century CE, by the translation teams around Ān Shìgāo 安世高 (fl. 148–170) (Zacchetti 2019: 630) and Lokakṣema (Zhī Lóujiāchèn 支婁迦讖, fl. 147–189) (Harrison 2019: 700). The evidence for Gāndhārī being the source language of these earliest

Chinese Buddhist texts is the use of phonetic transcriptions for technical Buddhist terms (Baley, Hill, and Caldwell 2023) that are better explained if the phonology of Gāndhārī is assumed.[1]

Because of its central role in the study of early Buddhism – both as the language of the earliest extant Buddhist texts and as the source language of the earliest Chinese Buddhist translations – the study of Gāndhārī's phonology – both synchronic and diachronic – plays an essential role in helping us understand these early texts and their transmission. Since the early forays into Gāndhārī studies, a large body of academic literature has dealt with its phonology and its development from Sanskrit, and a good introduction to a corpus-wide analysis can be found in Baums (2019).

One topic that has so far not received sufficient dedicated attention is the development of Sanskrit velar stops: beyond the uncontroversial treatment of aspirated stops (cf. for instance Brough 1962: 93, or Baums 2009: 143), the development of unaspirated velars is less well understood. This is in part due to the deficient nature of the Kharoṣṭhī script, but also because of the nature of the corpus: not only do different documents represent different periods of the development of the language (Fussman 1989: 455–65), but there is also a layering of Gāndhārī phonology with more Sanskritized elements (Salomon 2001: 245), resulting in individual manuscripts sometimes appearing inconsistent.[2] So far, the field has not reached a consensus, some considering that unaspirated velar stops have evolved into a velar fricative (usually the voiced [ɣ] because it is spelled with a modified <g>, transcribed g̱) (Burrow 1937: 6) while others think this spelling only reflected scribal idiosyncrasies, the velars supposedly having entirely disappeared from the language or remaining as a palatal approximant [j] (whether written <y> or absent from the spelling) (Baums 2009: 140).

In the rest of this article, we propose to solve this question of the development of Sanskrit velars into Gāndhārī. For completeness, we also include the uncontroversial development of aspirated velar stops. The structure of the article is as follows: first, we introduce our methodology, the decisions we have made, and the tools we have used. Then, we focus on the development of velars in different environments: word-initial, word-final, on either side of another consonant, and finally in intervocalic position. Since most of the difficulty lies in the intervocalic context, this section is itself divided into the analysis of aspirated stops on the one hand, and a longer analysis of the unaspirated ones on the other. In the process, we show that the development of unaspirated velar stops was clear and that the spelling practices found throughout the corpus reflect actual phonetic phenomena; the corpus itself is analysed in a diachronic fashion,[3] giving hints into the dating of the sound changes pertaining to velars. Finally, we summarize our discoveries and provide further phonetic evidence in other languages.

[1] Cf. Skt. *samādhi* > Gdh. *samasi* > Late Hàn Chinese (LHC) 三昧 /sɑm.məs/ (cf. P samādhi) with the Gāndhārī lenition of Skt. *dh* to *s* (Baums 2009: 144). Loanwords into Chinese also reflect Gāndhārī's preservation of the three Sanskrit sibilants: Skt. *śāriputra* > Gdh. *śariputra* > LHC 舍利弗 /śaᶜ.liᶜ.put/ (cf. P sāriputta), Skt. *suvarṇa* > Gdh. *suvaṃna* > LHC 羞桓 /su.ɣɑn/ (cf. P. *suvaṇṇa*), and Skt. *kāṣāya* > Gdh. *kaṣaya* > LHC 袈裟 /ka.ṣai/ (cf. P. *kāsāya*, *kāsāva*). Additionally see Skt. *śramaṇa* > Gdh. *ṣamana* > LHC 沙門 /ṣai.mən/ (cf. P. *samaṇa*) for the Gāndhārī sound change of Skt. sibilant + r to ṣ (Brough 1962: 102), also Skt. *akaniṣṭha* > LHC 阿迦貳吒 /ʔɑ.ka.ńis.tra/ for the preservation of pre-consonantal s-clusters in the donor language), and finally Skt. *rājā* > Gdh. *raya* > LHC 羅耶 /lɑi.ja/ (cf. P *rājā*) for the intervocalic lenition of palatals (Baums 2009: 140). Boucher (1998) considers it impossible to prove which was the source language, but in light of the examples above, we choose to follow the general consensus and take Gāndhārī as the source of these texts as the currently best available hypothesis.

[2] Examples abound, such as Skt. *bhagavant* "blessed" – which will be discussed later – occurring as *bhaga̱va* and *bhayava* in MSŪ^{B}.

[3] No attempt was made to discuss the corpus from a geographic angle.

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

[1]dhaṃma

(Skt. *dharman, dharma,* P *dhamma*) [dʰəmːə] *m.* (**1**) teaching; (**2**) law; (**3**) duty.

sg. dir. dharma, dhamu, dhamo, dhama, dharmo, dhramaṃ, dhaṃmo, dharmu, (*dhaṃmaṃ), dhaṃ(*ma), dhamra, dharme, dhraṃmaṃ, dhramo, dhrarma, *instr.* dharmena, dhameṇa, dharmeṇa, dhaṃmeṇa, *gen.* dhamasa, dhramasa, dhaṃmaṇa, dhaṃmasa, dha(*r)maṣa, *loc.* dhame, dharmi, dharme, dhrame, dhaṃmo, dhamami, (*dhami), dharma, dharmaṃmi, dharmo, *sg. abl.* dharmade, *pl. dir.* dharma, dhama, (*dha)me, (*dha)r(*ma), *instr.* dhamehi, dharmehi, darmehi, *gen.* dharmaṇa, dhamaṇa, dhama(*ṇa), dharmaṇaṃ, dharmana, *loc.* dharmeṣu, dhaṃmeṣo. ■ *Other:* dharma.

SHĀH IV 9 **dhrame**, 10 **dhramaṃ**; XII 6 **dhramo**, 10 **dhramasa**; XIII 10 **dhramaṃ**; MĀN IV 16 **dhrame**, 17 **dhramaṃ**; XII 6 **dhramaṃ**, 9 **dhramasa**; XIII 11 **dhraṃmaṃ**; ANAV^{L} 46 **dharmu**; AV^{L3} v3 **dharma**; AV^{L4} v226 **dharma**; KHVS 14 **dhaṃm(*o)**, 24 **(*dhaṃmaṃ)**, 29 **dhaṃmeṣo**; NIRD^{L2} aAa2 **dhama(*ṇa)**; NIRD^{L2} r22 **(*dhama)**, 27 **(*dhama)**, 29 **dhama**, 31 **(*dhama)** [2×], 35 **dhama**, 36 **dhama**, 108 **dhama**; CM^{L2} r44

Figure 1. Entry in Baums and Glass (2002) for the word "dhaṃma".

2. Methodology

Below, we briefly present the approach we have taken to build a dataset, annotate it, and analyse it.

2.1. *Corpus dataset*

Our study uses the online Gāndhārī dictionary found at Baums and Glass (2002). Baums and Glass rely on previously published text editions of theirs and others, and for each attested word form found in a Gāndhārī document (inscription, manuscript, coins, tablets, etc.), they list the position in which it is found in the document, the corresponding standardized Gāndhārī lemma, nominal or verbal inflection information (number, gender, case, etc.), and the corresponding Sanskrit and Pali words.[4] See Figure 1 for an example.

The data can be queried either through the standardized lemma, the specific attested form, the Sanskrit equivalent, or the English translation. For each of these modalities, regular expressions can be used, so that one can for instance search for all words having a <k> in Sanskrit and study their Gāndhārī outcomes.

As the focus of our study is the development of Sanskrit velar stops, we have retrieved all entries where a Sanskrit equivalent has been identified and where that Sanskrit equivalent contains a *k* or *g*, which naturally includes the aspirated series with *kh* and *gh*. These data have then been arranged in a table, with attested word form per line, along with its standardized Gāndhārī lemma, Sanskrit equivalent, and document of origin. This dataset contains some 5,000 entries.

2.2. *Dataset annotation*

To analyse the development of Sanskrit *k, g, kh*, and *gh* across the corpus, we annotate the outcome of that velar in each attestation. For instance, Skt. *dāraka-* "child" is attested in ANAV^{L} as *daraghasa* and in AV^{L6} as *daraga*, which we annotate respectively as *gh* and *g*, while Skt. *adhikam* is attested in SC2 as *a̱sio*, for which we annotate the outcome of the *k* as ∅ (i.e. an absence of character). The set of possible outcomes for the velar stops in our annotation scheme is {*k, k̲, kh, g, g̲, gh, h, y,* ∅}, where *k̲* and *g̲* transcribe the Kharoṣṭhī characters for *k*

[4] Other languages are also listed when relevant, such as Buddhist Hybrid Sanskrit, Chinese, or Tocharian.

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

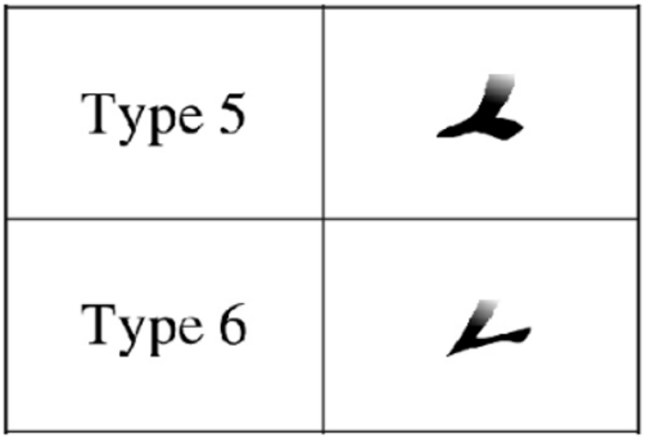

Figure 2. The footmark types found (amongst other characters) under *g* to denote *g̱*. Image taken from Glass's classification (Glass 2000: 56).

and *g*, to which the scribe had added one of a variety of foot marks, as discussed in Glass (2000: 56) and shown in Figure 2.

2.3. Presentation of the results

With the whole dataset annotated, we can then turn to analysing trends in the outcome of the velar stops in specific environments. In our tables, we use the document code names as found at *gandhari.org*, such as AV^{L1} for the Avadānas (BL 1.2). As the whole corpus available at *gandhari.org* contains dozens of documents, many of which are very short and might only contain a very small number of words of interest (i.e. derived from a Sanskrit ancestor having a velar stop), we choose to restrict our presentation of the outcome counts to the 30 documents containing the most words of interest. In the process, we also decided to group together the various families of documents indicated with a number suffix: for instance, for the Avadānas texts found in BL1.2, BL2, BL3A, BL4.2, BL12+14.2, BL16+25.2, BL21.2 and listed with the codes AV^{L1} through AV^{L7}, we group them under the heading AV^{L}. We apply this grouping approach to all documents having a number suffix,[5] except for the items using CKM (manuscripts), CKI (inscriptions), and CKD (documents) for which the code represents an otherwise unclassified item. We have found this approach to lead to data that is more legible and, thanks to the coherence of the groupings, better trends can be highlighted.

3. Non-intervocalic velars

Velar stops that are not in intervocalic position behave much more regularly than intervocalic ones and are generally uncontroversial. For completeness, we treat them first in the four possible contexts: word-initial, word-final, following or preceding a consonant.

3.1. Word-initial velars

In word-initial position, the four velar stops are nearly always preserved across the whole corpus, as has been widely noted. There are two main exceptions: first, the so-called "scribe

[5] This includes the Roman numbers used for the Shāhbāzgarhi Rock Edicts (SHĀH) and the Mansehra Rocks (MĀN) I through XIV.

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

1" of the EĀ^{L} (Allon 2001: 68), DHP-G^{L} (Lenz 2003: 41), and ANAV^{L} (Salomon 2008: 96) who never writes *g* in any context and instead writes *gh*. We consider this an idiosyncrasy and not indicative of a phonological difference, as seems to be implied by the three authors dealing with this scribe's texts above.

Second exception, Sanskrit word-initial *gh* is predominantly found as *g* in Gāndhārī, except in words inherited from Skt. *ghr-* or *ghṛ-*, which preserved the *gh-* in most attestations. This suggests a regular de-aspiration of the voiced aspirated velar stop, at least in initial position, that would have happened before the earliest extant Gāndhārī texts.[6]

Beyond the regular pattern of preservation of *k-*, *kh-*, and *g-*, the de-aspiration of *gh-*, and the scribal idiosyncrasies, we find a very small number of exceptions:

- Skt. *kula* "family, herd" is found as *khulana* in a few texts which otherwise show initial *k* reflexes.
- *kim* loses its initial velar and appears as *iṃ* in several texts, always in the pair "*ma iṃ*" or "*na iṃ*". This likely occurred because the word is an enclitic – the *k* behaves as if it were in intervocalic position.
- A single instance of Skt. *Kuṣāṇa-vaṃśa-saṃvardhaka* > Gdh. *guṣaṇavaśasaṃvardhaka* in CKI 149.
- Initial *k* appears as *ḵ* in four attestations in the SĀ^{S} documents: Skt. *kṛta* > *ḵiḍa* (SĀ^{S5}), Skt. *karaṇīya* > *ḵaraṇia* and Skt. *kim* > gen.sg. *ḵiṣa* (SĀ^{S6}), and Skt. *karaṇīya* > *ḵaraṇao* (SĀ^{S7}).
- Skt. *khalu* "indeed" is commonly eroded to shorter forms, some of which do not have *kh: ho, kho, o, khu, hu, gu, kha, u, ku, khalu*. This is likely due to *khalu* being an unstressed enclitic particle.

These exceptions are so few that they should be of no concern to us regarding the regularity of the rules established above, and summarized below:

$k > k \,/\, \#\_$

$k^h > k^h \,/\, \#\_$

$g > g \,/\, \#\_$

$$g^h > \begin{cases} g^h / \#\_r \\ g / \#\_ \end{cases}$$

3.2. *Word-final velars*

To our knowledge, word-final velar stops have not been discussed previously, probably due to their rarity. Of the four velar stops, only *-k* is found in Sanskrit, in a handful of adverbs and connecting words: *samyak* "properly", *yādṛk* "like", *tiryak* "across", *dhik* "fie (interjection)", and *viṣvak* "in all directions". In all of these, *k* disappears from the spelling in Gāndhārī, and we assume that this reflects the phonetic reality of its disappearance:

$k > \varnothing \,/\, \_\,\#$

3.3. *Velars following other consonants*

As the conventional wisdom in studies of Gāndhārī phonology is that most consonants lenite in intervocalic position (Baums 2019: 4), it is interesting to study the cases where the consonant is not in intervocalic position, so as to establish a baseline for the development of individual consonants. In this section, we study the outcome of velar stops preceded by a consonant.

[6] We believe this phenomenon has not been previously mentioned.

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

k is found after the following consonants: *l, r, s, t, ṃ, ṅ, ṣ*, and *ṛ*:[7]

- In *rk*, the cluster is preserved, although the spelling might suggest sporadic metathesis of the *r*. In the cluster *tk*, the *t* is dropped and *k* is preserved; this has usually been interpreted (e.g. Allon 2001: 89) as a case of regressive assimilation resulting in the gemination of *k*. Assuming this assimilation occurred, whether gemination was retained in Gāndhārī by the time of our corpus is uncertain. In *lk*, the preceding consonant is dropped and *k* is preserved, possibly also hinting at a geminated *k*, as found in other Middle Indo-Aryan languages.
- The cluster *sk* regularly yields *kh* (with *s* dropped), with the exception of Skt. *skandha* "factor" and its compounds, which all yield *k*.
- The cluster *ṣk* similarly drops the s, but the outcome is split equally between *k* and *kh*, and some documents show both outcomes. We have currently no explanation for this fluctuation, but perhaps dating or dialectal information could help understand this behaviour.
- *ṃk* yields *g* in all texts, and while the nasal is usually not written, Glass (2007: 106) comments that the lack of lenition is evidence of its preservation. This can be seen in the texts that normally preserve *k* (DHP^{K}) or yield *y* (SĀ^{S}) in intervocalic position, as these texts also show *ṃk* > *g*, which confirms that the nasal must have been preserved and caused the voicing of the velar stop.
- *ṅk* yields *g* in all texts. As with *ṃk*, we take this as evidence that the nasal was preserved and caused the voicing.
- There are few examples of cluster Skt. *ṛk*, whose outcome is similar to *k* in intervocalic position and suggest that the sound change ṛ > ri happened before the various changes happening to velars in intervocalic position and described further below.

g is found after the following consonants: *d, l, r, ḍ, ṃ, ṅ, ṛ*:

- As for *k* clusters, *rg* is preserved as a cluster. As for *tk* above, *dg* yielded *g*, interpreted as a geminate (Salomon 2000: 88). The outcome of *ḍg* as *rg* suggests that the development ḍ > ɽ – i.e. of a stop into a flap – happened earlier than the time the stop+velar clusters (*tk, ḍg*) assimilated to their stop.[8] The outcome of *lg* is *g*, perhaps also hinting at a geminated *g*.
- The clusters *ṃg* and *ṅg*, like *ṃk* and *ṅk*, preserve the *g* even in documents that otherwise show a lenition of intervocalic *g* to /j/ or ∅, bringing further evidence that – regardless of it being represented in spelling or not – the preceding nasal was preserved.
- The cluster *ṛg* is mainly found in the Skt. *word mṛga* "deer" and – like *ṛk* – also yields outcomes that suggest an early development ṛ > ri.

kh is found after the following consonants: *r, s, ḥ, ṃ, ṅ*. As opposed to *k* and *kh* behaving very differently in intervocalic position as will be seen later, *kh* behaves similarly to *k* when following a consonant:

- The cluster *ḥkh* yields *kh*, the *r* of *rkh* is preserved as in the development of *rk* and *rg*.
- The Skt. cluster *skh* loses its *s* and yields *kh*, like *sk*. The complete absence of any attested *sk* or *skh* in Gāndhārī tells us this change must have happened very early, but

[7] Although *ṛ* is traditionally classified as a vowel, we list it here due to its phonetic status as a syllabic consonant, and we will see later that this analysis is correct. The outcome of such clusters can reveal the timing of individual sound changes.

[8] Salomon (1999: 77) argues that *rg* is a spelling device for the gemination of *g*, but we find this suggestion unconvincing considering the absence of such spelling devices in other gemination contexts.

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

precludes any certainty regarding the development of these individual clusters. In the case of *sk*, we follow the suggestion of a reviewer of an early draft of this article, and propose the sequence sk > ^{h}k > k^{h}.
- *ṃkh* and *ṅkh* both yield *gh*, as *ṃk* and *ṅk* both yield *g*, adding further evidence of the retention of the nasal, at least until the stage where intervocalic unaspirated velar stops lenited.

gh is rather rare in clusters, found after the following consonants: *d, r, ṃ*. The development of these clusters is generally coherent with the previous sections:

- *dgh* drops its stop and yields *gh*. As for *tk* and *dg*, this has been interpreted as a sign of gemination.
- *ṃgh* is found in Skt. *saṃgha* "community" and yields *gh*, with the same conclusion that the nasal helped preserve the stop, since – as will be discussed later – *gh* in intervocalic position regularly yields *h*.
- *rgh* is found in Skt. *dīrgha*; all texts spell a metathesis of *r* as *dri-*, and the velar is variously spelled as *g, gh, kh*. We have no explanation to offer for the inconsistency in spelling of this word, found nine times throughout the whole corpus.

The changes can be summarized as such:
k^{h} > g^{ɦ} / [+nasal] _
k > g / [+nasal] _
sk^{h} > k^{h}
sk > ^{h}k > k^{h}

3.4. *Velars preceding other consonants*

A large part of the words showing a velar stop followed by another consonant are compound words, the velar stop being the last phoneme of the first part of the compound, and the other consonant the initial phoneme of the second part. In such cases, things behave as they usually do in compounds, i.e. the velar stop behaves as if it were at the end of a word and disappears. In the paragraphs below, we leave such cases out of the discussion. Overall, the scarcity of the material means that our present analysis should be considered tentative.

k is found before the following consonants: *l, r, t, v, ṛ*, and *ṣ*.

- *kl* is found in two words and their compounds: *kleśa* "defilement" always yields *kileśa*, while *śukla* "bright/light" always leads *śukra*. No specific behaviour can be inferred from these two sole examples.
- *kv* preserves its *v* in initial position and loses it otherwise.
- *kt* usually loses its *k*, interpreted as a gemination of *t*, although some *kt* are still found in spelling.
- *kr* is preserved very consistently.
- *kṛ* has two different outcomes, *kri* and *ki*. The outcome seems to be decided on a document basis; we can exclude the possible inability of some scribes to spell a *kr* cluster as the documents writing *ki* do have other *kr* clusters (cf. previous item). This leaves us with the tentative explanation that dialectal differences are at stake. If that is the case, and since *kr* itself is consistently preserved throughout the corpus, the route kṛ > kri > ki has to be excluded. Instead, since *ṛ* occasionally yields *ir* instead of *ri*, we propose kṛ > kir > ki/_ C, on the ground that r > ∅/_ C is a regular sound law in Gāndhārī (Baums 2019: 6).

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

- Skt. *kṣ* usually yields Gdh. *kṣ*, which is written with a specific ligature in Gāndhārī whose transcription as *kṣ* has been attributed on the ground of its usual correspondence with Sanskrit *kṣ*. The actual pronunciation of this ligature is still debated, the debate residing in whether the cluster is pronounced with an initial /ʈ/ (Bailey 1946: 774) or initial /k/ (Schoubben 2022: 150), a debate we do not aim to resolve here. What is interesting, however, and may contribute to the debate, is that beyond the more common outcome *kṣ*, we also find *kh* in a variety of texts and in a variety of environments,[9] a phenomenon we cannot yet explain but which would tend to favour a /k/ initial in the phonetic realization of the <kṣ> graph, at least in some contexts.

g is found before the following consonants: *n, r, ṛ*.

- *gn: agni* "fire" and its compounds all yield *g*; the fact that *g* is attested in MNRS (and not *ġ*), in BC (instead of *y*), and in K^{hvs} (instead of ∅), suggests that /n/ was pronounced at least until after the lenition of intervocalic *g*.
- *gr*: similar to *gn*, nearly always yields either *gr* or *g* (except in a few instances of *kr*), and *g* in $E\bar{A}^{L}$ suggests that the /r/ persisted at least until after the time of the lenition of intervocalic *g*, and was in all likelihood simply not written.
- *gṛ*: found in the various conjugations of the verb Skt. √*gr(b)h-* "take". Our interpretation of the evidence is here is similar to *kṛ* presented above.

The changes relevant to velar stops can be summarized as such:

$$\mathrm{k} > \varnothing \,/\, \_\,\mathrm{t}$$

$$\mathrm{k}\upsilon \;>\; \begin{cases} \mathrm{k}\upsilon/\#\_ \\ \mathrm{k} \end{cases}$$

4. Intervocalic velars

As the cases where velar stops are found with a consonant or a word boundary on either side have been covered, we now have a firm baseline to evaluate the specific development of velar stops in intervocalic position.

4.1. Velars at compound boundary

Previous literature has noted that intervocalic velar stops, and any obstruent in general, behaved differently when a compound boundary occurred on either side of the velar (Konow 1929: xcviii), effectively blocking the lenition.[10] This makes sense both if the compound was created after the period when the intervocalic lenition sound change was operating, and if the occurrences of the morpheme in unchanged positions provided sufficient analogical pressure for its preservation in other positions.[11]

In our dataset, we annotated each entry for which the velar is not word-initial or word-final (incl. word-final suffixes like *-ka*), marking whether a compound boundary was present before the velar. This affords us analysing the behaviour of intervocalic velars both in compound boundary environments and outside.

[9] For instance, CKD 540 *khoriṃta* < Skt. *kṣurati* "shaved", or $NIRD^{L2}$ *aśekhada* < Skt. *aśaikṣatā* "no longer being in training", perhaps under the influence of Pali *asekkhatā*, but the preservation of *ś* excludes a direct borrowing.

[10] Glass (2000: 82) notes, however, that some compounds might be perceived as single words, in which case lenition would operate normally, such as in *khargaviṣaṇagapo* < Skt. *khaḍgaviṣāṇa-kalpa* "like a rhinoceros".

[11] See Hill (2014) for a discussion of grammatically conditioned sound changes.

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

Overall, we reach the same conclusion as previous scholars: compound boundary blocks lenition and the velar stop is preserved, i.e. the second part of the compound behaves as if it were a separate word and the velar was in word-initial position. It is worth noting here that verbal and nominal prefixes (e.g. *abhi-* "towards", *adhi-* "over, besides", *ā-* "near, towards", *anu-* "after, near", *pra-* "before, forward", *prati-* "towards", *upa-* "towards, near to") themselves do not have this blocking effect, and we find lenition of the velar after them.

4.2. *Velars outside compound boundaries*

True intervocalic velars represent the most interesting segment of our study, because they are the ones where previous scholarship has shown the most confusion. We approach this in four parts: first, we deal with the uncontroversial aspirated velars; then, we study the behaviour of words with a *-ka* suffix and demonstrate that the velar there shows a very regular development; then, we look at k in environments other than this *-ka* suffix, and finally we study the behaviour of intervocalic *g* in light of our understanding of the development of intervocalic *k*.

4.2.1. *Aspirated velars* kh *and* gh

Previous scholarship notes that Skt. *kh* and *gh* in intervocalic position always yield a spelled *h* (for instance, Salomon 2000: 82; Glass 2007: 114; Schlosser 2022: 78).

We reach the same conclusion, but it is worth noting that some documents still spell *kh* in the words Skt. *mukha* "mouth, face, countenance", *sukha* "pleasant", and in the conjugation of the root √*likh* "to write". This is the case for the texts in SHĀH and MĀN, which generally display a very archaic phonology, but also in $EĀ^{BM}$, ASP and the AV^{L}. The outcome of *gh* follows the same pattern and yields *h* throughout the corpus, with a few exceptions: BHKS writes *gh* in its only attestation, DHP^{K} writes *k* twice, and $NIRD^{L2}$ writes nothing in all attestations of Skt. *pratigha* (while it writes *h* elsewhere), and so does $MSŪ^{B}$. The relative scarcity of data precludes any strong conclusion, as what we witness might be the result of sampling.

4.2.2. *The case of the* -ka *suffix*

Brough (1962: 91) commented that the weakening of *k* was found "with fair consistency" and was "especially common in stem-final position – frequent therefore for the suffix (*-ka*), although not limited to it". The particular tendency for the suffix *-ka* to lenite was also noted elsewhere, for instance in Salomon (2000: 82) and Allon (2001: 80).

Yet, other scholars comment that the variation in spelling is a cause for doubt. The evolution of scholarly positions on this topic is interesting. Brough (1962: 86) ascribes to *g̱* (then transcribed *ǵ*) the value of [ɣ] in some cases and [χ] in others and considers the matter "generally accepted",[12] but also notes that – for intervocalic single stops – "such variation [in spelling] is for the most part a matter of spelling".

Glass (2000: 56), in the first volume of the Gandhāran Buddhist Texts collection, comments that "these flourishes [such as the one at the bottom of g, transcribed g̱] do not have any phonetic value", and later in the volume Glass (2000: 82) notes "inconsistent treatment of original -k- is typical of Gāndhārī/Kharoṣṭhī texts in general and evidently shows that it was pronounced weakly, if at all", but also mentions "a diacritically marked variety of g, apparently represents a fricative pronunciation (/ɣ/)".

In a further volume of the collection, Lenz (2003: 118) remarks "a distinct character (transcribed ǵa), believed to have a fricative pronunciation /ɣ/ [...]. However, a distinct phoneme

[12] So much so that he does not provide any pointer to past literature on the topic.

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

Table 1. Distribution of the Gāndhārī outcomes for the *k* in Skt. *-aka#* (left) and *-ika#* (right)

Outcome document	k	k̲	g	g̲	y	Ø	Outcome document	k	g	g̲	y	Ø
SHĀH	7				1		SHĀH	7				
MĀN	6				1		MĀN	5				
CKD 511	4						CKD 511				1	
CKI 149	1		1			1	CKI 149				1	
DHP^{K}	24				1	2	DHP^{K}					6
CKI 249	1						CKI 249					3
CKI 48		2				2	CKI 48					
NIRD^{L}			24		1	3	NIRD^{L}		1			6
AV^{L}			21				AV^{L}		4			2
EĀ^{BM}			3				EĀ^{BM}		1			
DHP^{SP}			6	2	2		DHP^{SP}				2	
CKD 72				46			CKD 72					
ASP				5	1	5	ASP			19		6
CKD 76				7			CKD 76					
MSŪ^{B}				5			MSŪ^{B}					
CKD 502				5			CKD 502					
CKD 140				3		1	CKD 140					
CKD 842				2	2		CKD 842					
CKD 604	1			3			CKD 604					
CKD 14				3			CKD 14					
ANAV^{L}						12	ANAV^{L}					11
MĀ^{S}				1	3	3	MĀ^{S}					5
BHKS			1			6	BHKS		4			
SĀ^{S}						8	SĀ^{S}					2
BC						2	BC					7
MNRS						3	MNRS					5
K^{hvs}			2			1	K^{hvs}					4
CKD 345				1		4	CKD 345					
EĀ^{L}						5	EĀ^{L}					

is probably not indicated in our text".[13] Based on this comment, Lenz makes no attempt to transcribe this distinct letter, transcribing *g* where other volumes of the collection write *g̲*, an issue we discuss in Appendix B3.[14]

Further still, Glass (2007: 89) now notes "the rightward stroke at the foot of *k, g, ṭ, d, s* is a defining feature [...] rather than a foot mark", a very welcome development in the understanding of the Kharoṣṭhī script, but the phonetic realization is still considered uncertain,

[13] A reviewer on an earlier draft of this article suggested that Lenz (2003) might have been making a point about the phonemic status of the fricative phone. We think the use of "in our text" makes it unlikely this was the intended meaning, but this remains a possibility. In itself, the status of [ɣ] as a phoneme or a mere allophone of /g/ is not trivial, but like Fussman (1989: 467), we think it unlikely the scribes of that time had discovered the notion of allophone.

[14] In the rest of the article, we use Lenz's transcription.

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

saying it is "possible but not certain that this scribe used *k*/*g̲* to phonetically distinguish intervocalic *k* and *g* [...] the evidence is inconclusive on the basis of this manuscript alone" (Glass 2007: 107), and yet further down "these forms, and the developments of intervocalic *g*, suggest that the pair *g* and *g̲* were minimally distinguished" (Glass 2007: 119). In spite of Glass's remarks regarding the stroke at the bottom of *g̲* as a feature of the character and not a flourish, Lenz (2010) still does not distinguish it and transcribes all instances as *g*, in spite of being published in the same collection with the same editorial team. In the most recent volume of the collection, Schlosser (2022: 81) alludes to this issue, saying "When Gāndhārī manuscripts were initially being studied, *g* and *g̲* were not differentiated consistently by every editor because the two signs did not imply a difference in meaning. Since they do reflect a phonological difference, in this publication a distinction is maintained".

In light of the fluctuation in the recent scholarship, and answering calls by Brough (1962: 85) and Glass (2007: 107) to conduct an analysis of the distribution of the spellings for intervocalic velar stops, we first look at the outcome of the most commonly found intervocalic *k*, namely the suffix *-ka*. To this end, we tally the attestations of Gāndhārī words corresponding to Sanskrit words ending in *-ka*, and we group these attestations by the vowel that precedes *-ka* and by text.

Since the two main vowels preceding *-ka* are *a* and *i*, we first present these two in Table 1. The texts have been ordered to highlight the pattern: on the *-aka#* side (left), we can see four distinct groups: from the top to CKI 48, the outcome is predominantly *k*; from NIRD^{L} to DHP^{SP}, the outcome is predominantly *g*; from CKD 72 to CKD 14, *g̲*; finally, from ANAV^{L} to the bottom, the main outcome is an empty spelling ("∅", such as *uasao* in SĀ^{S6}, from Skt. *upāsaka* "lay follower"). On the *-ika#* side (right), however, and with the exception of the two oldest texts (SHĀH and MĀN), there is a very clear predominance of either an empty spelling or a *y*.[15]

Such a contrast between these two environments tells us several things:

- First, it cannot be regarded as correct that Gāndhārī scribes, on the whole, spelled randomly, sometimes transcribing the actual pronunciation (*y* /j/ or ∅), and sometimes somehow using *g* or *g̲* to indicate knowledge of an original *k* that would now be pronounced /j/. This is not to say that historical spellings do not exist in Gāndhārī, they certainly do, but on the whole "historical spelling" should be our last recourse when trying to explain seemingly chaotic spelling behaviours in Gāndhārī.
- Second, and now considering that Gāndhārī spelling represents pronunciation with a higher fidelity than previously assumed, the Gāndhārī corpus can be segmented in at least four groups, based on the outcome of *k*, and studies of other phonological or linguistic features can evaluate their material against this segmentation.
- Finally, at least for the "*k*" and "*g*" segments of the corpus, the outcome of *k* was dependent on the preceding vowel and led to at least two different phonemes: preceded by *i*, *k* underwent palatalization and yielded /j/, represented by *y* or a gap, and that /j/ might itself have disappeared. The outcome of *k* preceded by *a* depends on the document, with the *g* and *g̲* indicating a lenition of sorts, either through voicing to /g/ or fricativization (with or without voicing) to /x/ ~/χ/ or /ɣ/.

Combined with previous rules, such as compound boundaries (including directional prefix boundaries, most of the time), this conditioning environment explains most of the data, sometimes even at the text level: Salomon (2008: 108) comments that the outcome of *k*

[15] Information about the date and findspot of the documents in Table 1 can be found in Appendix A.

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

Table 2. Distribution of the Gāndhārī outcomes for the k in Skt. *-ek-* and *-uk-* (left) and *-ok-* (right)

Outcome document	k	g	g̲	y	Ø	Outcome document	k	g	g̲	y	Ø
SHĀH	4					SHĀH	3				
MĀN	2					MĀN	3				
CKD 511	2	1				CKD 511					
CKI 149						CKI 149					
DHP^{K}					4	DHP^{K}	13				2
CKI 249						CKI 249					
CKI 48						CKI 48					
NIRD^{L}		2			2	NIRD^{L}		12			
AV^{L}		10				AV^{L}		1			
EĀ^{BM}						EĀ^{BM}		5			
DHP^{SP}						DHP^{SP}		5			1
CKD 72						CKD 72					
ASP			1			ASP			1		
CKD 76						CKD 76					
MSŪ^{B}			1			MSŪ^{B}			2		
CKD 502						CKD 502					
CKD 140			1			CKD 140					
CKD 842						CKD 842					
CKD 604						CKD 604					
CKD 14						CKD 14					
ANAV^{L}					3	ANAV^{L}					
MĀ^{S}						MĀ^{S}			2		
BHKS						BHKS	1				
SĀ^{S}						SĀ^{S}		4	2		
BC					2	BC			1		3
MNRS						MNRS					
K^{hvs}						K^{hvs}		2			
CKD 345						CKD 345					
EĀ^{L}					6	EĀ^{L}	1				

"seems to be more or less arbitrary and unpredictable; compare, for example, the two different reflexes in the etymologically cognate aṇomia = *anavamikāḥ and aṇom[a]gha = *anavamaka-", providing a minimal pair where the difference in the *k*-preceding vowel entirely conditions the outcome. Similarly, although the data is not entirely clean, a look at NIRD^{L2} in Table 1 suggests that Baums' (2009) comment "[t]he distribution of the three spellings [*g*, *y* or Ø] does not follow any discernible pattern, and it has to be assumed that the same pronunciation is intended everywhere" (Baums 2009: 140) needs revision.

An important question emerges: for the *g* and *g̲* segments (and to a lesser extent for the *k* segment), what was the phonetic realization of the outcome of *k* preceded by *a*? As noted above, the general belief is that *g̲* represented a fricative, and Glass (2007: 119) says the distribution of *g* and *g̲* "suggest[s] that the pair *g* and *g̲* were minimally distinguished", which we can but agree with. Provisionally, we think adopting /ɣ/ as a realization of *g̲* is

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

most reasonable. What about the "*g* texts"? Because neither NIRD^{L2} nor EĀ^{BM} spells *ḡ* in any context, it is difficult to say whether the scribes simply lacked the means to represent the phonetic outcome of *k* and went with what sounded closest, grouping these texts with the *ḡ* ones, or whether the outcome was actually /g/ and they should be considered a group of their own. The fact that nearly all of the texts in the *ḡ* group were found in Niya could just as well suggest a distinct dialect or simply a local spelling innovation not found in the *g* group, and a comparison of other phonological developments would be necessary to help decide this issue.

4.2.3. *Intervocalic* k *in other environments*

The contrast between final *-aka* and final *-ika* established, we now turn to other environments, both non-final environments and other k-preceding vowels.

In the Gāndhārī corpus, we could find no real example of non-final *-aka* or *-ika*: all word-internal cases fall into one of the previously studied cases, namely compound boundary (NIRD^{L2} *kriṣakamo* < Skt. *kṛṣṇakarman* "dark action") and "word-final" in compounds. We do, however, find cases of *k* preceded by the other vowels *e, o*, and *u*.

It has been widely noted[16] that the word *eka* "one" is always spelled with a *k* in Gāndhārī, a phenomenon generally explained by saying that the word must have had a geminate velar – *ekka* – which would have prevented lenition. What is interesting and, to the best of our knowledge, has never been mentioned, is the fact that words based on this *eka* "one" such as Skt. *pratyeka* "one by one, separately" or Skt. *aneka* "many (lit. not one)" do lenite as one would expect, according to their group.[17]

Beyond this note, the outcome of *k* when preceded by either *e, o*, or *u* is the same as when preceded by *a*, as shown in Table 2. We note that there is an unexplained tendency in the "y/∅" segment of the table to spell Skt. *-ok-* words (right side of the table) with a *g* or *ḡ* instead of the expected *y* or ∅, a tendency not seen in *-ek-* and *-uk-* (left side). It is perhaps interesting that the conditioning of the outcome of Skt. *k* lies on a line between *i* on one side and all other vowels on the other side.

4.2.4. *Intervocalic*

With the outcome of Skt. *k* covered in intervocalic environments, a comparison of the outcome of Skt. *g* in similar environments is possible. As we find no "*i* vs all other vowels" contrast for *g*, the results are collapsed in Table 3.

Unfortunately, the results are somewhat difficult to interpret. The general shape of Table 3 resembles the previous tables, with the obvious difference that documents that had Skt. *k* preserved now have Skt. *g* preserved. Beyond this expected difference, which groups the so-called "*k* and *g* segments" together, the *ḡ* segment still shows distinctively.[18] As for the so-called "*y*/∅" segment, the inconsistency deserves some comment: for MĀ^{S}, Silverlock (2015: 174) notes that Skt. *bhagavant* is spelled *bhaḡava* in the body of the sutra, but *bhayava* in the standard closing formula, as is seen throughout the collection of manuscripts of the so-called Senior Scribe. He then proposes that the sutra was copied from an archetype that had the *ḡ* spelling – representing a by-then dated pronunciation – while closing formulae were fixed and "integral to the memorisation process of reciters". This spelling of *bhayava* also explains the 30 occurrences of *y* in the SĀ^{S} in Table 3.

[16] See for instance Brough (1962: 91), Salomon (2000: 82), or Allon (2001: 80).

[17] For instance, CKD 140 has *paḍiḡa* with a g, while ANAV^{L} has *praceabudhasa* with nothing, and NIRD^{L2} has *aṇegaruo* from Skt. *aneka-rūpa* while CKD 157 – also a document from Niya – has *aneḡa*.

[18] For MSŪ^{B}, Baums and Glass (2002) cite the outcome of Skt. *bhagavant* as *bhaḡava* seven times and as *bhagava* (with a plain g) once, after Strauch (2010: 28). As shown in Appendix B1, this is a transcription mistake, and *bhaḡava* is to be read everywhere. Likewise, in CKD 140, we choose to read *aroḡemi* instead of *arogemi* (see Appendix B2).

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

Table 3. Distribution of the Gāndhārī outcomes for the *g* in Skt. *-ag-*, *-eg-*, *-ig-*, *-og-*, and *-ug-*

Outcome document	k	k̲	g	g̲	y	∅
SHĀH			3			
MĀN			3			
CKD 511			15			
CKI 149						
DHP^{K}	44		8	1		9
CKI 249			3			
CKI 48		1				
NIRD^{L}			25			4
AV^{L}			25			
EĀ^{BM}			9			
DHP^{SP}			8	3		2
CKD 72						
ASP				25		
CKD 76						
MSŪ^{B}				8		
CKD 502						
CKD 140				3		
CKD 842						
CKD 604						
CKD 14				1		
ANAV^{L}						
MĀ^{S}			4	9	2	
BHKS			1			
SĀ^{S}	3	4	8	3	30	5
BC						2
MNRS				4		
K^{hvs}						
CKD 345			1			
EĀ^{L}					3	

Overall, the main takeaway from Table 3 is that the outcome of Skt. *g* is not spelled the same as that of Skt. *k*:

- As opposed to Skt. *-ik-*, Skt. *-ig-* does not commonly yield *y*/∅.
- As opposed to Skt. *-ak-*, *-ek-*, *-ok-*, and *-uk-*, Skt. *-ag-*, *-eg-*, *-og-*, and *-ug-* do not regularly yield *y*/∅ in the "*y*/∅ texts", and the distribution of attestations across a text – such as the distinction between the sutra and the closing formula mentioned above for the corpus of the Senior scribe (SĀ^{S} and MĀ^{S}) – is not sufficient to explain the distinction between the outcome of *k* and *g* preceded by these vowels.

This precludes k > g/V _ V, since otherwise the outcome of *k* and *g* would be identical. Instead, *k* must have fricativized first, with k > j/i _ V happening very early: this is the stage witnessed in the so-called "k documents". This was then followed by k > ɣ/V _ V (where the first V could not be /i/ any more). Then, some dialects must have undergone

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

ɣ > j/V _ V while others did not, leading to the distinction between the "*g* and *ġ* texts" on the one hand, and the "*y*/∅ texts" on the other, after both had undergone g > ɣ̄/V _ V.

5. Conclusion

In contrast with editions of individual texts, this study has approached the question of the development of Sanskrit velar stops in Gāndhārī from a corpus-wide angle. In doing so, we were able to highlight trends in the corpus and better understand the possible development of Gāndhārī phonology.

The development of velar stops is summarized below. While it is not possible to establish a relative chronology between every pair of sound changes listed below, we indicate chronological constraints when known.

- voiceless aspirated *kh*:

  $k^h > h$ / V _ V
  $k^h > g^ɦ$ / [+nasal] _
  $sk^h > k^h$

- voiced aspirated *gh*:

  $g^ɦ > h$ / V _ V

  $$g^h > \begin{cases} g^h/\#\_r \\ g/\#\_ \end{cases}$$

- voiceless unaspirated *k*:

  k > j / i _ V (in all but the very earliest texts)
  k > ɣ > j / V _ V (with texts showing the three stages)
  k > g / [+nasal] _
  sk > ʰk > k^h
  k > ∅ / _ #
  kt > t:

  $$kʊ > \begin{cases} kʊ/\#\_ \\ K \end{cases}$$

- voiced unaspirated *g*:

  g > g / # _
  g > ɣ > j / V _ V (only some texts show the second development; this chain visibly started after the corresponding k > ɣ > j / V _ V had completed)

On a historical phonology level, the main contribution of the article is the demonstration that the Sanskrit unaspirated stops *k* and *g* followed separate development chains in intervocalic contexts, and that in particular *k* followed a separate development depending on whether the preceding vowel was *i* or not (while the development of *g* shows no such conditioning). Such a discovery sheds light on why some Gāndhārī loanwords into Chinese are transcribed with characters with a final *-k* while others are not: Skt. *campaka* "of Campā" >

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

Gdh. *caṃpag(a) > Late Hàn Chinese 占匐 */tśam.bək/ with a final /-k/,[19] while Skt. *sugandhika* "some type water lily" > Gdh. *sugaṃdhi(a) > LHC 須揵提 */sio.gan.de/ without such a coda.

Beyond this specific phonological discovery, the main contribution of the article is to demonstrate that, overall, Gāndhārī scribes spelled with a much higher fidelity to pronunciation than is usually considered, and that attempting to discard "historical spelling" as a primary explanation is a promising approach that will yield new insights in Gāndhārī linguistics.

Acknowledgements. We would like to extend our thanks to Stefan Baums and Andrew Glass for permission to reproduce Figures 1 and 2, and the British Library for Figures 4, 5, 6, 7, 8, 9, 10, 11, and 12. For Figure 3, the original manuscript seems to have disappeared and its current location is unknown.

Funding. This work was funded by the Arts and Humanities Research Council (AHRC), UKRI, as part of the project "Han Phonology: When Chinese Became Chinese". Project Reference: AH/V008722/1. Principal Investigator: Nathan Hill, SOAS University of London, Trinity College Dublin.

References

Allon, Mark. 2001. *Three Gāndhārī Ekottarikāgama-Type Sūtras: British Library Kharoṣṭhī Fragments 12 and 14*. (Gandhāran Buddhist Texts 2.) Seattle, WA: University of Washington Press.

Bailey, Harold Walter. 1946. "Gāndhārī", *Bulletin of the School of Oriental and African Studies* 11, 764–97.

Baley, Julien, Nathan Hill, and Ernest Caldwell. 2023. "Chinese transcription of Buddhist terms in the Late Hàn Dynasty", *Journal of Open Humanities Data* 9/1, 10. https://doi.org/10.5334/johd.110.

Baums, Stefan. 2009. "A Gāndhārī commentary on early Buddhist verses: British Library Kharoṣṭhī Fragments 7, 9, 13 and 18". PhD dissertation, Seattle: University of Washington, Department of Asian Languages and Literature.

Baums, Stefan. 2019. "Outline of Gāndhārī Grammar". https://stefanbaums.com/baums_grammar_outline.pdf.

Baums, Stefan and Andrew Glass. 2002. "A dictionary of Gāndhārī". https://www.gandhari.org/dictionary.

Boucher, Daniel. 1998. "Gāndhārī and the early Chinese Buddhist translations reconsidered: the case of the Saddharmapuṇḍarīkasūtra", *Journal of the American Oriental Society* 118/4, 471–506. https://doi.org/10.2307/604783.

Boyer, A.M., E.J. Rapson, and E. Senart. 1920. *Kharoṣṭhī Inscriptions Discovered by Sir Aurel Stein in Chinese Turkestan, Part I. Text of Inscriptions Discovered at the Niya Site 1901*. Oxford: Published under the authority of H.M. Secretary of State for India in Council by the Clarendon Press.

Brough, John. 1962. *The Gāndhārī Dhammapada. Edited with an Introduction and Commentary by John Brough. [With Facsimiles of the Manuscript]*. (London Oriental Series, Vol. 7.) London: Oxford University Press.

Burrow, Thomas. 1937. *The Language of the Kharoṣṭhi Documents from Chinese Turkestan*. Cambridge: Cambridge University Press.

Burrow, Thomas. 1940. *A Translation of the Kharoṣṭhī Documents from Chinese Turkestan*. (James G. Forlong, Fund XX.) London: The Royal Asiatic Society.

Carling, Gerd and Georges-Jean Pinault. 2023. *Dictionary and Thesaurus of Tocharian A*. Wiesbaden: Harrassowitz Verlag.

Coblin, W. South. 1983. *A Handbook of Eastern Han Sound Glosses*. Hong Kong: The Chinese University Press.

Fussman, Gérard. 1989. "Gāndhārī écrite, Gāndhārī parlée", in *Dialectes Dans Les Littératures Indo-aryennes*, 55: 433–501. (Publications de l'Institut de Civilisation Indienne 8.) Paris: Institut de civilisation indienne.

Glass, Andrew. 2000. "Paleography and orthography", in *A Gāndhārī Version of the Rhinoceros Sūtra: British Library Kharoṣṭhī Fragment 5B*, 53–78. (Gandhāran Buddhist Texts 1.) Seattle: University of Washington Press.

Glass, Andrew. 2007. *Four Gāndhārī Saṃyuktāgama Sūtras: Senior Kharoṣṭhī Fragment 5*. (Gandhāran Buddhist Texts 4.) Seattle: University of Washington Press.

Harrison, Paul. 2019. "Lokakṣema", in *Brill's Encyclopedia of Buddhism*, 2, 700–06. (Handbook of Oriental Studies 2.) Leiden: Brill.

[19] No final /ɣ/ is available in Chinese.

Hill, Nathan W. 2014. "Grammatically conditioned sound change", *Language and Linguistics Compass* 8/6, 211–29. https://doi.org/10.1111/lnc3.12073.
Konow, Sten. 1929. *Kharoshṭhī Inscriptions with the Exception of Those of Aśoka*. Vol. II. (Corpus Inscriptionum Indicarum.) Calcutta: Government of India Central Publication Branch.
Lenz, Timothy. 2003. *A New Version of the Gāndhārī Dharmapada and a Collection of Previous-Birth Stories: British Library Kharosthi Fragments 16 + 25*. (Gandhāran Buddhist Texts 3.) Seattle: University of Washington Press.
Lenz, Timothy. 2010. *Gandhāran Avadānas: British Library Kharoṣṭhī Fragments 1–3 and 21 and Supplementary Fragments A–C*. Illustrated edition. (Gandhāran Buddhist Texts 6.) Seattle: University of Washington Press.
Lenz, Timothy. 2014. "The British Library Kharoṣṭhī fragments: behind the birch bark curtain", in Alice Collett (ed.), *Women in Early Indian Buddhism: Comparative Textual Studies*. Oxford: Oxford University Press.
Salomon, Richard. 1998. *Indian Epigraphy: A Guide to the Study of Inscriptions in Sanskrit, Prakrit, and the Other Indo-Aryan Languages*. New York: South Asia Research.
Salomon, Richard. 1999. *Ancient Buddhist Scrolls from Gandhara: The British Library Kharoṣṭhī Fragments*. London: British Library; Seattle: University of Washington Press.
Salomon, Richard. 2000. *A Gāndhārī Version of the Rhinoceros Sūtra: British Library Kharoṣṭhī Fragment 5B*. (Gandhāran Buddhist Texts 1.) Seattle: University of Washington Press.
Salomon, Richard. 2001. "'Gāndhārī Hybrid Sanskrit': new sources for the study of the Sanskritization of Buddhist literature", *Indo-Iranian Journal* 44, 241–52.
Salomon, Richard. 2008. *Two Gāndhārī Manuscripts of the Songs of Lake Anavatapta (Anavatapta-Gāthā). British Library Kharoṣṭhī Fragment 1 and Senior Scroll 14*. Seattle: University of Washington Press.
Schlosser, Andrea. 2022. *Three Early Mahāyāna Treatises from Gandhāra: Bajaur Kharoṣṭhī Fragments 4, 6, and 11*. (Gandhāran Buddhist Texts 7.) Seattle: University of Washington Press.
Schoubben, Niels. 2022. "The Son of the King: Iranistic notes on Gāndhārī Kṣabura", *Zeitschrift Der Deutschen Morgenländischen Gesellschaft* 172/1, 149–53.
Silverlock, Blair Alan. 2015. "An edition and study of the Goṣiga-Sutra, the Cow-Horn Discourse (Senior Collection Scroll No. 12): an account of the Harmonious Aṇarudha Monks". PhD thesis, Sydney: University of Sydney.
Stein, Aurel. 1907. *Ancient Khotan: Detailed Report of Archaeological Explorations in Chinese Turkestan*. Oxford: Clarendon Press.
Strauch, Ingo. 2010. "More missing pieces of early Pure Land Buddhism: new evidence for Akṣobhya and Abhirati in an early Mahayana Sutra from Gandhāra", *The Eastern Buddhist* 41, 23–66.
Zacchetti, Stefano. 2019. "An Shigao", in *Brill's Encyclopedia of Buddhism*, 2: 630–41. (Handbook of Oriental Studies = Handbuch Der Orientalistik. Section Two, India.) Leiden: Brill.

Appendix A. Date and findspot of the documents used in the article

Table 4 provides a reference to the dates and findspots attributed to the documents found in Table 1, 2, and 3 of the article, retrieved from Baums and Glass (2002).[20] The information can be used to study chronological and geographical trends in the development of the intervocalic unaspirated velars.

Appendix B. Proposed corrections to text transcriptions

B1. MSŪ[B]

Besides the seven attestations of *bhagava* for Skt. *bhagavant*, Strauch (2010: 28) transcribes the attestation of line 9 as *bhagava* (with a plain *g*). As shown in Figure 3, this is a mistake, and *bhag̱ava* should be read on line 9 like the two occurrences on line 12 (and *tasag̱adaśaśa-* on line 10). This means the scribe was consistent in the spelling of this word.

B2. CKD 140

One instance of Gdh. *aroga* (< Skt. *aroga* "healthy") is found in CKD 140 and transcribed in Boyer et al. (1920: 56) as *arogemi*. Based on the manuscript shown in Figure 4, and comparing with the other two instances of ***g̱*** (then transcribed ***ǵ***), we choose to read *arog̱emi*, although that reading is less certain than that of ***bhag̱ava*** in Appendix B.1.

[20] Where no information was available, the cells were left blank.

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

Table 4. Date and findspot of the documents from Tables 1, 2, and 3

Document	Date	Findspot
SHĀH	250 BCE	Shahbazgarhi, Mardan District, Khyber Pakhtunkhwa, Pakistan
MĀN	250 BCE	Mardan District, Khyber Pakhtunkhwa, Pakistan
CKD 511		Niya, Xinjiang, China
CKI 149	144 CE	Mankiala, Rawalpindi District, Punjab, Pakistan
DHP^{K}		Kok Marim Cave, Hotan, Xinjiang, China
CKI 249		
CKI 48		Mathura, Mathura District, Uttar Pradesh, India
NIRD^{L}		
AV^{L}		Haḍḍa, Bihsud District, Nangarhar, Afghanistan
EĀ^{BM}		
DHP^{SP}		
CKD 72		Niya, Xinjiang, China
ASP		
CKD 76		Niya, Xinjiang, China
MSŪ^{B}		Mahal, Mian Kaley, Dir District, Khyber Pakhtunkhwa, Pakistan
CKD 502		Niya, Xinjiang, China
CKD 140		Niya, Xinjiang, China
CKD 842		Niya, Xinjiang, China
CKD 604		Niya, Xinjiang, China
CKD 14		Niya, Xinjiang, China
ANAV^{L}		Haḍḍa, Bihsud District, Nangarhar, Afghanistan
MĀ^{S}		Haḍḍa, Bihsud District, Nangarhar, Afghanistan
BHKS		
SĀ^{S}		Haḍḍa, Bihsud District, Nangarhar, Afghanistan
BC		Mahal, Mian Kaley, Dir District, Khyber Pakhtunkhwa, Pakistan
MNRS		Mahal, Mian Kaley, Dir District, Khyber Pakhtunkhwa, Pakistan
K^{hvs}		Haḍḍa, Bihsud District, Nangarhar, Afghanistan
CKD 345		Niya, Xinjiang, China
EĀ^{L}		Haḍḍa, Bihsud District, Nangarhar, Afghanistan

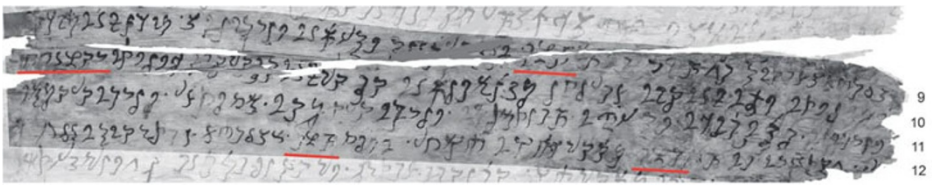

Figure 3. Lines 9–12 of the MSŪ^{B} manuscript, as found in Strauch (2010: 28)

B3. AV^{L6}

Lenz (2003: 118) – commenting on the spelling of *g* with a foot mark and on its usual attribution of the sound /ɣ/ – writes "a distinct phoneme is probably not indicated in our text". As a consequence, and in contrast with the other volumes of the Gandhāran Buddhist Texts collection, Lenz (2003) does not transcribe *g̲* separately from *g*.

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

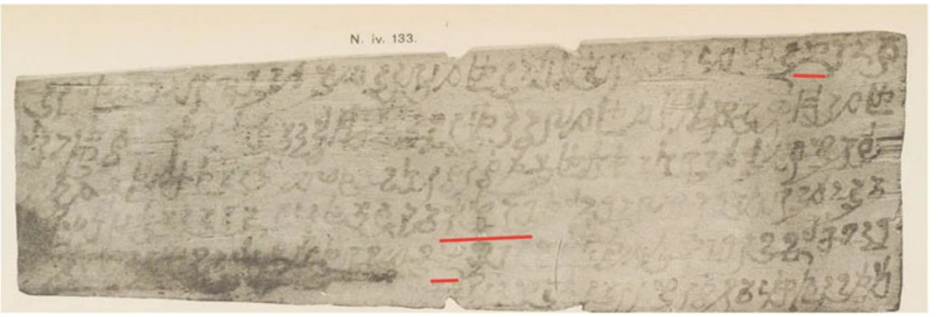

Figure 4. Lines 1–6 of the under-tablet obverse of CKD 140, as found in Stein (1907)

We would like to show that the spelling of this text is in fact very consistent with the phonological development of velar stops presented in this article.

Figure 5. The various shapes of the *ga* akṣara: (a) plain, (b) left tail, (c) right tail, (d) left-to-right stroke, (e) curl, (f) *rga*

We find the *ga* akṣara written with six different shapes (shown in Figure 5): with a straight vertical stem, with a left "tail", with a right "tail", with a stroke from left to right of the vertical stem, with a right curl continuing the vertical stem, and with a clockwise loop crossing the stem. Our analysis does not differ from the generally accepted palaeographic analysis of *ga* and its variations:

- The left tail is merely a flourish and represents the same sound as a plain vertical stem, as seen in Figure 6 in two words having *g* in word-initial position: *ga[do]* (plain stem, Skt. *gata* "gone") and *gilanag̱o* (left tail, Skt. *glāna*+*ka* "sick").
- The right tail and left-to-right stroke shapes represent /ɣ/, which we transcribe as *g̱*, as seen in Figure 7 in two words with expected /ɣ/: *nag̱are* (right tail, Skt. *nagara* "city") and *likhidag̱e* (left-to-right stroke, Skt. *likhita*+*ka* "written").
- In this text, the shape with a right curl represents *gr*, as seen in Figure 8: *anugrahartho* (Skt. *anugrahārtham* "as a favour").

Figure 6. Two words with an initial /g/: (a) *ga*[do], (b) ***gilanag̱o***

- The shape with a clockwise loop represents *rga*, see for instance Glass (2000, 62).

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

Figure 7. Two words with an expected /ɣ/: (a) *nag̲are*, (b) *likhidag̲e*

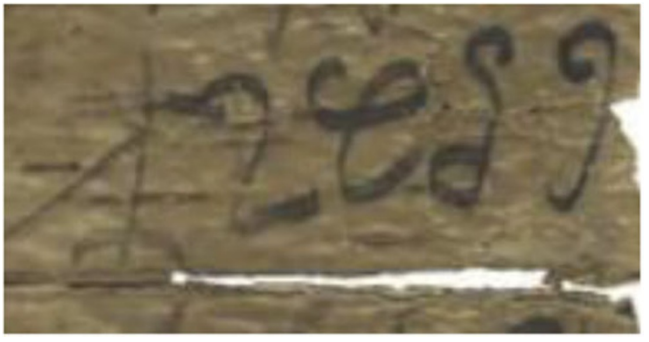

Figure 8. Word with an expected *gr: anugrahartho*

On this basis, we analyse all instances of *g* in the transcription of the text, and propose a revised transcription of the words. When the reading is uncertain, we transcribe according to the expected outcome and indicate our uncertainty with square brackets. The original transcription contains 54 instances of *g* across 51 words, listed in Table 5. For reference, the manuscript found in Lenz (2003) is reproduced here, with the relevant words underlined, in Figures 9, 10, 11, and 12.

The scribe's spelling shows a few peculiarities:

- Sanskrit words that have *g* or *k* preceded by a long *ā*, like Skt. *hastināga* "stately elephant", are spelled with a plain *g* with complete regularity (five occurrences). This seems to be a conditioning factor blocking further lenition, and this contrasts with the rest of the corpus where the length of the preceding vowel seems to make no difference to the outcome of the velar stop.
- The outcome of Skt. *eka* "one" is *eg[o]*, and not *eka*, as is found elsewhere in most of the Gāndhārī corpus.
- The name of the Saka people is spelled as *sago*. This form is probably not inherited from Sanskrit (Skt. *Śaka*) given the absence of *ś*, and the close proximity of the Saka people with Gāndhārī speakers may have led to a direct borrowing at a later date than the normal development of k > ɣ/a _ V.
- Except in *rayo* and *rayam* (< Skt. *rājā* "king"), expected *y* as the initial of the last syllable is never spelled *y*. Instead, the two instances of a locative (L27 and L28), the words in Skt. *-iya* and *-īya* are spelled with a *g*, while the words in Skt. *-ika, -ija*, and *-aya* are spelled with *g̲*. Since the scribe's spelling – as we will discuss below – is otherwise completely consistent with pronunciation, this spelling practice of rendering all expected *y* as *g* and *g̲* suggests a possible fortition of *y* /j/ to a palatal fricative, phonetically close to the realization of *g̲*.

These peculiarities aside, the scribe's spelling is extremely consistent with the expected spelling of a "*g̲* text":

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

Table 5. Revised transcription of words originally transcribed with g in Lenz (2003)

Line	Original	Revised	Mark	Sanskrit	Context	Expected outcome
16	nagare	nag̱are	right	nagara	VgV	g̱
17	samagen[o]	samagen[o]	damaged	samaya	ayV	y
17	dhaniga	dhani g̱a	full	dhanika	ikV	y
18	likhidage	likhidag̱e	full	likhita+ka	akV	g̱
18	bosisatvaprovayoge	bosisatvaprovayog̱e	right	Bodhisattva+pūrvayoga	VgV	g̱
19	vaniage	vaniag̱e	right	vāṇijaka	akV	g̱
19	(maha)///[sa]mudr[a]–[va]nige	(*maha)///[sa]mudr[a]–[va]nig̱e	right	mahāsamuddavāṇija	ijV	y
20	vaniaga	vaniag̱a	right	vāṇijaka	akV	g̱
21	avarnage	avarna[g̱]e	damaged	āpannaka	akV	g̱
21	[p](*ra)cagaran(*o)	[p](*ra)cag̱aran(*o)	right	pratyupakaraṇa	akV	g̱
21	[va]niaga	[va]niag̱a	full	vāṇijaka, vaṇijaka	akV	g̱
21	śpagam	śpag̱am	right	svayam	ayV	y
22	(*prova)///[yo]go	(*prova)///[yo]g̱o	right	pūrvayoga	ogV	g̱
24	hastinago	hastinago	none	hastināga	āgV	g̱
25	(dara-)ga	(dara-)g̱a	right	dāraka	akV	g̱
25	śilogano	śilogano	right	śloka	okV	g̱
27	p(*rova)yog(*o)	p(*rova)yog̱(*o)	right(?)	pūrvayoga	ogV	g̱
27	añage	añage	left	anya	loc.	y
28	[na]dig(*e)	[na]dig(*e)	left	nadī	loc.	y
28	praceg[a]sabudho	prace[g̱][a]sabudho	damaged	pratyekasaṃbuddha	ekV	g̱
28	ga[do]	ga[do]	none	gata	#g	g
31	upagarano	upag̱arano	right	upakaraṇa	akV	g̱

(*Continued*)

Table 5. (*Continued.*)

Line	Original	Revised	Mark	Sanskrit	Context	Expected outcome
32	anugrahartho	anugrahartho	curl	anugrahārtham	gr	gr
33	pracegabudho	pracegabudho	full	pratyekabuddha	ekV	g
35	pracegabudho	pracegabudho	right	pratyekabuddha	ekV	g
36	gilanago	gilana[g]o	left illegible	glāna+ka	#g agV	g g
37	pracegabudho	pracegabudho	right	pratyekabuddha	ekV	g
37	kalagado	kalagado	none	kālagata	#g	g
37	[ka]///(rma)///–vivago	[ka]///(rma)///–vivago	left	karmavipāka	ākV	g
37	pracegabudhano	pracegabudhano	right	pratyekabuddha	ekV	g
38	agada	agada	none	āgata	āgV	g
38	gina[ta]	gina[ta]	none	gṛ(b)hṇant	#g	g
40	[a]nagade	[a]nagade	none	anāgata	āgV	g
40	aragajo	aragajo	left	*ārāgayati	āgV	g
42	pru///(*vayo)///ge	pru///(*vayo)///ge	right	pūrvayoga	ogV	g
43	Gaṣabadhaga	Gaṣabadhaga	none right	Gāthābandhaka	#g akV	g g
43	pradeśige	pradeśige	left	prādeśika	ikV	y
44	Sabrudidrigo	Sabrudidrigo	left	Saṃvṛtendriya	iyA	y
44	S[a]b(*ru)didrig[o]	S[a]b(*ru)didrig[o]	damaged	Saṃvṛtendriya	iyA	y

(*Continued*)

Table 5. (*Continued.*)

Line	Original	Revised	Mark	Sanskrit	Context	Expected outcome
45	[pra]ce[geha]budhi	[pra]ceg̱[eha]budhi	full	pratyekabuddha	ekV	g̱
45	gatha	gatha	none	gātha	#g	g
47	marga[n]o	marga[n]o	“rg”	mārga	rgV	g
47	agachadi	agachadi	illegible	āgacchati	#g	g
47	gamo	g̱amo	full	grāma	gr	gr
47	Sago	Sago	left	Śaka	akV	g̱
48	istrigagana-pradivrud[o]	istrig̱ag̱ana-pradivrud[o]	right right	strīkagaṇaparivṛta	īk #g	y g
50	Sa[ga]///(*sa)	Sa[ga]///(*sa)	damaged	Śaka	akV	g̱
51	(*ka)///ran[i]go	(*ka)///ran[i]go	damaged	karaṇīya	īyV	y
51	Sago	Sago	damaged	Śaka	akV	g̱
51	e[go]	e[go]	none	eka	‘eka’	k
53	[ka]ranigo	[ka]ranigo	none	karaṇīya	īyV	y
53	Sago	Sago	none	Śaka	akV	g̱

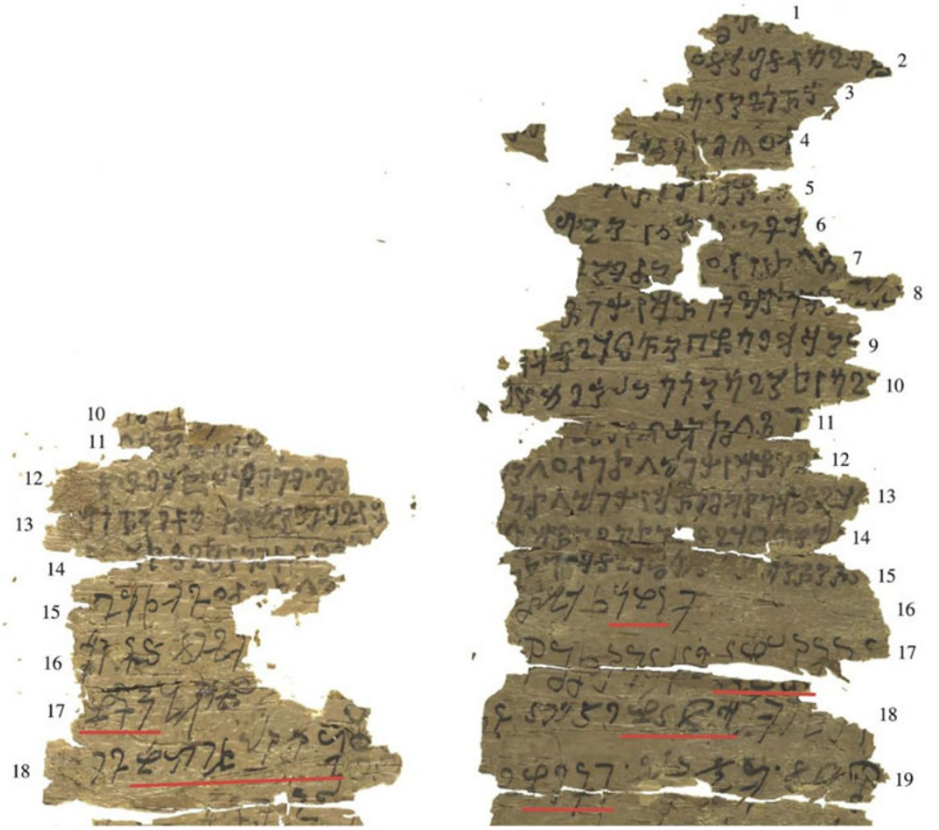

Figure 9. Av^{L6}, lines 1–19

- Out of 22 instances of expected *g̱*,[21] we find 19 instances of *g̱* and three cases where the manuscript is damaged right at the foot of the akṣara or the ink has faded too much; which translates to a complete consistency of the spelling for the cases where *g̱* would be expected.
- Of the nine instances of expected *g̱*, seven do indeed show *g̱*. The exceptions are *istrigag̱anapradivrud[o]* mentioned above, and *agachadi* (Skt. *āgacchati* "(he) comes"), for which the reading is uncertain, resulting in seven out of eight cases where we have enough confidence.

This conclusion, that Av^{L6} is a "*g̱* text", reduces the already small body of texts belonging to the "*g̱* texts", and suggests we need to re-visit the other transcriptions produced by Lenz following the same transcription conventions (Lenz 2010, 2014).

[21] One would normally expect 31 instances, but we find four instances of Sago and five words with -āg̱-, which have been discussed above.

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

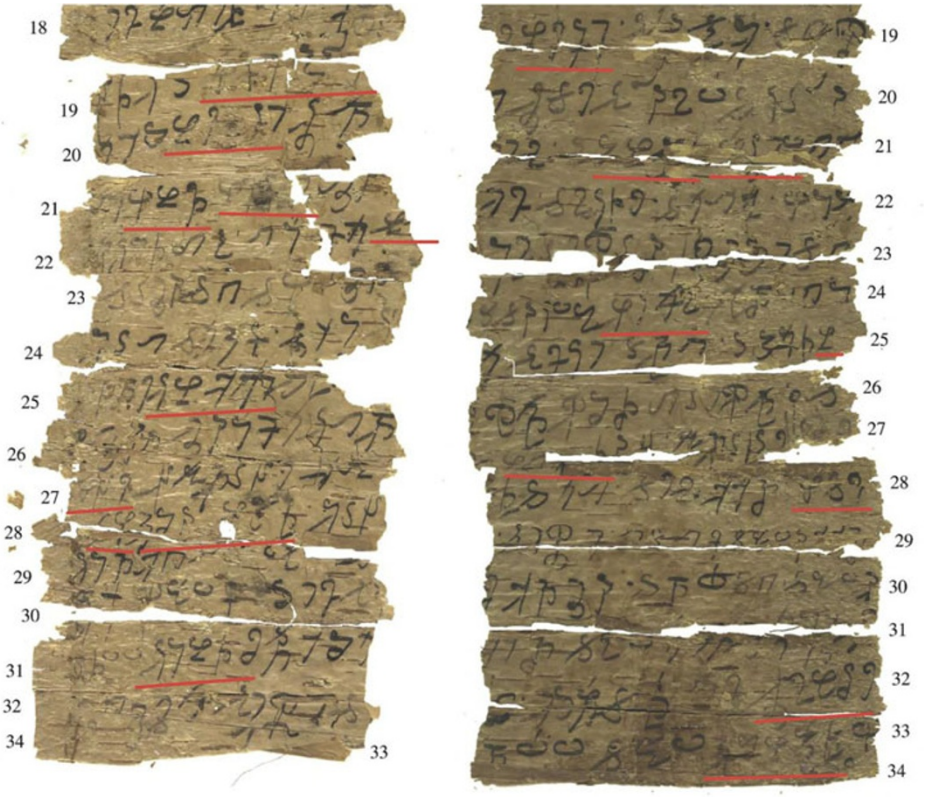

Figure 10. A_V^{L6}, lines 19–34

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

Figure 11. Av^{L6}, lines 35–48

https://doi.org/10.1017/S0041977X26101025 Published online by Cambridge University Press

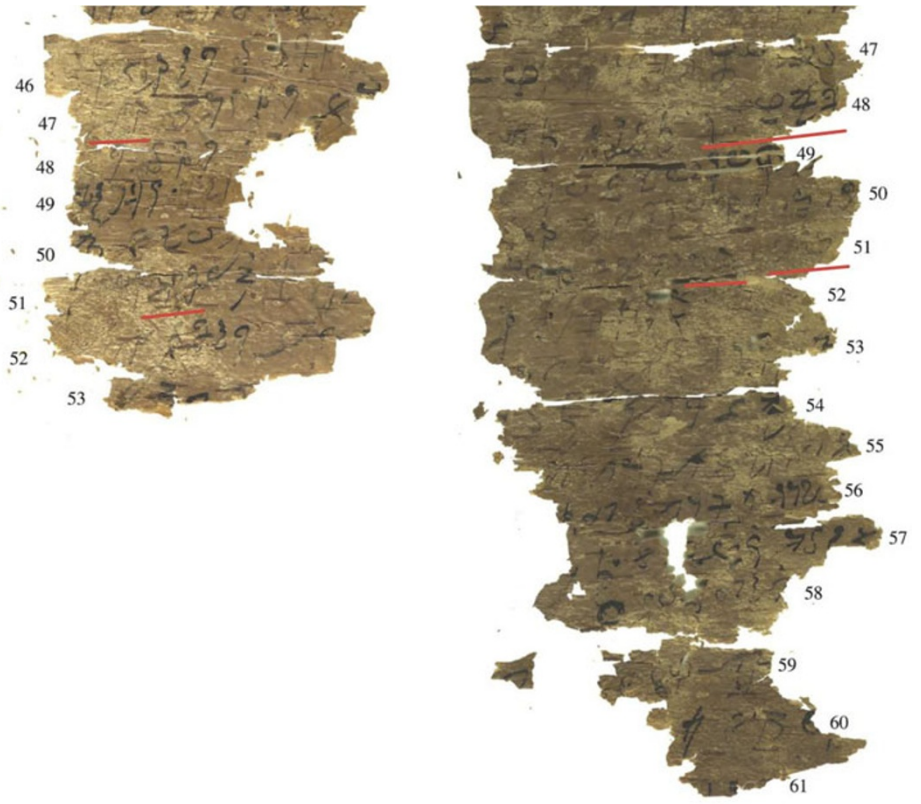

Figure 12. Av^{L6}, lines 47–61

Cite this article: Baley, Julien and Alan Avdagic. 2026. "The development of Sanskrit velars into Gāndhārī", *Bulletin of the School of Oriental and African Studies* 1–27. https://doi.org/10.1017/S0041977X26101025