



## **BuddhistRoad**

Dynamics in Buddhist Networks in Eastern Central Asia 6th–14th Centuries

### **BuddhistRoad Paper 1.7**

---

---

# **NE HĀḌE VAJRROPAMĀ VAŚĀRĀ: INDIC LOANWORDS IN THE KHOTANESE BOOK OF ZAMBASTA AND THE CHRONOLOGY OF THE SPREAD OF BUDDHISM TO KHOTAN**

DIEGO LOUKOTA





## BUDDHISTROAD PAPER

Peer reviewed

ISSN: 2628-2356

DOI: 10.46586/rub.br.274.246

BuddhistRoad Papers are licensed under the Creative Commons—Attribution NonCommercial—NoDerivates 4.0 International (CC BY-NC-ND 4.0).

You can find this publication also on the BuddhistRoad project homepage:

<https://buddhistroad.ceres.rub.de/en/publications/>

Please quote this paper as follows:

**Loukota, Diego, “Ne Hāḍe Vajropamā Vaśārā: Indic Loanwords in the Khotanese Book of Zambasta and the Chronology of the Spread of Buddhism to Khotan,” *BuddhistRoad Paper 1.7* (2023).**

### CONTACT:

Principal Investigator: Prof. Dr. Carmen Meinert

BuddhistRoad | Ruhr-Universität Bochum | Center for Religious Studies (CERES)

Universitätsstr. 90a | 44789 Bochum | Germany

Phone: +49 (0)234 32-21683 | Fax: +49 (0) 234/32- 14 909

Email: [BuddhistRoad@rub.de](mailto:BuddhistRoad@rub.de) | Email: [carmen.meinert@rub.de](mailto:carmen.meinert@rub.de)

Website: <https://buddhistroad.ceres.rub.de/>

BuddhistRoad is a project of



Centrum für Religionswissenschaftliche Studien  
Center for Religious Studies

RUHR  
UNIVERSITÄT  
BOCHUM

RUB

### SPONSORS:



European Research Council

Established by the European Commission



This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 725519).

NE HĀḌE VAJRROPAMĀ VAŚĀRĀ:<sup>1</sup>  
INDIC LOANWORDS IN THE KHOTANESE BOOK OF  
ZAMBASTA AND THE CHRONOLOGY OF THE SPREAD OF  
BUDDHISM TO KHOTAN

DIEGO LOUKOTA

*Abstract*

The Middle Iranian Khotanese language received copious amounts of loanwords from the Indic Gāndhārī and Sanskrit languages. These loanwords have not, however, been systematically studied, and although the present paper does not constitute an attempt to provide such a comprehensive survey, it makes a limited probe in the same direction through the examination of the borrowed vocabulary in the fifth century *Book of Zambasta*,<sup>2</sup> the longest extant and arguably most substantial Old Khotanese text. This focused examination of the Indic element in Khotanese highlights the fact that the form of Gāndhārī that informed the earliest Khotanese spread of Buddhism was probably not identical with the one represented in the sacred Buddhist texts from Gandhāra proper, as well as distinct chronological layers, with Gāndhārī providing the majority of the vocabulary for the most basic terms for Buddhist thought and practice and Sanskrit providing, instead, the more elaborate compound repertoire of terms associated with the Mahāyāna and with mature scholastic philosophy (Skt. *abhidharma*).

---

<sup>1</sup> This phrase, newly composed in Khotanese, would literally mean “indeed the diamond is not diamond-like,” which highlights the difference between the term ‘diamond’ (Kh. *vaśāra-*) derived from Gāndhārī *vayira* (cognate of Skt. *vajra*), and the term ‘diamond-like’ (Kh. *vajrropama-*) derived from Sanskrit *vajropama*.

<sup>2</sup> The Khotanese title of the text is unknown, not being preserved in any manuscript. Ernst Leumann gave it implicitly the sobriquet *Nordarische (sakische) Lehrgedicht des Buddhismus* in his publication *Das nordarische (sakische) Lehrgedicht des Buddhismus* (Leipzig: F.A. Brockhaus, 1933–1936,) but what really caught on was H.W. Bailey’s second sobriquet, the *Book of Zambasta*, on the basis of the official Ysambasta (d.u.) mentioned in the colophons of chapters 2, 13, and 19 as having commissioned the writing of the book (possibly, though, only of the individual manuscript); see H.W. Bailey, *Khotanese Texts VI—Prolexis to the Book of Zambasta* (Cambridge: Cambridge University Press, 1967). Here, he also lays forth the non-native Late Khotanese appellation ‘song of the buddhas’ (Kh. *baḷysānā āḷjsai*). *Ibid.*, vii.



## 1. Introduction

We do not know exactly why Gandhāra in particular and Indian culture in general were to become and remain stable cultural referents for the Iranian-speaking oasis Kingdom of Khotan (ca. 1st c.?–1006) and, for that matter, for the various peoples of the entire basin of the Tarim River in Eastern Central Asia throughout the pre-Islamic first millennium of the Common Era. Direct Kuṣāṇa (1st–mid-3rd c.) control at some point in the second century, although debated, may well have been a factor;<sup>3</sup> considerable movements of population from Gandhāra to the Tarim Basin are also a possibility to take into account.<sup>4</sup> Moreover, a process of reactive identity-formation by which the steady pressure of Chinese dynasties and politics from the east created an incentive to keep India, rather than China, as a dominant cultural model is likewise a possible factor.<sup>5</sup> In any case, the fact is that at least two languages of India, Gāndhārī, and Sanskrit, were to remain very stable points of reference for the systems of literacy of all the languages of the pre-Islamic Tarim Basin, including Khotanese.

Although both languages seem to have been used concurrently in the region during the first four centuries of the Common Era, since we can distinguish chronologically the periods in which Gāndhārī or Sanskrit would have been more prevalent in various functions, I suggest in this article a focused examination of a selection of loanwords from either language in a specific Khotanese text, the *Book of Zambasta*, with an eye to better understand the development of Khotanese literacy in tandem with diffusion of Buddhism in Khotan. In particular, I attempt here to examine the logic that underlies the adoption of Sanskrit or Gāndhārī loanwords for Buddhist technical vocabulary. Beyond the premise that loanwords from Gāndhārī are older than loanwords from Sanskrit, there have been no efforts at a more fine-tuned characterisation, and this study hopes to provide a further small step in that direction.

---

<sup>3</sup> The most spirited argument for Kuṣāṇa control of the basin is the paper by Doug Hitch, “The Special Status of Turfan,” *Sino-Platonic Papers* 186 (2009), last accessed August 11, 2022. [http://sino-platonic.org/complete/spp186\\_turfan.pdf](http://sino-platonic.org/complete/spp186_turfan.pdf).

<sup>4</sup> This would be the hypothesis favored by Valerie Hansen, *The Silk Road: A New History* (Oxford: Oxford University Press, 2017), 52–53.

<sup>5</sup> Diego Loukota, “Made in China? Sourcing the Old Khotanese *Bhaiṣajyaguruvaiḍūryaprabhasūtra*,” *Journal of the American Oriental Society* 139.1 (2019): 83–86.

The *Book of Zambasta* is a ca. fifth century compilation of Buddhist narrative and doctrinal matter, conveyed in a sophisticated literary form that seems to reflect conventions of indigenous Khotanese poetry such as moraic meter and concentric composition.<sup>6</sup> It is the longest extant Khotanese text, and it has been well edited and relatively well studied. The issue of the Indic loanwords in Khotanese is vast, but my goal here is to take this substantial text from the early literate period of Khotanese Buddhism and make a focused case study that may provide a template for broader explorations of the Indic element in Khotanese. This goal requires necessarily some broad observations on the linguistic history of the Khotanese language and its interaction with Gāndhārī and Sanskrit which allow an examination of the borrowed vocabulary and its implications for the chronology of the spread of Buddhism to Khotan.

## 2. A Note on Conventions and Phonological Models

Given the great inconsistency of the orthographies of Khotanese and Gāndhārī, a rather cumbersome but necessary system of notation is called for in order to distinguish the written expression of a word from its reconstructed phonetic value. Following regular linguistic notation, angled brackets (<>) will indicate the original orthography of sounds and words, and square brackets ([]) the presumed phonetic value of such written expression; the asterisk (\*) marks hypothetical forms. The notation <> [] will therefore equate a given written form with its reconstructed phonetic value as, for example, Khotanese <ātāśa ~ āgāśa> [a:ʔa:ʒa-] ‘sky’, from Gāndhārī <agaśa> [a:ja:çə-] (Skt. *ākāśa*). For the reconstruction of the values of Gāndhārī I have followed the model of phonological reconstruction of Stefan Baums, with the exception of the inherited long *-ā* feminine stems, which, unlike Baums, I prefer to

---

<sup>6</sup> The standard edition of the *Book of Zambasta*, from which the data presented here are taken, is Ronald E. Emmerick, *The Book of Zambasta: A Khotanese Poem on Buddhism* (London: Oxford University Press, 1968); the best recent survey on research, dating, and internal structure of the *Book of Zambasta* is the one provided by Mauro Maggi, “Khotanese Literature,” in *The Literature of Pre-Islamic Iran: Companion Volume to a History of Persian Literature*, ed. Maria Macuch and Ronald E. Emmerick (London: I. B. Tauris, 2009), 348–357. A valuable and comprehensive recent treatment of Old Khotanese metre is Nicholas Sims-Williams, *The Book of Zambasta: Metre and Stress in Old Khotanese* (Wiesbaden: Reichert, 2022).



hypothesise for Gāndhārī;<sup>7</sup> for Old Khotanese, I have followed the model of Doug Hitch,<sup>8</sup> and for Han Dynasty (206 BCE–220, 漢) Chinese, Axel Schuessler’s.<sup>9</sup>

### 3. Some Remarks on Khotanese and Gāndhārī Orthography

The orthography of the *Book of Zambasta* is relatively descriptive in that it distinguishes a number of sounds that are not distinguished in other, earlier texts. The *Book of Zambasta* is not the most archaic Khotanese text that we have.<sup>10</sup> The scanty texts that correspond to the earlier, so-called Archaic Old Khotanese, however, fail to distinguish certain sounds. For example, in the orthography of the Archaic Khotanese *Śūraṅgamasamādhisūtra*, the grapheme <ś> represents both the voiced sibilant [ž] and its unvoiced counterpart [š]: these sounds are represented, respectively, by <śś> and <ś> in the *Book of Zambasta* in intervocalic position. As an example of these different orthographies, both the *Śūraṅgamasamādhisūtra* and the *Book of Zambasta* write the word [ra:žā] ‘control’ as <rrāśa>, but only the *Book of Zambasta* distinguishes the unvoiced phoneme [š] with <śś> as in <harbiśśa-> [harbiša] ‘all’, whereas the same word is represented in the *Śūraṅgamasamādhisūtra* with the graphically ambiguous orthography <harbiśa> which, at face value, could map either onto [harbiša] or \*[harbižā]. On the downside of this more descriptive system of representing the language and of the richness and

<sup>7</sup> Stefan Baums, *Outline of Gāndhārī Grammar* (2019), last accessed September 20, 2022. [https://stefanbaums.com/baums\\_grammar\\_outline.pdf](https://stefanbaums.com/baums_grammar_outline.pdf).

<sup>8</sup> Doug Hitch, “The Old Khotanese Metanalysis” (PhD diss., Harvard University, 2016).

<sup>9</sup> Axel Schuessler, *Minimal Old Chinese and Later Han Chinese: A Companion to Grammata Serica Recensa* (Honolulu: University of Hawai’i Press, 2009).

<sup>10</sup> The following periodisation of the extant Khotanese corpus blends the basic model articulated in Prods Oktor Skjærvø, “Khotan, an Early Center of Buddhism in Chinese Turkestan,” in *Collection of Essays 1993, Buddhism across Boundaries—Chinese Buddhism and the Western Regions*, ed. John R. McRae and Jan Nattier (Sanchung: Fo Guang Shan Foundation for Buddhist & Culture Education, 1999) with the further subdivision of Old Khotanese proposed by Doug Hitch, “The Old Khotanese Metanalysis,” 11–12:

1. “Old Khotanese”

1.1. “Archaic Old Khotanese”: e.g. *Śūraṅgamasamādhī*, 4th–5th CE?

1.2. “Canonical Old Khotanese”: e.g. *The Book of Zambasta* 5th–6th CE?

2. “Middle Khotanese” 7th–8th CE?

3. “Late Khotanese” 9th–10th CE?

integrity of the text, it should be noted that the orthography of the *Book of Zambasta* already shows a layer of historical spelling and its ensuing inconsistencies, i.e., several written signs can represent a single sound of the language. For example, intervocalic <g>, <t>, <v> can all represent a hiatus / / or glottal stop /ʔ/, and <ä> [ě] and <i> [i] have started to coalesce and to be represented indiscriminately by either sign.

We should at this point take a look at the preceding period between the first and fourth centuries in which Khotanese was, as far as we can tell, not written and during which we have attestations of only three languages being used in Khotan and its environs: Chinese, Gāndhārī, and Sanskrit. The oldest records in the Tarim Basin are in fact in Chinese, and Chinese seems have to be used very early on for government and for court ceremonial among the indigenous polities of the basin, but discontinued for these purposes after the second century Kuṣāṇa incursions.<sup>11</sup> After the second century we see Chinese used exclusively for the administration of the Chinese military garrisons, stationed at different points in the Tarim Basin.

We have attestations of Gāndhārī from the first century in coin legends, which leads us to think that it was used in government and then between the third and fifth centuries we have abundant attestations of Gāndhārī being used as an administrative language in the Tarim Basin. Besides its administrative use, we also know that certain Buddhist texts circulated in Gāndhārī, like for example the *Dharmapada* of Khotan, and quotations of the *Udānavarga* and a *Prātimokṣasūtra* in the documents from Niya,<sup>12</sup> as well as fragments of *sūtra* literature from Kuča. Sanskrit, on the other hand, is attested from about the second century onwards. We do not have *sūtra* literature, but we have on the one hand *belles lettres*, like for example, theater plays and poetry, as well as scholastic philosophy (Skt. *abhidharma*).

It is useful to remember at this point that Gāndhārī and Sanskrit were conveyed in this region in two different scripts. Gāndhārī used the Kharoṣṭhī script, which is cursive and often ambiguous because of its generally fluid ductus. Sanskrit would be typically written in the Brāhmī script, a much clearer and careful book-script.

---

<sup>11</sup> Loukota, “Sourcing the Old Khotanese *Bhaiṣajyaguruvaiḍūryaprabhasūtra*,” 83–86.

<sup>12</sup> Respectively numbers 204 and 510 in the section “Documents” of the *Catalog of Gāndhārī Texts* by Stefan Baums and Andrew Glass, last accessed November 9, 2022. <https://gandhari.org/catalog>.





What exactly led the peoples of the Tarim Basin to write down their own languages around the end of the fourth century is still unknown. The demise of Gandhāra under the Hunnic invasions and the increase of military pressure from the Sinitic east, which may have triggered a process of reactive identity-formation may be factors to take into account for this period, in which the various languages of the region, such as the middle Iranian Śaka or Scythian languages, Tumšūqese and Khotanese, and the Tocharian Agnean, and Kučean languages (typically referred to, respectively, as Tocharian A and Tocharian B) were first written. How these languages began to be written is an issue about which we know very little, but during the period when this happened there was extensive experimentation with writing systems. Although eventually all of these languages would settle for scripts derived from Brāhmī, the script traditionally used for Sanskrit in the region, we have some tantalising evidence that some of the indigenous languages of the area may have initially been written with derivatives of the Kharoṣṭhī script. So for example, we have the corpus of the tentatively called Northern Tarim Para-Kharoṣṭhī script.<sup>13</sup> Represented by a handful of manuscript fragments and inscriptions, it records a yet undeciphered language, but the fact that the script derives from Kharoṣṭhī is generally agreed upon.<sup>14</sup> Moreover, the recently published wooden document BH5-7 from the National Library of China, although not yet fully deciphered, may also record a very early form of Khotanese in Kharoṣṭhī script (ca. 3rd or 4th c.).<sup>15</sup> This specific document records the earliest attestation of the diacritic, visually similar to a diaeresis (¨), used in all the aforementioned indigenous languages to represent the non-Indic ‘Fremdvokal,’ in the case of

---

<sup>13</sup> The corpus has been recently surveyed in Federico Dragoni, Niels Schoubben, and Michaël Peyrot, “The Formal Kharoṣṭhī Script from the Northern Tarim Basin in Northwest China May Write an Iranian Language,” *Acta Orientalia Academiae Scientiarum Hungaricae* 73.3 (2020): 335–373.

<sup>14</sup> Number 843 in the section “Documents” of the *Catalog of Gāndhārī Texts* by Stefan Baums and Andrew Glass, edited by Duan Qing 段晴, “Neirong buming Quluwen mudu: Guotu 内容不明佉卢文木牍: 国图 BH5-7 [A Kharoṣṭhī Wooden Tablet with Unidentified Content: BH5-7 from the National Library of China],” in *Zhongguo guojia tushuguan cang Xiyu wenshu: Fanwen, Quluwen juan* 中国国家图书馆藏西域文书: 梵文, 佉卢文卷 [Serindian Documents in the National Library of China: Manuscripts in Brāhmī and Kharoṣṭhī], ed. Duan Qing 段晴 and Zhang Zhiqing 张志清 (Shanghai: Zhongxi shuju, 2013), 203–205.

<sup>15</sup> See for example *māsta* ‘great’ (nom.pl.?), *ustama* ‘last’ (nom.pl.?), *śazdhā* = Brāhmī Khotanese *śaysdā* ‘snake’ (perhaps as a personal name, ‘[Born in the year of the] Snake’?).



Khotanese the vowel <-ä> [ě] but used originally in Kharoṣṭhī to express the Indic *visarga* (-ḥ), as well as the special diacritic used to express the Khotanese diphthong <-ei> ([aě]) in the shape of a superscript saltire (ˆ). The form of the Brāhmī script used for Khotanese would therefore not be a straightforward adaptation of the Brāhmī used for Sanskrit, but it would incorporate some features of Kharoṣṭhī such as the vowel diacritics just mentioned, as well as a complex system of digraphs whose phonetic logic is sometimes unclear, but which may actually be related to orthographic devices used by the Śaka/Scythians in India (the Indo-Scythians or Indo-Śakas) to record sounds from their language during an earlier era. Such is for example the case of the digraph *ys* representing the voiced dental fricative [z], attested in Indo-Scythian epigraphy, as in the coins of the Indo-Scythian Western Satrap ruler Caṣṭana, which mention the name of the king's father, Ysamotika ([zamoti:ka]). Similar Khotanese digraphs include <tc> [ts], <ts> [ts<sup>h</sup>], and <js> [dz]. A special mention is due to the phenomenon of orthographic gemination, which either distinguished voiced/unvoiced pairs (as in, <ś> [ʒ] vs. <śś> [ʒ̥]) or else distinguished the glottal stop [ʔ] from the original phonetic values of the signs (as in <g>, <t> [ʔ] vs. <gg> /g/, <tt> [t, t:]).

Regarding the Indic loanwords in the Khotanese *Book of Zambasta*, there are two points that we should take into account. On the one hand, Sanskrit loanwords often match Sanskrit orthography exactly as in *smṛtyupasthāna*, 'application of awareness,' but sometimes they have to adapt to the complex system of digraphs of Khotanese, as in *ttathāggattaggarbha-* for Sanskrit *tathagatagarbha*, '[doctrine of the] inner Buddha [nature]'. This tells us that the borrowing of this Sanskrit term did not happen on the basis of a written medium that would have privileged preserving the original Sanskrit orthography, but rather when the Khotanese writing system was mature enough and required that the original Sanskrit orthography was modified in order to preserve the original Sanskrit sound of the word. As a rule of thumb, on the other hand, Gāndhārī loanwords in general are far removed from Sanskrit orthography, reflecting not only the sound changes specific to Gāndhārī but also the adaptation of Khotanese orthography to reflect the sounds of Gāndhārī.

The whole question is complicated by the extreme orthographic oscillation of both Gāndhārī and Khotanese and their strong tendencies to historical spelling: Gāndhārī shows Sanskritising spellings from very early



on, and it is possible that Gāndhārī loanwords in Khotanese were later sanskritised too (much like Middle English *dette* and *iland* were orthographically latinised to *debt* and *island* in the Renaissance).

An extreme case of ‘bad behaviour’ in both languages is well illustrated by the term for the ‘ripening [of the cosmic requital of deeds],’ *vipāka* in Sanskrit and Pāli: Gāndhārī \*[viva:ɟə] is attested in the written forms <vivaa->, <vivava->, <vivaka->, and <vivaka->, whereas the Khotanese loanform [wiʔa:ʔa] is represented, only within the narrow confines of the manuscript tradition of the *Book of Zambasta*, as <vīvāga->, <vātāga->, <vivāta->, <vīvāta->, and <vāvāta->.

#### 4. The Form of Gāndhārī in Use in Khotan

This preamble also requires a brief notice on the dialectal position of the Gāndhārī source for the loanwords present in the *Book of Zambasta*. There are reasons to think that the form of Gāndhārī from which some words were loaned into the *Book of Zambasta* constituted, in the main, a different dialect than the one attested in Gandhāra proper, but very close instead to the specific language of the so-called Khotan *Dharmapada* in Gāndhārī (the later Shanshanese documents in Gāndhārī are of limited help in this regard on account of their heavily sanskritising spelling).<sup>16</sup> Some of these features, that oppose phonetic features in the Khotan *Dharmapada* and the *Book of Zambasta* (‘Tarim Gāndhārī’) against what we could term ‘Metropolitan Gāndhārī,’ are the following:

(1) ‘Tarim Gāndhārī’ shows Sanskrit  $d^h, t^h/V\_V > h$  vs. ‘Metropolitan Gāndhārī’  $> z$  (already attested in the Aśokan inscriptions), e.g. compare *Book of Zambasta* <praṇāhāna-> [praṇēhāna] against “Metropolitan Gāndhārī” <praṇiṣana->=\*[prəniza:nə], and Khotan *Dharmapada* <nihai> ‘having deposited’ (~ Sanskrit *nidhāya*) against ‘Metropolitan Gāndhārī’ <ṇiṣae> [niza:e].

(2) ‘Tarim Gāndhārī’ displays progressive assimilation of Sanskrit nasal-occlusive clusters, absent in ‘Metropolitan Gāndhārī,’ e.g. *Book of Zambasta* <Kośāmā->, <Kauśāmā->, <Kauśśāmā-> [Koʒa:mē] (with, perhaps, a sanskritised lengthened middle vowel) against Sanskrit

<sup>16</sup> The language of the Khotan *Dharmapada* is excellently described in John Brough, *The Gāndhārī Dharmapada* (London: Oxford University Press, 1962).

Kauśāmbī, Pāli Kosambi ‘the city of Kauśāmbī’, and Khotan *Dharmapada gammira* ‘deep’ against ‘metropolitan’ Gāndhārī *gaṃbhira* and Sanskrit *gaṃbhīra*.

(3) ‘Tarim Gāndhārī’ shows a shift from Sanskrit *ṃs* to *>ṃts*, also absent in ‘Metropolitan Gāndhārī,’ e.g. *Book of Zambasta* <saṃtsāra> [sants<sup>h</sup>a:ra] siding with Khotan *Dharmapada* <satsara> against Gāndhārī <saṃsara->, Niya *maṃtsa* ‘flesh’ against Sanskrit *maṃsa*, and Khotan *Dharmapada* <ahitsā> ‘in non-violence’ (feminine oblique singular) against Pāli *ahiṃsāya*.

### 5. Some Diagnostic Features to Assess a Likely Gāndhārī Origin

#### 5.1. Voicing of Intervocalic *s*

Intervocalic Sanskrit *s* became voiced in Gāndhārī. Since the orthography of the *Book of Zambasta* distinguishes clearly between unvoiced \*/s/ (= <s>) and voiced \*/z/ (= <ys>), this is a case in which Sanskrit loanwords with intervocalic <s> such as <bodhisatva-> ‘bodhisattva’, <sarvasatva-> ‘all beings’, and <asāra-> ‘insubstantial’ contrast with likely Gāndhārī loanwords such as the following:

<aysura-> [azura] ‘*asura*/demigod’.

<Aysāta-> [azēʔa-] ‘The seer Asita’.

<Kailāysa-> [kaila:za] ‘Mount Kailāsa’.

<Bimbāysāra-> [bimbēza:ra-] ‘King Bimbisāra’.

<Vyāysa-> [wja:za-] ‘The seer Vyāsa’ (traditional author of the *Mahābhārata*).

#### 5.2. Monophthongisation of Open-syllable Diphthongs and Raising of the Resulting Mid Vowels

A sound change characteristic of Khotanese concerns the Old Iranian diphthongs *\*ai* and *\*au*, which become, respectively, *ī* and *ū* in Khotanese. For example, Old Iranian *\*staura* ‘beast of burden’ (compare Av. *staōra*) becomes *stūra* in Khotanese, and Old Iranian *\*daiṇu* ‘cow’ (compare Av. *daēnu* and Skt. *dhenu*) becomes Khotanese *dīnu*. An intermediate stage of monophthongisation to the mid vowels *e* and *o*, analogous to the development in Indic, is attested as a loanword in the Gāndhārī documents



of Shanshan/Nuava, where *\*staura* appears as *stora*. This process affected Indic loanwords too, such as:

<kūla->, possibly from Gāndhārī <koḍi-> [ko:ɽi] ‘crore’.

<Jīyavana-> [ji:jawana-], from Gāndhārī <jedavana> [dze:ðəvənə], ‘the Jetavana garden’.

<Śśādūtana-> [šēdūʔana], from Gāndhārī *\*<Śudhodana->* (\*[eud<sup>h</sup>:o:ðənə], compare Sanskrit Śuddhodana.

<Sumīru> ‘Mount Sumeru’.

The last step of the process, the closing of *e, o* into *ī, ū*, cannot be too old, because, again, it affected also Indic loanwords. Yet, the Indic loanwords that display this behavior conform more closely to the phonology of Gāndhārī.

### 5.3. Orthographic Gemination

As mentioned above, orthographic gemination of <t>, <g>, and <ś> serves to preserve their original values, as the non-geminate representation had come to represent different values, namely *\*[ʔ]* for the first two and *\*[ʒ]* for the third. Transparent Sanskrit loanwords often display such geminate norm:

<āyattana-> [a:yatana-], compare Sanskrit *āyatana* ‘base of the senses.’

<kuśśalamūla->\*[kuš:alamu:la-] compare Sanskrit *kuśalamūla* ‘roots of good’.

<tathāggattagarbha->, compare Sanskrit *tathāgatagarbha* ‘containing a buddha.’

<nayutta-> [nayuta-], compare Sanskrit *nayuta* ‘10<sup>4</sup>’

<prītīsukha-> [pri:tisukha-] Sanskrit *prītisukha* ‘pleasure and joy [that arise from meditation]’.

<viśśeṣa-> [wišeṣa-], compare Sanskrit *viśeṣa* ‘distinction, variety’.

Conversely, the absence of gemination hints at Gāndhārī phonology, as in the following examples:

<Aysāta-> [azēʔa-] ‘The seer Asita’.

<Jaṃbvīviya->, <Jaṃbutīva->, <Jaṃbvīya->, <Jaṃbutīta-> [dzambuʔi:ʔa], from Gāndhārī <Jaṃbudiva-> [dzəmbud:i:və] or

[dzəmbuði:və] ‘Jambudvīpa, the Indian subcontinent’, compare Sanskrit Jambudvīpa.

<nātāma-> [nēʔa:ma-], from Gāndhārī <ṇiama-> [niʔa:mə] ‘established course or method’, compare Buddhist Sanskrit *nyāma* and *niyama* as well as Pāli *niyama* and *niyāma*.

<vīvāga->, <vātāga->, <vīvāta->, <vivāta->, <vāvāta-> [wiʔa:ʔa], from Gāndhārī <vivaa->, <vivaka->, <vivava->, <vivaka-> [viva:jə] ‘ripening [of the cosmic requital of deeds]’, compare Sanskrit and Pāli *vipāka*.

<vyātarāta-> [wja:ʔarēʔa] ‘prophesied’, compare Gāndhārī <vagarana-> [va:jərəŋə] ‘prophesy’, from Sanskrit *vy-ā-√kr* ‘to prophesy’ (in the Buddhist idiom). In spite of the medial *a*, the Khotanese form is likely to stem from Gāndhārī <vyagrida> [vja:grəðə].

A special case concerns *initial* gemination, in which Sanskrit consonants would typically have been sheltered from phonetic change in Gāndhārī, as in the following:

<ttāryasūni-> [tērjažu:ni] ‘animal’ (literally ‘of oblique birth’), analogous to Sanskrit *tiryagyoni*. The compound is not attested in Gāndhārī, but the elements <tiya-> and <yoni-> are. The conjectural Gāndhārī would have been \*<tiyajoni-> [tij:əɖzoni], with fortition of medial -y-.

<Śśādūtana-> [šədu:ʔana]. The Gāndhārī form is unattested but the expected outcome would have been \*[ɕud<sup>h</sup>:o:ðənə], admittedly almost identical to Sanskrit Śuddhodana: compare, however, the different outcomes of intervocalic *ddh* and *d*.

The whole situation is further complicated by the fact that orthographic gemination sometimes represents actual phonetic gemination. Doug Hitch has shown that <tt> represents both \*[t] and \*[t:].<sup>17</sup> In the case of <anicca-> [anič:a] ‘impermanence’ from Gāndhārī <anica-> [ənite:ə] we may have another case of orthographic gemination expressing phonetic gemination.

<sup>17</sup> See Doug Hitch, “Tt in Old Khotanese,” *Journal of the American Oriental Society* 135.4 (2021): 663–687.



#### 5.4. Voicing and Fricativisation of Intervocalic *p* into *v*

Voicing and fricativisation of intervocalic *p* into *v* took place independently and in parallel in Gāndhārī (as in most Middle Indic languages) and in Khotanese. Theoretically, if the change in Khotanese is recent relative to the beginning of the written record of Khotanese, Sanskrit loanwords with intervocalic *p* may have undergone this phonetic change. However, the words that display this specific feature seem to stem from Gāndhārī as they display other gandhārisms, like <dīva-> [di:wa] ‘island’ from Gāndhārī <diva-> [di:və] ‘island, continent’: compare Sanskrit *dvīpa* and note the Middle Indic progressive assimilation *dv* > *d*. In a similar vein, the verbal root <upev-> ‘produce’ is conjectured to come from unattested *p*-causative to the Sanskrit verb *ut-√pad* ‘arise’.<sup>18</sup> The expected Gāndhārī form would have been \*[up:əðəv-]; compare to Pāli *uppādeti* and Sanskrit *utpādayati*. Note the Middle Indic regressive assimilation *tp* > *p*:

#### 5.5. Deaspiration of Voiced Stops

Khotanese lacked the aspirated voiced stops of Indic languages (*g<sup>h</sup>*, *j<sup>h</sup>*, *ḍ<sup>h</sup>*, *d<sup>h</sup>*, *b<sup>h</sup>*), and so it is entirely unsurprising that the orthography of the *Book of Zambasta* very often omits the written aspiration. However, a consistent pattern is that the loanwords more prone to a deaspirated orthography tend to show also traces of Gāndhārī phonology:

<bhūma ~ būma-> [bu:ma-], compare Gāndhārī *bhumi* and Sanskrit *bhūmi* ‘earth, stage’. Note the thematicisation of the stem, on which see the following section.

<avidharma-> [abidarma-], compare Sanskrit *abhidharma* ‘scholastic philosophy’. Note the *v* representing the expected Gāndhārī phoneme <vh> [v<sup>h</sup>], from Sanskrit intervocalic *bh*.

<Śśādūtana-> [šēdu:ʔana], from Gāndhārī \*<Śudhodana-> [çud<sup>h</sup>oðəna], essentially identical with Sanskrit Śudhodana.

<badratalpiya- ~ bhadrratalpiya-> [badraʔalpija-], compare Sanskrit *bhadrakalpika* ‘related to the fortunate æon.’ The conjectural Gāndhārī would be along the lines of \*<bhadrayapiya> [b<sup>h</sup>ədɾəp:iə].

<sup>18</sup> See Ronald E. Emmerick, *Saka Grammatical Studies* (London: Oxford University Press, 1968), 14.

<irdi> [irdi], from Gāndhārī <irdhi-> ‘supernatural power’, compare Sanskrit *ṛddhi*. See however also the Sanskrit form within the compound *riḍdhāpāta* corresponding to Sanskrit *ṛddhipāda* ‘idem’.

### 5.6. Lowering of *i* into *a*

Several words show an Indic *i* lowering to *a* in unstressed syllables, as in the following examples:

<pārāmatā-> [pa:ra:maʔa:] possibly from Gāndhārī <paramida-> [pa:rəmiða:].

<candāvani-> [čanda:wani], from Gāndhārī or Sanskrit *cintāmani* ‘wish-granting jewel’. Intervocalic *m* > *v* is attested in the Gāndhārī of the Khotan *Dharmapada*.<sup>19</sup>

<Kālāsundhari-> [ka:š:asundari], in all likelihood from *Kāśīsundarī* ‘the beauty of Benares’ (probably through a Gāndhārī form \*[ka:ç:isund<sup>h</sup>ari].

It is in stem-final position that Indic *i* is most likely to open into *a*, e.g. Indic *bhūmi* > *Book of Zambasta* <bhūma ~ būma>, leading to such nominal stems to become effectively thematicised.

Distinguishing thematic *-a* stems and *-i* stems is difficult because the nominative singular case ending for both types—by far the most common form—are identical (*-ä*). However, it is not the case that all Indic *-i* stems become ‘thematicised’ in this way: *dhāraṇi* (> Skt. *dhāraṇī*) ‘incantation’ remains an *i*-stem as shown by the genitive plural form *dhāraṇānu*, which would have been *\*dhāraṇānu* had *dhāraṇī* been borrowed as *\*dhāraṇa*- and not as *\*dhāraṇi*. Once again, the words that feature this shift seem in general to be closer to the phonology of Gāndhārī.

### 5.7. Weakened *-ka* Stems

The *-ka* suffix is common to Indic and Iranian and very productive in both language families. Both Gāndhārī and Khotanese suffered a weakening of the medial intervocalic *k*, yielding, respectively, <-aka ~ -a<sub>k</sub>a ~ -aga ~ -aya> [-əjə] and *-aa*: therefore, stems in *aa*- in Khotanese can be either

<sup>19</sup> See Oskar von Hinüber, *Das Ältere Mittelindisch im Überblick* (Vienna: Verlag der Österreichischen Akademie der Wissenschaften, 1986), §210.





indigenous Khotanese words or Indic loanwords. As in the preceding cases, however, Indic loanwords in *-aa* tend to conform more closely to Gāndhārī phonology, e.g. *nālaa*- ‘theatre play’, related to Sanskrit *nāṭaka* but showing Middle Indic  $t > d > l$ .

### 5.8. Intervocalic $d > l$ Oscillation

Although there is some oscillation throughout Sanskrit and Middle Indic between  $d$  and  $l$ , Middle Indic shows generally a preference for  $l$  in words in which such oscillation is attested, with some hypercorrect forms in Hybrid Buddhist Sanskrit with  $d$ .<sup>20</sup> Intervocalic  $d$  (often from a Sanskrit intervocalic  $t$ ) developed in Gāndhārī into a retroflex flap  $*[ɽ]$  that was still written  $<d>$ . Intervocalic  $d$  in the Middle Indic loanwords of the *Book of Zambasta* shows consistently  $l$ : it should be kept in mind, though, that Old Iranian rhotic clusters such as  $*-rd-$  yielded  $l$  in Khotanese (e.g., *kamala*- ‘head’, from Old Iranian  $*kamrda$ , see Avestan *kamərəda*), whereas the related Tumšūqese preserved them: conceivably, the intermediate stage may have been a retroflex rhotic sound close to Gāndhārī  $*[ɽ]$ . If this was the case, loanwords containing  $*[ɽ]$  would have suffered the same fate as the indigenous Khotanese words. Some likely Gāndhārī loanwords are the following:

$<kūla->$  ‘crore’ from Gāndhārī  $<koḍi->$   $[ko:ɽi]$ , compare Sanskrit *koṭi*.

$<nālaa->$  ‘theatre play’, akin to Sanskrit *nāṭaka*, but displaying Middle Indic intervocalic  $t > d > l$ .

$<vārūlaa->$  ‘beryl’, close enough to Gāndhārī  $<veḍurya->$   $[vəɽu:rijə]$  but actually closest to Ardhamāgadhī *veruḷiya* and the Greek loanword *béryllos*, in any case further away from Sanskrit *vaiḍūrya*.

Characteristic of the hypercorrective tendency of Buddhist Hybrid Sanskrit in this point would be instead  $<vaittāḍa->$   $[vetāḍa]$ , Classical Sanskrit *vetāla*, but Buddhist Hybrid Sanskrit *vetāḍa* ‘revenant, vampire’.

<sup>20</sup> See Franklin Edgerton, *Buddhist Hybrid Grammar and Dictionary, Volume I: Grammar* (New Haven: Yale University Press, 1953), §246.

### 5.9. Gemination of Palatals Prefixed with *l*

Ronald E. Emmerick has shown that palatals preceded by a written *l*, which may have represented a palatal lateral (\*[λ]), often represent geminates.<sup>21</sup> The origin of this specific orthographic device is unclear, but a further puzzling element is the fact that these are typically words in which the palatal quality stems from a Sanskrit cluster with *y-*, at least in two cases showing the trademark Gāndhārī development, *ṣy* > *-ś*. The words spelled in this manner show traces of Gāndhārī phonology:

<Kālśasundhari-> [ka:ś:asundari], >Kāśīsundarī ‘the beauty of Benares’, perhaps reflecting a Gāndhārī \*<-ś-> [-ç:-] from *-śy-*.

<Kālśava-> [ka:ś:ava], from Gāndhārī <kaśava-> [kaç:əvə], compare Han Chinese rendering \*Kaijap (迦葉) and Sanskrit Kāśyapa.

<puśa-> [puš:a], from Gāndhārī <puśa-> [puç:ə], compare Sanskrit *puśya* ‘the asterism Puśya’.

<puñā-> [puñ:a-], from Gāndhārī <pu[ṃ]ñā> [puṃ:ə] ‘merit’. This spelling is found only once (11.37), otherwise *puñā-* (still with geminate ñ, \*[puñ:a-]?) is the norm.

### 5.10. Intervocalic \**t<sup>h</sup>*, \**d<sup>h</sup>* > *h*

As mentioned above, Sanskrit *t<sup>h</sup>* and *d<sup>h</sup>*, which in ‘Metropolitan Gāndhārī’ regularly yield \*[z], fail to do so in many—not all—of the presumed Gāndhārī loanwords in the *Book of Zambasta*, showing instead *h*. This feature is shared with the dialect of the Khotan *Dharmapada*. For example:

<Anāhapiṇḍiya->, compare Pāli and Buddhist Hybrid Sanskrit Anāthapiṇḍika (opposed to later Buddhist Sanskrit Anāthapiṇḍada). *-iya-* from Gāndhārī \*[-ijə-]>*-ika* seems likely here.

<karmapaha->, compare Buddhist Hybrid Sanskrit *karmapatha* ‘path of action’.

<prañāhāna-> [prañēhāna]. The attested Gāndhārī form is <praniṣana-> [prəniza:nə], but the outcome of the Sanskrit gerund *nidhāya*

<sup>21</sup> Ronald E. Emmerick, “The Dunhuang MS. Ch 00120: Its Importance for Reconstructing the Phonological System of Khotanese,” in *Turfan and Tun-huang: The Texts: Encounter of Civilizations on the Silk Route*, ed. Alfredo Cadonna (Florence: Leo S. Olschki Editore, 1992), 167.



‘having put down’ differs in the Gāndhārī *Khadgaviṣāṇasūtra* (*ṇiṣae*) from the Khotan *Dharmapada nihaī*.

<samāhāna->, compare Sanskrit and Pāli *samādhāna* ‘concentration’, instead of the ubiquitous Sanskrit *samādhi* ‘idem’ which occurs in *Zambasta* (10.16–27); contrast with Gāndhārī <*samaṣi*-> [səma:zi].

### 5.11. Contraction

A few Indic loanwords show fairly drastic processes of contraction such as the following:

<bilsaṃgga-> [bilsaṅga-], from Sanskrit or Gāndhārī *bhikṣusaṃgha* ‘monastic assembly’, likely through contraction of the unaccented second syllable \*[b<sup>h</sup>ikṣu-] > \*[b<sup>i</sup>ṣə-] > \*[b<sup>i</sup>zə-] > <bil->.

<praśnātarāṇa-> [praśnā<sup>?</sup>arāṇa], compare Buddhist Sanskrit *praśnavyākaraṇa* and Pāli *pañhavyākaraṇa* ‘elucidation of questions’, as opposed to the Gāndhārī of the Khotan *Dharmapada* in the form <praśāṇa>, which scans disyllabically and must be presumably reconstructed to \*[prəṣṇə].

<prāmūkṣa-> and <prātāmūkṣa-> [pra:(<sup>?</sup>ə)mukṣa], compare Sanskrit *prātimokṣa* ‘binding [monastic rules]’ Khotan *Dharmapada pradimukha*-.

<prārhālya->, <prāhālya->, compare Pāli *pāṭihāriya* Buddhist Hybrid Sanskrit *prātihārya* ‘miracle.’ The Sanskrit prefix *prati* manifests in Gāndhārī variously as <prati-> ~ <paḍi> ~ <praḍi> [p(ɾ)əṭi]. The Khotanese form may represent a contraction from a conjectural Gāndhārī \*[p(ɾ)əṭ[i]ha:rijə] with dissimilation of *r* into *l*.

The extreme transformations of these words make it difficult to establish their origin: however, specific Gāndhārī developments such as *śn*>*śñ* and Khotanese changes that seem to have affected only Gāndhārī loanwords, such as *o*>*ū*, make it likely that forms experiencing this level of contraction be old Gāndhārī loanwords as well.

## 6. Thematic Analysis

### 6.1. Toponyms

The toponyms identified in this survey as showing traces of Gāndhārī phonology (e.g., Kavilavāstu=Kapilavastu, Kośāmi=Kauśāmbī, Jīyavana=Jetavana, Bārāṇyasi=Vārāṇasī) are all closely related to central events in the basic life-narrative of the Buddha. The Sanskrit toponyms feature instead the typical backdrops for the Buddha’s preaching and may therefore be linked with the *sūtra* genre. It is interesting to note, in fact, that although the amount of Gāndhārī and Sanskrit remains of Buddhist literature in the first four centuries of the Common Era is substantial, the *sūtra* genre is almost unattested in this period: the spread of the genre seems to be connected, precisely, to diffusion through Sanskrit. It is also interesting that the name of the abode of Buddha Maitreya, the Future Buddha, Kettumati (=Ketumatī), displays also a Sanskrit form.

#### 6.1.1. Gāndhārī

<Kavilavāstu-> ‘Kapilavastu.’

<Kośāmi-> [koʒa:mi], compare Sanskrit Kauśāmbī, Buddhist Hybrid Sanskrit Kośāmbī, Pāli Kosambi. The conjectural form in the dialect of the Khotan *Dharmapada* would be \*[kojam:i-]. See section 4 above.

<Jīyavana-> [ji:jawana-], from Gāndhārī <jedavana> [dze:ðəvəṇə], ‘the Jetavana garden’.

<Bārāṇyasi-> [ba:ra:ṇazi-] ‘Benares’ (Skt. Vārāṇasī).

#### 6.1.2. Sanskrit

<Kettumati-> ‘[Maitreya’s abode] Ketumatī’.

<Rājagrha->

<Vaiśāli-> [vaiša:li], ‘Vaiśāli’.

<Śrāvasti-> ‘Śrāvastī’.

### 6.2. Cosmology

The vast bulk of the vocabulary related to cosmology in this survey shows traces of a Gāndhārī origin. This includes the names of the various realms of the cosmos, the abodes of the gods, and the various types of living creatures that populate the universe. This vocabulary, although not infrequent in the *sūtra* genre, seems particularly akin to Buddhist narrative, which must have circulated through oral and written media.



Tellingly, one of the few Sanskrit terms of note within this category, *kaliyuga* ‘the cosmic age of the demon Kali’ occurs, within the realm of Buddhist literature, only in Mahāyāna texts.

### 6.2.1. *Gāndhārī*

<ātāsa ~ āgāśa> [a:ʔa:ʒa-] ‘sky’, from Gāndhārī <agaśa> [a:j:açə-]; contrast with Sanskrit *ākāśa*.

<aysura-> [azura-] ‘*asura*/demigod’ with Gāndhārī intervocalic *s* > *z*.

<avitsara-> [awits<sup>h</sup>ara-] ‘*apsaras* nymph’. The Gāndhārī outcome of Sanskrit *apsaras* is unattested, as are Middle Indic forms with anaptyxis; Pāli has *accharā*.

<Avīśa->\*[awi:ʒa-] ‘[the hell] Avīcī’ from Gāndhārī <aviya->\*[əvi:jə].

<upala-> ‘lotus’, from Gāndhārī <upala> [up:ələ-], see also Tocharian B *uppāl* against Sanskrit *utpala*.

<Kailāysa-> [kela:za] ‘Mount Kailāsa’.

<kūla-> ‘crore’ from Gāndhārī <koḍi-> [ko:ɽi], compare Sanskrit *koṭi*.

<candāvani-> [čanda:wani], from Gāndhārī or Sanskrit *cintāmani* ‘wish-granting jewel’.

<jambvīviya- ~ jambutīva- ~ jambvīya- ~ jambutīta-> [jambuʔi:ʔa] ‘The Indian subcontinent’ from Gāndhārī <jambudiva-> [dʒambuði:və].

<tāvātrīśa-> [taʔatri:ʒa-] or [tawatri:ʒa-] ‘[related to the heaven] of the thirty-three [gods]’, from Gāndhārī <traetriśa- ~ tritriśa->\*[trəjətɽjə]; contrast with Sanskrit *trāyastriṃśa*, but note the interesting similarity to Pāli *tāvatiṃsa*.

<tāryaśūni-> [tērjaʒu:ni] ‘animal’, Sanskrit *tiryagyoni*. The compound is not attested in Gāndhārī, but the elements <tiya-> and <yoni-> are. The conjectural Gāndhārī would have been \*[tj:ədʒoni], with fortition of medial -y-.

<dīvātā-> [di:waʔa:] ‘deity’ from Gandhārī <devada ~ devata>\*[de:vəða:], compare to Sanskrit *devatā*.

<nāta-> [na:ʔa-] ‘*nāga* snake spirit’, from Gāndhārī <naga-~ṇaga-> [na:jə].

<prārhālya->, <prāhālya->, compare Pāli *pāṭihāriya* Buddhist Hybrid Sanskrit *prāṭihārya* ‘miracle.’ The Sanskrit prefix *prati* manifests in Gāndhārī variously as <prati- ~ paḍi ~ praḍi> [p(r)əɽi]. The Khotanese form may represent a contraction from a conjectural Gāndhārī *p(r)əɽ[i]ha:rijə* with dissimilation of *r* into *l*.

<prīya-> ‘*preta* ghost’ from Gāndhārī <preta-> [pre:ðə-].

<badratalpiya- ~ bhadrratalpiya> [badraʔalpija], compare Sanskrit *bhadrakalpika* ‘related to the fortunate æon.’ The conjectural Gāndhārī would be a form along the lines of \*[b<sup>h</sup>ædrəjəp:ijə].

<magara-> [maʔara] ‘aquatic monster, crocodile’. In spite of the spelling, the form seems to show traces of Gāndhārī intervocalic *k > g > j* and compare with Sanskrit *makara*.

<lova-> ‘world’ and <loviya-> ‘worldly’, from Gāndhārī <loka-~ loa> [lojə] and <loia- ~ logiga>\*[lojjə-]; contrast with Sanskrit *loka*, and *laukika*.

<vaśāra-> [waʒəra] ‘mace of Indra, diamond’ from Gāndhārī <vayira> [vəjjirə]; contrast with Sanskrit *vajra*.

<värūlaa-> ‘beryl’, close enough to Gāndhārī <vedurya-> [vəru:rijə] but actually closest to Ardhamāgadhī *veruḷiya* and the Greek loanword βήρυλλος; in any case, further away from Sanskrit *vaidūrya*.

<śśuddhāvāysa-> ‘pure abode [of the highest category of gods]’, Sanskrit *śuddhāvāsa*. The intervocalic *s > z* is expected for Gāndhārī.

<Sumīra-> ‘[Mount] Sumeru’.

### 6.2.2. Sanskrit

<Kalāyugga-> ‘the æon of the demon Kali’, Sanskrit *Kaliyuga*.

<Asāpattravana-> ‘[the hell of the] forest of knife blades’, Sanskrit *asipattravana*.

<vaittāḍa-> [vetāḍa] ‘revenant, vampire’ Classical Sanskrit *vetāla*, but Buddhist Hybrid Sanskrit *vetāḍa*

<pāśśāca->\*[pəša:ča-] ‘*piśāca* demon’, as opposed to Gāndhārī <*piṣaya*>.

### 6.3. Anthroponyms and Theonyms

Personal names in the *Book of Zambasta* show diverse origins, and no clear pattern emerges. For example, the names of the disciples of the Buddha, Kātyāyana and Maudgalyāyana, show clearly Sanskrit phonology, but Kāśyapa is termed with a form certainly loaned from Gāndhārī (*Kāśava*). It is interesting that various personal names related to the two Sanskrit Epics *Mahābhārata* and *Rāmāyaṇa* that occur in an extraordinary passage (5.1–10) do show traces of Gāndhārī phonology. Of particular interest, but highly puzzling, is the form *Sīysā* ([si:za:]) for Sītā, the heroine of the *Rāmāyaṇa*. The form is undoubtedly related—and possibly ancestral—to Late Khotanese *Sīysā*, Tocharian A *Sīsā*, and Old



Uyghur *Siza*. Again, in ‘Metropolitan Gāndhārī’ [z] is, among others, the outcome of intervocalic Sanskrit [th] and [dh], e.g., Sanskrit *bodhi* ‘awakening’ > Metropolitan Gāndhārī <*bosi*> [bozi] and Sanskrit *śamatha* ‘quiet’ > Metropolitan Gāndhārī <*śamasa*> [çəmæzə]. The Khotanese form could be made sense of by postulating a source form \**Sīthā* (or \**Sīdhā*) for a hypothetical Metropolitan Gāndhārī \*<sisā> [si:zɑ:], which seems not to be attested. Turner however mentions a form *sihā* ‘furrow’ (i.e., the primal meaning of Sanskrit *sītā*) from the Bhalesi language spoken in central Kashmir.<sup>22</sup> If the hypothesised etymon \**Sīthā* is behind these forms, in Bhalesi we would have [th, dh] > [h] instead of Metropolitan Gandhārī [th, dh] > [z]. The *h*-form of Bhalesi would be most closely related to the dialect of Gāndhārī distinct from ‘Metropolitan Gāndhārī’ described in section 4 above.

It is also interesting to note that the names of the mahāyānic celestial Bodhisattvas Ākāśagarbha and Kṣitigarbha (spelled <Ākāśaggarbha> and <Kṣāttāggarbha>) appear in their Sanskrit forms.

### 6.3.1. Gāndhārī

<Anāhapiṇḍiya->, compare Pāli and Buddhist Hybrid Sanskrit Anāthapiṇḍika (opposed to later Buddhist Sanskrit Anāthapiṇḍada). *-iya-* from Gāndhārī \*[-ijə-] > *-ika* seems likely here.

<Aysāta-> [azēʔa-] ‘The seer Asita’.

<Kālodāti-> [ka:loda:ʔi], compare Sanskrit Kālodāyin, Pāli Udāyi, conjectural Gāndhārī \*[ka:lōḍa:ji].

<Kālsāva-> [ka:š:ava], from Gāndhārī <kaśava-> [kaç:əvə], compare Han Dynasty Chinese rendering \*Kaijap (迦葉) (in all likelihood a loanword from Gāndhārī) and Sanskrit Kāśyapa.

<Kaṃsadāyisa-> [kansada:za], only attested in the instrumental singular form *kaṃsadāyisna* (5.2), equivalent to Sanskrit Kaṃsadāsa ‘slave of Kaṃsa’ an epithet of Kṛṣṇa (compare *Mahābhārata* 9.60.27 and *Viṣṇupurāṇa* 5.27.13), possibly showing Gāndhārī *s*>*z*, unless the \*[z] is the outcome of voicing in contact with the instrumental ending *-na*.

<Ggoviyā- ~ Ggaupiyā-> [gowija:ʔ], compare Buddhist Hybrid Sanskrit Gopikā (one of the wives of the future Buddha). The conjectural Gāndhārī form would be \*[govija:].

<sup>22</sup> Ralph Lilley Turner, *A Comparative Dictionary of Indo-Aryan Languages* (London: Oxford University Press, 1962–1985), §13428.



<Daśagrīva-> [daʒagri:wa], compare Sanskrit *Daśagrīva* ‘He of the Ten Necks’ (epithet of Rāvaṇa). The conjectural Gāndhārī form would be \*[dəʒəgrī:və], and the lack of written gemination (<śś> [ʃ]) favors slightly a Gāndhārī origin.

<Dīvaṃggara-> [diwaŋgara-], from Gāndhārī <Dhivhakara-> [di:vəŋgərə] ‘[the past buddha] Dīpaṃkara’.

<Bimbāysāra-> [bimbəza:ra-] ‘[king] Bimbisāra’.

<Vaśārapāna-> [vaʒəpa:na-] ‘[the *yakṣa*] Vajrapāṇi’, compare Gāndhārī <vaṃira> [vəjirə], but with preservation of intervocalic *p*. The more Sanskritic form Vajrapāṇa- also occurs, though, in 4.8.

<Vyāyasa-> [wja:za-] ‘The seer Vyāsa’ (traditional author of the *Mahābhārata*).

<Śśāya-> [ʃa:ya] ‘Śākya’, from Gāndhārī <śahia> [ʃa:jiə], but see also Sanskrit Śśākyamuni- elsewhere.

<Śśādūtana-> [ʃədūʔana], from Gāndhārī \*Śśudhodana- (\*[eud<sup>h</sup>:o:ðənə]), compare Sanskrit Śuddhodana.

<Sīysā-> [si:za:], Sītā (see discussion above).

### 6.3.2. Sanskrit

<Ākāśśaggarbha-> ‘[the mahāyānic *bodhisattva*]’ Ākāśagarbha.

<Ekaśrṅga-> ‘[the seer] Ekaśrṅga’

<Kātyāyana->, ‘[the disciple] Kātyāyana’. Contrast with Gāndhārī <Kacana-> [kəte:a:nə] and Pāli Kaccāna.

<Kṣāttāggarbha-> [kṣītigarba], ‘[the mahāyānic *bodhisattva*] Kṣītigarbha’

<Mudgalyāyana->, ‘[the disciple] Maudgalyāyana’ (the forms in Mudg~ and Modg~ are attested in Buddhist Hybrid Sanskrit). Contrast with Gāndhārī <Mogalana-> [mo:gəl:a:nə], loaned into Han Chinese as \*Muklian 目連, and Pāli Moggallāna.

<Valmīki-> ‘[the sage] Valmīki’ (traditional author of the *Rāmāyaṇa*).

<Sarasvati-> ‘the goddess Sarasvatī’. The shift *sv* > *sp* would have been expected for Gāndhārī.

### 6.4. Monasticism

The vocabulary of Buddhist monasticism in the *Book of Zambasta* shows clear traces of an early transmission through a Gāndhārī medium. This would be consistent with an introduction of Buddhism into Khotan concurrent with the institution of Buddhist monasticism. In the



neighbouring Shanshan/Nuava during the third and fourth centuries, Gāndhārī-speaking monks (Gand. *ṣamaṃṇa*) were employed as clerks, scribes, and agents for the state, and there is a possibility that during the time of strongest Kuṣāṇa influence on Khotan, Buddhist monks may have also been associated with the administration of a state under the Kuṣāṇa aegis. Within the terminology of everyday monasticism, transparent Sanskrit loanwords seem rather rare in the sample examined here: perhaps only the compound *kāṣāyavastra*- ‘monastic robes’ fits this category. The exception are perhaps technical terms having to do with meditation: <prītisukha-> [pri:ti.suk<sup>h</sup>a] > Sanskrit *prītisukha* ‘pleasure and joy [that arise from meditation]’ and <vajropama-> > Sanskrit *vajropama* ‘in the likeness of a diamond’, in reference to absorption, a usage found typically in certain Mahāyāna *sūtras*, in particular in the *Prajñāpāramitā* [Perfection of Wisdom] literature.

#### 6.4.1. Gāndhārī

<arahanda-> ‘enlightened person’, from Gāndhārī <arahada ~ arahamta> [əɾəhəðə], compare Sanskrit *arhat* (nasal-infix strong stem *arhant*) and Pāli *arahant*.

<ggāṭhaa-> [gaṭ<sup>h</sup>:aa-] ‘householder (=non-monastic)’, from Gāndhārī <ghahaṭha-> [g<sup>h</sup>əhəṭ<sup>h</sup>ə(jə)]? ‘, compare Sanskrit *grhastha*.

<jāna->, from Gāndhārī <jana-> [dʒ<sup>h</sup>a:nə] ‘meditation’, compare Sanskrit *dhyāna*. However, the compound *dhyānapārāmatā* (10.8, 10.17. 10.27) for Sanskrit *dhyānapāramitā* shows the Sanskrit form.

<trvīlaa-> [triwi:laa-], from Gāndhārī <trepiṭaka-> [treviṭəjə] ‘knower of the Three Baskets [of canonical scripture]’, compare Sanskrit *traipitaka* and Pāli *tepiṭaka*.

<ttārthiya-> [tirt<sup>h</sup>i:ja-], from Gāndhārī <tirthiga-> [tirt<sup>h</sup>ijə] ‘non-Buddhist ascetic,’ compare Buddhist Hybrid Sanskrit *tīrthika* and Pāli *tīthiya*.

<pāṇḍāvāta-> [pəṇḍəwa:ʔa] ‘alms collection’ from Gāndhārī <piṇḍavada-> [piṇḍəva:ðə], compare Buddhist Hybrid Sanskrit and Pāli *piṇḍapāta*.

<bilsaṃga-> [bilsaṅga-], from Sanskrit or Gāndhārī *bhikṣusaṃgha* ‘monastic assembly’, undoubtedly through contraction of the unaccented second syllable \*[b<sup>h</sup>ikṣu-] > \*[biṣə-] > \*[biṣə-] > <bil->.

<riṣaya-> [riʒaja], from Gāndhārī <[r]iṣea-> [[r]iʒe:ə-] ‘[brahmanical] holy man, compare Sanskrit *ṛṣi* and Pāli *isi*. See however the Sanskrit form within the name Riṣādatta- (=Rṣidatta).

<līna-> [li:na-], compare Pāli *leṇa* and Sanskrit *layana* ‘monastic cell’. A conjectural Gāndhārī form would be \*[le:nə].

<vaysambatā-> [wazambaða:], from Gāndhārī <vaṣapaḍa-> [vəzəpəða:] ‘monastic ordination’.

<vūsāta- ~ pūysāta-> [wu:zēʔa], from Gāndhārī <posadha-> [pozədʰə] ‘day of religious observance’, compare Pāli *uposatha* and Buddhist Hybrid Sanskrit *poṣadha*, *poṣatha*, *uposatha*, *upoṣadha*, Sanskrit *upavasatha*.

<saṃkhārama- ~ saṃkhārma->, compare Pāli and Buddhist Sanskrit *saṃghārāma* ‘retreat of the monastic assembly, monastery’, although the form *saṃghārama* with a short penultimate *a* is attested sporadically in Pāli and in Apabhraṃśa, which would more easily explain the syncopated Khotanese form. The devoicing (with retention of aspiration) of *gh* into *kh* is extraordinary and unusual.

<śākṣāvata-> [śikṣa:waʔa] ‘basis of [monastic] training’, from Gāndhārī <śikṣavada> [çikṣa:va:ðə], compare Sanskrit *śikṣāpada*.

<ṣamana-> [ṣamaṇa], from Gāndhārī <ṣamana-> [ṣəmənə] ‘ascetic’, compare Sanskrit *śramaṇa*.

<ṣāmaṇa-> [ṣa:maṇ:a], from Gāndhārī <ṣamaṇña-> ṣa:məj:ə ‘asceticism’, compare Sanskrit *śrāmaṇya* and Pāli *sāmaṇña*.

<samāhāna->, compare Sanskrit and Pāli *samādhāna* ‘concentration’, instead of the ubiquitous Sanskrit *samādhi* ‘idem’ which does occur in *Zambasta* (10.16–27), and ‘Metropolitan Gāndhārī’ <*samaṣi*-> [səma:zi].

<sthīra-> [stʰi:ra-], from Gāndhārī <thera ~ sthāira> [(s)tʰe:rə] ‘monastic elder’, compare Sanskrit *sthavira* and Pāli *thera*.

<cīya-> [çi:ya-] ‘religious monument, shrine’ from Gāndhārī <cediga- ~ cetiga- ~ cetiya-> [tee:ðijə], compare Sanskrit *caitya*.

<parṣā-> ‘monastic assembly’ from Gāndhārī <pariṣa> [pəriʒa:], compare Buddhist Hybrid Sanskrit *parṣā* and *pariṣā* along with Sanskrit *pariṣad*.

#### 6.4.2. Sanskrit

<prītīśukha-> [pri:tisukha] > Sanskrit *prītīśukha* ‘pleasure and joy [that arise from meditation].



<vajropama-> Sanskrit *vajropama* ‘in the likeness of a diamond’ (typical attribute if certain forms of meditation in the *Perfection of Wisdom* literature).

### 6.5. Doctrine

A large proportion of the borrowed terms for doctrinal terms in the *Book of Zambasta* are taken from Sanskrit and would be too numerous to list here. A remarkable pattern is, however, that whereas the Sanskrit vocabulary in general is represented by elaborate technical compounds, the Gāndhārī element consists of widespread *simplices* that express core concepts of Buddhist doctrine, such as those that pertain to the description of the ripening of the cosmic requital of deeds (Skt. *karmavipāka*), interdependent causation, the realms of rebirth, the marks of existence, etc. Of especial interest here is also that vocabulary that is virtually exclusive to the Mahāyāna genre, such as ‘restraints of the *bodhisattva*’ (Skt. *bodhisattvasamvara*) and ‘vehicle of the solitary buddhas’ (Skt. *pratyekabuddhayāna*).

#### 6.5.1. Gāndhārī

<anica-> [anič:a] ‘impermanence’ from Gāndhārī <anica-> [ənite:ə]; contrast with Sanskrit *anitya*.

<avāya-> ‘destruction’, from Gāndhārī <avaya->; contrast with Sanskrit *apāya*.

<upev-> ‘produce’, conjectured to come from unattested *p*-causative to the Sanskrit verb *ut-√pad* ‘arise’.

<pārāmatā-> [pa:ra:maʔa:] ‘perfection’ possibly from Gāndhārī <paramida-> [pa:rəmiða:]; compare Sanskrit *pāramitā*. The form <pāramī> also exists.

<pulña-> [puñ:a:], from Gāndhārī <pu[m]ña> [puɲ:ə] ‘merit’. This spelling is found only once (11.37), otherwise *puña-* (still with geminate ñ, \*[puñ:a-]?) is the norm.

<pracaa-> ‘element’, from Gāndhārī <p(r)acea-> [p(r)ətɛejə-]; contrast with Sanskrit *pratyaya*.

<phāṣṣa-> ‘comfortably’, from Gāndhārī <phaṣa->, compare also Buddhist Hybrid Sanskrit *phāṣa*; the etymological Sanskrit cognate *sparśa* ‘touch’ lacks this meaning.

<mamaṅkāra-> ‘sense of ownership’, closest to Pāli *mamaṅkāra*; contrast with Sanskrit *mamakāra*.

<rūva-> ‘form’ in the compounds *buddharūva-* (*buddhrūpa*), *arūvaddāta-* (*ārūpyadhātu*), *rūvakāya-* (*rūpakāya*), from Gāndhārī <ruva>.

<viñāna- ~ vāñāna-> ‘consciousness’ from Gāndhārī <viñāna>, compare Sanskrit *viññāna*, which is attested too (5.82).

<vimūha-> [wimu:ha], from Gāndhārī <vimoha-> [vimohə] ‘liberation’; compare Sanskrit *vimokṣa*. Gāndhārī tends to retain the *kṣ*: in this case we are likely faced with the Gāndhārī weakening of a form \**vimokkha* inherited from a Gangetic Middle Indic language.

<vīvāga->, <vātāga->, <vīvāta->, <vivāta->, <vāvāta-> [wiʔa:ʔa], from Gāndhārī <vivaa->, <vivaḱa->, <vivava->, <vivaka-> [viva:jə] ‘ripening [of the cosmic requital of deeds]’, compare Sanskrit and Pāli *vipāka*. The Sanskrit form *vipāka* is however also used side by side with the Gāndhārī form throughout Chapter 4 of the *Book of Zambasta*, with no apparent difference in meaning.

<śūṇa-> ‘empty’, from Gāndhārī <śumña> [çuŋ:ə]; contrast with Sanskrit *śūnya*.

<ṣaddā-> ‘faith’, akin to Khotan *Dharmapada* <ṣadha-> [ṣəd<sup>h</sup>:a:]; contrast with Sanskrit *śraddhā* and note the trademark Gāndhārī shift *śr>ṣ*.

<ṣāvaa-> expected Gāndhārī \*<ṣavaya> [ṣəʋəjə] ‘disciple of the Buddha,’ Sanskrit *śrāvaka*.

<saṃñā-> [sañ:a:] ‘perception’, from Gāndhārī <saṃña> [səŋ:a:]; contrast with Sanskrit *saṃjñā*.

<saṃtsāra> [sants<sup>h</sup>a:ra] ‘cycle of rebirth’ akin to Khotan *Dharmapada* <satsara> against ‘Metropolitan Gāndhārī’ <saṃsara-> and Sanskrit *saṃsāra*.

<samai-> [same] ‘correct’, from Gāndhārī <same->; contrast with Sanskrit *samyañc* (*samyak-* and *samyag* in composition).

<vīra-> ‘hatred’, from Gāndhārī <vera->; contrast with Sanskrit *vaira*.

### 6.5.2. Sanskrit

<āyattana-> ‘base of the senses’, from Sanskrit *āyatana*.

<ahaṃkāra-> ‘sense of ego’.

<upekṣa-> ‘indifference’, from Sanskrit *upekṣā*.

<klaīśa ~ kleśa> ‘defilement’, from Sanskrit *kleśa*.

<tathāggattagarbha-> ‘[doctrine of the] inner Buddha [nature]’, from Sanskrit *tathāgatagarbha*.

<nairātma-> ‘egolessness’, from Buddhist Sanskrit *nairātma*.



- <paramārtha-> ‘supreme truth’.
- <pratyekayāna-> ‘vehicle of the *pratyekabuddhas*’, from Sanskrit.
- <bodhisatvasaṃvara-> ‘restraints of *bodhisattvas*’.
- <upāsakasamvara-> ‘restraints of *upāsakas*’.
- <bodhyāṅga-> ‘component of enlightenment’.
- <smṛtyupasthāna-> ‘application of mindfulness’.
- <svabhāva-> ‘own nature’.

## 6.6. Text Titles

One widespread pattern on which there will be further occasion to comment is that although the vocabulary of the Mahāyāna in the *Book of Zambasta* typically displays Sanskrit features, the *Book of Zambasta* references several Mahāyāna texts by name, and these show instead traces of Gāndhārī phonology. Most of the text titles and names of textual genres do display, in fact, traces of a Gāndhārī origin. Below follow some examples:

<avidharma-> ‘Abhidharma/Scholastic philosophy’.

<praśnātarāṇa-> [praśnāṭaraṇa], compare Buddhist Sanskrit *praśnavyākaraṇa* and Pāli *pañhavyākaraṇa* ‘elucidation of questions’; the Khotan *Dharmapada* has <praśāṇa>, which scans disyllabically and must correspond to \*[præɟnə].

<prāmūkṣa-> and <prātāmūkṣa-> [pra:ʔ(ē)mukṣa], compare Sanskrit *prātimokṣa* ‘binding [monastic rules]’.

<Bārata-> [ba:raʔa-], ‘the [*Mahā*]bhārata epic’.

<Buddhavalamṭsaa->, Sanskrit *Buddhāvataṃsaka*. The *-valamṭsa* segment is closely akin to the Middle Indic form attested in Pāli *vaṭamṣa* ‘ornament’ and not to Sanskrit *avatamṣa*, showing Middle Indic *ṭ>ḍ>l*. Prolexis 251

<Ratnakūla->, Sanskrit *Ratnakūṭa*, with Middle Indic *ṭ>ḍ>l*.

<Rāmāyana->, ‘the Rāmāyaṇa epic’.

<Vibhāṣa->, ‘the [*Mahā*]vibhāṣa’ [i.e. the Great Commentary on scholastic philosophy of the Sarvāstivādins].

<Vināa-> [winaʔa] ‘*vinaya*’ [i.e. monastic discipline], but also see the element *vinaya* in the mention of the text *Vinayavinīścaya* (= *Upālipariṣcchā*?) in 10.33.

<Sandāvāta-> [sandēwa:ʔa], compare Sanskrit [*Mahā*]saṃnipāta. Mauro Maggi is very probably correct in seeing here a hypercorrect form, where a geminate *-n:* was restored to *-nd-* from a conjectural form \*[sən:iɪva:ðə] where the progressive assimilation of nasal-occlusive clusters proper to the Gāndhārī in use in Khotan but not to ‘Metropolitan Gāndhārī’ would have led to the impression that the geminate [n:] had originated in one of such clusters.<sup>23</sup>

### 6.7. Minimal Pairs of Contrast between Mainstream Buddhist Vocabulary and Mahāyāna Vocabulary

The shift from Gāndhārī to Sanskrit as a source of Buddhist technical vocabulary is that whereas *simplices* often display a Gāndhārī guise, longer compounds are borrowed from Sanskrit instead. An interesting subset of these pairs of *simplex* and compound is the one of those in which the former is a mainstream item of Buddhist vocabulary, but the compound is specific to Mahāyāna literature, as in the examples below:

<ātāśa-> [a:ʔa:ʒa] ‘sky’ vs. <Ākāśśaggarbha-> [A:ka:śagarb[h]a] ‘Ākāśagarbha’ (‘He Whose Chamber is the Sky’, name of a Mahāyānic celestial bodhisattva).

<vaśāra-> \*[wažēra] from Gāndhārī <vayira> [vəjirə] ‘mace of Indra, diamond’ vs. <vajropama-> ‘resembling a diamond’ (frequent epithet of absorption in the *Perfection of Wisdom* Mahāyāna texts).

<śṣavaa-> expected Gāndhārī \*<śavaya> [ʒa:vəjə] ‘disciple of the Buddha’ vs. śrāvakayāna ‘vehicle of the disciples’ (disparaging term used in Mahāyāna texts).

## 7. Conclusions

At this point we can sum up the most significant finds from this preliminary probe. The first is a point of caution regarding the dialect of Gāndhārī that must have been in use in Khotan during Kuṣāṇa times: it seems to differ rather markedly from the one attested in the textual body of Gandhāra proper. Once this dialectal divide is investigated, it may be necessary to distinguish this variant of ‘Tarim Gāndhārī’ from

<sup>23</sup> Mauro Maggi, “L’importanza del manoscritto T III S 16 per la storia della letteratura cotanese,” *Litterae caelestes* 1.1. (2005): 162.





‘Metropolitan Gāndhārī’: we may perhaps end up using a Sanskrit neologism along the lines of *Saitī*, from the presumed Sanskrit name of the Tarim River (Skr. *Sītā*).

Having made this caveat, it is evident from this examination that the core of the borrowed Buddhist vocabulary in Khotanese derives from Gāndhārī: the most essential vocabulary regarding proper names, cosmology, and the structure of monasticism show the clear phonetic imprint of the language. Doctrine is a realm in which this is less clear: whereas some core tenets are expressed through borrowings from Gāndhārī, the more elaborate technical terms show instead a Sanskrit form: this highlights also the parallel rise of the Mahāyāna, on the one hand, and of mature scholastic Buddhist philosophy on the other.

One related point that remains to be explored is the relationship between translated and borrowed terms. As in the Sinitic context, translated terms freely alternate with direct borrowings from the Indic languages, and often the same concept appears in both forms within the single text: in the *Book of Zambasta*, the Indic term *saṃskāra* ‘mental formation’ (4.1, 4.69) alternates with the much more common native Khotanese coinage *ṣkaum̃gya* for the very same concept, variously spelled. Where do translated terms fall within the continuum between older Gāndhārī borrowings and later Sanskrit ones is still a question to be investigated.

The single most significant pattern in the vocabulary of the *Book of Zambasta* for the chronology of Buddhism in Khotan is that the core of the mainstream technical Buddhist vocabulary comes from Gāndhārī, while most of the vocabulary of the Mahāyāna and of scholastic philosophy comes from Sanskrit. This find anchors the spread of Buddhism to Khotan to the heyday of Gāndhārī literacy in Kuṣāṇa times, but also anchors the famous efflorescence of the Great Vehicle in Khotan to the shift from Gāndhārī to Sanskrit literacy that took place in Gandhāra after the demise of the Kuṣāṇas. The spectacular recent finds of Mahāyāna texts in Gāndhārī have complicated the old notion that the origin of the Mahāyāna movement was tied to the shift from the Middle Indic languages to Sanskrit.<sup>24</sup>

---

<sup>24</sup> See item 363 in the section “Inscriptions” and 390 in the section “Documents” of the *Catalog of Gāndhārī Texts* by Stefan Baums and Andrew Glass.

In the specific case of Khotan, however, as far as we can glean from this probe into a substantial early text, it seems that the spread of the Mahāyāna is indeed connected with the spread of Sanskrit. In fact, the lack of interest in the Mahāyāna throughout the Tarim Basin in the first four centuries of the Common Era stands in contrast with the literature from Kuṣāṇa Gandhāra that arrived in Han and Early Medieval China: the traces of a Gāndhārī substratum in this period of Buddhist literature in Chinese translation are well known, and Mahāyāna texts feature prominently within the oldest body of translations. One should, however, take into account the fact that the Gandhāran diasporas in the Tarim Basin and in China may have been quite different: whereas in the Tarim Basin Gandhārāns constituted the military and administrative elite, in China they seem to be more closely associated with the mercantile community. How exactly this difference correlates with the issue of the spread of the Mahāyāna is still uncertain, but the contrast between the two areas in this early period is striking. Except for the two epigraphic occurrences of the term *mahāyāna* from Shanshan/Nuava,<sup>25</sup> no features of the mature Mahāyāna are traceable in the early historic period in text or art in the Tarim Basin.

All this being said, the shift from Gāndhārī to Sanskrit in the context of the Mahāyāna is generally taken to be a mere reflex of cultural changes that took place in Gandhāra itself, but the peoples of the Tarim Basin may have been involved too in making Sanskrit the sole language of scripture and doctrine: Sanskrit was widely cultivated in the region from early on, and the fact that Tarim Basin recensions of Buddhist Sanskrit texts show, from the point of view of Classical Sanskrit grammar, a more standard language norm than the later Gilgit and Nepalese recensions has already been noted by Schopen.<sup>26</sup>

The most significant exception to the pattern outlined above concerns the text titles of Mahāyāna *sūtras*, which do show Middle Indic features and point towards the existence of Gāndhārī language Mahāyāna texts such as those recently discovered from Gandhāra proper. Once again,

---

<sup>25</sup> See for example items 57, 114, 265, and 371 in the section “Manuscripts” of the *Catalog of Gāndhārī Texts* by Stefan Baums and Andrew Glass.

<sup>26</sup> See Gregory Schopen, “On the Absence of Urtexts and Otiose Ācāryas: Buildings, Books, and Lay Buddhist Ritual at Gilgit,” in *Écrire et transmettre en Inde classique*, ed. Gérard Colas and Gerdi Gerschheimer (Paris: École française d’Extrême-Orient, 2009), 191.



however, the agency of the Khotanese, who after all adopted the Brāhmī script associated with Sanskrit to write their own language, should be taken into account: it may be the case that Mahāyāna texts that started their career in Gāndhārī became prevalent only when they took on a Sanskrit guise. No matter whether the sanskritisation happened in Khotan or in Gandhāra, the finds of this article highlight the role of Sanskrit as a vehicle of diffusion for the Mahāyāna, no longer in the sense that mahāyānic doctrine and thought would have been initially expressed in Sanskrit, but rather that Sanskrit provided a medium that, for reasons that we are only beginning to comprehend, retained unequalled prestige and stability for centuries.

### Symbols

*	hypothetical form
<◇	original orthography of a word
[]	presumed phonetic value of a written expression
<◇ []	given written form with its reconstructed phonetic value

### Bibliography

#### Resources

- Baums, Stefan. *Outline of Gāndhārī Grammar*. 2019.  
[https://stefanbaums.com/baums\\_grammar\\_outline.pdf](https://stefanbaums.com/baums_grammar_outline.pdf).  
 Baums, Stefan, and Andrew Glass, *Catalog of Gāndhārī Texts*. [gandhari.org/catalog](http://gandhari.org/catalog).

#### Secondary Sources

- Bailey, H.W. *Khotanese Texts V—Prolexis to the Book of Zambasta*. Cambridge: Cambridge University Press, 1967.  
 Brough, John. *The Gāndhārī Dharmapada*. London: Oxford University Press, 1962.  
 Dragoni, Federico, Niels Schouppen, and Michaël Peyrot. “The Formal Kharoṣṭhī Script from the Northern Tarim Basin in Northwest China May Write an Iranian Language.” *Acta Orientalia Academiae Scientiarum Hungaricae* 73.3 (2020): 335–373.  
 Duan Qing 段晴. “Neirong buming Quluwen mudu: Guotu BH5-7 内容不明佉卢文木牍: 国图 BH5-7 [A Kharoṣṭhī Wooden Tablet with Unidentified Content: BH5-7 from the National Library of China].” In *Zhongguo guojia tushuguan cang Xiyu wenshu: Fanwen, Quluwen juan* 中国国家图书馆藏西域文书: 梵文、佉卢文卷 [Serindian Documents in the National Library of China: Manuscripts in Brāhmī and Kharoṣṭhī], edited by Duan Qing 段晴 and Zhang Zhiqing 张志清, 203–205. Shanghai: Zhongxi shuju, 2013.  
 Edgerton, Franklin. *Buddhist Hybrid Grammar and Dictionary, Volume I: Grammar*. New Haven: Yale University Press, 1953.  
 Emmerick, Ronald E. *The Book of Zambasta, A Khotanese Poem on Buddhism*. London: Oxford University Press, 1968.  
 Emmerick, Ronald E. *Saka Grammatical Studies*. London: Oxford University Press, 1968.  
 Emmerick, Ronald E. “The Dunhuang MS. Ch 00120: Its Importance for Reconstructing the Phonological System of Khotanese.” In *Turfan and Tunhuang: The Texts: Encounter of Civilizations on the Silk Route*, edited by Alfredo Cadonna 145–70. Florence: Leo S. Olschki Editore, 1992.  
 Hansen, Valerie. *The Silk Road: A New History*. Oxford: Oxford University Press, 2017.  
 Hitch, Doug. “The Special Status of Turfan.” *Sino-Platonic Papers* 186 (2009). Last accessed August 11, 2022. [http://sino-platonic.org/complete/spp186\\_turfan.pdf](http://sino-platonic.org/complete/spp186_turfan.pdf).  
 Hitch, Doug. “The Old Khotanese Metanalysis.” PhD diss., Harvard University, 2016.

*BuddhistRoad Paper* 1.7. Loukota, “Ne Hāḍe Vajjropamā Vaśārā”



- Hitch, Doug. “Tt in Old Khotanese.” *Journal of the American Oriental Society* 135.4 (2021): 663–687.
- von Hinüber, Oskar. *Das Ältere Mittelindisch im Überblick*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften, 1986.
- Leumann, Ernst. *Das nordarische (sakische) Lehrgedicht des Buddhismus*. Leipzig: F.A. Brockhaus, 1933–1936.
- Loukota, Diego. “Made in China? Sourcing the Old Khotanese *Bhaiṣajyaguruvaidūryaprabhāsūtra*.” *Journal of the American Oriental Society* 139.1 (2019): 67–90.
- Maggi, Mauro. “L’importanza del manoscritto T III S 16 per la storia della letteratura cotanese.” *Litterae caelestes* 1.1 (2005): 159–64.
- Maggi, Mauro. “Khotanese Literature.” In *The Literature of Pre-Islamic Iran: Companion Volume to a History of Persian Literature*, edited by Maria Macuch and Ronald E. Emmerick, 330–417. London: I. B. Tauris, 2009.
- Schopen, Gregory. “On the Absence of Urtexts and Otiose Ācāryas: Buildings, Books, and Lay Buddhist Ritual at Gilgit.” In *Écrire et transmettre en Inde classique*, edited by Gérard Colas and Gerdi Gerschheimer, 189–219. Paris: École française d’Extrême-Orient, 2009.
- Schuessler, Axel. *Minimal Old Chinese and Later Han Chinese: A Companion to Grammata Serica Recensa*. Honolulu: University of Hawai’i Press, 2009.
- Sims-Williams, Nicholas. *The Book of Zambasta: Metre and Stress in Old Khotanese*. Wiesbaden: Reichert, 2022.
- Skjærvø, Prods Oktor. “Khotan, an Early Center of Buddhism in Chinese Turkestan.” In *Collection of Essays 1993, Buddhism across Boundaries—Chinese Buddhism and the Western Regions*, edited by John R. McRae and Jan Nattier, 265–344. Sanchung: Fo Guang Shan Foundation for Buddhist & Culture Education, 1999.
- Turner, Ralph Lilley. *A Comparative Dictionary of Indo-Aryan Languages*. London: Oxford University Press, 1962–1985.