



Universiteit
Leiden
The Netherlands

Traces of language contact in Niya Prakrit: Bactrian and other foreign elements

Schoubben, N.

Citation

Schoubben, N. (2024, November 6). *Traces of language contact in Niya Prakrit: Bactrian and other foreign elements*. Retrieved from <https://hdl.handle.net/1887/4108454>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4108454>

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 1 Introduction

This monograph examines traces of language contact in Niya Prakrit. The focus lies on Bactrian influence, but I will also study contact with other languages. The present chapter introduces the topic of foreign influence on Niya Prakrit (§1.1); outlines the research questions to be addressed (§1.2); and presents the languages studied in their historical context (§§1.3–1.5).

1.1 Foreign influence on Niya Prakrit

In the 3rd and 4th centuries AD, in oases on the southern edge of the Tarim Basin in NW China, scribes in the Shanshan kingdom were relentlessly writing accounts, letters, and other official documents. Remarkably, the language used was neither their own nor Chinese; they used an Indian language and Kharoṣṭhī, an Indian script. We know this because archaeologists have, since the early 1900s, discovered hundreds of documents which reveal that a dialect of Gāndhārī, a Middle Indo-Aryan (MIA) language related to Sanskrit, spread from Pakistan to the deserts of Xinjiang to function as “a language not of literature, but of affairs” (Thomas 1921: 278). This dialect of Gāndhārī was baptised Niya Prakrit after the find spot of most documents.¹

Even a cursory glance shows that Niya Prakrit has been subject to the effects of linguistic contact: its phonology is marked by “puzzling phonetic peculiarities”, not all of them explainable by language-internal sound changes, while its vocabulary is, although containing “official designations familiar from ancient Indian usage” and some “purely Indian names”, interspersed with foreign words, e.g. “some of the titles borne by officials”.²

1.1.1 Iranian loanwords

The presence of Iranian loanwords had been apparent even before Burrow (1934; 1935a) devoted two articles to the topic (see also §3.1.2). In his view, many loanwords “had been taken into the Prakrit in N.W. India before it came to be used in Central Asia” (LKD: vii). For reasons to be given in §4.2, I accept Burrow’s view, as do Weber (1997: 31) and Wright (1998: 237). However, “Iranian” is not one language but a language family. Before asking *where*

¹ For convenience, I will sometimes use “Niya” as a shorthand for “Niya Prakrit” (following Jamison 2000: 65 fn. 16).

² The quotations are from Stein (1903: 399f.).

these loanwords were adopted, one should first investigate *from which language(s)* they were borrowed. Yet, the latter question has only rarely been addressed (Weber 1997 is one important exception).

The past decades have nonetheless seen progress on this point, as it became clear that one source of Iranian loanwords in Niya Prakrit is Bactrian, a Middle Iranian language once spoken in present-day Afghanistan but hardly known when Burrow wrote his articles (cf. §1.4.1). Three by now familiar examples of Bactrian loans are (i) *aviṃdama* ‘fine, recompense’ ← αβινδαμο /əvindāmə/; (ii) *kṣuna* ‘date, reign, regnal year’ ← χρουνο /xšunə/; and (iii) *lastana* ‘dispute, lawsuit’ ← λαστανο /lāstānə/ (BD II: 184; 281; 226; §3.2.1 s.vv.).

Even so, Bactrian loanwords in Niya Prakrit have not yet been explored in depth, with the result that we still do not know how many previously known Iranian loans came from Bactrian and if there are as yet unrecognised loanwords from this language.

Many more questions remain largely unanswered. For instance, were the loanwords all adopted together, or can we stratify them? Do these loanwords come from more than one dialect of Bactrian? Is there evidence that these loanwords were not restricted to Niya Prakrit but also present in other varieties of Gāndhārī, perhaps even other Indo-Aryan languages? What information do these loanwords provide on the pronunciation of Bactrian and other donor languages? To what extent and how were loanwords adapted to the inflectional system of Niya Prakrit? Which semantic domains are covered by the Bactrian loanwords? Do loans from other Iranian languages span different domains?

1.1.2 Other types of Iranian influence

Depending on the type and intensity of the contact situation, loanwords are not the only type of borrowing possible: languages can also adopt foreign phonemes and morphemes; they can calque words from another language (loan translations); and their syntax can be influenced by corresponding patterns in neighbouring languages.

Niya Prakrit is no exception: Burrow (1935a: 789f.) already discovered “Iranian idioms in Niya Prakrit”. However, Burrow presented his discovery only as an appendix to his work on the loanwords, and there have hardly been any subsequent studies on the non-lexical influence of Iranian

languages on Niya Prakrit (see further §5.1.2).³ Consequently, we are even less well informed on this topic than we are on the loanwords. Which grammatical elements were adopted? Was the donor language Bactrian or another Iranian language? Which mechanism is involved: substrate interference or borrowing from a prestige language?

1.1.3 A more controversial hypothesis: the Shanshan substrate

In addition to the Iranian influence, scholars have hypothesised substrate interference from the undocumented language of the local inhabitants of the Shanshan kingdom to account for (i) a corpus of words and personal names in Niya Prakrit neither inherited nor immediately recognisable as Iranian loans and (ii) some foreign-looking grammatical features.

Burrow (1935b; LKD: vii–ix) proposed that the language spoken by the Shanshan people was an otherwise unknown third Tocharian language, nowadays referred to as “Tocharian C”. He tried to substantiate this claim with some loan etymologies and a phonological examination of the supposedly non-Iranian loanwords and foreign personal names; the latter would, amongst other things, prove that the substrate language was, like Tocharian, “devoid of voiced stops and aspirates”.⁴

Were Burrow’s theory correct, it would not only have repercussions for the linguistic interpretation of Niya Prakrit but, for instance, also for our understanding of “the prehistorical migrations of the Tocharians” (Peyrot 2013: 2; cf. also Mallory 2015). However, despite its possible significance, Burrow’s hypothesis has “not prompted further study” (Hansen 2012: 44), and a re-examination in light of the continued advances in Gāndhārī and Tocharian studies since the 1930s is long overdue.⁵

In Schoubben (2021), I argued that some Niya Prakrit words for which Burrow suggested a Tocharian etymology were unrecognised loans from

³ A recent exception is Barchi & Peschl (2022), on which see §5.4.

⁴ According to Thomas (1925; 1926b; 1927; 1944), the assumed substrate language would have been a Tibeto-Burman language distinct from Old Tibetan, while Lüders (1940b: 34f.) thought it to be “eine ostasiatische Sprache”, historically related to, but still different from, Chinese. The evidence for either theory remains weak, and Tremblay (2001: 35 fn. 54) seems to have been the only scholar to have approved of them in recent decades. Therefore, I will not discuss these alternative substrate theories.

⁵ The only in-depth review of Burrow’s argument is an unpublished handout by Banti (2000); see §6.1.1.

Bactrian. Thus, the question arises whether substrate influence from the undocumented Shanshan language—be it “Tocharian C” or something else—has been overestimated and the generally acknowledged influence from Bactrian and other Iranian languages on Niya Prakrit not yet appreciated to the extent which it deserves.

1.1.4 The historical context behind the linguistic contact

As foreign influence on a language does not come about in a vacuum, there is a historical context behind every linguistic contact. Yet, the historical context responsible for the Bactrian and other influences on Niya Prakrit is another underexplored topic. Here are three examples of issues needing further study: (i) were the loans adopted in a specific sociolinguistic setting; (ii) did speakers of Bactrian transfer words and grammatical features of their native language to Niya Prakrit because they had only imperfectly learnt the latter, or was the prestige of Bactrian responsible for the borrowings; (iii) do these loanwords support the idea of a Kuṣāṇa domination over the Tarim Basin (§5.4.7, fn. 852; Chapter 8), or are they consistent with different scenarios of how Niya Prakrit spread to the Tarim Basin?

1.2 Research questions and aims

In light of the above, this monograph seeks to address four overarching research questions:

- (i) Which Iranian languages contributed loanwords to Niya Prakrit, and can we identify more loanwords, especially Bactrian ones?
- (ii) Was Iranian influence limited to loanwords, or did Iranian languages, in particular Bactrian, also exert a structural impact on Niya Prakrit’s grammar?
- (iii) Is the assumption of Tocharian substrate influence on Niya Prakrit justified?
- (iv) Which sociolinguistic scenario accounts best for the foreign influence on Niya Prakrit?

For each question, I shall now briefly elaborate on the exact topics studied, their wider relevance, and the methods applied to study them. The introductions to the pertinent chapter(s) provide more detailed discussions on these matters. Here and throughout the book, the way of presentation is data-driven rather than theory-heavy: concepts and methodologies from

(historical) contact linguistics are applied, but the use of overly technical terminology is avoided to ensure accessibility to a wider audience.⁶

1.2.1 Lexical influence from Bactrian and other Iranian languages

The first research question has led me to investigate the lexical influence of Bactrian and other Iranian languages on Niya Prakrit. First, I aimed to compile an up-to-date corpus of the Iranian loanwords in Niya Prakrit in which the words discussed are classified according to their most likely donor language and the reliability of the etymology (Chapter 3). Secondly, I examined the thus accumulated material to determine which Iranian languages have been in contact with Niya Prakrit; how loanwords were phonologically and morphologically integrated; and which semantic domains are covered by them (Chapter 4). Both types of research require one to analyse Niya Prakrit and Bactrian words phonologically—not always an easy task given the deficiencies of the scripts used to write these languages. Therefore, the investigation starts with an introductory chapter on the phonology of both Niya Prakrit and Bactrian (Chapter 2).

When compiling the loanword corpus, I first tried to collect all Niya Prakrit words for which an Iranian source had already been proposed. For each word, I examined the arguments to consider it a loanword and not an inherited lexeme; whether its semantics were safely deducible from the context; whether they matched the meaning of the proposed donor form; and whether there were any phonological problems with the etymology, e.g. an unusual and unparalleled sound correspondence. Based on this examination, I rejected 24 unconvincing proposals (§3.7). The remaining proposals were divided into four categories depending on their reliability, i.e. “secure”, “plausible”, “arguable”, and “tentative” etymologies (see §3.1.7). With these proposals, I also aimed at identifying the most likely donor language by investigating whether (i) the etymology requires the donor form to have undergone a sound change specific to one relevant Iranian language or (ii) the semantics fit only one particular donor form because its cognates have a different meaning.⁷

⁶ For discussions on the application of contact linguistic theory to corpus languages (both regarding loanwords and structural influence), see e.g. Langslow (2002); Adams (2003: 1–110); Mullen (2013: 3–143); McDonald (2015: 36–62); Bianconi (2019: 82–95); Noonan (2019: 8–20); Boyd (2021: 48–113); Dickey (2023: 1–19); Meyer (2023: 215–235).

⁷ For the classification system which resulted from this procedure, see §3.1.4.

Another aim of my research was to enlarge the corpus with newly suggested Iranian loanwords, especially Bactrian ones. Here, the first step was to draft a list of all etymologically obscure words in Niya Prakrit and lexemes whose currently assumed derivation from Old-Indo-Aryan (OIA) requires irregular phonological changes and/or ad hoc semantic developments. Ideally, the next step would then have been to look in the glossary of Sims-Williams' edition of the Bactrian documents (BD II: 179–286) and in texts not excerpted in this glossary to find Bactrian words (i) which have a clear Iranian etymology; (ii) whose semantics are transparent; (iii) which form a semantic fit with a probably not inherited Niya Prakrit word; and (iv) which also phonologically match with the same Niya word. If one finds a word satisfying all four criteria, one can be confident to have discovered a possible Bactrian loanword in Niya Prakrit, although one still has to investigate if the Niya word could also be a loan from a cognate form in another Iranian language (cf. *supra*).

Regrettably, strictly following this methodology leads one into a blind alley: the Bactrian text corpus may have been significantly enlarged over the past decades (cf. §§1.4.1–1.4.2), but we nonetheless know only a tiny proportion of the vocabulary of this language. For this reason, looking at the words currently available in Sims-Williams' glossary and elsewhere does not yield much more than the occasional minor addition to the already established loanword corpus. Did we know as much about Bactrian vocabulary as we do about the lexicon of languages like Sanskrit or Middle Persian, it would probably be far less challenging to discover further Bactrian loanwords in Niya Prakrit.

As it would be difficult to make much progress otherwise, I also reconstruct possible Bactrian sources of Niya Prakrit words by applying Bactrian sound changes to word forms and morphemes known from other Old and Middle Iranian languages. For a fuller justification of this method, see §3.1.6.⁸ Needless to say, loan etymologies whereby the source form is not actually attested are less reliable than those where the source does occur,

⁸ A not strictly academic parallel is provided by an anecdote from the life of J.R.R. Tolkien (taken from Garth 2003: 16). When as a young school boy Tolkien wanted to address the annual Latin debate at King Edward's School (Birmingham) in Gothic, the fact that Gothic is not so well attested obliged him to apply his knowledge of Gothic historical phonology to cognate forms in other Old Germanic languages to reconstruct unrecorded Gothic words!

and I have taken account of this point when classifying the etymologies according to their reliability (see §3.1.7).

As a corollary of my investigations into Bactrian and other Iranian loanwords in Niya Prakrit, it has additionally been possible to clarify the semantics of some words, correct a few erroneous readings, and refine the translation and interpretation of certain passages.

1.2.2 Non-lexical influence from Bactrian

Whether, as asked in the second research question, there is also evidence for grammatical borrowings from Bactrian will be dealt with in Chapter 5, which is a collection of case studies. One reason to study this topic has been given in §1.1.2: it is an underexplored research terrain because previous studies have been concerned almost exclusively with lexical influence from Bactrian and other Iranian languages. In addition, looking at grammatical borrowings is relevant when trying to determine the type and intensity of the contact between Bactrian and Niya Prakrit because the combination of lexical borrowings and structural influence typically points to more than casual contact.

The investigation in Chapter 5 will start with two appurtenance suffixes, which I propose to have been borrowed from Bactrian into Niya Prakrit. Here, the methodology is largely the same as with the loanwords. For a borrowing scenario to be plausible, four requirements need to be met: (i) the Niya suffix cannot be inherited from OIA; (ii) this suffix is especially frequent in loanwords; (iii) the proposed Bactrian source descends from a clear Old Iranian preform; and (iv) the Niya suffix and its suggested donor form match both formally and semantically.

The remainder of Chapter 5 will investigate the possibility that Bactrian exerted a structural influence on the morphosyntax of reciprocals and reflexives in Niya Prakrit, on the one hand, and preterite formations, on the other. The methodology follows again a similar pattern as with the lexical influence. I begin by synchronically examining the morphosyntax of the relevant grammatical item in Niya Prakrit. In so far as possible, I then compare other MIA languages to try to determine which of the morphosyntactic features observed are inherited and which could be due to contact. Finally, I draw a comparison with a corresponding formation in Bactrian to see if there are significant parallels that could point to Bactrian influence on Niya Prakrit. When necessary, this comparison is preceded by a synchronic study of the Bactrian formation.

1.2.3 A re-examination of the “Tocharian C” substrate

In keeping with the third research question, Chapter 6 re-examines Burrow’s hypothesis of a “Tocharian C” substrate in Niya Prakrit (§1.1.3). This chapter takes the form of a critical review of (I hope) all arguments which have been adduced in support of a Tocharian substrate. Commenting on these arguments one by one, I try to show that hardly any of them stands up to closer scrutiny. Complementing the findings of Chapters 3–5, I also argue that some linguistic features of Niya Prakrit that have been ascribed to the hypothetical “Tocharian C” could instead come from an attested language already known to have influenced Niya Prakrit, i.e. Bactrian.

1.2.4 A sociolinguistic scenario

To address the fourth research question, Chapter 7 briefly studies against which historical background the foreign influence on Niya Prakrit is best understood. The aim is to find out when, why, and in which circumstances the Bactrian (and other foreign) elements entered Niya Prakrit.

The examination will proceed in three steps. First, I will take stock of the evidence for foreign influence on Niya Prakrit assembled in other chapters. In a second step, I shall use Thomason and Kaufman’s oft-cited borrowing scale (1988: 74–76) to determine the intensity of contact required to result in the Bactrian influence visible in Niya Prakrit. The third step will consist of finding a parallel, better-researched example of language contact in the ancient world that can inform us about the type of contact between Bactrian and Niya Prakrit.

Proceeding in this way, I shall argue that we should distinguish between two levels of contact. On the one hand, there seems to have been casual contact between speakers of Bactrian and most, perhaps all, dialects of Gāndhārī because of the geographical proximity between the two languages and the sociopolitical influence of the Kuṣāṇa empire on Gandhāra. This influence may, for instance, be responsible for the Bactrian loanwords which Niya Prakrit shares with South Asian Gāndhārī, sometimes also with other Indo-Aryan languages.

The second level of contact—seemingly more intense—would be that between Bactrian and Niya Prakrit itself. Based on a comparison with the contact between Old Persian and Achaemenid Official Aramaic, I shall hypothesise a scenario whereby speakers of Kuṣāṇa Bactrian adopted the immediate ancestor of Niya Prakrit as a chancellery language; this type of

contact would explain administrative and economic loanwords not found in other variants of Gāndhārī and traces of structural influence exerted by Bactrian on Niya Prakrit.

It should be obvious that the sociolinguistic scenario to be proposed is a hypothesis that will need further testing in subsequent research. Yet, even when the sociolinguistic scenario will need revision, it illustrates how a strictly linguistic and philological examination can contribute to the wider historical study of Gandhāra, Bactria, and the Tarim Basin.

1.3 Niya Prakrit and other varieties of Gāndhārī

The rest of this chapter introduces in somewhat more detail Niya Prakrit (§1.3), Bactrian (§1.4), and other languages that (may) have been in contact with Niya Prakrit, either in Gandhāra or the Tarim Basin (§1.5).

1.3.1 Niya Prakrit

Niya Prakrit is the best-attested variety of Central Asian Gāndhārī. It is alternatively known as Niya Gāndhārī, Kroraina Prakrit, or Shanshan Prakrit.⁹ The chancellery idiom of the Shanshan kingdom at the southern edge of the Tarim Basin,¹⁰ Niya Prakrit is documented in more than 900 administrative documents in Kharoṣṭhī script written on wood and leather and dating back to the 3rd–4th centuries AD.¹¹

Most Niya documents were discovered by the explorer-cum-philologist Aurel Stein during four archaeological expeditions to the southern part of the Tarim Basin (1900–01; 1906–07; 1913–14; 1930). Editing them required, as Stein anticipated (1903: 398; 1907: 363), “much time and patient labour”; they were eventually published between 1920 and 1929 in three volumes by Augustin-M. Boyer, Émile Senart, Edward J. Rapson, and Peter S. Noble (Boyer et al. 1920–29 = KD).¹² Rapson’s student Thomas Burrow contributed

⁹ As pointed out by Jamison (2000: 64 fn. 7), Fussman (1989: 440; cf. also 2012: 191 fn. 5) went too far in claiming that Niya Prakrit should not be considered a variety of Gāndhārī due to the high amount of foreign influence.

¹⁰ Some of the more recent studies on the history and Buddhist culture of the Shanshan kingdom are Atwood (1991); Padwa (2007); Hansen (2004; 2012: 25–55); van Schaik (2014); Høisæter (2020; 2023); and Jiang & Loukota (2023).

¹¹ For the documents on leather, see Stein (1907: 344–347), and for the date of the Niya documents, cf. Høisæter (2020: 91–103).

¹² Before the publication of the edition, preliminary results had been made known in Rapson (1905); Boyer (1905); and Boyer, Rapson & Senart (1918).

a grammar of Niya Prakrit (LKD), an edition of additional documents (1937), and a translation of the whole corpus (TKD). Nowadays, the whole corpus, including more recently discovered documents, is online available as Baums & Glass (2002b–).

1.3.2 Central Asian Gāndhārī

The spread of the Gāndhārī language from its heartland in present-day Pakistan to the Tarim Basin in NW China is “the result of the unifying force of the vast Kuṣāṇa empire, in which Gāndhārī served both as a vehicle for the Buddhist message and as an administrative language” (Salomon 2002: 120; cf. also Salomon 2011b). Nevertheless, the exact relationship between the various forms of Central Asian Gāndhārī and their ancestor dialects remain hard to establish due to the scantiness of the available evidence (Burrow 1936; Caillat 1992; Kogan 2017).

The Khotan area was the scene of the discovery of arguably the most famous text in Central Asian Gāndhārī, i.e. the *Khotan Dharmapada*, a local version of a Buddhist verse collection partially preserved on a birchbark manuscript discovered in 1892 by Jules-Léon Dutreuil de Rhins (Brough 1962). From Khotan, there is further one, possibly two, administrative documents (CKD 661 and perhaps 843) in a regional variety of Gāndhārī called Khotan Prakrit.¹³ Likewise often associated with Khotan are the Sino-Kharoṣṭhī coins, bilingual coins in Gāndhārī and Chinese the earliest of which date back to the 1st century AD (cf. now Cribb 2023); such coins have nevertheless also been discovered in places like Kashgar and Kuča (Cribb 2020).

The territory of the ancient Shanshan kingdom saw not only the discovery of the Niya documents but also of a few other Gāndhārī texts: in Endere, a stone inscription has been discovered (CKI 363; Salomon 1999b); two of the famous frescos at Miran are accompanied by inscriptions (CKI 443; 445; Boyer 1911: 414–418); and an inscribed piece of silk hails from the same find spot (CKM 406; Boyer 1911: 418–422).

At the northern fringe of the Tarim Basin, a few literary fragments were discovered now housed in Paris as part of the Pelliot collection (CKM 90–97;

¹³ On CKD 661, see §5.4.7 and Appendix 3 (p. 525). Based on evidence from the *Prophecy of the Li Country* (Emmerick 1967), Yoshida (*apud* de la Vaissière 2014: 87 fn. 14) tentatively dates this document to ca. 320 AD. For a preliminary study of CKD 843, see Duan (2013b).

Salomon 1998a). In addition, there are some administrative documents in Kuča Prakrit, including bilinguals with Tocharian B on the other side (CKD 829–838; 844–845; 870–877; 883; 891–896). Their exact dating and interpretation remain to be established with certainty.¹⁴

1.3.3 South Asian Gāndhārī

Central Asian Gāndhārī is to be distinguished from, yet closely related to South Asian Gāndhārī. As the vernacular Middle-Indo-Aryan (MIA) language spoken in the Peshawar valley (Pakistan), the latter form of Gāndhārī is attested for more than six centuries: from the 3rd century BC onwards in the form of the Aśokan Rock Edicts at Shāhbāzgarhī (SHĀH) and Mānsehrā (MĀN) until the 3rd or 4th century AD when Gāndhārī got so much Sanskritised as to become Sanskrit eventually (Strauch 2012).

As the result of spectacular manuscript finds over the past decades,¹⁵ the present-day corpus consists not only of the long-known coin legends and inscriptions but also of a substantial corpus of Buddhist manuscripts on birch bark and palm leaf.¹⁶ In addition, there are two administrative documents comparable in form and function to the Niya documents: a silk fragment from the Gilgit area (CKM 430; Falk 2021; Schoubben fthc.) and a loan contract preserved as Bajaur collection fragment 15 (CKM 278; to be published by Gudrun Melzer).¹⁷

¹⁴ See mainly Schmidt (2001) and Ching (2013). Some Kuča Prakrit sources from the Pelliot collection remain unpublished. An offshoot of the Central Asian Gāndhārī writing tradition is seen in a small corpus of fragments in what has been named “Formal Kharoṣṭhī”. While still largely undeciphered, this script may write an Iranian language related to Khotanese and Tumshuqese (Dragoni, Schoubben & Peyrot 2020 *contra* Schmidt 2018).

¹⁵ Cf. e.g. Salomon (1999a; 2003); Strauch (2008); Falk (2011); Falk & Strauch (2014).

¹⁶ For the historical development of the writing materials and formats used, cf. e.g. Baums (2014; 2021). For an introduction to Gāndhārī literacy, with a translation of selected fragments, see Salomon (2018).

¹⁷ For preliminary remarks on the Bajaur loan contract, see Strauch (2008: 127) and Falk & Strauch (2014: 72). There is also one specimen of a monastic ledger in Gāndhārī, published in Allon (2019).

1.4 Bactrian

1.4.1 The language

Bactrian is the name given by Henning (1960: 47f. = 1977 II: 545f.) to the Middle Iranian language spoken in the region of ancient Bactria and neighbouring territories, approximately corresponding to present-day (northern) Afghanistan. The language is attested in documentary sources spanning most of the first millennium AD and mainly written in a modified form of the Greek script (see §2.2.1).¹⁸

Before the discovery of a substantial inscription at Surkh Kotal, there was only a small evidential basis to study Bactrian, the corpus being limited to coin legends.¹⁹ It took the discovery of the Rabatak inscription and a corpus of more than 150 administrative documents and letters in the 1990s before our knowledge of Bactrian grammar reached a standard comparable to that of some other Middle Iranian languages (cf. Sims-Williams 1997a; 2018).²⁰ Bactrian's vocabulary is still only imperfectly known (cf. §1.2.1; §3.1.6).²¹

1.4.2 The text corpus

The Bactrian corpus nowadays consists of legends on coins and seals; some longer inscriptions; administrative documents and letters; some fragments of Buddhist literature (BD II: 173–177; Sims-Williams 2010b: 204f.); the Hephthalite fragments from the Berlin Turfan collection (Sims-Williams 2004b); and a unique specimen of a Manichaean Bactrian folio written in an

¹⁸ For the Manichaean Bactrian fragment, see below. Falk (2023) has argued that the Issyk-Kushan script (also known as the “Unknown Script”) was also used to write Bactrian, whereas Bonmann et al. (2023) propose that the language recorded in this script is a previously unknown Middle Iranian language. As the conclusions of this book are not affected by it, I will leave this question aside here.

¹⁹ First publication of the Surkh Kotal inscription in Maricq (1958) and Benveniste (1961: 114–140). Synoptic edition in Davary (1982: 53–64), with English translation in Sims-Williams (2012b: 78f.).

²⁰ *Editio princeps* of the Rabatak inscription in Sims-Williams & Cribb (1995–96); more recent edition and translation in Sims-Williams (2008c; 2012b: 77f.). The documents have been edited by Sims-Williams in BD I–III. For their chronology, see Sims-Williams & de Blois (2018 = SCBD).

²¹ Julian Kreidl (Bloomington) is preparing a Bactrian-German dictionary and kindly put a preliminary version of this dictionary at my disposal.

adapted form of the Manichaean script (M 1224; Sims-Williams 2009a; 2011a).²²

The corpus is divided into two parts: an older part dating back to Kuṣāṇa times and written in a more archaic variant of the language, i.e. Kuṣāṇa Bactrian (1st–3rd cent. AD), and a more recent part composed in the linguistically more innovative but better-known post-Kuṣāṇa Bactrian language (3rd–9th cent. AD). The earliest longer text in Bactrian is DN₁, found in Dasht-e Nawur and dated to 105 AD (Halfmann et al. 2024: 7). The last tangible evidence of Bactrian literacy is the Tochi inscriptions (ITAB; ITSB; ITB), which are datable to the middle of the 9th century (SCBD: 85ff.). By then, Arabic had established itself as the new language of culture in the region (cf. Khan 2007).

1.4.3 The collateral tradition: Bactrian loans into other languages

Because of the prestige which the Bactrian language enjoyed during the Kuṣāṇa (1st–3rd cent. AD) and Hephthalite periods (5th–8th cent. AD), Bactrian words were not only borrowed into Gāndhārī, but also into (Buddhist) Sanskrit (Schoubben 2022b); Tocharian (Schwartz 1974; Tremblay 2005a: 435–437); Khotanese (Weber 1985); Sogdian (Sims-Williams 1996a: 51; 2004a: 541 fn. 8); Persian (Lurje & Yakubovich 2017); and Nuristani languages (Halfmann 2023).²³ Some of the established Bactrian loanwords in these languages do not occur in the currently available Bactrian texts, but they can be reconstructed based on words in other Old and Middle Iranian languages (see also §3.1.6). We can therefore speak of a Bactrian collateral tradition, comparable in significance, albeit less well researched than the Old Iranian *Nebenüberlieferung* in, for instance, Elamite and Aramaic and the Parthian one in Armenian.

1.5 Other languages in contact with Niya Prakrit

This section finally introduces which languages other than Bactrian (could) have been in contact with (the ancestor of) Niya Prakrit. The structure is at the same time chronological and geographical. First, I give an overview of languages in contact with Niya Prakrit during the prehistory of the language

²² One still hopes for an online edition of the currently available Bactrian materials, comparable to what Baums & Glass (2002b–) provide for Gāndhārī.

²³ For loanwords into Bactrian, cf. Sims-Williams (2002) and Gholami (2018).

in Gandhāra when various foreign people invaded this area.²⁴ These languages are Old Persian, Greek, “Indo-Scythian”, and Parthian. Secondly, I introduce languages that could have influenced Niya Prakrit in the Tarim Basin, viz. Khotanese, Tocharian A/B, Sogdian, and Chinese.

1.5.1 Contact languages in Gandhāra

The earliest identifiable stratum of loanwords in Niya Prakrit comes from Old Persian (cf. §4.2.2). These loanwords were borrowed by the ancestor of Niya Prakrit when the Achaemenid empire ruled over Gandhāra from the 6th until the 4th century BC (Vogelsang 1990; Ruffing 2021), and Aramaic literacy was introduced to the area (cf. §2.1.1).

From the 3rd to 1st century BC, the Graeco-Bactrians and Indo-Greeks left their mark on large parts of northern India (Mairs 2020). A part of their legacy is a number of Greek loanwords in Gāndhārī, some of them attested in Niya Prakrit as well (cf. §3.2.1 s.v. *kṣuna*; §3.2.2 s.v. *milima*; Appendix 1, p. 515).

Concurrently, other Iranian people started to invade Gandhāra and influence the local language and culture. First among them were the Indo-Scythians or Indic Śakas, who began to penetrate the South Asian subcontinent in the 2nd century BC and whose influence lasted in some parts of India until the 4th century AD (see e.g. Neelis 2007: 56–79). While their Middle Iranian language has not been put to writing, it has been argued to be closely related to Khotanese (Lüders 1913 = 1940a: 236–255) and to have left loanwords in Gāndhārī (e.g. KI: xx) and other Indo-Aryan languages, including Sanskrit (Burrow 1975). “Indo-Scythian” must thus also be taken into account as a possible donor language of loanwords in Niya Prakrit (cf. §4.2.7).²⁵

A second group of Iranians occupying parts of NW India were the Indo-Parthians (e.g. Fröhlich 2020). Their reign started at the beginning of the 1st century AD when Gondophares, their first king, declared independence from the Parthian empire. It lasted until the first half of the 3rd century AD. Although the Parthian heartland in present-day NE Iran and Turkmenistan was geographically remote from Gandhāra, one accordingly expects there

²⁴ For recent overviews of the various invasions and their role in the development of a Buddhist culture in the region, cf. Falk (2013b) and Albery (2020: 21ff.).

²⁵ But see §3.1.6 for the methodological problems faced when working with loanwords from reconstructed languages.

also to have been some Parthian influence on the predecessor of Niya Prakrit (cf. §4.2.4).

1.5.2 Contact languages in the Tarim Basin

Its spread to the Tarim Basin brought Niya Prakrit in contact with still other languages. In the southwestern part of the Tarim Basin, around the modern city of Khotan, a Middle Iranian language was spoken, which we now call Khotanese (sometimes also Khotan Saka), and which is documented by way of religious texts (Maggi 2009) and secular documents (e.g. Zhang 2016). We know from the Niya documents that the Shanshan kingdom frequently exchanged with the Khotan area (cf. Høisæter 2020: 357–363; 2023: 419). Therefore, Khotanese is one of the languages that can have contributed loanwords to Niya Prakrit (cf. §4.2.5).²⁶

In the northern part of the Tarim Basin, Tocharian A and B were spoken, two closely related Indo-European languages known from Buddhist literature and administrative documents.²⁷ Tocharian A (also “Agnean” or “East Tocharian”) was indigenous to the kingdom of Agni/Yānqí in the northeast of the Tarim Basin; Tocharian B (also “Kuchean” or “West Tocharian”) was spoken more to the west, in the kingdom of Kuča (Lévi 1913). Since the Niya documents testify to encounters between the Shanshan kingdom and the Tocharian-speaking areas (cf. Høisæter 2020: 367–369), linguistic contact between Niya Prakrit and Tocharian may have existed too (see Chapter 6, also on the issue of “Tocharian C”).

Besides these locally spoken languages, Niya Prakrit can also have been in contact with languages that played a role in the Tarim Basin for different reasons. Such a language is Sogdian (cf. §4.2.6), a Middle Iranian language originally home to the region known as Sogdiana and located in present-day Uzbekistan and Tajikistan. Sogdians were living in the territory of the Shanshan kingdom because of their role as traders on the Silk Road (Høisæter 2020: 370–374; 405–413), and some of the oldest Sogdian texts have been discovered in the Tarim Basin, including in Niya itself (Sims-Williams & Bi 2018; Sims-Williams 2023: 71ff.).

²⁶ It seems geographically and historically unlikely that Tumshuqese, a sister language of Khotanese spoken in the NW corner of the Tarim Basin (Konow 1935), influenced Niya Prakrit.

²⁷ Secular documents in Tocharian A have been discovered only recently (Ogihara 2014).

Due to various military conflicts, there was additionally a Chinese influence on the Tarim Basin for most of the first millennium AD, although in different degrees of intensity according to time and place. As concerns the Shanshan kingdom, the non-linguistic evidence for encounters with Chinese civilisation has recently been described by Høisæter (2020: 120–124; 363–367).²⁸ Potential linguistic influence from Chinese on Niya Prakrit has yet to receive a systematic investigation, and will only be tangentially touched upon in subsequent chapters.²⁹

²⁸ Chinese people living in the Shanshan kingdom left records of their presence in the form of Chinese documents contemporaneous with the Niya Prakrit ones (Chavannes 1913; Conrady 1920; Maspero 1953; Hou & Yang 1999).

²⁹ To my knowledge, there is so far only one secure example of a loanword directly adopted from Chinese, i.e. the political title *jīṭumgha* ← Chin. *shì zhōng* 侍中 ‘serve, wait upon’ + ‘centre’ > ‘palace attendant’, (E)MC *dzyiH trjuwng* (Brough 1965: 600ff.; 1970a = 1996: 295ff.; 351–357). Recently, Duan (2016a: 183–188) has proposed that the administrative term *hastalekha*, literally ‘hand-document’, would be a calque of Chin. *chǐ dú* 尺牘 ‘tablet measuring one foot’. I find this proposal unlikely because Gandh. *hastalekha* (dir.sg. *hastalekho*) is also attested in a loan contract found in Bajaur (CKM 278), i.e. outside the Chinese sphere of influence.