

# 1. INTRODUCTION

## 1.1. PRELIMINARIES AND AIMS

This study investigates the linguistic contacts between Khotanese and Tumshuqese on the one hand and Tocharian A and B on the other. Its main objective is to detect and analyse the Tocharian lexicon of Khotanese and Tumshuqese provenance. The longest chapter (Chapter 2.) presents and discusses Tocharian lexical items possibly or probably borrowed from Khotanese and Tumshuqese and rejects several unlikely borrowing etymologies that have been proposed. The corpus determined in Chapter 2. is then subject to a phonological (Chapter 3.) and a semantic (Chapter 4.) analysis. Chapter 5. contains a preliminary assessment of the Tocharian component in the lexicon of Khotanese and Tumshuqese, and Chapter 6. summarises the results of the investigation.

The research questions that are at the basis of this study can be summarised as follows:

1. Is it possible to expand the corpus of Khotanese and Tumshuqese loanwords in Tocharian already known from the scientific literature?
2. What are the phonological and morphological features of these loanwords?
3. Is it possible to classify the loanwords chronologically? From which stages of Khotanese and Tumshuqese did the borrowing take place?
4. Which semantic areas of the lexicon were subject to borrowing from Khotanese and Tumshuqese?
5. Which type of linguistic contact took place between Tocharian and Khotanese and Tumshuqese?

Chapter 2. is concerned with the first research question, Chapter 3. with the second and the third, and Chapter 4. with the fourth. Chapter 6. summarises the most important conclusions and provides possible answers to the fifth question.

In chapters 4. and 5., and in the discussion of some of the lexical items in §2.1., I have sketched some possible socio-historical scenarios explaining the intensity and quality of language contact between Tocharian and Khotanese and Tumshuqese. It should be stressed that none of these scenarios has been sufficiently explored. Therefore, the historical conclusions summarised in Chapter 6. still have the character of hypotheses that await a more detailed investigation. It is hoped that such research may be carried out in the not-so-distant future, as it might reveal a great deal about the cultural history of the Tarim Basin.

After a brief introduction to the Tocharian languages (§1.2.) and Khotanese and Tumshuqese (§1.3.), this chapter defines the research problem (§1.4.) and offers an overview of the scientific literature on the subject (§1.5.). Further, it describes the methodology employed (§1.6.) and, finally, the structure of the entries in §2.1. and Chapter 5. (§1.7.).

## 1.2. TOCHARIAN AND ITS CONTACT LANGUAGES

‘Tocharian’ is the conventional designation of two extinct Indo-European languages once spoken in the northern part of today’s Xinjiāng Uyghur Autonomous Region in Northwest China. These two languages are referred to as Tocharian A (TA), originally from Agni/Yānqí (also called ‘East Tocharian’ or ‘Agnean’), and Tocharian B (TB), originally from Kuča (also called ‘West Tocharian’ or ‘Kuchean’). The designation goes back to the beginning of the 20<sup>th</sup> century when the first Tocharian manuscripts were unearthed from the sands of the Tāklimakan desert (Sieg and Siegling 1908).

The manuscripts written in Tocharian B can be dated approximately from the 5<sup>th</sup> to the 10<sup>th</sup> c. CE. Tocharian A is attested in manuscripts dating from the 7<sup>th</sup> to the 10<sup>th</sup> c. CE (Pinault 1989a: 7–10). Following the standard chronological periodisation by Peyrot (2008), Tocharian B can be divided into an archaic, a classical, and a late phase. Further, a ‘colloquial’ type is distinguished (Peyrot 2008: 190). As for Tocharian A, the language attested in the extant manuscripts seems to be more uniform. Ogihara (2014) has shown that, besides its use as a religious language, it was also employed as an administrative language in the monasteries. Both languages are written in the so-called ‘North-Turkestan’ variant of the Indian Brāhmī script.

Tocharian A and B are genetically related. It is possible to reconstruct their ancestor language, which is conventionally termed ‘Proto-Tocharian’ (PT). The dating of Proto-Tocharian is debated, but it can be estimated between the 10<sup>th</sup> and 5<sup>th</sup> c. BCE (see further §6.2.2.1.).

Language contact played an important role in the historical development of Tocharian. Neighbouring languages have left extensive traces in all language levels, i.e. phonology, morphology, and the lexicon. In prehistoric times, Tocharian was probably in contact with ‘Old Steppe’ Iranian (OSr.), an otherwise unattested Old Iranian language (Peyrot 2018),<sup>2</sup> and with Uralic (Peyrot 2019). More recent contacts involve Old and Middle Chinese, Old Uyghur, Sogdian, Bactrian, and Parthian. With the expansion of Buddhism in the Tarim Basin, a significant part of the lexicon was borrowed from Buddhist Hybrid Sanskrit and Middle Indic languages, chiefly Gāndhārī. The precise dating and extent of language exchange with Khotanese and Tumshuqese (see §1.3.) is unknown because no comprehensive studies are available. This work shows that Khotanese and Tumshuqese influence on Tocharian was much more intense than expected and spanned almost two millennia.

## 1.3. KHOTANESE AND TUMSHUQESE

Khotanese and Tumshuqese are two Middle Iranian languages once spoken in the southwestern and northwestern parts of today’s Xinjiāng Uyghur Autonomous Region in Northwest China. At the beginning of the 20<sup>th</sup> century, following their discovery, the two languages were named after the two cities Khotan (today’s 和田 Hétián) and Tumshuq (today’s 图木舒克 Túmùshūkè).

Khotanese is richly documented. The literature includes literary and religious (Buddhist) texts and many secular documents (Maggi 2009a). The oldest manuscripts are plausibly dated to the 5<sup>th</sup> c. CE on palaeographical grounds (Maggi 2016, 2022a) and the language may have

---

<sup>2</sup> The contact with Old Steppe Iranian is the subject of the PhD research of my colleague Chams Bernard (Leiden University), from whom I adopt this preliminary language label (see §1.5.)

been spoken roughly until the Qarakhanid conquest of Khotan at the beginning of the 11<sup>th</sup> c. CE. Two main stages of the language are conventionally distinguished: Old and Late Khotanese.<sup>3</sup> In this work, I reconstruct a pre-stage which I term ‘Pre-Khotanese’ (PK). Whereas manuscripts written in Old Khotanese were mainly found within the Khotan area, Late Khotanese is also documented through manuscripts from the Dunhuang area, where a Khotanese community was residing. The extant manuscripts are either Chinese book rolls or Indian-type pustaka books. They are written in the southern variant of Turkestan Brāhmī (Dragoni 2017: 396). Old Khotanese is one of the most conservative Middle Iranian languages. It preserves six of the eight Proto-Iranian cases, shows traces of neuter gender, and has four moods (with traces of an injunctive).

Tumshuqese is known only from a handful of documents (Maue 2009) dated approximately to the 8<sup>th</sup> c. CE (Ogihara and Ching 2017: 467–69). As far as can be gathered from the scanty material, Tumshuqese was heavily influenced by Tocharian B. Traces of this influence can be found in the script, a northern variant of the ‘Turkestan Brāhmī’ also used for Tocharian, the lexicon, and the literature.<sup>4</sup> The so-called ‘Fremdzeichen’, or ‘foreign signs’, are a particular feature of the Tumshuqese writing system. Some are original inventions, and some are shared with Tocharian, Sogdian, and Old Uyghur Brāhmī. The manuscript of the Tumshuqese *Karmavācana* (Emmerick 1985a) might be earlier than the rest of the documents, as only one of the Fremdzeichen ( $x_3$ ) was used in this text. Still, no exact dating can be proposed with certainty.

The importance of Tumshuqese lies in the fact that it is genetically related to Khotanese, but it is far more conservative. As an example, one may compare Tq. *rorda-* ‘given’ and OKh. *hūḍa-* ‘id.’, both from PIr. \**fra-brta-*. As in the case of Tocharian A and B, the comparison between Khotanese and Tumshuqese may allow the reconstruction of a common ancestor that I will term ‘Proto-Tumshuqese-Khotanese’ (PTK) following Peyrot (2018: 272–74).

#### 1.4. THE SIGNIFICANCE OF KHOTANESE AND TUMSHUQESE LOANWORDS IN TOCHARIAN

Why is it important to study Khotanese and Tumshuqese loanwords in Tocharian? In the first place, little is known about the linguistic prehistory of the Tarim Basin. The analysis of the loanword corpus may shed light on the age and significance of the first contacts between Khotanese and Tocharian. Through the comparative method (see Campbell 2020: 140–93), it is possible to reconstruct the pre- and proto-stages of Khotanese and Tumshuqese. This allows to establish whether the phonological features of the loanwords into Tocharian are to be dated to the Proto-Tumshuqese-Khotanese period (see §1.3.) or to the historically attested stages. The relative chronology of the loanwords and a thorough semantic analysis may show precisely which parts of the lexicon were most extensively borrowed at what stage in the history of the languages under investigation.

<sup>3</sup> This is only a conventional definition that must be refined in the future. Skjærvø (KMB: lxx), in addition to Old and Late Khotanese, distinguishes a Middle Khotanese stage.

<sup>4</sup> If the identification of the language of the so-called ‘Formal Kharoṣṭhī’ fragments proposed in Dragoni, Schoubben, and Peyrot (2020: 357–58) is correct, it may be an earlier form of Tumshuqese. Significantly, the fragments were found as far east as Kuča, Šorčuq, and Tuyuq in the vicinity of Turfan, i.e. in Tocharian-speaking territory.

As loanwords can provide essential insights into the social interactions among different groups in the past (Epps 2015: 585–86), the analysis conducted in this study contributes to a better understanding of the dynamics of interactions among the population groups of the prehistoric Tarim Basin. The results of this analysis can be employed to address more complex questions related to power relations, language dominance, and ancient population movements in the Tarim Basin. Furthermore, the study of more recent loanwords may shed light on the same dynamics in historical times.

As an example, the results of this study deliver relevant material for the study of the spread of Buddhism among the people of the Tarim Basin by contributing to the ongoing discussions on the circulation of texts and ritual practices in the area. As many of the loanwords discussed here belong to the medical language, this study also contributes to the study of the circulation of medical knowledge in the Tarim Basin, both before and after the introduction of Ayurvedic texts and practices along with the spread of Buddhism in the region (Dragoni 2021). Medical loanwords from prehistoric stages of Tumshuqese and Khotanese shed new light on the Pre-Buddhist medical practices in the Tarim Basin. Determining the borrowing directions of medical terminology of Indic origin into the vernacular languages of the Tarim Basin, on the other hand, contributes to a better understanding of the main routes of circulation of Indian medical knowledge in the region.

On a different note, this study may also be considered a contribution to Tocharian and Khotanese lexicography. Although the Tocharian situation is slightly better than the Khotanese one (Pinault 2019, Emmerick and Maggi 2001), the lexicography of the two languages is still in the preliminary phase. As Bailey's dictionary (DKS, 1979) is outdated, Khotanese lacks a comprehensive, up-to-date lexicographical tool. Scholars must rely on the glossaries of the edited texts and combine them with the three volumes of *Studies in the Vocabulary of Khotanese* (SVK I-III). On the Tocharian side, Adams' dictionary (DoT), Carling's first volume of the *Dictionary of Tocharian A* (DTTA), and the online *Comprehensive Edition of Tocharian Manuscripts* (CEToM) are the most important lexicographical tools available. However, as many texts in both languages are still unedited, it is often necessary to provide new translations of the text passages under investigation. A direct examination of the text passages in which a lexeme occurs is essential to determine its correct meaning and phonological shape. Thus, some of the results of this investigation can also be read as a contribution to the philological study of Tocharian, Khotanese, and Tumshuqese texts.

### 1.5. PREVIOUS STUDIES

The problem of the linguistic contact between Tocharian and Khotanese has always been inextricably connected to the problem of Iranian loanwords in Tocharian. A detailed analysis of previous studies on this subject is found in Bernard (2023: 12–14). Only the studies directly concerned with Khotanese and Tumshuqese will be examined in this context.

Hansen (1940) is the first attempt at a systematic overview of the Iranian loanwords in Tocharian. Fifty-one items are analysed and commented upon. In Hansen's view, twenty-seven lexemes can be traced back to Khotanese. This analysis is now outdated because of its lack of consideration of the Gāndhārī, Bactrian and Old Steppe Iranian influence on Tocharian: of his twenty-seven items, only four can now be considered as borrowed from Khotanese (see §2.1. s.v. *aṅkwaṣ(t)*, *pissaṅk*, *tvāṅkaro*, *yolo*).

Cursory allusions to the Tocharian material can be found in some of H.W. Bailey's articles and, most notably, in the *Dictionary of Khotan-Saka* (DKS) and the *Prolexis to the Book of Zambasta* (KT VI).<sup>5</sup> However, only one section of an article by H.W. Bailey deals exclusively with the contacts between Khotanese and Tocharian. In 'Recent work in "Tocharian"' (Bailey 1947: 149–50), the author briefly lists ten lexemes that, in his opinion, may have been borrowed from Khotanese. In this work, I show that only three of these ten items can be considered loanwords from Khotanese (see §2.1. s.v. *añkwaṣ(t)*, *tvāñkaro*, *ṣpakīye*).<sup>6</sup>

Isebaert's (1980) unpublished dissertation is the only comprehensive monograph on Iranian loanwords in Tocharian. However, as for the Middle Iranian data, it is now outdated. Moreover, its continuous resorting to a general label of 'Middle Iranian' without specifying the donor language is problematic. A significant contribution that excluded a Khotanese origin for a group of Tocharian lexemes by arguing for a Bactrian provenance instead is Schwartz (1974). A solid confirmation of his hypothesis came from the recent discovery of the Bactrian documents (Sims-Williams 1997: 23). Other repertoires of loanwords are the more recent Tocharian A and B lexicographical works, i.e. Adams' dictionary of Tocharian B (DoT) and Carling's Tocharian A Thesaurus (DTTA).

Schmidt (1985) was the first scholar to recognise an ancient layer of Old Iranian provenance in the group of loanwords distinguished by the correspondence Ir. \*a ~ TB e, TA a. Further studies (Pinault 2002: 245, Peyrot 2015, Peyrot 2018: 280, Bernard 2023) confirmed that this layer is to be attributed to an otherwise unattested Old Iranian language, possibly sharing some affinities with the 'Scythian' group of Iranian steppe dialects, hence the designation by Chams Bernard of 'Old Steppe Iranian'.

Tremblay (2005) challenged this hypothesis by identifying this Old Iranian layer with the ancestor of Khotanese and Tumshuqese, a reconstructed 'Old Sakan' (Tremblay 2005: 422). The main argument for this identification is his interpretation of the Tocharian word for 'iron', TB *eñcuwo* A *añcu\**, which shows the exclusively 'Old Sakan' outcome \**sw* of the Proto-Iranian cluster \**čw* and contains the Iranian vowel \**a* in the donor language. In my opinion, TB *eñcuwo* A *añcu\** is more likely to contain an original \**e* in the donor language, the product of an early 'trajected umlaut' of original \**a* (see §2.1. s.v. *eñcuwo* and Peyrot, Dragoni, and Bernard 2022). Therefore, this word did not belong to the early layer of loanwords in which Old Iranian \**a* corresponded to TB e A a. Another argument against Tremblay's suggestion has been put forward by Peyrot (2018). His discovery that the Tocharian B word for 'mule', TB *etswe*, corresponds to PIr. \**ačwa-* 'horse' and does not show the palatal outcome observed in the Tumshuqese-Khotanese branch separates the Old Steppe Iranian loanwords from the Tumshuqese-Khotanese branch.

Without this older Old Steppe Iranian layer, the Khotanese loanwords into Tocharian amounted to no more than fifteen items, according to Tremblay's (2005) list. The Khotanese and Tumshuqese people were historically the oldest neighbours of the Tocharians, so the

<sup>5</sup> Both in the *Dictionary* and in the *Prolexis*, the quotations of the Tocharian material are mostly cursory and without an in-depth analysis of the borrowing paths involved.

<sup>6</sup> Bailey (1947: 150) concludes that 'The Annals of Khotan and the Krorayina documents show that the Khotanese had close connexions with the cities of Kashghar, Kuci, Argi and Krorayina in political matters. Linguistic interchange was inevitable.' However, it should be noted that, whereas allusions to Kashghar are pretty evident in the *Li yul lung bstan pa*, the same cannot be said about some alleged references to Tocharian-speaking towns in the North. Bailey's hypotheses on the origin of *er mo no* (KT VII: 18-9) and *o sku* (Bailey 1947: 147) need more detailed research.

number appeared to be very low. This observation constituted the starting point of this research. Two possible explanations exist for these data: either the intensity of lexical borrowing was minimal, or the corpus can still be enlarged through a more detailed analysis of the Tocharian lexicon. The first explanation considers that geographical proximity, even over a long period, does not always result in heavy borrowing from one language to another. It is entirely possible that language contact between Tocharian and Khotanese resulted only in very moderate lexical borrowing. This hypothesis may be backed by the fact that the majority of the already known Khotanese loanwords in Tocharian belong to the technical language of medicine (Dragoni 2021) and are part of the nonbasic vocabulary, the first to be borrowed in a situation of casual contact (Thomason and Kaufman 1988: 77, Thomason 2010: 41).<sup>7</sup> On the other hand, however, it can also be argued that centuries of proximity, if not more than one millennium, could have resulted in more intense contact. Given that the subject is understudied, more Khotanese loanwords may be found in the Tocharian lexicon.

#### 1.6. KEY CONCEPTS AND METHODOLOGY

As outlined in the preceding section, this study was born out of the necessity to determine whether the corpus of Khotanese loanwords in Tocharian was limited to fifteen items. The first step of the investigation involved an in-depth critical assessment of the already-known corpus of Khotanese loanwords. This preliminary analysis aimed to determine which phonological features distinguished the already-known Khotanese loanwords from loanwords from other languages.

Based on this initial corpus of fifteen items, I could establish that the Tocharian B ending nom. sg. *-o* was quite widespread among loanwords from Khotanese.<sup>8</sup> As a consequence, the focus of the research became a re-examination of all Tocharian B lexemes with nom. sg. *-o* and obl. sg. *-a* or *-ai* with unclear etymology. This methodology revealed a new set of prehistoric loanwords from Pre-Khotanese and the ancestor language of Khotanese and Tumshuqese. This study contains a detailed investigation of this new set of loanwords. In the analysis, only ca. half of the possible loanwords examined were classified as reliable. Many etymologies were rejected or considered doubtful (see §2.2.).

Before entering into the subject, some key concepts from current research on language contact need to be defined and explained.<sup>9</sup> In this study, a *loanword* is defined as a word that entered the lexicon of a language at a certain point in its history as the result of a *borrowing* process (or *transfer*, *copying*, see Haspelmath 2009: 36). The term *borrowing* broadly refers to the transfer or copying process in which any linguistic feature of a language (the *donor* or *source language*) is transferred to another language (the *recipient language*).<sup>10</sup> Following Haspelmath (2009: 50–51), I distinguish between two types of borrowing. If the borrowers are native speakers, one can speak of *adoption*. On the other hand, if they are non-native

<sup>7</sup> On the problems connected with the notion of ‘basic’ vocabulary, see Tadmor, Haspelmath, and Taylor (2010).

<sup>8</sup> I believe this ending can be interpreted as the Tocharian B adaptation of the Khotanese acc. sg. ending *-u* of the source form (see §3.4.3.2. and §3.4.10.).

<sup>9</sup> For the possibility of applying modern language contact theories to the study of ancient languages, cf. the discussion in Boyd (2021: 91–94), focused on the ancient Near East.

<sup>10</sup> Following a common habit in the scientific literature, I also use *borrowing* to refer metonymically to a borrowed element, i.e. a *loan* (Haspelmath 2009: 37).

speakers, the process is called *imposition*.<sup>11</sup> This distinction is not directly relevant to this study, as the type of contact investigated here involves most likely an adoption situation, i.e. native speakers of Tocharian borrowing from speakers of Khotanese and Tumshuqese (§6.2.3.).

Another important distinction is between *material* and *structural* borrowing (Haspelmath 2009: 39). This study is primarily concerned with lexical borrowing (i.e. *loanwords*), a type of material borrowing. Structural borrowing (e.g. *calques*) is not systematically investigated here. A loanword can undergo a process of *adaptation* in the recipient language, which may involve phonological, morphological, syntactic, or orthographic changes aimed at making the loanword fit better into the recipient language. If no adaptation process occurs, one should speak more precisely of a *foreignism* rather than a *loanword* (Haspelmath 2009: 41–42). An example of adaptation in the corpus analysed in this study is the Khotanese acc. sg. ending *-u*, adapted as nom. sg. *-o* in Tocharian B. As Tocharian B has no nom. sg. ending *-u*, the ending *-o* was chosen as its phonologically closest equivalent within the Tocharian B morphological system (see §3.4.10.).

As for the causes of borrowing, an important distinction can be made between *cultural* and *core* borrowings (Haspelmath 2009: 46–49). Cultural borrowings are loanwords for new concepts from the outside, whereas core borrowings duplicate existing words of the recipient language. It is common to refer to cultural borrowings as due to ‘necessity’ and to core borrowings as due to ‘prestige’ (see Carling *et al.* 2019).

Identifying a loanword is often a complex process. In the case of the present study, the procedure is even more difficult because it involves fragmentarily attested languages with no direct continuants in the present day (see §1.4.). Once a suspect pair of lexemes have been identified, the first step involves thoroughly examining the occurrences to determine their correct meaning and phonological shape.

The second step aims at excluding any alternative explanation to borrowing (Haspelmath 2009: 44). Therefore, the etymological proposals available in the literature for every Tocharian lexeme under scrutiny have been analysed according to the principles of the comparative method (Campbell 2020: 140–77) and the traditional check-list by Hoffman and Tichy (1980).

If, after this analysis, the Proto-Indo-European etymology of the Tocharian word appears impossible or highly uncertain, a preliminary borrowing path from Khotanese or Tumshuqese can be proposed.

The third step involves the examination of the proposed Khotanese and Tumshuqese source forms. Combining the comparative method with internal reconstruction (Campbell 2020: 194–209) allows the reconstruction of the linguistic stages of the Khotanese and/or Tumshuqese form before its historical attestation (PTK and PK, see chapter 3.). For a proposed borrowing path to be plausible, the phonological shape and the meaning of the Tocharian word should be compatible with at least one of the five linguistic stages of Khotanese and Tumshuqese considered in this study (PTK, PK, OKh., LKh. or Tq.).

The fourth step involves the determination of the *direction* of borrowing. In this study, the criteria listed by Haspelmath (2009: 45) have been adopted: a. morphological analysability in the donor language, b. signs of phonological adaptation in the recipient language, c. attestation of the lexeme in a language closely related to the donor language but spoken

<sup>11</sup> For a slightly different terminology, see Thomason and Kaufman (1988: 20–21).

outside the sphere of influence of the recipient language, d. semantic plausibility. The direction of borrowing may be difficult to establish in the case of a *Wanderwort*, i.e. ‘a borrowed word diffused across numerous languages, usually with a wide geographical distribution’ (Campbell and Mixco 2007: 220). However, as the concept of *Wanderwort* is extremely vague (De Vaan 2008a), I have tried to avoid its use as an explanatory device as much as possible. A special effort has been put into determining the most plausible borrowing directions, even if a lexeme does not reveal any recognisable Iranian etymology.

In §2.2., a classification of the examined items into three categories (reliable, less reliable/doubtful and rejected loanwords) is attempted. The checklist for the inclusion of an item into any of these three categories involves the following three criteria:

- Phonological correspondence.
- Semantic identity.
- Occurrence of the source form either in Khotanese or in Tumshuqese.

If a loanword satisfies all three criteria, it is placed in the first category (‘reliable loanwords’). Cases like TB *cowo\** ‘robbing’ violate the third principle only superficially. For TB *cowo\**, the Khotanese form is attested in a derivative with *ka*-suffix not present in Tocharian. It can be argued that a form without *ka*-suffix existed at the time of borrowing into Tocharian. This assumption is not problematic given the ample spread of the *ka*-suffix in Middle Iranian. Therefore, *cowo\** has been classified as reliable.

The second category (less reliable/doubtful loanwords) contains all the etymologies that fully satisfy two of the above criteria but only partially the third one. For instance, cases like TB *kontso\** and TB *kompo\** have an excellent phonological correspondence in an attested Khotanese lexeme, but their meaning in Tocharian is unclear. However, the contexts in which they occur may justify a translation very close to the meaning attested for the Khotanese words. In the case of TB *wicuko* ‘cheek, (jaw)bone’, the nominal formation is not attested in Khotanese. However, the verb from which it could be derived is attested, so the existence of this lexeme cannot be ruled out. Therefore, these etymologies cannot be rejected and are classified as doubtful. I have rejected all the etymologies that violate at least one of the abovementioned criteria.

#### 1.7. STRUCTURE OF THE ENTRIES IN §2.1. AND CHAPTER 5.

§2.1. constitutes the central part of this work. In this section, I discuss items that I consider potential loanwords from Khotanese and Tumshuqese into Tocharian. This part is structured as a dictionary of borrowed lexical items. The lexemes are listed according to the Devanāgarī-based order customary in Tocharian studies (DoT: xii). Both the structure of the single entries and, by extension, the structure of this work as a whole follows the tradition of studies in the loanword corpus of the Hebrew bible (Ellenbogen 1962, Mankowski 2000, Noonan 2019). Each entry has the following structure:

- Tocharian occurrences
- Khotanese and/or Tumshuqese occurrences of the source form
- Discussion
- Results



If a lexeme is well-known and very well-attested or if its occurrences have already been treated in full in recent scientific publications, the Tocharian and/or the Khotanese or Tumshuqese lists of occurrences might be omitted if they do not bear any relevance to the discussion.

The discussion includes a critical assessment of the previous literature on the lexeme (when available) and an in-depth analysis of its phonology and semantics. The results briefly recapitulate the conclusions of the discussion and establish a borrowing scenario. A complete reference list of the examined lexical items is given in §2.2. Chapter 5. discusses potential Tocharian loanwords in Khotanese and Tumshuqese. Its entries are structured on the above mentioned model, but the list of Khotanese and Tumshuqese occurrences precedes the Tocharian one.

