Watañi lāntaṃ: Khotanese and Tumshuqese loanwords in Tocharian
Dragoni, F.

1. INTRODUCTION

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 (ch. 2.) presents and discusses possible and probable Tocharian lexical items borrowed from Khotanese and Tumshuqese, and rejects a number of unlikely borrowing etymologies that have been proposed earlier. The corpus determined in ch. 2. is subject to a phonological (ch. 3.) and a semantic (ch. 4.) analysis.

1.1. TOCHARIAN

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

The manuscripts written in Tocharian B can be dated approximately from the 5th to 10th c. CE. Tocharian A, on the other hand, is attested in manuscripts dated from the 7th to 10th c. CE (Pinault 1989a: 7-10). Following the standard chronological periodisation by Peyrot (2008), Tocharian B can be further divided into an archaic, a classical and a late phase. Further, a 'colloquial' type is distinguished (Peyrot 2008: 190). As for Tocharian A, on the other hand, the language attested in the extant manuscripts seems to be more uniform. Ogihara (2014) has shown that, beside 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 before the split, which is conventionally termed 'Proto-Tocharian'. The dating of Proto-Tocharian is debated, but it can be estimated between the 10th and 5th c. BCE (see further §5.2.2.1.).

Language contact has played an important role in the historical development of Tocharian. In fact, neighbouring languages have left sometimes extensive traces in all levels of the language, i.e. phonology, morphology and the lexicon. In prehistoric times, Tocharian was probably in contact with Old Steppe Iranian, an otherwise unattested Old

Iranian language (Peyrot 2018)[1] 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 from Middle Indian dialects, chiefly Gāndhārī. As no comprehensive studies on this subject are available, the precise dating and extent of language exchange with Khotanese and Tumshuqese (see §1.2.) is not known. This study aims at filling this gap.

1.2. KHOTANESE AND TUMSHUQESE

Khotanese and Tumshuqese are two Middle Iranian languages once spoken in the southwestern and northwestern part of today's Xīnjiāng Uyghur Autonomous Region in Northwest China. At the beginning of the 20th 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).

Tumshuqese is known only from a handful of documents (Maue 2009), which can be dated approximately to the 8th c. CE (Ogihara and Ching 2017: 467-9). A particular feature of the Tumshuqese writing system are the so-called 'Fremdzeichen', or 'foreign signs'. Some of them are original inventions and some are in common with Tocharian, Sogdian and Old Uyghur Brāhmī. As no trace of these Fremdzeichen is to be encountered in the manuscript of the Tumshuqese Karmavācana (Emmerick 1985a), this text may be earlier than the rest of the documents, but no exact dating can be proposed with certainty. As far as can be gathered from the scanty material at our disposal, Tumshuqese was heavily influenced by speakers of neighbouring Tocharian B. Traces of this influence can be found in the script, a Northern variant of the 'Turkestan Brāhmī' used also for Tocharian, in the lexicon, with a significant number of loanwords, and in the literature.[2]

Khotanese, on the other hand, is much more richly documented. The literature includes literary and religious (Buddhist) texts and many documents (Maggi 2009a). The oldest manuscript is plausibly dated to the 5th c. CE on palaeographical grounds (Maggi 2004) and the language may have been spoken roughly until the Qarakhanid conquest of Khotan at the beginning of the 11th c. CE. Two main stages of the language are conventionally distinguished: Old and Late Khotanese.[3] Additionally, for the purposes of

---

[1] The contact with Old Steppe Iranian (OSIr.) is the subject of the PhD research of my colleague Chams Bernard (Leiden University), from whom I take over this provisional language label (cf. §1.4.)

[2] If the identification of the language of the so-called 'Formal Kharoṣṭhī' fragments proposed in Dragoni, Schoubben and Peyrot (2020: 357-8) is correct, this could be an earlier form of Tumshuqese. It is significant that the fragments concerned were found as far East as Kuča, Šorčuq and Tuyuq, in the vicinity of Turfan, i.e. in Tocharian speaking territory.

[3] This is undoubtedly only a conventional definition which will need to be refined in the future. Skjærvø (KMB: lxx), in addition to Old and Late Khotanese, distinguishes also a Middle Khotanese stage.

this work, I reconstruct a pre-stage which I will term 'Pre-Khotanese' (PK). Whereas manuscripts written in Old Khotanese were mainly found within the Khotan area, Late Khotanese is mostly 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 (see recently Dragoni 2017). Old Khotanese is one of the most conservative Middle Iranian languages. It preserves six of the eight Proto-Iranian cases, shows traces of a neuter gender and has preserved four moods (with traces of an injunctive) and three tenses (present, preterite and pluperfect).

The importance of Tumshuqese lies in the fact that it is genetically related to Khotanese, but it is far more conservative with regard to the phonology. As an example, one may compare Tq. *rorda-* 'given' and OKh. *hūḍa-* 'id.', both from PIr. **fra-bṛta-*. As in the case of Tocharian A and B, the comparison between Khotanese and Tumshuqese may allow the reconstruction of a common ancestor, which I will conventionally term 'Proto-Tumshuqese-Khotanese' (PTK) (Peyrot 2018: 272-4). In the case of Tq. *rorda-* 'given' and OKh. *hūḍa-* 'id.', the reconstructed form would be PTK **hra-wurda-*.

1.3. 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 prehistory of the Tarim basin. The linguistic analysis of the loanword corpus may shed light on the age and significance of the first contacts between Khotanese and Tocharian. In fact, through the comparative method (Campbell 2013) it is possible to reconstruct the pre- and proto-stages of Khotanese and Tumshuqese and determine whether the phonological features of the loanwords into Tocharian are to be dated to the Proto-Tumshuqese-Khotanese period (cf. §1.2.) or to the historically attested stages. Therefore, the relative chronology of the loanwords, together with a thorough semantic analysis, may determine precisely which parts of the lexicon were most extensively borrowed at what stage in the history of the languages under analysis.

As loanwords can provide important insights into the social interactions among different groups in the past (Epps 2015: 585-6), the analysis conducted in this study is a fundamental step towards a better understanding of the dynamics of interactions among the ancient population groups of the prehistoric Tarim basin. It is hoped that the results of this analysis may be employed in the future to address more complex questions related to power relations, prestige and language dominance and ancient population movements within the Tarim basin. On the other hand, the analysis of more recent loanwords may significantly contribute to a better understanding of the same dynamics in the historical times. As an example, the results of this study may deliver relevant materials for the study of the spread of Buddhism among the people of 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 may also contribute to a better understanding 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). As such, medical loanwords from prehistoric stages of Tumshuqese and Khotanese may shed new light on the Pre-Buddhist medical practices in the Tarim basin. The determination of the main borrowing directions of Indian medical terminology, on the other hand, may contribute to a better understanding of the main routes of circulation of medical knowledge in the region.

On a different note, this study may also be seen as 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 a preliminary phase. As Bailey's dictionary (DKS) is now definitely outdated, Khotanese lacks any comprehensive, up-to-date lexicographical tool. Accordingly, one has to make extensive use of 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* 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 examination. It follows that, in order the determine the correct meaning and phonological shape of a lexeme, it is often necessary to examine directly the text passages in which it occurs. Accordingly, some of the results of this investigation may be also read as a contribution to the philological study of Tocharian, Khotanese and Tumshuqese texts.

1.4. PREVIOUS STUDIES ON THE LINGUISTIC CONTACTS AMONG KHOTANESE, TUMSHUQESE AND TOCHARIAN

The problem of the contacts among Khotanese, Tumshuqese and Tocharian has always been inextricably connected to the problem of Iranian loanwords in Tocharian in general. A detailed analysis of previous studies on this subject is to be found in Bernard (Forthc.). In this context, only the studies directly concerned with Khotanese and Tumshuqese will be examined.

Hansen (1940) is the first attempt at a systematic overview of the Iranian loanword material in Tocharian. 51 items are analysed and commented upon. Of these 51 lexemes, a considerable number (27 items) are traced back to Khotanese. Hansen's analysis, however, is now outdated because of its lack of consideration of the Gāndharī, Bactrian and Old Steppe Iranian (see *infra*) influence on Tocharian. Accordingly, of his 27 items, only 4 can now be safely considered as borrowed from Khotanese (cf. s.v. *aṅkwaṣ(ṭ)*, *pissaṅk*, *tvāṅkaro*, *yolo*).

Except for numerous short allusions to the Tocharian material in some of his articles and, most notably, in the Khotanese Dictionary (DKS) and in the Prolexis to the Book of

Zambasta (KT VI),[4] there is only a section of one article by H.W. Bailey that deals exclusively with the contacts between Khotanese and Tocharian. In 'Recent work in 'Tocharian'' (Bailey 1947: 149-50) the author briefly lists a series of 10 lexemes which in his opinion may have been borrowed directly from Khotanese. As I show in ch. 2., of these 10 items, only 3 can be now safely considered as loanwords from Khotanese (see s.v. *aṅkwaṣ(ṭ), tvāṅkaro, ṣpakīye*).[5]

An important 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 hypotheses came from the recent discovery of the Bactrian documents (Sims-Williams 1997: 23). Isebaert's (1980) unpublished dissertation is the only comprehensive monograph on the Iranian loanwords in Tocharian. However, with regard to the Middle Iranian data, it is now unfortunately outdated. Moreover, its continuous resorting to a general label of 'Middle Iranian' without further specifying the donor language is problematic. Other useful repertoires of loanwords are the more recent Tocharian A and B lexicographical works, i.e. Adams' Tocharian B dictionary (DoT) and Carling's Tocharian A Thesaurus (DTTA).

As for the group of loanwords distinguished by the correspondence Ir. **a ~ TB e*, TA *a*, Schmidt (1985) first recognized in it a very old layer of Old Iranian provenance. Further studies (Pinault 2002: 245, Peyrot 2015, Peyrot 2018: 280, Bernard Forth.) 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 conventional designation by Chams Bernard of 'Old Steppe Iranian'.

Tremblay (2005) tried to challenge this theory 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 the interpretation of the word for 'iron', TB *eñcuwo* A *añcu**, which shows the exclusively 'Old Sakan' outcome **św* of the Proto-Iranian cluster **ćw*, and contains the Iranian vowel **a* in the donor language. I cannot agree with this hypothesis. 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 ch. 2. s.v. and a forthcoming article by Peyrot, Dragoni and Bernard). 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 that speaks against Tremblay's theory has been put forward by Peyrot (2018). His discovery that the Tocharian B word for

---

[4] Both in the Dictionary and in the Prolexis, the quotations of the Tocharian material are mostly cursory and no in-depth analysis of the borrowing paths involved is usually attempted.

[5] 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 in passing that, whereas allusions to Kashgar are quite evident in the *Li yul lung bstan pa*, the same cannot be said with regard to some alleged references to Tocharian speaking towns in the North. In fact, Bailey's hypotheses on the origin of *'er-mo-no* (KT VII: 18-9) and *'o-sku* (Bailey 1947: 147) are in need of a more detailed research.

'mule', TB *etswe*, corresponds to PIr. **aćwa-* 'horse' and does not show the palatal outcome observed in the Tumshuqese-Khotanese branch clearly separates the Old Steppe Iranian loanwords from the Tumshuqese-Khotanese branch.

Without this older layer, the Khotanese loanwords into Tocharian, according to the scientific literature, amounted to no more than 15 items. Given that the Khotanese and Tumshuqese people were historically the oldest neighbours of the Tocharians, the number appeared to be very low. This observation constituted the starting point of this research. In fact, there are two possible explanations for these data. On the one hand, geographical proximity, even through a long period of time, does not always result in heavy borrowing from one language to another. It is well 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 belonged to the technical language of medicine (Dragoni 2021) and were therefore part of the nonbasic vocabulary, the first to be borrowed in a situation of casual contact (Thomason and Kaufman 1988: 77, Thomason 2010: 41).[6] On the other hand, it can also be argued that centuries, if not more than one millennium, of proximity could have resulted in more intense contact. Given that the subject is definitely understudied (cf. *supra*), it is possible that more Khotanese loanwords may be found in the Tocharian lexicon.

The first explanation offers a possible solution to the problem of the scarcity of Khotanese loanwords in Tocharian but, to be thoroughly demonstrated, one needs to verify whether more Khotanese loanwords are to be found in Tocharian or not. The best method to do this is by establishing which phonological features distinguish the already known Khotanese loanwords from loanwords from other languages. Therefore, the set of already known items became the object of a thorough investigation. On the basis of this initial corpus, I was able to establish that the Tocharian B ending nom. sg. *-o* was quite widespread among loanwords from Khotanese.[7] As a consequence, the focus of the research was shifted to all Tocharian B lexemes in *-o*, *-a* and *-ai* with unclear etymology. This methodology revealed a whole new set of prehistoric loanwords from the ancestor language of Khotanese and Tumshuqese (PTK) and from Pre-Khotanese (PK). This study contains a detailed investigation of this new set of loanwords.

1.5. AIMS

As outlined in the preceding section, this study is concerned with the linguistic description and analysis of the Khotanese and Tumshuqese loanwords in Tocharian. Therefore, its aim is twofold. First, it aims at determining a corpus of Khotanese and Tumshuqese loanwords in Tocharian (ch. 2.). Second, it seeks to analyse this loanword

---

[6] On the problems connected with the notion of 'basic' vocabulary, see Tadmor, Haspelmath and Taylor (2010).

[7] In this study, this ending is interpreted as the Tocharian B adaptation of the Khotanese acc. sg. ending *-u* of the source form (cf. §3.4.3.2.).

corpus from the phonological (ch. 3.) and semantic (ch. 4.) point of view. The main 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. If yes, 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?

Ch. 2. is concerned with the first research question, ch. 3. with the second and the third and ch. 4. with the fourth. Ch. 5. summarises the most important conclusions and answers to the fifth question.

In ch. 4. and 5., and within the discussion of some of the lexical items in ch. 2., I have attempted to sketch some possible socio-historical scenarios that may explain the intensity and quality of language contact between Tocharian and Khotanese and Tumshuqese. However, it should be stressed that none of these scenarios has been sufficiently explored and, therefore, the historical conclusions summarised in ch. 5. still have the character of hypotheses that await a thorough investigation. It is hoped that such investigation may be carried out in the not so distant future, as it might potentially reveal a great deal about the cultural history of the Tarim basin.

1.6. KEY CONCEPTS AND METHODOLOGY

As oulined in §1.4., the starting point of this study was an in-depth critical assessment of the already known corpus of Khotanese loanwords, even if its dimensions were quite small. Once the vowel correspondences Khot. *a* ~ TB *a* and Khot. *-u* (acc. sg. of *a*-stems) ~ TB *-o* (nom. sg.) were established, the corpus could be expanded considerably. In the course of the analysis, only ca. half of the possible loanwords examined was considered as assured. A significant number of etymologies were rejected or considered doutful (see §2.2.).

For this procedure to be effective, some key concepts from current research on language contact need to be defined and explained.[8] 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* 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

---

[8] For the possibility to apply modern language contact theories to the study of ancient languages, cf. the recent discussion in Boyd (2021: 91-4), focusing on the ancient Middle East.

(the *recipient language*).[9] Following Haspelmath (2009: 50-1), 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 speakers, the process is called *imposition*.[10] This distinction is not directly relevant for 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 (§5.2.3.).

Another important distinction is that between *material* and *structural* borrowing (Haspelmath 2009: 39). This study is mostly concerned with lexical borrowing (i.e. *loanwords*), which is a type of material borrowing. Structural borrowing (e.g. *calques*) has not been 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: 41-2). An example of adaptation in the corpus analysed in this study is the Khotanese acc. sg. ending *-u*, which was 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.

As for the causes of borrowing, an important distinction can be made between *cultural* and *core* borrowings (Haspelmath 2009: 46-9). Cultural borrowings are loanwords for new concepts coming from the outside, whereas core borrowings duplicate already existing words of the recipient language. It is common to refer to cultural borrowings as due to 'necessity' and core borrowings due to 'prestige' (see recently 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 (see §1.3.) with no direct continuants in the present day. Once a suspect pair of lexemes has been identified, the first step always involves a thorough examination of the occurrences to determine their correct meaning and phonological shape. The second step aims at excluding any alternative explanation to borrowing (Haspelmath 2009: 44).[11] Therefore, the etymology of every Tocharian lexeme under scrutiny has been analysed according to the principles of the comparative method (Campbell 2013: 107-158) and the traditional principles listed e.g. by Hoffman and Tichy (1980). If, after this analysis, the etymology of the Tocharian word appears impossible or highly uncertain, a preliminary borrowing etymology can be proposed. The third step involves the examination of the proposed Khotanese and Tumshuqese source forms. A combination of comparative method and internal reconstruction (Campbell 2013: 211) allows the reconstruction of the linguistic stages of the Khotanese and Tumshuqese form prior to its historical attestation

---

[9] 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).

[10] For a slightly different terminology, cf. Thomason and Kaufman (1988: 20-1).

[11] For criticisms to this approach, see Mailhammer (2013 and 2014).

(PTK and PK, see ch. 3). For a proposed borrowing etymology to be plausible, the phonological shape and the meaning of the Tocharian word should be compatible with at least one of the five linguistics 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 sister language of the donor, which could not have been influenced by the recipient, 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 language, usually with a wide geographical distribution' (Campbell and Mixco 2007: 220). However, as the concept o*f Wanderwort* is extremely vague (De Vaan 2008a), I have tried to avoid its use as an explanatory device in this study as much as possible. In the majority of the cases examined, therefore, a special effort has been put into determining the most plausible borrowing directions, even if a lexeme does not reveal any recognizable Iranian etymology.

In §2.2., a classification of the examined items into three categories (reliable, less reliable/doutful and rejected loanwords) is attempted. The checklist for the inclusion of an item into any of these three categories involves the following three criteria: a. good phonological correspondence, b. good semantic identity, c. occurrence of the source form either in Khotanese or in Tumshuqese. If a borrowing etymology 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 with the addition of a *-ka-*suffix not present in Tocharian. It can be argued that, at the time of borrowing into Tocharian, a form without *-ka-*suffix existed. Given the ample spread of the *-ka-*suffix in Middle Iranian, this assumption is not problematic at all and seems rather quite trivial. Therefore, *cowo** has been classified as reliable. The second category (less reliable/doutful loanwords) contains all the etymologies for which the adherence to only one of the three criteria is problematic, but not to be excluded completely. Therefore, cases like TB *kontso** and TB *kompo** have a good phonological correspondence in an attested Khotanese lexeme, but their meaning in Tocharian is not clear. 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 actually attested, so the existence of this lexeme cannot be completely ruled out. Therefore, these etymologies cannot be completely rejected and are classified as doubtful. I have rejected all the etymologies that fully violate at least one of the criteria listed above.

1.7. STRUCTURE

The loanword corpus (§2.1.) is structured as a dictionary of borrowed lexical items (*Lehnwörterbuch*) in alphabetical order. Both the structure of the single entries, and, by

extension, the structure of this study as a whole, follows in the main the tradition of studies in the loanword corpus of the Hebrew bible (Ellenbogen 1962, Mankowski 2000, Noonan 2019).[12] Each entry has the following structure:

1. Tocharian occurrences
2. Khotanese/Tumshuqese occurrences of the source form
3. Discussion
4. Results

The Tocharian and/or the Khotanese/Tumshuqese lists of occurrences could be occasionally omitted if they are not deemed useful for the discussion, i.e. if the word is well-known and very well-attested. The discussion includes a critical assessment of the previous literature on the word (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, if possible. A full reference list of the examined lexical items is given in §2.2.

Ch. 3. is a description of the phonological correspondences that govern the adaptation of Khotanese loanwords in Tocharian. The correspondences are arranged chronologically, according to the linguistic stages of Khotanese and Tumshuqese from which they were borrowed. Moreover, ch. 3. also contains a classification of the loanwords according to their morphological patterns, part of speech and gender in Tocharian.

Ch. 4. classifies the loanwords according to their semantic areas. In addition, it puts forward some preliminary proposals on possible historical scenarios.

Ch. 5. is a summary of the main conclusions of this study.

---

[12] A similar structure is also to be found e.g. in Brust's (2005) volume on Iranian and Indic loanwords in Greek. Pronk-Tiethoff (2013) is only concerned with loanwords into a reconstructed stage of a language family and is therefore quite different in scope from the present study.