Like Dust on the Silk Road

On the Earliest Iranian and BMAC Loanwords in Tocharian

Chams Benoît Bernard

LEIDEN STUDIES IN INDO-EUROPEAN 27

BRILL

Like Dust on the Silk Road

Leiden Studies in Indo-European

Series editors

Alexander Lubotsky
Alwin Kloekhorst
Tijmen Pronk

VOLUME 27

The titles published in this series are listed at *brill.com/lsie*

Like Dust on the Silk Road

On the Earliest Iranian and BMAC Loanwords in Tocharian

By

Chams Benoît Bernard

BRILL

LEIDEN | BOSTON

This is an open access title distributed under the terms of the CC BY-NC-ND 4.0 license, which permits any non-commercial use, distribution, and reproduction in any medium, provided no alterations are made and the original author(s) and source are credited. Further information and the complete license text can be found at https://creativecommons.org/licenses/by-nc-nd/4.0/

The terms of the CC license apply only to the original material. The use of material from other sources (indicated by a reference) such as diagrams, illustrations, photos and text samples may require further permission from the respective copyright holder.

This research has been funded by the Netherlands Organisation for Scientific Research (NWO, project number 276-70-028). Open Access of this publication was made possible by additional funding of NWO.

The Library of Congress Cataloging-in-Publication Data is available online at https://catalog.loc.gov
LC record available at https://lccn.loc.gov/2025017343

Typeface for the Latin, Greek, and Cyrillic scripts: "Brill". See and download: brill.com/brill-typeface.

ISSN 0926-5856
ISBN 978-90-04-73252-0 (hardback)
ISBN 978-90-04-73253-7 (e-book)
DOI 10.1163/9789004732537

Copyright 2025 by Koninklijke Brill BV, Plantijnstraat 2, 2321 JC Leiden, The Netherlands.
Koninklijke Brill BV incorporates the imprints Brill, Brill Nijhoff, Brill Schöningh, Brill Fink, Brill mentis, Brill Wageningen Academic, Vandenhoeck & Ruprecht, Böhlau and V&R unipress.
Koninklijke Brill BV reserves the right to protect this publication against unauthorized use.
For more information: info@brill.com.

This book is printed on acid-free paper and produced in a sustainable manner.

Mācri pācri

∴

Contents

Acknowledgments XI

1 Introduction 1
1.1 Tocharian and Iranian 2
1.2 State of the Art 4
1.3 Research Issues 6
1.4 Methodology 7
1.5 Structure 8
1.6 Spelling of Tocharian Stress 9
1.7 Alphabetic Order 10

2 Old Steppe Iranian Loanwords in Tocharian 11
2.1 Introduction 11
2.1.1 *Old Iranian Loanwords and the History of Their Research* 11
2.1.2 *Linguistic Criteria for the Identification of the Correct Layer* 12
2.1.3 *Naming the Language* 14
2.2 Old Steppe Iranian Borrowings: Plausible Cases 14
2.2.1 *TB* entse, *TA* eṃts *'Envy, Greed'* 14
2.2.2 *TB* epiyac, *TA* opyāc *'Memory'* 17
2.2.3 *TB* etswe *'Mule (?)'* 18
2.2.4 *TB* ainake, *TA* enāk *'Common (Vulgar); Evil, Bad'* 22
2.2.5 *TB* -aiwenta *'Units of'* 24
2.2.6 *TB* kertte *'Sword'* 27
2.2.7 *TB* keṣe, *TA* kaṣ *'Arms' Span (Measurement Unit)'* 31
2.2.8 *TB* kraupe, *TA* krop *'Group, Crowd'* 31
2.2.9 *TB* newiya *'Canal or Channel'* 34
2.2.10 *TB* pāke, *TA* pāk *'Part, Portion, Share'* 36
2.2.11 *TB* peret, *TA* porat *'Axe'* 38
2.2.12 *TB* perne, *TA* paräṃ *'Fortune, Good Fortune'* 43
2.2.13 *TB* perpente *'Burden, Load'* 51
2.2.14 *TB* mañiye *'Servant'* 51
2.2.15 *TB* -maññe, *TA* -māññe *'Belonging to the House; Garden'* 52
2.2.16 *TB* melte* *'Pile', TA* malto *'On Top'* 53
2.2.17 *TB* maiyya, *arch.* meyyā, *etc. 'Power, Might'* 56
2.2.18 *TB* yentuke* *'India(n)'* 59
2.2.19 *TB* yetse, *TA* yats *'Skin'* 61

2.2.20 *TB* retke, *TA* rätäk *'Army'* 61
2.2.21 *TB* raitwe *and TA* retwe *'Combination, Mixture' and TB* eraitwe *'In Connection with'* 62
2.2.22 *TB* raimo*, *TA* rem *'Dust'* 68
2.2.23 *TB* weretem* *'Debt; Surety'* 70
2.2.24 *TB* waipecce *'Property, Possessions'* 71
2.2.25 *TB* śāte, *TA* śāt *'Rich'* 72
2.2.26 *TB* speltke, *TA* spaltäk *'Zeal, Effort'* 73
2.2.27 *TB* tsaiññe *'Ornament'* 75
2.2.28 *TB* tsain *'Arrow'* 75
2.3 Old Steppe Iranian Borrowings: Possible Cases 76
2.3.1 *TB* ākteke, āktike *adj. 'Wonderful, Astonishing; Astonished', Noun 'Wonder'* 76
2.3.2 *TB* ekṣinek* *'Dove'* 79
2.3.3 *TB* epastye *'Skillful', TA* opäśśi *'Id.'* 79
2.3.4 *TB* epe *'Or'* 80
2.3.5 *TB* ñyātse ~ ñātse, *TA* ñātse *'Distress, Calamity; Danger'* 92
2.3.6 *TB* twere* *'Door'* 96
2.3.7 *TB* mətstsa-, *TA* nätswā- *'To Starve, to Die of Hunger'* 96
2.3.8 *TB* mot *'Alcoholic Beverage'* 98
2.3.9 *TB* wertsiya *'Assembly', TA* wartsi *'Assembly, Crowd'* 99
2.3.10 *TB* welke*, *TA* waltäk* *'Stamen'* 100
2.3.11 *TB* tsetke *'A Hundred,* centaine *(?), Century (?)'* 102
2.3.12 *TB* tsere *'A Measure of Liquid Volume of Half a* lwāke *(?)' or 'Cup (?)'* 104
2.4 Old Steppe Iranian Borrowings: Difficult Cases 105
2.4.1 *TB* ecce, *TA* aci *'Hither, From'* 105
2.4.2 *TB* eprete *'Resolute, Steadfast'* 107
2.4.3 *TA* kump *'Cauldron'* 108
2.4.4 *TB* kercci *'Palace'* 111
2.4.5 *TB* leṅke *'Valley, Cleft'* 118
2.4.6 *TA* wankā- *'To Joke, Chatter'* 120
2.4.7 *TA* nātäk *'Lord', fem.* nāśi *'Lady'* 123
2.4.8 *TB* tˢəw- *'Attach Oneself to, Stick to', TA* tˢäw- *'To Put Together'* 127
2.4.9 *TB* tseriteke *'Young Grass (?), Lizard (?), Turtle Shell (?)'* 129
2.5 Old Steppe Iranian Borrowings: Rejected Cases 131
2.5.1 *TB* atiya*, *TA* āti *'Grass; Straw'* 131
2.5.2 *TB, TA* āp *'Water; River; Flood'* 132
2.5.3 *TB* ām *'Silence'* 133

2.5.4 *TB* kāswo *'Skin Disease, in Particular Leprosy and Alphos'* 135
2.5.5 *TA* kāwṣtā- *'Strike, Kill'* 141
2.5.6 *TB* tāno *'Seed, Grain'* 143
2.5.7 *TB* waipte *'Separately, Apart'* 145
2.5.8 *TB* waimene *'Difficult'* 146
2.6 Old Steppe Iranian Calques in Tocharian 147
2.6.1 *TB* ñäkte, *TA* ñkät *'God, Lord'* 147
2.6.2 *TB* yok, *TA* yok *'Hair, Body Hair, Color'* 149
2.6.3 *TB* śaiṣṣe *'World'*, *TA* śoṣi *'People'* 150
2.7 Discussion of the Features of Old Steppe Iranian 151
2.7.1 *Old Steppe Iranian–Tocharian Vowel Correspondences* 154
2.7.2 *Phonetic Changes of Old Steppe Iranian* 155
2.7.3 *Lexical Isoglosses of Old Steppe Iranian* 184
2.7.4 *Semantic Innovations of Old Steppe Iranian* 188

3 BMAC Words in Tocharian (a Selective Survey) 192
3.1 Introduction 192
3.2 Analysis of Potential BMAC Loanwords in Tocharian 193
3.2.1 *TB* iṣcäke *'Clay'*, *TB* iścem *'Clay'* 193
3.2.2 *TB* eñcuwo, *TA* añcu* *'Iron'* 200
3.2.3 *TB* etre, *TA* ätär *'Hero'* 202
3.2.4 *TB* kenek, *TA* kanak *'Cotton Cloth, Cotton Fabric'* 203
3.2.5 *TB* kercapo *'Donkey'* 206
3.2.6 *TB* kroṅkśe *TA* kroṅśe *'Bee'* 209
3.2.7 *TA* lepäś* *'Jackal'* 213
3.2.8 *TB* witsako *'Root'* 214
3.2.9 *TB* śerwe, *TA* śaru *'Hunter'* 217
3.2.10 *TB* ṣecake, *TA* śiśäk *'Lion'* 219
3.3 Other Possible BMAC Loanwords in Tocharian 221
3.3.1 *TB* kanti *'Naan' or 'Groats'* 221
3.3.2 *TB* peñiyo, *TA* pañi *'Splendor'* 225
3.3.3 *TB* mewiyo *'Tiger'* 225
3.4 Discussion 226
3.4.1 *Presentation of the Loanwords and Their Phonetics* 226
3.4.2 *The Semantics of the BMAC Loanwords in Tocharian* 228
3.4.3 *Western and Eastern BMAC Varieties* 232

4 Conclusion 236
4.1 Results 236
4.2 The Tocharian Way 242

Appendix 1: The Tocharian Word for 'Parrot' and Its Origin 253
Appendix 2: On the Etymology of TB *patstsāṅk* and TA *pātsaṅk* 'Window' 260
Bibliography 264
Index 288

Acknowledgments

My first acknowledgment is for the Lord of Wisdom, from Whom all knowledge, all blessings and all life arises.

The second, for my dear parents, who birthed me and raised me and encouraged me. For my dearest grandmother as well, whose eyes could not see the completion of it, but whose soul will.

The third, for my dear teachers, in particular Prof. Georges-Jean Pinault and Prof. Michaël Peyrot, who bathed me in Tocharian studies until I could swim, and encouraged me to always go forward, as well as Prof. Lubotsky for his patient teaching and readiness to help and answer all of my questions.

The fourth, for my dear colleagues and friends, who sustained me mentally and emotionally while I was writing this book, in particular Dr. Romain Garnier and Philippe Hattat as well as the two great Tocharologists Dr. Athanaric Huard and Alessandro Del Tomba (from afar), Dr. Kate Bellamy, Dr. Stefan Norbruis, Dr. Benjamin Suchard, Dr. Xander Vertegaal, Dr. Marijn van Putten and (soon to be) Dr. Aljoša Šorgo (from near), and Dr. Johnny Cheung, the living encyclopedia who remained my friend through thick and thin (both from near and afar). Dr. Athanaric Huard and Dr. Niels Schoubben, a particular word of thanks for carefully reading my manuscript and providing so many useful suggestions, and to Dr. Ruixuan Chen for his constant collaboration.

The fifth, for my non-Academic friends who always supported me during all those years, in particular my dear friends Nikoo (Zara) Qorbanseresht, Vincent van Strien, Laure Protat, Niloufar Rashedi. For my high-school teacher, Éric Blaisse, who woke up my yearning for etymology. For my teacher, Ardeshir Farahmand.

Thank you all.

CHAPTER 1

Introduction

I wish to invite the reader to visit a wondrous meadow, below magnificient mountain chains, that of language contact, and those of the Tarim Basin. Since their discovery, the Tocharians have never ceased to capture our imagination: who were they exactly, when did they leave the Indo-European homeland, how and why? What did they encounter on their way? This book will address the last point only, or rather: which Iranians did first meet their way and with which BMAC speaking people did they maintain relationships before reaching the Tarim Basin. This topic has direct consequences on our knowledge of the prehistory of the Tocharians, and on that of the regions they crossed on their more or less straight line from the Yamnaya steppes to the Tarim Basin, migrating throughout the Siberian Steppes and Dzhungaria.

This book is the enlarged and improved version of my PhD thesis, successfully defended on the 20th of April 2023 at Leiden University. It includes the original text of the dissertation, two appendices constituted by related studies written together with my dissertation, an index verborum, and many corrections and additions, often inspired by scholarly comments made by respected colleagues.

Among the additions: I have added a few word studies, many existing word studies were improved and lengthened, I added a sub-chapter on Old Steppe Iranian calques in Tocharian, and greatly updated the literature, in particular in terms of archaeological and genetic data. This is notably reflected in the conclusion. I have deleted some paragraphs, which were incorrect or too speculative. And in many parts, I have used the newest possible research in order to make the presented arguments more compelling. A study published after my doctoral defense mentions Old Steppe Iranian as possibly having a descendent in the newly deciphered "unknown Kushan script" (Bonmann, Halfmann, Korobzow & Bobomulloev 2023: 325). Whether this hypothesis is correct or not, it is apparent that the study of Old Steppe Iranian, the oldest reconstructed Iranian language of the region, has consequences for our understanding of prehistory and history of, not only Iranian languages, but also archaeology and migration history in Central Asia and Siberia. Hence, I felt the need to publish this book, as the topic continues to gain interest and attract attention. The study of these small human groups is important to understand the history of Asia.

The topic of this book is the investigation of contacts between the Tocharian languages and an Iranian language I tentatively named Old Steppe Iranian,

© KONINKLIJKE BRILL BV, LEIDEN, 2025 | DOI:10.1163/9789004732537_002
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

as well as BMAC languages. In this introduction, I will first introduce both the Tocharian and the relevant Iranian languages, mainly Old Steppe Iranian (§ 1); then discuss the state of the art of the research in this particular field (§ 2) and introduce and delimit the problems of Tocharian–Iranian contacts to be addressed and present my research questions (§ 3). After this, I will discuss the methodology of the present work (§ 4), describe the structure of the book (§ 5) and finish with a short presentation of the notation of stress in Tocharian B (§ 6) and a short note on the Tocharian alphabetic order used throughout this book (§ 7).

1.1 Tocharian and Iranian

Tocharian is a branch of the Indo-European language family. This branch is known through two extinct languages: Tocharian A and Tocharian B, which were spoken in the north of the Tarim Basin, today part of the Xinjiang region in the Northwest of China. The Tarim Basin is mostly desert, with few oases, among which those where the Tocharian culture flourished. Tocharian A and B are known to us from texts found (and sometimes bought) in the beginning of the 20th century by British, German, French, Russian and Japanese expeditions (cf. Pinault 1989: 7; Fellner 2007). The speakers of Tocharian, during the historical period and as far as we know, were Buddhist.

A large number of texts in multiple languages were found in the Tarim Basin, among them the Tocharian manuscripts (see map p. 243). The Tocharian texts were found along the northern rim of the Taklamakan desert, in several places around Kuča, Yānqí and the famous Turfan oasis. About 10,000 Tocharian B manuscript fragments and about 2,000 Tocharian A manuscript fragments have been preserved (Peyrot 2015a). Canonical religious texts are mostly Sarvāstivādin, some are Mūlasarvāstivādin, that is for both, Śrāvakayāna schools. Non canonical Tocharian texts demonstrate a strong influence of the Mahāyāna school, although in general their sectarian affiliation is less important as they focus on reaching a lay audience (Pinault 2016b: 170–171).

Linguistically, Tocharian B can be divided into three chronological stages: Archaic Tocharian B, Classical Tocharian B and Late Tocharian B (Peyrot 2008). Tocharian A texts do not show such internal variation (see, however, Itkin 2002). While only Tocharian B fragments have been found in the region of Kuča, Tocharian B and Tocharian A fragments have been found side by side in the regions of Yānqí and Turfan (Peyrot 2010: 133). No unanimously recognized Tocharian A borrowing could be identified in Tocharian B, but Tocharian A has many Tocharian B loanwords, usually, if not exclusively, from (later) Classical Tocharian B (cf. Peyrot 2010: 139, with references).

Most of the Tocharian texts are of Buddhist content, but a trove of documents, that is, non-literary texts, were also found, including everyday documents such as laissez-passers for caravans (Pinault 1987), and also inscriptions on walls of grottoes (Ching & Ogihara 2020), etc. The texts can be dated to the 5th to the 9th or 10th centuries A.D. for Tocharian B, and to the 7th to the 10th centuries for Tocharian A. There is also a Manichean Tocharian B text, in Manichean Uyghur script, written a in non-standard form of the language typical of late manuscripts from the Turfan oasis (cf. Pinault 2008b). It is thought that Tocharian A and B were not mutually intelligible (Pinault 2002a: 245).

Iranian languages are a branch of Indo-European that was spoken in a huge area, spanning from Bulgaria to Siberia, including parts of China, the entire Western part of Central Asia, the Iranian plateau, and Eastern Anatolia. Nowadays, Iranian languages are still found from Eastern Anatolia and the Caucasus to Pakistan and China. For more information on the major Iranian languages, see for example the *Compendium Linguarum Iranicarum* (Schmitt 1989). Three chronological stages of the Iranian languages can be distinguished:

1. Old Iranian: Avestan (documented), Old Persian (documented), Scythian (indirect, scarce documentation);
2. Documented Middle Iranian: Middle Persian, Parthian, Bactrian, Sogdian, Khwarezmian, Khotanese, Tumshuqese, and a newly deciphered Middle Iranian language (cf. Bonmann, Halfmann, Korobzow & Bobomulloev 2023);
3. New Iranian: New Persian, Balochi, Kurmanji, Sorani, Yidgha, Munǰī, Pashto, Yaghnobi, Wakhi, etc.

The diversity and dialectal features of the Middle Iranian languages show that there must have existed more than the three Old Iranian languages listed above. Likewise, New Iranian languages show that there must have existed more than seven Middle Iranian languages. Most, if not all of the documented Middle Iranian languages were spoken in Central Asia. Many of these languages were spoken in the West of Central Asia, in regions too far removed from the Tarim Basin for our purposes. However, some, such as Bactrian and Khotanese, were spoken in regions close to the Tocharian languages.

In the present book, I will look at the oldest contacts between Tocharian and Iranian, which can be dated to the Old Iranian stage. I will also investigate Tocharian words possibly borrowed from the so-called BMAC language, which may form an even earlier layer of the Tocharian lexicon.

1.2 State of the Art

Meillet (1913) was the first to recognize an Iranian loanword in Tocharian: the word *kuñcit* 'sesame'. Lidén (1916) repeated the observation by discussing a few more words, at a very early stage of Tocharian studies.

The first systematic study was by Hansen in 1940, who brought together a number of Tocharian words which he considered to be of Middle Iranian origin. Many of the words are indeed of Middle or Old Iranian origin, although some others were incorrectly identified, and some were based on Tocharian ghost words, due to wrong readings of the texts. Even though it was not followed by any major work for some time, Hansen's study opened the road to the critical analysis of Iranian loanwords in the Tocharian corpora.

After this, an important article was published by Werner Winter (1971). On the basis of the analysis of common phonological features, Winter suggests that a number of Iranian words in Tocharian are borrowed from Bactrian, such as TB *perne*, TA *paräṃ* 'glory, dignity; rank'. As Bactrian was only fragmentarily known at the time, the Bactrian source forms were not attested but had to be reconstructed. As Schwartz (1974) demonstrated, Winter's attribution of this layer of loanwords to Bactrian was incorrect. Although there are Bactrian loanwords in Tocharian, like TB *akālk*, TA *ākāl* 'wish' and TB *mālo* 'alcohol', words like *perne* 'glory, dignity; rank' cannot belong to the Bactrian layer.

Schmidt (1985) showed that the words that Winter had misidentified as Bactrian are in fact of Old Iranian origin, and not Middle Iranian, as had been assumed up to that point. However, he did not provide a specific identification of this "Old Iranian" language, and limited himself to listing sound correspondences between Old Iranian and the Tocharian borrowings, thus expanding the correspondences established by Winter.

Although it has never been published, Lambert Isebaert's 1980 PhD thesis, which discusses a great number of Indic and Iranian loanwords in Tocharian, deserves special mention. Based on a large corpus of words, and adding a considerable number of new ones, it systematically discusses the rendering in Tocharian of Indic and Iranian *a* and *ā*, which he considers to be "diaphones".

Based on the new discoveries in Bactrian made by Sims-Williams (2002), Pinault (2002a) studied various strata of Iranian–Tocharian contacts, focusing on Bactrian–Tocharian contacts. He posits the following stages of Iranian loans: 1. Old Iranian, 2. Bactrian, 3. Sogdian and Khotanese. He qualifies Old Iranian as "unidentified languages, differing from Avestan and perhaps attributable to the ancestors of the Ossetes" (Pinault 2002a: 245).

A major systematic study following Isebaert, Schwartz, Schmidt and Pinault was that of Tremblay (2005). Tremblay tried to systematically classify all Ira-

nian loanwords in Tocharian known to him. He also tried to identify the Old Iranian language discovered by Schmidt (1985), and suggested that it is "Old Sakan", that is, the ancestor language of Khotanese. He also assumed various stages of Middle Iranian loanwords in Proto-Tocharian, and of later borrowings.

An advantage of Tremblay's theory that the Old Iranian elements in Tocharian derive from "Old Sakan" is that earlier stages of Khotanese are known to have been spoken in the region, and we thus would not need to add a new language to explain the Iranian elements in Tocharian. However, he did not take Schmidt's (1985) findings into full account, and considered Proto-Iranian **ǰ* to be reflected as **z* in Old Sakan, whereas Schmidt had already shown that the source of the Tocharian borrowings must have had the reflex **dz*. Peyrot (2018a) adduced further counterarguments against Tremblay's identification, arguing that the Old Iranian language cannot be closely related to Khotanese. Another shortcoming of Tremblay's paper is that he assumes Middle Iranian loanwords in Proto-Tocharian, but this is chronologically flawed, because Proto-Tocharian is certainly much older than Middle Iranian. Despite these shortcomings, Tremblay (2005) is usually used as a reference work in scholarship (e.g. Peyrot 2018a).

After Tremblay's article came two studies by Peyrot (2015a; 2018a) and one by Isebaert (2019). Peyrot (2015a) follows Pinault's (2002a: 245) classification of the loanword stages, and he discusses examples of each stage. Peyrot (2018a) discusses the discovery of the Iranian loanword in Tocharian B *etswe* 'mule', which derives from Old Iranian **at^{s}u̯a-* (s.v. *etswe* and section 2.7.4.6). This latter article is very important, as it is the first seriously argumented discussion on the identification of the Old Iranian language in contact with Tocharian, claiming that it is not closely related to Khotanese, but is rather closer to the rest of Iranian languages. Isebaert (2019), based on the publications mentioned above, proposed two new Iranian etymologies, which I also discovered independently and presented at a conference, and are now to be found in this book (s.v. *entse*, *ñyātse*).

One can also cite Kim (1999), a work that tries to establish an absolute and relative chronology of Tocharian sound changes based on loanwords, some of these uncertain, some others well-known.

A related issue is that of BMAC loanwords, as they are also of a great antiquity, perhaps even older than Old Steppe Iranian loanwords. The study of BMAC loanwords is required to understand better the chronology and history of Old Steppe Iranian loanwords as well, and to understand Old Steppe Iranian vocabulary. Indeed, as both BMAC and Old Steppe Iranian represent a very archaic layer of borrowings in Tocharian, it is necessary to study both to know which is

which, notwithstanding the fact that Iranians as well had contact with BMAC speakers, perhaps at various stages (see § 3.4.2 and § 3.4.3 of the present book). BMAC, namely, the Bactriana-Margiana Archaeological Complex, also commonly called Oxus Civilization, denotes a group of coherent archaeological findings from a vast region in Northern Afghanistan and Turkmenistan. The BMAC language (or languages) are unknown, or at least, have not been directly documented. Lubotsky (2001) has demonstrated that a great number of BMAC loanwords can be identified in Indo-Iranian. Pinault (2002, 2003 and especially 2006) further applied this observation to the Tocharian vocabulary.

1.3 Research Issues

The previous scholarship mentioned above answers many questions but raises a number of new ones, because not all the discoveries and not all the findings have led to clear, defined results, and many facts remain unclear or doubtful. The most important research issues seem to be:

1. Are the words belonging to the Old Iranian stage from one single language, as Schmidt thought, or multiple ones, as found elsewhere in the literature?
2. If the Old Iranian layer is one single language, can its features be established? Is it homogeneous? How can we identify the words belonging to this stage and layer?
3. Are there any BMAC loanwords in Tocharian? Did they come directly from a BMAC language, or through, for instance, Iranian?

The goal of my book is to address these questions. The first research question will be answered in chapter 2: I will start with the working hypothesis that there was one single language, and analyze all previously identified loanwords in order to test this hypothesis. I will try to establish features of this language or these languages, if possible, and see if we can draw a coherent picture of the source language. This will permit me to answer the second research question, and establish a specific methodology in order to identify Old Iranian loanwords in Tocharian (see § 2.7).

Chapter 3 will be dedicated to the third research question: are there any BMAC loanwords in Tocharian? I will start with the working hypothesis that there were indeed BMAC borrowings in Tocharian. I will analyze a selection of such possible loanwords, among which some rejected Old Iranian loanwords, and I will investigate whether they share common features. This will lead me to conclusions about the specific BMAC language or dialect in contact with Tocharian, and on the type of contact these two might have had on each other.

1.4 Methodology

In order to address the two first questions above, concerning the Old Iranian stage of Tocharian loanwords, I will base myself on the following methodology.

Working along similar lines as Schmidt (1985) and Koivulehto (e.g. 2001), I will first select the words which were already recognized as being of Old Iranian origin. If these words have specific traits that are coherent within their subgroup, I will take these traits as representing either features of the source language, or as Tocharian adaptations of these features, thus, building on the feature analysis by Schmidt (1985).

Since the donor language is not otherwise known, I always compare any of these features with what we know of Proto-Iranian, and with other Iranian languages. In this way, the donor language(s) of these Iranian words in Tocharian is surrounded by Proto-Iranian on one side, and by Tocharian on the other, and it becomes more easy to both reconstruct the source words and understand some of its features. In order to identify what is specific to the source language, or source languages, I will focus on where it differs from Proto-Iranian. The features that are identified will be used throughout the etymological study part of chapter 2, in order to judge existing and to find new etymologies.

This will lead to a double process: first, starting with a smaller group of secure words, namely, a group of evident etymologies recognized in scholarship, I propose a series of sound correspondences. These sound correspondences will lead me to check and control those already established etymologies, in order to secure them even further. They also permit me to add new etymologies, which, after being checked against those correspondences, might yield new correspondences. These new correspondences will permit me to reject some etymologies, consider some as doubtful, and add new ones. Thus, *ainsi de suite*. The resulting etymologies are subdivided into four categories: 1) plausible cases in 2.2; 2) possible cases in 2.3; 3) difficult cases in 2.4; and finally 4) rejected cases in 2.5.

For the BMAC words in Tocharian (chapter 3), the method will be somewhat different since it will in a certain sense be based on negative results: I discuss here words which according to my analysis cannot be of Old Iranian origin, but nevertheless appear to be old, for instance because they show regular correspondences between Tocharian B and Tocharian A and are thus reconstructible for Proto-Tocharian, alongside words which were already proposed to be of BMAC origin in previous scholarship (e.g. in Pinault 2006). My goal will be to analyze these words and to determine if they share common features. Based on the coherence of these features, I hope to be able to gather more BMAC words,

primarily among the words that I discard as Old Iranian loanwords, but which still show features of early borrowing. In most cases, these words resemble suggested BMAC loanwords in Indo-Iranian.

Based on my analysis of the features of loanwords from Old Iranian and the so-called BMAC language, it will become easier to determine to which of these two sources a given loanword more plausibly belongs. Also, it will become easier to recognize Middle Iranian loanwords, since these do not conform to the features identified for the oldest layers of borrowings. In this book, I do not discuss Middle Iranian loanwords systematically, but I have added a few discussions on Middle Iranian loanwords where this seemed relevant to the discussion of an Old Iranian loanword. Further, I discuss some Middle Iranian loanwords that have been proposed to be of Old Iranian origin, or could perhaps be argued to be so.

Of course a very crucial element in the analysis is presented by the semantics: the meaning of the word should not be too far removed from the meanings we find in Iranian languages, in the sense that there should not be too many steps between an attested Iranian meaning and the attested Tocharian meaning. There is no objective method to determine whether an etymology is acceptable or not, apart from one very important element, which is the existence of parallels permitting to confirm that a specific semantic directionality is possible. Whether or not parallels exist, it is important as well to ensure that a specific semantic shift makes sense, both in general terms and in the specific cultural setting in which it is supposed to have taken place.

1.5 Structure

The book is structured as follows. First, in chapter 2, Old Iranian loanwords and calques in Tocharian are discussed, and then, in chapter 3, possible BMAC words. The book ends with a conclusion, followed by two appendices on a Chinese and a Bactrian loanword in Tocharian.

In chapters 2 and 3, I will follow a similar structure for most loanwords: first a more or less short presentation of the Tocharian meaning, and then a discussion of some etymologies previously proposed in the literature, if there are any. Then follows a critical analysis of the etymologies, and, if need be, I will propose my own etymology. At the end, I present a short conclusion summarizing all the elements described above.

At the end of these two chapters I add a linguistic discussion of the findings based on the etymological discussions of the relevant chapter. The discussion of chapter 2 will consider all the sound changes, the isoglosses and the

semantic shifts that can be attributed to the Old Iranian stage of loanwords. These will be compared to a number of selected Iranian languages, in order to refine the conclusions one can make on that Old Iranian stage. For chapter 3, on BMAC loanwords, I will follow a similar structure, although somewhat shorter.

My work is exclusively of a linguistic nature, that is, I base both my research and my conclusions on linguistic observations. However, beyond linguistics, there is much more to say, on the basis of many elements described in this book. I thus chose to venture beyond pure linguistics in the conclusion of the book (p. 242 f.), and I propose a (pre-)historical scenario of how and where Tocharians and ancient Iranians came into contact.

1.6 Spelling of Tocharian Stress

A very important part of the discussion in the next chapters is on questions relative to Tocharian B stress and the reconstructed accent of Proto-Tocharian. I will depart from the assumption that Tocharian B preserves Proto-Tocharian stress as a rule, and that this stress was lexical, and try to verify this assumption. For this, it is necessary to understand how Tocharian B stress was spelled.

While there is no clear spelling of the stress in Tocharian A and in Archaic Tocharian B (Peyrot 2008), stress is indicated in Classical and Late Tocharian B in the following way. When stressed, the Tocharian B phoneme /a/ is written as ‹ā› (there is no phonological length in Tocharian B), but as ‹a› when unstressed. The Tocharian B phoneme /ə/ is written as ‹ä› when unstressed, and as ‹a› when stressed. There is thus sometimes a spelling ambiguity between unstressed /a/ and stressed schwa /ə/, which are both spelled ‹a›. This ambiguity is usually solved by either etymology or by variants of the same word. Indeed, either an archaic spelling or a suffixed or inflected form, such as the plural ending, can confirm that the vowel was originally a schwa /ə/ or /a/. For example Tocharian B *yasar* ‘blood’ is phonologically /yə́sar/ rather than /yasə́r/, as can be deduced from its plural *ysāra* /ysára/. Besides, Tocharian B words could never be accented on the final syllable (Krause 1971: 11), so that there is no doubt about the stressed syllable in disyllabic Tocharian B words.

In some cases, we find the phonemes *i* and *u* spelled as ‹ī› and ‹ū›. This can sometimes indicate stress, for example in TB *kamartīke* ‘ruler’, TB *kuñcīt* ‘sesame’, where the stress is also known from other data. This is however not always the rule, for instance in the Tocharian B spelling *āktīke* ‘wonderful, astonishing’, where the stress can obviously not be on both the first and the second syllable. As a rule, stress is only systematically marked on the Tochar-

ian B phonemes *a* and *ə*, but this will be enough to be able to make a number of deductions concerning both Tocharian and Iranian stress.

1.7 Alphabetic Order

The alphabetic order followed in the present book is the Brahmi Tocharian one, as is customary in scholarly literature on Tocharian. The letters are:

> A, Ā, Ä, I/Ī, U/Ū, Ṛ, E, AI, O, AU, K, KH, G, GH, Ṅ, C, CH, J, JH, Ñ, Ṭ, ṬH, Ḍ, ḌH, Ṇ, T, TH, D, DH, N, P, PH, B, BH, M, Y, R, L, LY, V, W, Ś, Ṣ, S, H, TS.

CHAPTER 2

Old Steppe Iranian Loanwords in Tocharian

2.1 Introduction

2.1.1 *Old Iranian Loanwords and the History of Their Research*

In this chapter I will discuss all the Iranian loanwords in Tocharian (A and B) imputable to an Old Iranian stage. The fact that there was a "mystery" Old Iranian layer of contact with Tocharian has been known since Schmidt 1985, but the quantity of words attributable to this stage has been increasing ever since. One important question to consider is whether the loanwords of this Old Iranian stage belong to one single language or to multiple varieties. I have collected all examples that I consider to be secure, adding a few (section 2.2). I will discuss those which are more problematic in 2.3, and those which are less likely or which could also come from other sources in 2.4. Finally, those which I rejected, or which cannot be considered as of Old Iranian provenance at this stage, are discussed in 2.5.

Since Schmidt (1985), a number of discoveries have been made concerning this layer of Old Iranian words. Specific features were established already by Schmidt (1985), such as Proto-Tocharian *t^s representing the reflex of Proto-Iranian *$ʃ$ and *$ć$ in the source language, Proto-Tocharian *e representing the reflex of Proto-Iranian *a, Proto-Tocharian *a representing the reflex of Proto-Iranian *$ā$, etc. Since his seminal study, new words have been found, and I have endeavored to establish a complete list of all features, both features discovered by and after Schmidt, and newly found ones. All these features are discussed in detail in § 2.7.

Despite these features being better known than before, no study on loanwords in Tocharian or on Tocharian historical phonology has gathered them systematically. On the contrary, many studies done on Tocharian loanwords have ignored the systematicity and coherence of the earliest layer of Old Iranian loanwords in Tocharian. Tremblay (2005), notably, although he aimed at systematicity, has not been able to analyze all the data coherently, did not understand the sound correspondences, and confused various layers of contact. This has led to the inclusion of words which are not of Iranian origin such as TB *kenek*, TA *kanak* 'cotton cloth, cotton fabric', s.v. and TB *witsako* 'root', s.v. Because this category of words was included in research on Iranian loanwords in Tocharian, I have dedicated a chapter to the discussion of another group of old loanwords which are not of an Iranian origin: the so-called substratum or BMAC loanwords in Tocharian.

© KONINKLIJKE BRILL BV, LEIDEN, 2025 | DOI:10.1163/9789004732537_003
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

In fact, once this layer of oldest words is separated from the other borrowings, it can inform us on Iranian languages. Armenian has yielded a trove of useful information, and one can think that this trove has not been exploited to its fullest extent: Iranian loanwords in Armenian provide insight into Middle Iranian languages, the chronology of changes, and dialectological data, alongside cultural and historical information. The present study aims to prove that similar results can be obtained with Iranian loanwords in Tocharian, in particular, that the unearthing of an otherwise unattested ancient language will be made possible.

I will depart from the assumption that all these ancient loanwords in Tocharian, sharing a set of features, are from one single language. I provide a justification for this method and discuss the consequences of both possible methodologies in the conclusion (§ 4.1). In my view this is more practical than, for example, starting with the idea that these loanwords are from multiple languages. While the latter is possible, naturally, if one assumes that there is one single language with definite features, it will be easier to recognize when some of the features analyzed in those words do not fit, and if more than one language is to be assumed. On the other hand, assuming multiple languages makes us ask, with every word, the question whether this word belongs with this one, or with that one. With a wide set of loanwords, this is thinkable, but there are only about 48 words that I deem of Old Iranian origin discussed in this chapter.

In this chapter I will thus discuss the etymology of each of the recognized Old Iranian loanwords in Tocharian, then discuss possible calques. A discussion will ensue, where I will try to determine to which group of Iranian languages this Old Iranian language belonged, and if it is one single language. This discussion will be based on the isoglosses found in the words mentioned here, with, naturally, a greater emphasis on the first parts of the list, the more secure loanwords, but occasionally using the latter part of the list to support specific points. A conclusion will follow.

2.1.2 *Linguistic Criteria for the Identification of the Correct Layer*

The methodology I developed over the years in order to determine to which strata Tocharian loanwords belonged permitted me to establish a list of identification criteria in order to tell if a word belongs to the oldest stage of Iranian–Tocharian contact or not. This methodology I combined with that developed by my colleagues working on Iranian, Niya Prakrit and Tocharian. In turn, we refined each other's criteria list, and this let us reach some level of certainty which was not reached in previous scholarship.

One of the major criteria is the presence of final vocalism in Tocharian B or not: if an Iranian loanword in Tocharian ends in a consonant, it belongs to

a Middle Iranian language. The three Iranian sources from which Tocharian B words ending in a vowel derive are: 1. Old Steppe Iranian (the present work); 2. Proto-Tumshuqese-Khotanese, Pre-Khotanese or Khotanese (Dragoni 2023); 3. Pre-Bactrian (Bernard & Chen 2022; Schoubben 2024). Words ending in a consonant can theoretically be of Bactrian, Sogdian, Parthian origin, or from any other Middle Iranian language.

Another major criterium is the rendering of Iranian **a* by Proto-Tocharian **e* (TB *e*, TA *a*). This criterium was discovered by Schmidt (1985), and it is specific to the oldest stage of loanwords: neither words of the Khotanese group (which have **a* = **a*, cf. Dragoni 2023: 213), nor of any type of Bactrian, nor of any other Iranian language have such a sound substitution. However, an equivalence between Iranian **a* and Proto-Tocharian **e* can also be observed in BMAC loanwords in Tocharian, but other criteria are useful in order to differentiate both groups (see § 3.4.1 in particular).

Another useful tool to determine the layer of loanwords is the preservation of Proto-Iranian diphthongs. These were preserved as **aɨ̯* and **aṷ* in the oldest stage, and rendered as **ey* and **ew* respectively in Proto-Tocharian (sometimes reflected as *ey* and *ew* in archaic Tocharian B). These diphthongs had already become **ē* and **ō* in Proto-Tumshuqese-Khotanese (cf. Dragoni 2023: 213). Examples for Pre-Bactrian are lacking, but even if they were preserved, these would not yield **ey* and **ew* in Proto-Tocharian or archaic Tocharian B, but **ai* and **au*.

The oldest Iranian layer had **dᶻ* and **tˢ* as reflexes of Proto-Indo-European palatals **ǵ* and **ḱ* respectively, both reflected in Proto-Tocharian as **ts* (cf. already Schmidt 1985). All the other Iranian languages in contact with Tocharian, Proto-Tumshuqese-Khotanese (except in front of **ṷ*) and Pre-Bactrian included had /z/ and /s/, reflected in both Tocharian languages as *s*.

These are the most basic criteria. The identification and rejection of words on these bases permitted to establish more criteria, and led to the discovery of a series of sound laws proper to the earliest layer of Iranian loanwords in Tocharian, which I discuss in detail in § 2.7.2. Naturally, there are semantic arguments surrounding each example, but these are secondary to the formal elements, which are the ultimate criterion in historical linguistics. These latter, secondary elements are stronger when combined with technological and archaeological data. Excursively, one can also deduce technological information from the formal analysis of the loanwords. For instance, given that the word for 'iron', Tocharian B *eñcuwo*, Tocharian *añcu** 'iron' derives from a Pre-Khotanese form, as revealed by formal analysis and internal reconstruction (Peyrot, Dragoni & Bernard 2022), one can infer that Old Steppe Iranians did not yet possess the iron technology, or not prominently enough for it to be borrowed from them by the ancestors of the Tocharians.

2.1.3 *Naming the Language*

A new language—no matter how old—deserves a new name. If indeed this Old Iranian language is one single language, we cannot know how these Iranians called themselves or their language. Possibly, they called themselves 'Aryans', that is, Iranians, or 'Scythians', whatever reality this notion might have covered for them. Perhaps they called themselves 'men', as did the Ossetes (Digoron *ir*, Iron *iræ* 'Ossetic, Ossetian' ultimately derives from **u̯īrāh* 'men', cf. Bielmeier 1987: 103; Cheung 2002: 193), and so many other people. This we will never know for sure, until we have found a designation of that people in Tocharian A or B, or in another source.

In the absence of a clear ethnonym, or auto-ethnonym, I propose to name this Old Iranian language "Old Steppe Iranian". This is based on the supposition that the speakers of this language came from the eastern Eurasian steppes, possibly as part of the culture associated with roller pottery, which massively replaced previous Andronovo cultures (cf. Parpola 2022). In these Eurasian steppes, they were possibly in contact with speakers of Proto-Tocharian, perhaps in the plains to the north of the Tarim Basin. I dedicate a section of the conclusion of this book (§ 4.2) to thoughts on the prehistory of this people, and on possibilities surrounding their contact with Tocharians.

If Old Steppe Iranian belonged to the Old Iranian linguistic stage, and was probably spoken at some point in the steppic areas, this does not necessarily mean that it is identical to all the other Old Iranian varieties spoken in the Steppes, and in particular not to those spoken in the Western part of the Steppes (cf. § 2.7). I hope the reader will forgive the vagueness of this name of convenience, which is the best I could offer. Working with dead languages of bygone peoples has its difficulties, but it also certainly has its own beauty. It is comparable, in a way, to discovering a fascinating text in a palimpsest; or to cleaning an old frame and discovering a splendid painting hidden under the stuff. And so I hope the reader will accept to enter the story of this language, following it through the traces it left, like dust on the Silk Road.

2.2 Old Steppe Iranian Borrowings: Plausible Cases

2.2.1 *TB* entse, *TA* eṃts *'Envy, Greed'*

Tocharian B *entse*, Tocharian A *eṃts* 'envy, greed' go back to Proto-Tocharian **entse*, with Tocharian A demonstrating regular sound change: Pre-A **antsa* > **aynts* > *eṃts* (cf. Hilmarsson 1986: 282). The etymology of PT **entse* is debated. In the following, I will first discuss Hilmarsson's etymological proposal, and then put forward my own proposal, which is largely similar to that of Isebaert (2019).

According to Hilmarsson (1986: 282), TB *entse* and TA *eṃts* are inherited and derive from Proto-Indo-European *$*h_2emǵ^hos$-, compare Sanskrit *áṃhas-* 'fear, anxiety', Av. *ązah-* 'constriction; distress, peril'. In addition, Hilmarsson reconstructs an Indo-European "holokinetic" masculine *s*-stem *$*h_2emǵ^hōs$, "preserved in Lat. *angor* (*-ōris*) "anguish; compression of the throat"." For more cognates, s.v. *ñyātse*. More exactly, Hilmarsson proposes that TB *entse* and TA *eṃts* go back to an *o*-stem derivative of the *s*-stem, i.e. *$*h_2omǵ^h$-*s*-*o*-:

> [a]n abstract *o*-stem derivation with accentuated *o*-vocalism of the root [...] i.e. I.-E. *$*h_2omǵ^hso$ [my notation], would in Tocharian yield **enkse* [my notation], which would result in the attested B *eṃtse*, A *eṃts* [...]. (1986: 282).

There are three main difficulties linked to this etymology: first, it presupposes a Proto-Indo-European *o*-stem derivative *$*h_2omǵ^h$-*s*-*o*- formed from the *s*-stem *$*h_2emǵ^hos$-, of which there is no trace. Second, it is based on a sound law, namely Pre-Proto-Tocharian **nks* > Proto-Tocharian **n*(*t*)*s*, that has no parallel.[1] Third, the meaning of this word would, in any case, be far removed from that of 'greed'. The root, which means 'to tie (a knot), to restrain', often takes the meaning of 'distress', due to the idea of narrowness, of a tight throat and of difficulty to breathe (cf. *ñyātse*), but it does not normally designate 'greed'.

Because of these three difficulties with Hilmarsson's derivation from Proto-Indo-European, it seems much more probable that Proto-Tocharian **entse* was directly borrowed from Old Steppe Iranian **and^{z}a*(*h*), cf. Avestan *ązah-* 'constriction; distress, peril' < PIr. **anǰah-*, compare also Ved. *áṃhas-* 'id.',[2] as was

1 While the sound change **nkt* > *nt* occurred in Tocharian A (e.g. PT **p'ənkte* 'fifth' > TA *pänt* but TB *piṅkte* 'id.'), there is no trace of it in Tocharian B (cf. DTB: 644). The example TB *wäntalyi* 'bow(-string)' adduced by Blažek & Schwarz (2017: 97) can be explained much more easily by a connection to TB *wənta-* 'to surround' (cf. DTB: 644; Peyrot 2013: 538 f.; cf. Huard 2022: 186 for the meaning of the verb which he convincingly renders as 'ceindre, entourer', although it was translated before as 'to cover') than by a relationship to Lithuanian *vìngis* 'bow, bending' (Van Windekens 1976: 556; Blažek & Schwarz 2017: 97).

2 According to Morgenstierne (1942: 265), the derivation of New Persian *hōšāz* 'excessive thirst of cattle' from **haų̯ša-and^{z}ah-* 'drought distress' is slightly problematic. Indeed, **an-* is not expected to yield **ā-* in Persian. A compound **haų̯ša-ād^{z}i-* would mean 'drought desire; drought greed' which at first does not seem very convincing. However, one could interpret the compound as meaning 'greed (caused by) drought', which does not seem out of place here. I thus believe an etymon **haų̯ša-ād^{z}i-* to be more fitting than Morgenstierne's 'drought distress'. The word is obviously not a genuine Persian one, as one would expect †*xōšād* instead.

also proposed independently by Isebaert (2019). The assumption of a borrowing from Old Steppe Iranian accounts neatly for the form of the Tocharian word, with the regular correspondences between OSIr. **a* and PT **e*; OSIr. **dᶻ* < PIr. **ǰ* and PT **ts*; and between final **-a(h)* and PT **-e*. For the last correspondence, cf. TB *perne* < PT **perne*, borrowed from OSIr. **farna(h)*.

However, there is a discrepancy in meaning between PIr. **anǰah-* 'constriction, distress, peril' and PT **entse* 'greed'. Possibly, the Old Iranian word at the source of this borrowing developed the meaning 'greed', perhaps under the influence of the similar-sounding and perhaps related **āzi-* 'greed', a cognate word. Or perhaps the shift in meaning occurred in Tocharian, due to the influence of the verb TB *eṅk-* / TA *ents-* 'to seize, to grab' cf. TB *eṅkäl* 'passion' and *eṅkalñe* which translates *upādāna* 'attachment, clinging, grasping'. Here, the difference in the last steps could be explained as the notion of "need", motivated by external circumstances, changing to that of "greed", caused by an internal (usually mental) motivation.

An alternative etymology of Proto-Tocharian **entse* has been suggested to me by Georges-Jean Pinault (p.c.), whereby the word goes back to (PIE) **n̥-d(h₃)-o-* 'absence de don, qui ne donne pas'. This etymology is possible formally, if **d* developed to **ts* before **H*, but no Indo-European cognate comes to confirm the existence of the projected **n̥-d(h₃)-o-*. In DToA: 70, Carling and Pinault have a different view: they suggest TA *eṃts* to be related to the verb *entsa-* 'seize, grasp, take for oneself'. This, however, would hardly account for TB *entse* as the verb in TB is *eṅk-* 'seize'.

In conclusion, Tocharian B *entse*, Tocharian A *eṃts* 'envy, greed' are probably of Iranian origin and borrowed from OSIr. **andᶻa(h)*. This also implies that the sound law **antsa* > **aynts* of Pre-Tocharian A is posterior to the borrowing, and it can be used as an argument in favor of the hypothesis that at least some Old Steppe Iranian words were borrowed into Proto-Tocharian.

Interestingly, Tocharian B *entse* is probably the word in which the Tocharian B suffix *-sse* has arisen. This suffix is found in *śāmpasse* 'with anger' (meaning suggested by Huard, p.c.), *waikesse* 'with falsehood', *werässe* 'with hate', and similar words, according to Winter denoting 'provided with the negative property X' (cf. Winter 1979: 991).

Winter (1979: 991–992) argues that this suffix spread from *entsesse* < **entsetstse* 'greedy, of greed', where *-tstse* was dissimilated to *-sse* after the *-ts-* of the base word. It spread both to inherited words such as *waike* (DTB: 666) and to other loanwords, such as *wer*, which is borrowed from Prakrit, and *śāmpo**, which is borrowed from Khotanese (cf. Dragoni 2023: 174–176). Thus, indirectly, the Tocharian B suffix *-sse* arose due to the borrowing of the Old Steppe Iranian word **andᶻa(h)*.

2.2.2 *TB* epiyac, *TA* opyāc *'Memory'*

Tocharian B *epiyac* and Tocharian A *opyāc*, both meaning 'memory', go back to Proto-Tocharian **epiyac*. The Tocharian A *o-* of *opyāc* is the result of the *p/m*-umlaut of Pre-TA **a* < PT **e*, as in TA *porat* (< PT **peret*) 'axe' s.v. TB *peret*, TA *porat*. TB *epiyac* and TA *opyāc* have been recognized as Iranian loanwords for a long time (e.g. Hansen 1940: 151).

The Iranian source of the borrowing clearly belongs to the group of words of Khotanese *byāta* 'memory', and Middle Persian *ayyād*, New Persian *yād* 'memory', etc. These words reflect a formation with PIr. **abi-*, and the root is set up as **HiH-* by Cheung (EDIV: 175–176), i.e. PIr. **abi-HiHta-* (or **-ti*). Cheung translates **HiH-* as 'to remember' and supposes that it is related to Skt. *yā-* 'drive'. We could perhaps think of **abi-HiHta-* as 'the thing that came around (the mind)' that is, the memory.

The phonological correspondences to be observed between the Iranian words cited and PT **epiyac* point to borrowing from Old Steppe Iranian, for which a form **abiiāta-* can be set up. The correspondences are then mostly regular: OSIr. **a* is rendered as PT **e* and OSIr. **ā* is rendered as PT **a*. The prefix **abí-* presumably had fixed accent on the second syllable (as in Indic), which is reflected in Tocharian, i.e. TB *epiyac* /epə́yac/. Naturally, this is the normal expected Tocharian B accent, so the specificity of this accentuation can only appear by comparing this word to other Old Steppe Iranian words.

However, the final *-c* of the Tocharian forms can absolutely not derive from OSIr. **-ta*, and it was already mentioned as problematic by Hansen (1940: 151). Nonetheless, most researchers have presented it as being the result of a palatalization of **t* in front of **i*, as if from **abi-iāti-*, e.g. Isebaert (1980: 103); Hilmarsson (1986: 56); Klingenschmitt (2000: 199); Tremblay (2005: 424); Pinault (2008: 451); etc.

Adams observes, correctly in my view, that nothing permits us to posit a proto-form **abi-iāti-* at any stage of Iranian (DTB: 95). Nevertheless, he too suggests that a "Middle Iranian" **abi-iāti-* was borrowed early enough for it to undergo the Tocharian reflexes of PIE *ti*-stems, with an alternation *-t* ~ *-c*.

This is completely *ad hoc*, as there is not solid evidence that **ti* is reflected as *c* in Tocharian words borrowed from Old Steppe Iranian, and this one single example is not convincing enough. If one believes my tentative derivation of TB *epastye*, TA *opäśśi* (s.v.) < PT **epəstəye* from OSIr. **abí-st(H)-iia-*, this would show that Old Steppe Iranian **-ti-* was rather reduced to **-tə-*, at least after **-s-*.

A few hypotheses can be set up to explain the *c* of TB *epiyac*, TA *opyāc*, but none so far is convincing. An idea which I abandoned consisted in suggesting that, since *epiyac* is frequently used together with the verb *yam-* 'to do', with the meaning 'to remember', a palatalization of **epiyat yam* (< **epiyate yam*),

after which the phrase would have been analyzed as *epiyac* + *yam*, could hypothetically have occurred. There is, however, no explanation for the supposed loss of final *-e*, and, more importantly, no reason to consider an inner Tocharian palatalization of *-t* + *y-* > *-c* + *y-*. This proposal is thus not attractive in order to explain the palatal *-c* of *epiyac* 'memory' and other solutions must be put forward.

One could, for instance, imagine a derivation from a participle **i̯ānt-* to which the Old Iranian **abi-i̯āta-* could go back. It would thus be **i̯ānt-* → **i̯āϑi̯a-* (cf. Avestan *hant-* → *haϑi̯a-*), but this is no longer a productive derivation process in Iranian. Another possibility is an ancient gerundivum: **abi-i̯ā-tii̯a-* 'what needs to be remembered', but this word too is not found anywhere.

Yet another possibility to explain the final *-c* in TB *epiyac*, TA *opyāc*, consists in viewing the word as an ancient allative form PT **epiya-c* or **epiyate-c* (> **epiyatc*) 'towards the memory'. Memory, indeed, represents a moving process, cf. the possible etymon **HiaH-* 'to go' of Old Iranian **abi-i̯āta-*. Memory can be conceptualized as a motion (of the mind) towards the past. One could think that this notion of movement would have been expressed as an allative in Proto-Tocharian. Either the allative was added to a bare form **abi-i̯ā-* instead of **abi-i̯āta-*, or one can conceive of a simplification **epiyate-c* > **epiyatc*, but it has no parallels in our corpus. To support the proposal of an allative, Dr. Federico Dragoni (p.c.) informs me that Khotanese *byāta* 'memory' is strictly used in collocation with *yan-* 'to do', and that it is generally assumed that it is originally a frozen instrumental.

To conclude, the Proto-Tocharian word **epiyac*, TB *epiyac*, *epyac*, TA *opyāc* was borrowed from Old Steppe Iranian. The etymology is evident, and it has many cognates among Iranian languages. However, the final *-c* of the Tocharian forms remains difficult to explain.

2.2.3 *TB* etswe *'Mule (?)'*

The Tocharian B word *etswe* was first identified and recognized as a loanword from Iranian by Peyrot (2015: 222–223), and then discussed extensively in Peyrot (2018a).[3] It appears in a bilingual Tocharian B–Old Uyghur text, and corresponds to Old Uyghur *katır* 'mule'.[4]

3 If the word were attested in Tocharian A, which is not the case, we would expect †*atsu*.

4 Interestingly enough, the Old Uyghur word *katır* was borrowed from Sogdian *γrtr'k* 'mule' < **xara-tara* 'a sort of donkey' (cf. Henning 1946: 723). This made me think that, perhaps, Old Steppe Iranian **atˢu̯a-* is a shortened form of **atˢu̯a-tara-* 'a sort of horse' (thus, a mule, cf. Persian *astar* 'mule', which has the same etymology). Although this would perhaps solve things semantically, it is an arduous idea, as no falsifiable parallels come to comfort it.

There is no doubt that this word must be a borrowing, as the Tocharian reflex of the same Indo-European etymon (*$h_1e\acute{k}_{\underset{\cdot}{u}}o$-) is found in TB *yakwe*, TA *yuk* 'horse'. It was already known that Proto-Iranian *$\acute{c}$, *$\acute{j}$ were reflected as *t^s, *d^z in Old Steppe Iranian, see e.g. TB *entse* or TB *tsain*. Since *etswe* must derive from Old Steppe Iranian *$at^s{}_{\underset{\cdot}{u}}a$- < PIr. *$a\acute{c}_{\underset{\cdot}{u}}a$- < PIE *$h_1e\acute{k}_{\underset{\cdot}{u}}o$- 'horse' (cf. TB *yakwe*, TA *yuk* 'horse', cf. Peyrot 2015: 223), this borrowing shows that the Proto-Iranian cluster *-$\acute{c}_{\underset{\cdot}{u}}$- was reflected as *-$ts_{\underset{\cdot}{u}}$- in this language (see Peyrot 2018a: 271–272). In Proto-Khotano-Tumshuqese the cluster *-$\acute{c}_{\underset{\cdot}{u}}$- remained as such, and is regularly reflected as -*śś*- in Khotanese, as in Khot. *aśśa*- 'horse' (DKS: 11). It has been suggested that the Iranian word for 'horse' had been borrowed by other languages as well, for instance apparently Ugaritic *s*, *ṣs*, *ṣw* m. 'horse' (cf. Del Olmo Lete & Sanmartín 2003: 772).[5]

If, as Peyrot suggests (2018a: 271), the word was borrowed from Old Steppe Iranian together with a horse or a horse-like animal, it seems that the speakers of Old Steppe Iranian had a different kind of horse than the "inherited" horses that the speakers of Proto-Tocharian called *$yək^we$. Or else, the word should have undergone the semantic specialization within Tocharian that caused it to be rendered in Uyghur as *katır* 'mule'. Archeological data seem to indicate that Old Steppe Iranians—not necessarily Old Steppe Iranian speakers, of course, but Iranians from the Steppe around the second and first millennium BCE, were skilled horse-riders who used horse riding as a fast transportation method, for instance in raids (see p. 246). Whatever contact between them and Tocharians took place, it is obvious that Tocharians would not call their horses 'mules'. Either this semantic change occurred over time, or the translation 'mule' needs to be specified, which is why I write an interrogation mark after 'mule'.

An important scientific breakthrough was made by Trautmann, Frînculeasa, Preda-Bălănică & al. (2023). They have looked at alterations in human skeletons, due to horse-riding habits, and have shown that such alterations can be observed in five specific Yamnaya individuals from the 3rd or 2nd millennium BCE, in Romania, Bulgaria and Hungary. They are the oldest human beings identified as riders. This contrasts with a recent archaeogenetic research on horses (Librado, Khan, Fages & al. 2021) which has suggested that, contrary to

5 However, Wikander (2022), convincingly argues for a Luwian origin of the Ugaritic word and its Semitic cognates (cf. also Hebrew *sūs*, Akkadian *sīsū*(*m*)). More specifically, he suggests they derive from a Luwian plural **azzuinzi* /aʦwinʦi/. If, however, it were uncorrect, an Indo-Aryan origin is much more likely than an Iranian origin, if only for historical and geographic reasons (cf. Wikander 2022: 428–429[6-8-9] for a comprehensive discussion on conclusions found in the scholarship concerning the etymology of these Semitic words, as well as a discussion of the Indo-Aryan origin of the Hurrian word for 'horse').

the commonly held belief, the Indo-European expansion is not directly connected to horse and horse-riding. The ancient horses of Yamnaya were domesticated, or rather "managed and herded" (Librado, Khan, Fages & al. 2021: 638) but they were not used in the Indo-European expansion. The general use of horses for large-scale transportation would be later, and date back to the 3rd and 2nd millennium BCE; and spoke-wheeled chariots emerged around 2000 to 1800 BCE in the Trans-Ural Sintashta culture, which is further to the East (Librado, Khan, Fages & al. op cit.). On the other hand, literary motifs coupled with etymological analyses seem to project horse-riding and horse-chariots back to the earliest Indo-European migrations (and even conquests) in the Ancient Near East (cf. Wikander 2022: 452–453), perhaps earlier or about the same time as the date suggested by Librado, Khan, Fages & al. for the emergence of spoke-wheeled chariots.

Regarding Tocharians, the latest archaeogenetic study suggests they left the PIE homeland around 3400 BCE (cf. Solans 2024: 268 f.). This date is coherent with all other genetic and archaeological data, and such a high date is also coherent with the linguistic argumentation in favor of an early departure of Tocharians from the IE homeland. This corresponds to the higher end of the horseriding chronology given by Trautmann, Frînculeasa, Preda-Bălănică & al. (2023), which means that either Tocharians left shortly after horse-riding was introduced (prompted by it?) or before it started being practiced more widely in the Indo-European homeland. The methodology of Trautmann, Frînculeasa, Preda-Bălănică & al. (2023) has been criticized by Hosek, James & Taylor (2024), who suggested ways to improve it, notably by integrating more detailed work on the various pathologies caused by different types of horse transport, by looking at more individuals, including populations for whom horse transportation is effectively known, and by looking at the effect of riding non-horse animals. The refined strategy Hosek, James & Taylor (2024) suggest will certainly lead to more efficient research and more secure results in the years to come.

I expect more research to also confirm and perhaps even expand our knowledge on early PIE riding, at the present stage, I believe I cannot say much more than that it is possible, but not certain, that Tocharian speakers were a horse-riding population when leaving the homeland around 3400 BCE. Besides, it is unlikely that horses were raised by the ancestors of Tocharians and brought with them, as horse-breeding started at a much later date (e.g. Librado, Khan, Fages & al. 2021). It is conceivable that they practiced horse-riding without practicing active horse-breeding. Wikander (2022: 442) poetically suggests that "[h]orses were demonstrably present in the [Indo-European] homeland, and the concept of it demonstrably highly present in the descendants of their lan-

guage, and in their religious traditions." He adds that Indo-Europeans "[…] brought the *idea* of the horse, the concept of the horse, the poetic character of the horse".

In the case of the ancestors of the Tocharians, I would suggest that they used wild horses (for riding, at some point charioting, eating) that they found in their environment. Indo-Iranians left the homeland later, and they were more familiar with horse-breeding. Old Steppe Iranians were certainly very apt riders. This would provide two different backgrounds for the zoological reality behind 'horse' in Iranian on the one hand and Tocharian on the other. To note, the world's oldest saddle was recently discovered precisely in the Yanghai cemetery site in the Turfan Basin (Wertmann, Yibulayinmu, Wagner & al. 2023). It has been radiocarbondated to 727–396 BCE. Very interestingly, the authors write that:

> Despite the generally very good preservation of leather, only two saddles were discovered in Yanghai which makes them an exception rather than the norm and raises the question of whether these saddles were acquired from more specialized horse breeders, riders, and saddlers in the North. (Wertmann, Yibulayinmu, Wagner & al. 2023: 1).

They also say that

> The woman in tomb IIM205, who was bestowed the saddle for her afterlife, might have come from a community of mounted pastoralists living in the steppes or mountains in the near or distant western or northern neighborhood or her saddle was among the items acquired through contact with them. (Wertmann, Yibulayinmu, Wagner & al. 2023: 14).

We are thus bound to wonder whether these mounted pastoralists who lived in the North of the Tarim Basin around the first half of the 1st millennium BCE are not either the Proto-Tocharian speakers or the Old Steppe Iranian speakers—or perhaps both.

This entire discussion is new, and the archaeological research on the topic is on-going. In particular, I think more discoveries linked to the Afanasievo animal remains and horses could be critical. At this point, all evidence I have stated remains a tool and no conclusive thesis can be built on the reality of the *etswe* vs. the reality of the *yakwe*. Speculatively, one can suggest that Old Steppe Iranians had a different name for 'horse' as a ridden animal, and **at^{s}u̯a-* was dedicated to bearing weighty things, hence the equation in the Tocharian text to Old Uyghur *katır* 'mule'.

In conclusion, Tocharian B *etswe* 'mule (?)' is a borrowing from Old Steppe Iranian *$at^{s}u̯a$- 'horse'. Semantic details concerning the difference between TB *etswe* and TB *yakwe*, TA *yuk* 'horse' are not completely clear yet.

2.2.4 *TB* ainake, *TA* enāk *'Common (Vulgar); Evil, Bad'*

Tocharian B *ainake* and Tocharian A *enāk* mean 'common (vulgar); bad, mean'. There is only one occurrence of the word in Tocharian A, which I cite below.

A 343 b5
tu enāk wraso(*m*) ///
"you [are] a mean person" (DToA: 70).

The meaning is not obvious from the context, but it is very well possible, and nothing comes to contradict it.

It has long been assumed that these two words are of Iranian origin (most recently DToA: 70). Hansen (1940: 146) saw a connection of this word to MP ‹ʾynykyh› *ēnīkīh* 'Gewalttat'. The Middle Persian word—also cited as a cognate by Schmidt (1985: 766[25])—is not in the main Middle Persian dictionaries. It is, however, discussed in Bartholomae (1906: 93–94), who proposed the reading *ēnīkīh*. I could find it as seen in Figure 1 on the next page (Vendidad 2:5).

As it appears, the word should be transcribed as ‹ʾyndgyh› (because of the diacritic over the ‹y/d›, although this can be a scribal mistake). It is translated as آزار(*āzār*) 'great pain, suffering' in Persian below, and it notably translates Avestan *axti-* 'disease (or pain)' in the phrase *nōit axtiš nōit mahrkō* (Vd. 2:5) "neither disease/pain, nor death". A similar transcription is needed for other manuscripts, such as G10 (‹ʾyndkyh›, ‹ʾyrdkyh› or ‹ʾywdkyh›). I suppose that the reason why the scribe wrote this as ‹ʾyndgyh› is that he understood the word as *āyandagīh* 'retribution' in the sense of pain (the word is not attested in Middle Persian to my knowledge, but it is easily formed and its meaning is evident). Even if the meaning were 'evil' and the word *ēnīgīh*, it would not come regularly from Proto-Iranian *$ai̯naka$- *vel sim*. In fact, besides assuming a stem *$ai̯ni$- suffixed with *-ka-*, there would be no way to explain *ēnīgīh*. In any case, the problems surrounding this Pahlavi word are too complicated to let it be used in the discussion surrounding the Tocharian words.

Tremblay (2005: 425) proposes to connect Tocharian B *ainake* and Tocharian A *enāk* to what he calls a "Middle Iranian formation" *$ai̯naha$-*ka-*. However, a formation such as *$ai̯naha$-*ka-* cannot be Middle Iranian: what Middle Iranian language would have preserved such an archaic form? He perhaps meant to consider this as a *Transponat*, but it is not clear. Let us look at both the preservation of *-aha-* (even though probably contracted to *$ā$ in the Iranian

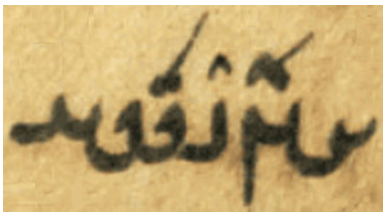

FIGURE 1
Avestan manuscript F10, Vendidad 2:5

source word) and the word-final *-ka* (→ *-ke*), neither of which are preserved in any known Middle Iranian language. Also, Tremblay claims that the phonetics of the Tocharian forms are clearly Proto-Tocharian, that is, that borrowing occurred at the Proto-Tocharian period (Tremblay 2005: 425). This is chronologically impossible, as there are no parallels for Middle Iranian loanwords in Proto-Tocharian.

Isebaert (1980: 115) proposes to derive *ainake* from OIr. **aiⁿnāka-* 'misdadig, gemeen, lasterlijk'. Although this would work phonetically, it is only a transposition of the Tocharian form back into Old Iranian, with no evident basis in Iranian. Schmidt (1985: 763) proposed to connect the Tocharian words to Avestan *aēnah-* m. 'Übeltäter', similarly to Isebaert, and considered rightly that the preservation of the *aⁱ-* diphthong indicated a borrowing of Old Iranian age. For the meaning 'evil', a derivation from the neuter stem is possibly preferable.

The preform posited by Isebaert is a plausible etymon. Indeed, since the Iranian form is too archaic to be linked to any known or unknown Middle Iranian language, it seems sound to ascribe the origin of the Tocharian words, which go back to Proto-Tocharian **eynake*, to an Old Steppe Iranian form **aiⁿnāka-* < Pre-OSIr. **aiⁿnaha-ka-*, cf. CSogd. *'yn'ɣwc* adj. 'abusive' (cf. Sims-Williams 2020: 45). A contraction of **aha* to **ā* would explain the remarkable **ā* in the Old Steppe Iranian form, for which Isebaert had offered no explanation.

This word is particularly interesting because it is, to my knowledge, the only secure ancient **-ah* stem (like *perne*, *entse*) among Old Steppe Iranian loanwords in Tocharian that is suffixed with *-(a)ka*.

Iranian languages have two strategies with regard to *-ka* suffixation of *h* stems: **ah-* + *-aka-* > *-aka-*, like in Avestan and most Iranian languages (e.g. Avestan *vaβžaka* 'wasp' ← **vaβžah-* < **u̯ebʰ-so-s*), and **ah-a-ka-*, as is reflected in Sogdian (e.g. *sāk* 'number' < **sahaka-*, SD: 453), and a few other languages, such as Yidgha-Munǰī (e.g. *sīɣ* 'hare' < **sahaka-*, cf. Morgenstierne 1938: 57). These two strategies are morphological in nature, rather than phonetic, and thus relevant in order to determine isoglosses within the Iranian phylogenetical tree. The Iranian source language of this Tocharian form was closer to Sogdian, Yidgha-Munǰī, and a few others, in this respect, than to the rest of the Iranian languages.

Since **aḭnaha-ka-* would have been rendered in Proto-Tocharian as **eyne'eke* > TB †*aineke*, it appears that the change **aha* > **ā* (then borrowed as PT **a*) happened in the source language of the borrowing, not in Proto-Tocharian. In other words, there is no contraction of *e.e* to *a* in Tocharian, while *a.a* in Iranian would certainly yield *ā*, as in the examples cited above from other Iranian languages. For a more detailed discussion on this sound change, see § 2.7.2.3.

Another interesting point concerning these words is that the archaic form of Tocharian B *ainake* is *eynāke* (cf. Peyrot 2008: 58). Together with *meyyā* (s.v. *maiyya*) and *krewpe* and *newiya*, this shows the rendering of Old Steppe Iranian **a* + *yod* or *waw*. Apparently, the Old Steppe Iranian diphthong sounded to Tocharians like **e* (= OSIr. **a*) + *yod* or *waw* rather than like the native Tocharian diphthong **ay*.

In *ainake* we can also notice that it had initial stress, i.e. /áynake/, which is characteristic of Old Steppe Iranian loanwords, see § 2.7.2.7. However, if we take Vedic *énas-* 'crime, mischief' into account, the initial stress in this word might be old.

One problem remains: the Tocharian A form *enāk* did not undergo vowel weakening, the rule being PT **ay _ *ā* > TA *e_a* (see for example Pinault 1989: 45; Kim 2007: 1). If vowel weakening had applied, one would expect a Tocharian A form †*enak*. One can propose, very cautiously, that TA *enāk* was, in fact, borrowed from the posited Sogdian form *ēnāk**, which is not attested as such, but can be deduced from its derived form, *ʾynʾqwc* 'abusive' (cf. Sims-Williams 2020: 45). This assumption would explain the absence of vowel weakening in the TA word. I have made a similar proposal for TA *paräṃ* 'glory, rank' below, in order to explain the lack of *p/m*-umlaut of the *a* of the first syllable, which did occur in *porat* 'axe' < pre-TA **parat*, as well as in inherited words, such as TA *ñom* 'name' < pre-TA **ñam*.

In conclusion, Tocharian B *ainake* 'common (vulgar); evil, bad', Archaic Tocharian B *eynāke*, derives from Proto-Tocharian **eynake* 'id.', itself a borrowing from Old Steppe Iranian **aḭnāka-*, with short **a* being rendered as PT **e* and long **ā* as PT **a*. Tocharian A *enāk*, though, might either derive from the same source-word, or from Sogdian.

2.2.5 *TB* -aiwenta *'Units of'*

It is difficult to translate Tocharian B *-aiwenta*. I translated it in the title as 'units of', but it is in fact an individualizing plural suffix. For the translation of this suffix, see Winter (1962) and Pinault (2008: 462), who treats it as a "distributive plural" or a "plurative".[6]

6 "Le tokh. B possède une formation spéciale de pluriel distributif ou "pluratif" [...], par uni-

The element *-aiwe* immediately reminds one of the Iranian individualizing suffix *-*aiu̯e* (e.g. Middle Persian *-ēw*, New Persian *-ē*).

In Tocharian B texts, it is found in various words, such as:

THT 368 a4
täṅwaṃñana eśaiwentasa
"with lovely eyes"

IOL Toch 188 b1
śle yasar misaiwentasa
"with blood and pieces of flesh"

PK AS 16.1 a5
naś kwalyiṃne yokaiwenta tesare
"… they put the individual drinks in the *kwalyiye**"

It is also found with the word for house/habitation: *ostuwaiwentane* (multiple texts), which can be translated as "in the individual groups of houses".

Since *misaiwenta* is built on *misa*, formally a plural, Winter (1962: 116) notes that it is likely that *ostwaiwenta* is built on the plural stem *ostwa* and not on the singular *ost*, which would have yielded **ostaiwenta*. He thus translates it as "several groups of houses" and, accordingly, translates *eśaiwenta* as "[many] individual pairs of eyes". He rightly corrected Krause's analysis of the suffix as "*-īwe-*": *eśaiwenta* for instance clearly shows the morphological division to be *eś-aiwenta* and not **eśa-iwenta*.

Krause (1954) explained the suffix (which, according to him, was used for a "Plurativ" number of Tocharian) as deriving from PIE *$o_{i}^{\cdot}u̯o$- 'one, alone': "Hat man das hieraus zu erschließende Element *īwe* vielleicht etymologisch mit idg. *$o_{i}^{\cdot}u̯o$- „ein", „allein" zu verknüpfen?" (1954: 6[1]). The second element is the Tocharian B plural suffix *-nta*.

---

verbation d'un pluriel massif avec un terme exprimant la pluralité d'unités individuelles: par exemple, B *mīsa* "viande" (plur. tantum): *misaiwenta* "morceaux de viande"; B *ost*(*u*)*wa* "maisons (vrai plur. de ost "maison"): *ostuwaiwenta-ne* "dans des maisons prises une par une" (une pour chaque moine). Le nom univerbé, B *-aiwe-nta*, remonte au pluriel individualisant de *aiwe** < **aiwæ* < **oi-wo-*, cf. lat. singuli, angl. singles. La formation de pluriels individualisants correspond à une tendance sur le long terme: elle rend compte aussi en partie de la multiplicité des suffixes de pluriel. […]" (Pinault 2008: 462). Winter (1962: 117) argued against the existence of a "plurative" as a morphological category of inflection in Tocharian. In my view, he is right, as data is too scarce to indicate a real paradigmatic morphological use of the suffix *-aiwenta*, which only occurs in a limited number of texts and with a limited quantity of nouns.

No forms ending in *-*aiwe* are attested, and it seems that there was only a plural *-aiwenta* in Tocharian B. However, given that plurals in *-nta* were very productive, it is possible that the suffix *-nta* was added much later to a suffix *-*aiwe* which was, of itself, individualizing.

As to the etymology, this suffix could theoretically be inherited, as proposed by Krause (cf. also Pinault 1988: 202). A series of arguments stand against this proposal. For one, Tocharian does not have any other reflex of PIE *$o_{\underline{i}}$- 'one'. Most Indo-European languages have a reflex of this root, for which they use different suffixes (it is never or almost never unsuffixed): *$o_{\underline{i}}$-*no*- in Latin, Celtic, Germanic, Baltic, *$o_{\underline{i}}$-*ko*- in Indo-Aryan, *$o_{\underline{i}}$-$u_{\underline{u}}$*o*- in Greek and Iranian. Tocharian uses a different root altogether for 'one': TA *sas*, obl. *ṣom*, TB *ṣe*, obl. *ṣeme*. Besides, of all Indo-European languages, only (Indo-)Iranian and Greek have forms that go back to **Ho*$i_{\underline{i}}u_{\underline{u}}$*o*- (e.g. Avestan *aēuua*- 'one'). In Sanskrit it perhaps became a particle, *evá*, but that is uncertain (cf. EWAia[1]: 270). Furthermore, it is only in Iranian that *$o_{\underline{i}}u_{\underline{u}}$*o*- has become a productive suffix, found in several languages such as Middle and New Persian, Balochi, some Eastern Iranian languages, etc.[7] It would seem very coincidental that this suffix also arose in Tocharian, which has not only no trace of inherited *$o_{\underline{i}}u_{\underline{u}}$*o*- as an independent word, but also no other reflex of *$o_{\underline{i}}$- whatsoever.

Pinault (1988: 201) had proposed the meaning of a new form TB *iwār*: 'at once', which he translates as "'aussitôt', 'd'un seul coup, tout à coup' ('auf einmal')", etymologically constituted of *$\dot{i}$*w*- and of the adverbial suffix *-ār* (1988: 201 f.). He wondered if *$\dot{i}$*w*- could come from "*(*e*)$i_{\underline{i}}u_{\underline{u}}$*o*-" while **aiwe* would derive from **Ho*$i_{\underline{i}}u_{\underline{u}}$*o*- or from *-*a-*$\dot{i}$*we*, the latter hypothesis being disproved by Winter (1962, see above).

Pinault's etymology is difficult formally, as there is no other evidence for ablaut in *$o_{\underline{i}}u_{\underline{u}}$*o*- in Indo-European. Pinault (p.c.) no longer believes this etymology to be true: indeed, the suffix of distributive numerals *-ār* (in fact *-*arə*) can no longer be considered a cognate of the adverbial suffix *-ār* (< *-*ōr*), cf. Pinault (2008: 560–561). Further, the form *iwār* is likely to be a ghost word, and *i*(*s*)*t*(*ā*)*r* 'forthwith, right away' should be reconstituted instead (Athanaric Huard, p.c.).

There is thus no other trace of PIE *$o_{\underline{i}}$- 'one' in Tocharian. One could even perhaps consider it as a post-Tocharian core IE innovation, in the framework of the Tocharian first hypothesis.[8] Because of this, and because *$o_{\underline{i}}u_{\underline{u}}$*o*- is rather rare in Indo-European, being only found in Indo-Iranian and Greek, it seems to me that *-aiwenta* can more straightforwardly be derived from Old Steppe

7 On this suffix in Middle Persian, see Josephson (2011).

8 I thank the anonymous reviewer of the book for that suggestion.

Iranian *-*aiu̯a*- by means of borrowing. This *-*aiu̯a*- could have been an individualizing suffix, possibly also indicating indefiniteness, as in Persian.

The word **aiu̯a*- has many cognates among Iranian languages, at least one in every language, and is the basic word for 'one' in Proto-Iranian. It became a suffix for instance, in Middle and New Persian (Josephson 2011: 25). This suffix possibly initially indicated individuality, but in attested texts, it is an indefinite suffix, thus not entirely similar in function to the one in Tocharian. There also exists a Persian individualizing plural suffix, namely Contemporary Persian: *-ihå* (< Cl. Pers. *-ē-hā*), which has the same function as the Tocharian suffrix *-aiwenta*, and which is also built on *-*aiu̯a*- (> *-ē* > *-i*) + plural suffix *-hā*, e.g. *åbi*(*-*)*hå* 'individual bodies of water' (individual waters + plural). Added after a plural, the individualizing suffix is used to mean 'some X': *åbhå-yi* 'some waters', *še'rhå-yi* 'some poems', *gušthå-yi* 'some pieces of meat' etc.

However, both the individualizing and the indefinite function can derive from a primary meaning 'one' or, secondarily, 'single'. It is, however, remarkable that Proto-Iranian **aiu̯a*- seems to have become a suffix in Old Steppe Iranian too. Nonetheless, I do not think this grammaticalization process can be used as an argument for the classification of Old Steppe Iranian among Iranian languages, as it is possible that it occurred independently in various Iranian languages.

In conclusion, the Tocharian individualizing suffix *-*aiwenta* which could be rendered as meaning 'individual pieces of', composed of *-*aiwe* and of the plural suffix *-nta*, has been viewed as deriving from Proto-Indo-European **Hoiu̯o*- 'single; alone'. However, this Indo-European word is rare, being only found in two branches: Indo-Iranian and Greek, and being otherwise completely absent from Tocharian. Because of this, I suggest that this suffix originates from Old Steppe Iranian *-*aiu̯a*-, with a probably similar if not identical meaning.

2.2.6 *TB* kertte *'Sword'*

The Tocharian B word *kertte* 'sword' was recognized as being of Iranian origin for the first time by Van Windekens (1963: 486). The Iranian etymon is clearly **karta*-, cf. Avestan *karəta*- 'knife', but there has been discussion about the exact source language of the borrowing.[9]

9 Some scholars have considered that *kertte* could be inherited from Indo-European (Isebaert 1980: 89), but this is improbable. Although an Indo-European form **kor-to*- would indeed yield TB *kert*(*t*)*e*, no other Indo-European branch, not even Indo-Aryan, reflect such a form. As Professor Lubotsky informed me (p.c.): the *nomen agentis* **kortó*- 'cutter (?)' is never found outside of Iranian. In his forthcoming etymological dictionary of Proto-Indo-Iranian, Profes-

Winter (1971: 218) has claimed that the source was Bactrian, but Schwartz (1974: 409) has shown that this is not possible. Bactrian does not preserve (short) word-final vowels of Old Iranian in polysyllabic words, which proves problematic to explain the final *-e* of *kertte*. One could theoretically oppose that Pre-Bactrian loanwords may display a final vowel in Tocharian, as in the case of *mālo* and *koṣko* (see Del Tomba 2023: 130; Bernard & Chen 2022). However, a certain and definitive argument against a Bactrian origin of this word is that Tocharian loanwords from Bactrian do not show *e* for Bactrian *a* internally, so that the vocalism of *kertte* certainly cannot be explained from Bactrian. Last but not least, the Bactrian word for 'sword' is χαγγαρο (Sims-Williams 2007: 276),[10] and the meaning 'sword' of the Tocharian word is well assured from the attestations, while the bare Old Iranian noun **karta-* generally means 'knife' in most Iranian languages.[11] So, to sum things up, neither the phonetics, nor the documentation of Bactrian, nor the semantics of *kertte* support Winter's hypothesis of a Bactrian borrowing.

Once Bactrian has been eliminated, it is clear that this word must be from another Iranian dialect. No other Middle Iranian language seems to fit with the form (cf. Sogdian *ākartē* 'sword', Khotanese *kāḍara-* 'id.') and the vocalism *e_e* indicates a likely Old Steppe Iranian origin. An Old Steppe Iranian **karta-* would regularly have yielded Proto-Tocharian **kerte*. The Tocharian A outcome should have been †*kart*, but apparently this lexeme was replaced in Tocharian A by *kāre* (see below).

One needs to explain the discrepancy in meaning between Proto-Iranian **karta-* 'knife' and Tocharian B *kertte* 'sword'. There are two solutions. One would be to suggest that the meaning derives from a suffixed form **karta-tara-* 'knife-like, sword' at the source of, for example, Khotanese *kāḍara-* 'sword'. This **-tara-* suffix would then perhaps have either been removed in Old Steppe Iranian, while the meaning remained. It is very speculative.

The other solution, which is much more sound, consists in looking at the Ossetic evidence: Ossetic has a word *kard* (‹a› in Ossetic stands for a long

---

sor Lubotsky reconstructs **krti-* 'knife' for Indo-Iranian, and considers **kartá-* as a properly Iranian innovation.

10 Compare Christian Sogdian *xnγr* (var. *xγr*) 'sword' (Sims-Williams 2020: 216), New Persian *xanǰar* 'dagger'.

11 For example in THT 79 a2 (*sto*)*rmeṃ kertteṃ oṅkor malkānte śle-yärke lāntaś weskeṃ* "(stand)ing they put [their] swords [in the] sheaths [and] speak with deference to the king" (my translation, on the basis of CEToM); THT 404 b8 *wes rano ñake kerteṃ yamamtär ścirona ṣñārä* "We will now also make the swords of each of us sharp." (Peyrot 2013: 658).

vowel, *ā*) from **karta-* with vowel lengthening. This word means 'knife; sabre; [epic] sword' (cf. Cheung 2002: 196), and it is the meaning one can attribute to this word in Alanic in general: a cognate of the Ossetic form was also borrowed in Hungarian *kard* 'sword' (cf. Holopainen 2019: 124). It thus appears that Ossetic, and Alanic in the wider sense, is the only Iranian language displaying the meaning 'sword' for the unsuffixed word **karta-*. Given the quantity of similarities between Old Steppe Iranian and Ossetic, and the possibility, in my view, to link them to a common "Scytho-Steppic" ancestor, it is not unnatural to assume a meaning 'knife; sword/sabre' (in any case a cutting war weapon) for the proto-form of Alanic *kard* and Old Steppe Iranian **karta-*. Interestingly, this would show that Ossetic **ard* > **ārd* is posterior to the separation of both languages. Further, the Old Steppe Iranian could also have meant 'knife', but the Tocharians, who were probably either under their rule or in any case in a type of martial relationship towards them, would have borrowed the meaning 'sword' exclusively.

In conclusion, Tocharian B *kertte* 'sword' derives straightforwardly from Old Steppe Iranian **karta-* 'sword'.

2.2.6.1 A note on TA *kāre* 'Sword'

No etymogical cognate of Tocharian B *kertte* is so far attested in Tocharian A. It would have been, to all probability, †*kart*. Instead, the Tocharian A word for 'sword' is *kāre* (pl. *kāreñ*) 'sword' (cf. DToA: 109). No convincing etymology had been proposed for this word, to my knowledge. I suggest that it derives from Khotanese, where the word for 'sword' is *kāḍara-* (DKS: 58). This, in turn, ultimately derives from PIr. **karta-* too. This etymology of mine, cited *apud* Dragoni's dissertation (cf. now Dragoni 2023: 82–83), is viewed as "probable" by the DToA (DToA: 109).

For **-ḍ-* → *-r-*, a parallel development can be found in TAB *kor* (besides TB *koṭ* and TAB *koṭiśvar* 'millionaire') 'ten million' from Middle Indic *koṭ* 'idem', among other examples. One can also cite Manichean Tocharian B *mukur* 'diadem', ultimately from Sanskrit *mukuṭa-* through Middle Indic (cf. Pinault 2008b: 115). There are also other forms, such as Tocharian A *prop-mahur* which designates a diadem, Tocharian B *prāp-mahur*, and also TB *mahūr*, var. *maur* 'diadem, crown' (cf. Pinault 2011: 157). The second element of these words goes back to Buddhist Sanskrit *makuṭa-* 'diadem, crown' (cf. Pinault 2011: 158).

Nonetheless, concerning this rendering of Sanskrit *-ḍ-* as *-r-*, a *caveat* is in order. The intervocalic *-ḍ-* of Sanskrit was lenited to *-ḭ-* /ɽ/ already in Gāndhārī, cf. Schoubben (2024: 21), which was likely a flap [ɽ] as in New Indo-Aryan languages. This very /ɽ/ would have naturally been rendered in Toch. as *-r-* (and in Sogdian too) and not the retroflex stop articulation of Sanskrit. Thus *kor* would

be a Gāndhārī loanword and *koṭ* a Sanskrit one.[12] However, to this I would object that we do not know with certainty how the *-ḍ-* of Khotanese was pronounced, nor do we know for certainty how it would have been perceived by Tocharian speakers. There is a phonetic continuity between the [ɽ] flap and the plosive retroflex [ɖ] so that, if the Khotanese phoneme was also a plosive retroflex, it is unclear to which degree a phenomenon of friction would have been auditorily perceived and then reproduced as such by Tocharian speakers, according to their own phonological system. Until more examples are found, this remains an open question.

Two different pathways are available to explain the Tocharian A word on the basis of the Khotanese. An enlarged form **kāḍaraka-* would have undergone a weakening of the medial syllable: **kartaraka-* > **kāḍaraa-* > **kāḍäraa-*, which would have yielded **kāḍärei*/**kāḍärai* in Old Khotanese.[13] This could have been borrowed as **käräre* in Tocharian A, then becoming *kāre* through syncope.

Another solution is possible. Khotanese *kāḍara-* 'sword' has generally been explained as deriving from Proto-Iranian **karta-* 'knife, dagger' with an unexplained suffix *-ra-* (e.g. DKS: 58). The suffix *-ra-* is found in verbal adjectives in Khotanese (cf. Degener 1989: XXXIV), and it would also not make much sense at the Proto-Iranian stage in this specific word. It seems more reasonable to consider that *kāḍara-* should derive from **karta-*, suffixed with *-tara-* 'sort of', 'a sort of dagger/cutting weapon' > 'a sword', and haplology (**karta-tara-* > **kartara-*). Tocharian A *kāre* could theoretically also derive from the unsuffixed (and unattested) form **kāḍa-* < **karta-*, with the addition of a *-ka*-suffix: **kartaka-* > OKh. **kāḍaa-*, with OKh. nom.sg. **kāḍei*/**kāḍai* → (regularly) TA *kāre*. Unfortunately, these forms are completely unattested in Khotanese, and the first hypothesis is thus somewhat more likely, but this second hypothesis is formally more straightforward.

In conclusion, I propose that Tocharian A *kāre*, pl. *kāreñ* is likely borrowed from Khotanese, it ultimately either goes back to Khotanese *kāḍara* 'sword', suffixed with *-ka-*, or to an unattested Khotanese **kāḍa-*, also suffixed with *-ka-*. The weakening of the ending *-aka-* to **ei* or **ai* already in Old Khotanese led to the final *-e* of the Tocharian form, while the Khotanese *-ḍ-* was rendered as Tocharian A *-r-* as in Middle Indic loanwords.

12 I thank Dr. Niels Schoubben for pointing out this to me.

13 I thank Dr. Federico Dragoni for confirming these facts to me.

2.2.7 *TB* keṣe, *TA* kaṣ *'Arms' Span (Measurement Unit)'*

Tocharian B *keṣe* and Tocharian A *kaṣ* 'arms' span (measurement unit)' or 'fathom', corresponding to Sanskrit *vyāma-* 'id.' (cf. DToA: 100), derive from Proto-Tocharian **keṣe*. Proto-Tocharian **keṣe* in turn was borrowed from Old Iranian **kaša-*, compare Av. *kaša-* 'armpit' (see already Isebaert 1980: 84–85; DTB: 213). The word must have been borrowed from Old Steppe Iranian in view of the vowel correspondence OIr. **a*: PT **e*.

Hilmarsson (1996: 137) found it "advisable not to seek a non-Indo-European origin" for this word: he viewed a connection of this word with Vedic *hásta-* 'hand' as uncertain and favored a connection with TB *keś* TA *kaś* 'sequence, number' (1996: 138). His semantic derivation departs from a meaning 'pertaining to sequences or calculations' > 'measure' > 'fathom'. TB *keś* TA *kaś* has recently been derived from Proto-Tumshuqese-Khotanese (cf. Dragoni 2023: 100–103), which would formally invalidate Hilmarsson's etymology, which is rather weak from the semantic point of view. It is better to assume a straightforward etymology, as did Isebaert (op. cit.): Old Steppe Iranian **kaša-* borrowed as Proto-Tocharian **keṣe*.

This word **kaša-* could already have designated a measurement unit, as semantic convergence of body part-terms and measurement units is extremely common cross-linguistically. The semantic change from 'armpit' *vel sim.* to 'arms' length' could have occurred in Old Steppe Iranian itself, as a measurement unit is much more easily borrowable than a body-part term. However, given how essential words (such as TB *epe* 'or', s.v.) were borrowed from Old Steppe Iranian into Proto-Tocharian, it is not impossible either that the semantic shift occurred within Proto-Tocharian itself.

In conclusion, Tocharian B *keṣe* and Tocharian A *kaṣ* derive from Proto-Tocharian **keṣe* 'arms' length', itself a borrowing from Old Steppe Iranian **kaša-* 'arm' probably with a meaning 'arms' length' (the length of two arms stretched together, Dutch *vadem*, English *fathom*) already present in Old Steppe Iranian.

2.2.8 *TB* kraupe, *TA* krop *'Group, Crowd'*

The Tocharian B word *kraupe* 'group, crowd' (Archaic TB *krewpe*) and its Tocharian A equivalent *krop* 'crowd, heap, herd, flock' (DToA: 162), adv. *okrop* 'in a crowd' are usually viewed as deriving from the verb TB *krawp-*, TA *krāwp-* 'gather, amass' (so Adams, DTB: 238) < PT **krawp-* (cf. Peyrot 2013: 741). However, it appears that this verb does not have a convincing etymology.

The etymology of this verb as given in Adams is unsatisfactory: he claims cognacy of this word with Greek κρύπτω 'hide', but there is no evident semantic link between "hiding" and "gathering". Obviously, one can pile things up onto something in order to hide that thing, but the semantic stretch is too far in my view.

According to Adams (also Hilmarsson 1996: 179), *kraupe* is further cognate to Lithuanian *kráuju* 'pile up', *krūvà* 'pile', Old Church Slavonic *kryjǫ* 'cover, hide' *krovъ* 'roof'. However, it is doubtful whether any of these words can be connected to Tocharian B *kraupe* and Tocharian A *krop*, simply because we cannot account for the *-p-* in Tocharian. In some Buddhist texts, *kraupe* can also mean '*saṅgha*, Buddhist community' (cf. Pinault 1995: 20); otherwise it translates Sanskrit *gaṇa-* and *varga-* 'troop, group' (cf. Pinault 1995: 20–21). The semantic connection between *kraupe* to *kráuju* is all but evident, and formally these forms are much too far apart. It can be concluded that the cognates in the set proposed by Adams are most probably not related.

Tocharian B *krewpe*, *kraupe*, Tocharian A *krop* 'group, crowd' is more easily derived, both for the meaning and for the form, from Old Iranian **graṷpa-* (or **graṷfa-*). Neither this etymology nor the Old Iranian word have hitherto been recognized. The basis for the reconstruction of this Old Iranian word is found in a few Iranian languages, namely, Middle Persian, New Persian, Balochi and Dawāni.

The facts are these: Middle and New Persian *grōh* 'group, company', borrowed into Armenian *groh* 'Nation, Volk, Truppe' (Hübschmann 1897: 132); Balochi *grōp* 'assembly, group'; Middle Persian **grōhag* seen in the Syriac borrowing ‹grwhqʾ› 'small round cakes' (cf. Ciancaglini 2008: 146), and New Persian *gurōha* 'globe, bowl, or any spherical figure; ball for a cannon; a kind of sweetmeat; an assembly, multitude of people' (cf. Steingass 1892: 1085); Dawāni *gorufa* 'ball of fiber or rope'.

Multiple etymologies were provided for these words, none of which is satisfactory.[14] Balochi *grōp* can go back to either **graṷpa-* or **graṷfa-* (cf. Korn 2005), but the forms with *-h* can only reflect an earlier **-f*, and it is difficult to accept that the Dawāni form *gorufa* shows a sound change **-p-* > *-f-*, although more information about this language would be welcome.

There is a difficulty with the form with *-h-* in Middle Persian: *-f* only becomes *-h* in New Persian, for example: Middle Persian *kōf* 'mountain' > New

14 See Hasandoust (2014: 2385 f.) for an enumeration of them. One can cite Horn's very cautious proposal of connecting this word to Germ. *Leute* etc., through perhaps a hypothetical **u̯i-raṷda-* (Horn 1898–1901: 186) which is impossible because of the *g-* in the Middle Persian form (*u̯i-* only becomes *gu-* in New Persian, and should have remained as *wi-* in Middle Persian). Furthermore **u̯i-raṷda-* is not fitting semantically: in Sanskrit *virodha-* means 'opposition, enmity, struggle'. Another attempt, by Nyberg (1931: 84), consisted in setting up a proto-form **gravaϑva-* which, besides being completely *ad hoc*, does not explain the Balochi and Dawāni forms. Monchi-Zadeh (1990: 79) proposes a proto-form in *-ϑ-* (namely **graṷϑa-*), which he derives from the Indo-European root **gel-* 'to group, to amass' (the root at the basis of Eng. *cloud*). This proto-form **graṷϑa-* is obviously impossible, for

Persian *kōh* 'idem'. Here though the *-h* is more ancient, as shown by the Syriac borrowing ‹grwhq'› (see above). We thus have no other choice but to assume borrowing from a variety that underwent the sound change *-f* > *-h* earlier than "Standard Middle Persian".

No matter whether it was **graṷpa-* or **graṷfa-*, my hypothesis is that an Old Iranian antecedent of the forms cited above was borrowed in Proto-Tocharian as **krewpe*, hence Tocharian B *kraupe* (archaic Tocharian B *krewpe*) and Tocharian A *krop*. A weak point in my hypothesis is the limited attestation of the word in Iranian in combination with the lack of a cognate beyond Iranian, i.e. in Indo-Aryan or in Indo-European. We should thus assume that the word entered Iranian at some point, likely by means of borrowing, but we cannot be sure about the dating. Since the semantic and phonological correspondences with Tocharian are exact, with Old Iranian **aṷ* being reflected as Proto-Tocharian **ew* as in TB *newiya* ← Old Steppe Iranian **naṷiiā-* (and cf. further OSIr. **aị* → PT **ey*), I assume that the word has a longer history within Iranian and was borrowed into Tocharian from Old Steppe Iranian.[15]

The fact that the attestation in Iranian is limited may be explained by the circumstance that the lexicon of Middle Persian is so much better known than that of other varieties. The same applies to OSIr. **rataka-* 'army line', which is also only known from Middle Persian.

In Tocharian, a verb was derived from the noun **krewpe*, as for example, PT **rəytwa-* was derived from PT **reytwe* (s.v. *raitwe*). The reconstruction of the verbs in TB and TA (cf. Peyrot 2008: 150; Peyrot 2013: 741), with two different presents, fits very well with the verbs being derived from the Proto-Tocharian noun **krewpe*. For instance, Tocharian B *kraupe* would not easily be derived from a stem *krawpa-*, etc., especially in view of the fact that the older vocalism of the noun is *-ew-*. This *-ew-* may have been subject to *a*-umlaut in stems like *krawpa-*, but the reverse development is not possible.

While the verb *krawpa-* displays an intricate set of patterns, including an athematic present I (TA *kropäntär*) patterning with a preterite I / subjunctive V (cf. Peyrot 2013: 741), which might instill doubt as to its denominal status, I shall add that the verb PT **spaltka-* (TB *spalk-*, TA *spāltkā-*) 'make an effort' which is doubtlessly denominal (§ 2.2.26), also has irregular patterns. The complex

---

the reasons evoked in this discussion. Even if one were to accept the implausible derivation from Indo-European **gel-*, Proto-Iranian **graṷpa-* or **graṷfa-* would still require an opaque suffixation in **-pa-* or **-fa-*.

15 Since there are no cognates of this word outside of Iranian, we should obviously reject the option that the Tocharian and Iranian words are both inherited from Proto-Indo-European, instead of being related by borrowing.

appearance of a Tocharian verb is not in and of itself a sufficient argument to assert that it is inherited or not denominal.

In conclusion, the Tocharian B *kraupe*, Tocharian A *krop* words meaning 'group, crowd' are straightforward borrowings from Old Steppe Iranian **grau̯pa-* or **grau̯fa-* (more likely **grau̯fa-*) which had the same meaning. The verb TB *krawp-*, TA *krāwp-* is derived, within Tocharian, from the Proto-Tocharian noun **krewpe*.

2.2.9 *TB* newiya *'Canal or Channel'*

TB *newiya* 'canal' is attested only twice, in the passage I cite below. It is a contract for the sale of a land (Pinault 1998: 364), and this contract discusses the boundaries of that land.

Tokyo National Museum, Shiryokan Archives n. 174 4–5[16]

> *omotruññaiṣṣe yateññe ckeṣṣe ārte sim kom kläskomeṃ orotsa newiya sim oṣṣalemeṃ armokiññe ckeṣṣe ārte sim orotsai newiyai täṅtsi*
> "of the south, the boundary is the arm of the river Yāte; on the west the boundary is the great canal; on the north, the boundary is the arm of the river Armoki up to the great canal inclusively." (Pinault 1998: 364–365)

The meaning 'canal' is probably based on the likely etymology of this word. Adams (DTB: 364) derives it from Old (Steppe) Iranian **nau̯i̯ā-*, which is shortened from PIr. **nāu̯i̯a/ā-* according to the sound law discussed section 2.7.2.9 (p. 173 f.). The original meaning of the Iranian formation is disputed, see for instance a review of the literature in Utas 1965–1966; more recently see Filippone 2017, with a review of the more recent literature as well as a thorough analysis of relevant passages—she concludes that the precise semantics of the Avestan cognate *nāuuiia-* have not yet been securely determined.

A word of caution, on the semantics of TB *newiya*: the Tocharian A word *new** (n., masc.) 'flow, stream' has a better known meaning, and is known to correspond to Sanskrit *ogha-* 'stream, flood' (cf. DToA: 255 with references to previous literature). It seems far from unlikely that both words should be related, and so the meaning of *newiya* could theoretically be closer to the Tocharian meaning. However, *new** is not a regular correspondent of *newiya*, so the meanings do not have to correspond neither. If TB *newiya* had been borrowed into Tocharian A, one would have expected a form ***newi*, and if it were inherited from Proto-Tocharian, one would have expected a form ***nawi*. Since it is not

16 Alternatively known as Otani 19.1.

clear to me what the etymology of TA *new** is, and whether it is inherited or borrowed, I prefer to leave it out of the present discussion.[17]

As per Filippone 2017, one can see that there are two possible ways to understand Proto-Iranian **nāu̯i̯a/ā-*. It can be analyzed either as meaning 'navigable (canal)' or 'channel'. Either meaning could be ascribed to TB *newiya*. One can hope that further attestations of this word in Tocharian might help us determine the meaning of the Old Iranian cognates with more exactness. It seems to me, in any case, quite straightforward to see TB *newiya* as the Old Steppe Iranian feminine gender substantivization of an adjective meaning *'relative to *nau̯i̯a-* rivers', the word for 'river' being a feminine noun, thus 'navigable (river)' (see for instance Utas 1965–1966: 127 f.) or 'channel' (cf. Filippone 2017 for references to previous literature). This 'navigable (river)' or this '(navigable?) channel', made into a noun, could have designated canals, as they are (artificial) navigable rivers, made precisely for the purpose of navigation. Nonetheless, one should not ignore the possibility that the canals mentioned in the Tocharian text are irrigation canals, not made for navigation. As I wrote above, the meaning of this word and related words in Old Iranian is extremely difficult to ascertain, however, it is beyond doubt that they designate running waters, and it is also beyond doubt that the first syllable contains—originally in any case—a long **ā*.

Although the derivation of TB *newiya* from Old Steppe Iranian seems straightforward, an alternative interpretation has been proposed by Widmer (2007). He also sees TB *newiya* as a borrowing from Iranian, but his explanation differs in crucial points. According to him, the meaning of Old Persian *nāviya-* is not connected to the notion of 'boat' but to that of 'flowing'. He supports his view with the Akkadian translation of the Old Persian passage, which has "the Tigris river was in flood". He also finds the meaning '(stark) fliessend, strömend' fitting for the Avestan cognates. Widmer (2007: 224) also mentions the Avestan word *nāuuiia-*, for which he proposes the possible meanings 'Fluss, Bach'.[18]

Widmer argues that the source of the Tocharian borrowing was a Middle Iranian form **neu̯(i)i̯a*. In my opinion, this is impossible, since it presupposes an *ad hoc* metathesis of the original **nāu̯i̯a-* to **nāi̯u̯a-*, followed by a restoration of a suffix *-(i)i̯a*, which was no longer available in Middle Iranian, as far as I know. Also, I do not know any Middle Iranian language in the region where **-āi̯-* would become **ē* or **e*, as usually OIr. **-āi̯-* remains *āy*.

17 Theoretically, TA *new** could come from PIE **noi̯u̯o-*.

18 Widmer (2007) surprisingly does not cite Utas' seminal work of 1965–1966; in which work the translations of the Old Persian passage are examined, and where the meaning of *nāviya-* is discussed.

Here, the data offered by Tocharian clearly contradicts Widmer's hypothesis, and the traditional meaning 'navigable' usually found in scientific literature is to be upheld. Furthermore, Akkadian and other versions of the text are not literal translations of the Old Persian, but simply different versions written in different languages.

The word *newiya* is of special interest since it seems to show that shortening of PIr. **āu̯* to **au̯* occurred in Old Steppe Iranian (see 2.7.2.9, p. 173 f.). Moreover, it gives a precious insight into the culture of the people speaking Old Steppe Iranian. Apparently, these people knew what canals were. If these are not simple irrigation canals, it means that Tocharians knew of navigation. Did they practice it in the Balkhash Lake, which is far south, but one of the closest lakes nowadays in the region, or in another, smaller body of water? Or, perhaps, was there much more water in the region?[19] Naturally, the Tocharian texts reflect a different sort of civilization from the original Old Steppe Iranian one, and the original meaning of *newiya* might have been different from the English definition of 'canal'—if we do not take it to have meant 'channel', as many scholars did for **nau̯i̯a-*.

In conclusion, Tocharian B *newiya* 'canal or channel' is borrowed directly from Old Steppe Iranian **nau̯i̯ā-*, a substantivized feminine adjective of a feminine noun 'river', meaning 'thing relative to navigation; canal', with shortening in the first syllable from PIr. **nāu̯i̯ā-*.

2.2.10 *TB* pāke, *TA* pāk *'Part, Portion, Share'*

The Tocharian B noun *pāke* 'part, portion, share' (derived noun *pakeññe* 'member, partner'), plural *pakenta*, Tocharian A *pāk* 'idem', reconstructible as Proto-Tocharian **pake* 'id.', has been considered as either of Iranian origin or inherited (DTB: 389). In both Tocharian A and B there is a set phrase *pāke yam-* 'to participate in' (cf. Meunier 2013: 149 f.).

In order to determine whether it is more likely that TB *pāke* and TA *pāk* are inherited or borrowed, I will first present the generally admitted Indo-European etymology of these two words, and then discuss their Iranian etymology.

I will present below the two main Indo-European reconstructions which have been proposed in order to derive Tocharian B *pāke*, Tocharian A *pāk* from Proto-Indo-European. The first one is exemplified in Pinault (2008), and the second one in Adams (DTB).

19 The Elamite phrase *kurtaš numakaš* could have designated irrigation workers (very hypothetically, see Henkelman 2018: 233 with references to earlier literature). Possibly, if this were true, those equally hypothetical Tocharian **kercïye* 'house servant' (s.v. *kercci*) could

Pinault (2008: 30) considers both B *pāke* and A *pāk* 'part, portion, share' as straightforwardly derived from Proto-Indo-European **bʰag-os* "de la racine **bʰag-* «(se) partager» (LIV², p. 65)". This is naturally only possible if one accepts the existence of a vowel **a* in Proto-Indo-European. It is to be noted that the reconstruction with **a* in the LIV²: 65 is based on the Tocharian forms. The reasoning here could be seen as circular: one reconstructs a proto-form in order to explain two Tocharian words, which are themselves only explained through that reconstruction.

The second explanation can be found in Adams (DTB: 389), who suggests among others to derive it from **bʰeh₂góm* (or **bʰeh₂go-*), while noting a problem with the *-nta* plural. Indeed, as Adams rightly notes, one would expect a plural ***pāki* if the word was inherited from Proto-Indo-European **bʰeh₂go-*, and not *pakenta*. However, there is yet another problem with this reconstruction: **bʰeh₂go-* or **bʰeh₂góm* would yield TB †*poke* and TA †*pak*. Adams (DTB: 389) does not completely exclude a borrowing from Iranian.

The verbal root **bʰe(h₂)g-* 'to divide (a portion)' is supposedly attested in Indo-Iranian (cf. LIV²: 65; EDIV: 2; Beekes 2010: 1543), in the Greek verb φαγεῖν 'to eat, consume, swallow', and in Armenian *bekanem* 'to break' (see below). If inherited and cognate to these forms, Tocharian B *pāke*, Tocharian A *pāk* 'part, portion, share', which go back to Proto-Tocharian **pake* 'id.', can only come from PIE **bh₂go-*.

For the reconstruction with a laryngeal **h₂* for this word, see Lubotsky (1981: 134 f.; see also EDIV: 2). A reconstruction **bh₂go-* would explain one Old Persian form, *bigna*, which could contain an *-i-* < **h₂*. In the Iranian etymon **baga-* 'god, lord' < **bʰeh₂go-* the loss of the laryngeal would have resulted from generalization from contexts where it would be deleted by Lubotsky's law. Nonetheless, this word *bigna* is only found in one personal name, *Bagābigna* < **baga-* 'god, lord' + *ā-bigna-* (cf. Kent 1953: 199). This word has been explained otherwise variously, notably as 'having the attacking power of God' (cf. Tavernier 2007: 14, with references to other scholars). In my view, this example, of unclear meaning, is not enough to warrant the reconstruction of this root with a laryngeal.

If one connects, as is often done, this root **bʰeg-* 'to divide (a portion)' with **bʰeg-* 'to break' (LIV²: 66), then the reconstruction **bʰeg-* without a laryngeal is warranted. In that case the Tocharian form—if inherited—could only derive from **bʰōgo-*. This would more easily account for the Indo-Iranian forms.

---

have been building canals for the Old Steppe Iranians. This, however, implies a type of society and empire that we can only hypothesize for the Old Steppe Iranians.

Only Greek φαγεῖν requires the reconstruction with *h_2 (cf. Lubotsky 1981: 134), but it is semantically distant from 'to divide (a portion)'. Even if we were to connect it to the other words, it could derive from *b^hng-, as suggested by Professor Lubotsky (p.c.). On the other hand, Armenian *bekanem* 'break', *bekor* 'fragment, piece', which fits the semantics of the Indo-Iranian much better than that of the Greek, does seem to be cognate of what I believe we should reconstruct as *b^heg- 'to divide' (cf. Olsen 1999: 524).

If we derive Proto-Tocharian **pake* from Old Steppe Iranian **bāga*- 'portion, share' by means of borrowing, the problems mentioned above are solved: no laryngeal *h_2 or PIE **a* is needed. Many Iranian cognates have exactly the right meaning, and the assumption of borrowing would also neatly account for the *-nta* plural *pakenta*, as it is a very common plural often used for loanwords, including Old Steppe Iranian ones (TB *waipeccenta* 'possessions'). Old Steppe Iranian **bāga*- 'portion, share' in its turn is from the Iranian root **bag-* / **baǰ*- 'to bestow, divide, have a share' (EDIV: 1–2). The expected correspondences between Old Steppe Iranian and Proto-Tocharian pose no problem here, with OSIr. **ā* corresponding to Proto-Tocharian **a* and OSIr. **a* → PT **e*. As can be seen from Cheung (EDIV: 1–2) almost all Iranian languages, Ancient, Middle and Modern share very similar meanings to the Tocharian words. Suffice it to cite Bactrian βαγο 'share, part, portion, lot, fate' (Sims-Williams 2007: 200); Sogdian *βʾγ* 'allotment, garden' (SD: 97), Christian Sogdian *bʾγ* 'garden, vineyard' (Sims-Williams 2020: 50), Wanjī *weǰ* 'part, portion'.

In conclusion, Proto-Tocharian **pake* 'part, portion, share', yielding Tocharian B *pāke* and Tocharian A *pāk* 'id.', was borrowed from Old Steppe Iranian **bāga*- 'id.'. The phonological correspondences observed are regular for Old Steppe Iranian, while attempts to derive **pake* from Proto-Indo-European must cope with insuperable phonological problems.

2.2.11 *TB* peret, *TA* porat *'Axe'*

The Tocharian B word *peret*, corresponding to TA *porat*, both 'axe', is among the first Tocharian words correctly identified as being of Iranian origin (Lidén 1916: 16 f.; see also Hansen 1940: 154 f.). The vowel of the first syllable of the Tocharian A word shows *p/m*-umlaut of the Pre-TA **a* to **o*, a phenomenon which also affects inherited words (compare for example TA *cmol* 'birth' < Pre-TA **cmal*, TA *ñom* 'name' < Pre-TA **ñam*). Thus, the predecessor of the Tocharian A and B word can be reconstructed as Proto-Tocharian **peret* or **peretu*. The rendering of Iranian **a* with PT **e* in Tocharian shows that the Tocharian word was borrowed from Old Steppe Iranian. However, the Iranian data, but also more generally the Indo-Iranian data is somewhat complex. I will discuss below how an Old Steppe Iranian form **paratu* can be justified.

2.2.11.1 Why Proto-Iranian **paraću-* Cannot Be Reconstructed

The Iranian antecedent of the Tocharian form **peret* has often been compared to Vedic *paraśú-* 'hatchet, axe', cf. also Ashkun *pōs*, Kati *pɛċ* 'large axe', and to Greek πέλεκυς 'axe, double axe, hatchet' (e.g. DKS: 203; Beekes 2010: 1166 f.; Carling 2020: 135). The usually reconstructed Proto-Indo-Iranian form is **paraću-*, and, together with the Greek word, a Proto-Indo-European form **peleḱu-* has been posited. As I shall demonstrate, this reconstruction is illusory, and only rests on a *tour de magie*: the invention of a convenient Old Persian form that seemingly solves everything.

First, let us consider the Iranian evidence for apparent cognates of Vedic *paraśú-*. There are three culprits: Old Khotanese *paḍa-* (acc. sg. *paḍu*) 'axe' < **paratu-*, with transfer to *a*-stems and syncope (cf. Canevascini 1993: 269), Khwarezmian *pdyk* '(large) axe' < **partaka-* (Benzing 1983: 520), and Ossetic *færæt* 'axe'. Cheung (2002: 74) considers the Ossetic word etymologically unreliable since "it cannot go back to a clear Iranian proto-form". A clear insight, as usual. His hesitance is probably due to the fact that the Vedic cognate *paraśú-* 'hatchet, axe' does not correspond regularly to the Iranian forms so that an accurate Indo-Iranian reconstruction for this word is difficult.

This is the heart of the matter. Indeed, the comparison of Old Indic *paraśú-* 'hatchet, axe' yields an evident problem, in that we would expect PIr. **paraću-* (corresponding to the Old Indic form) to be reflected as Khot. †*parsa*. Bailey (1982: 55) explains this as dialectal variation, but this is impossible: Khotanese is clearly a language that reflects PIIr. **ć* as *s* (or *śś*, cf. Sims-Williams 1998: 136; Peyrot 2018a: 271–272[6]; Peyrot, Dragoni & Bernard 2022: 417–418), and not as **t* as would be required for the attested form *paḍa-*.

In order to solve the problem of the discrepancy between the Iranian and Indic forms, Mayrhofer (EWAia[2]: 87) and others suggest a borrowing from Old Persian **paraϑu-* into the rest of Iranian. This solution has been adopted by the consensus, despite some silent disagreements such as Cheung's. Although this could theoretically work for Ossetic, as it could indeed derive from **paraϑu-* as well as from **paratu-* (cf. Cheung 2002: 20 f.), it is an unlikely solution to this issue.

This Old Persian borrowing hypothesis needs to be rejected for multiple reasons, the first being that this Old Persian **paraϑu-* is very unlikely to have existed. Indeed, cognate forms are absent from the entire Western Iranian territory (a hundred languages or more), so that Eastern Iranian languages would have borrowed a Western word that is unknown and for which there are Western Iranian equivalents, for a concept they also have. This in itself makes such a scenario unlikely. I should add that, at the time Old Steppe Iranian existed as such, Old Persian was most probably a small Iranian language, not yet that of an empire.

The second reason is phonetic: even if the word **paraϑu-* had ever existed in Old Persian, there was a phoneme *ϑ* in Khotanese and Khwarezmian. This phoneme only later yielded *h* in Khotanese, and never even completely ceased to exist in Khwarezmian. Therefore, the supposed Persic **ϑ* is unlikely to have been rendered as **t* in Pre-Khotanese and in Khwarezmian.[20]

The third reason is that a word for 'axe' is attested throughout the Persic languages: Old Persian **tapara-* 'axe', indirectly found in Elamite *dabara* (cf. Tavernier 2007: 322) and reflected in New Persian *tabar* 'axe'. It thus seems that not only assuming a borrowing from Old Persian **paraϑu-* is unlikely for several reasons, but even reconstructing such a preform is unwarranted.[21] The Sogdian word for 'axe' is *tš-*, cf. Avestan *taša-*. Sims-Williams (2002: 239), noting the rather unexpected final **-t*, suggested that these words, or in any case, the Tocharian form TB *peret* TA *parat*, were borrowed from one undetermined Scythian language, which is of course theoretically possible.[22]

In conclusion, the scientific method forces us to reject the possibility of an Old Persian word, unattested in all of Iran and the Western Iranian world, in the face of an actual Old Persian word with the same meaning also reflected in later languages, only attested in three documented and one reconstructed Iranian languages, two of which were probably spoken in Southern Siberia at the time, at a time when the Achaemenid Empire did not even exist.

To set records straight: there was a direct cultural line of artefact borrowings, which could also imply the existence of loanwords: it stretched from Persia to the Scythians through 1. Bactria → 2. Sogdiana and/or Khwarezm (Francfort 2023: 60). Here, the word **paratu-* is found in Old Steppe Iranian / Scythian (Alanic) → Khotanese → Khwarezmian, and it stops there. Taking things at face value, it is a word that spread in the East at a relatively later period and its spread stopped around Khwarezm.

Put together, these facts suggest there was no Proto-Iranian **paraću-*. Its limited attestation, relatively late with the exception of Old Steppe Iranian and Scythian, and in languages of various Iranian branches, indicates that this is a later borrowing. By later I mean post Proto-Iranian. In no way can we safely project **paratu-* back to Proto-Iranian, because of its geographic distribution

20 After *r*, Old Iranian **ϑ*, does not yield Khot. *-h-*, but rather *-ṭh-*, even if the cluster **rϑ* arose secondarily due to syncope. Examples are Khot. *paṭhu-* 'to burn' < **pari-ϑaṷaịa-*, or Khot. *baṭha-* 'cuirass' (cf. Avestan *varəϑa-*).

21 Yazghulami *parus* is completely isolated, even among Pamir languages, and thus likely to be a borrowing from an Indo-Aryan language.

22 In passing, the division of the Iranian word between the **tapara-*, the **taša-* and the **paratu-* Iranian languages could be an interesting aspect to look at for future studies on isoglosses, language contact and, perhaps, phylogeny.

and as no trace of it is found beyond that zone where it spread. It is not absurd to propose that the Khwarezmian form is borrowed from a (Pre-)Khotanese form, which would reduce the actual **paratu-* zone even more.[23]

2.2.11.2 Why Proto-Indo-European **peleḱu-* Cannot Be Reconstructed

As Proto-Iranian **paraću-* never existed, and since one cannot safely reconstruct Proto-Iranian **paratu-*, Proto-Indo-Iranian **paraću-* is necessarily a one-legged reconstruction, only based on Indo-Aryan and Nuristani **paraću-*. There is no evidence that the Indo-Aryan and Nuristani forms are not borrowed at some stage, independently or commonly. The only element that would make this form appear inherited is Greek πέλεκυς 'axe, double axe, hatchet'.

First, the Greek accent is in disagreement with that of Vedic *paraśú-*. The syncope in pre-Khotanese (**paratu-* > **parta-*) and pre-Khwarezmian (**paratu-* > **parta-*, with syncope and secondary thematization, as in Khot.) suggests that the accent was on the first syllable, that is PIr. **páratu-*. However, as I argued in § 2.2.11.1, this Iranian word is unrelated and the fact remains that only Vedic *paraśú-* with final accent and its cognates face Greek πέλεκυς.

From the Indo-European point of view, it is unclear what **peleḱu-* would be: is it a stem **peleḱ-* with an *u*-ending? **pele-* with a **ḱu-* ending? Is it a borrowing from an ancient language, that happens to only be reflected in Greek and Indo-Aryan and Nuristani, not even Iranian nor any other European language? None of these solutions are satisfying.

It is much more sensible to consider the Greek word as borrowed from a different source, namely, from a Semitic language, cf. Watson (2013: 170), who derives it from Aramaic/Syriac *plq* m. 'axe' (*p-l-q* 'to split'). Akkadian *pilaqqu* means 'spindle' so it is excluded as a source, although the word from which it is based should be a cognate thereof. In that case, one could suggest that the Greek borrowing occurred in the second millennium BCE from a North West Semitic language, and that the *-u* of the Greek could be a trace of the ancient nominative case ending.[24]

---

23 Very speculatively, I would like to suggest that the Middle Persian name *prdk-y*, Parthian *prdk* (cf. Back 1978: 243), rendered in Greek as Φαρρεκ, could go back to **paratu-ka-* 'small axe' (perhaps originally a mocking name, referring to somebody with "a small axe"). However, if my etymology is correct, it should befrom a Scythian language, because of its initial *f-* (as can be seen in the Greek version), cf. Scythian Φαραδου 'PN'. Interestingly, a name with a similar meaning is also found in Sogdian: *tš'kk* (cf. Lurje 2010: 391). For other, phonologically unconvincing suggestions concerning this name, see Back (1978: 244), with references.

24 I thank Dr. Benjamin Suchard for his help concerning this point.

Thus, if the Greek word is from Semitic, and the "Iranian" word is from some other language, what should the Indo-Aryan/Nuristani word be from? It clearly cannot continue some Indo-European form of which it would be the only descendant. Although I cannot definitely answer this question, I would like to suggest a scenario. In § 3.4.3 I have noted formal (and semantic) differences in the treatment of so-called BMAC loanwords in Tocharian vs. Indo-Iranian. My conclusion is that the BMAC or Oxus Civilization source (whatever we might want to call it) was different in both cases, and I have termed them Western and Eastern BMAC varieties. In the present case, we might have an example of a survival of the so-called Eastern BMAC variety at a later stage, probably about the end of the second millennium BCE and the beginning of the first (a *terminus ante quem*, as it is the date at which I place Old Steppe Iranian–Tocharian contact, cf. § 4.2).[25] Indo-Aryan and Nuristani **paraćú-* would have been borrowed from a more southern or western variety; the form **páratu-* would have been borrowed at a later stage by some Iranian languages, from a more northern or eastern variety. There are sound archaeological arguments for why the 'axe' would have been borrowed from the Oxus Civilization, as it was a very important cultural object in the BMAC area (cf. Francfort 2018a: 129 f. and Parpola 2022: 26).

2.2.11.3 More on Proto-Tocharian **peretu-*

The fact that the **-u* of the Old Steppe Iranian form is not reflected in Tocharian is expected as there was no final *-u* in Proto-Tocharian (and barely any /u/ at all). We also observe this phenomenon with *tsain* 'arrow' ← OSIr. **dᶻainu-*. Since the *-wa* plural (seen in *tsainwa* 'arrows') might reflect the original *u*-stem (cf. Peyrot 2018a: 270), it has been assumed that the form **tˢeynu* and, by extension, **peretu*, existed at some stage of the borrowing process (see also OSIr. **natsu-* → Proto-Tocharian **nətswa-* > TB *mətstsa-* and TA *nätswā-* 'to starve'). In this case, the *-u* was still there in Old Steppe Iranian, and the initial Proto-Tocharian form is **peretu-*, later reduced to **peret-*, probably within Proto-Tocharian itself. There perhaps was a plural **peretwa* at some stage (perhaps even in Tocharian B, although it is not attested).

The syncope that occurred in Khotanese and Khwarezmian (§ 2.2.11.2) had not taken place in the Old Steppe Iranian form from which Proto-Tocharian

25 Dr. Svenja Bonmann (ongoing research) is working on this very topic, and has found traces of a later BMAC substrate influencing the syntax of a series of Central Asian Iranian languages.

**peretu* was borrowed. It is conceivable that a similar syncope occurred in Old Steppe Iranian after the borrowing into Tocharian. In this regard it may be interesting to compare Old Uyghur *baltū* < *baltō* 'axe' (cf. Clauson 1972: 333; Wilkens 2021: 141), which seems to have been borrowed from an Iranian language that had undergone syncope but had not yet lost its final vowel (cf. Clauson 1972: 333, cf. also Wilkens 2021: 141). Khotanese *paḍa-*, for example, has undergone the syncope but has not preserved the word-final *-u* (cf. DKS: 203), so that it does not at first sight seem to qualify as a possible source for the Turkic word. The Turkic word could however have been borrowed from the Khotanese accusative singular *paḍu*, cf. Dragoni (2023: 159), as the Old Uyghur sequence *lt* rendering a retroflex dental occurs in at least one other word (*koltı* ← Skt. *koṭi*, cf. Sims-Williams apud Tremblay 2005: 424[15]). Otherwise, the Turkic cluster *-lt-* could underlie an Old Iranian dialect with **lt* < **rt* (possibly **rt* > **rd* > **lt*). Naturally, both the syncope and the cluster *-lt-* seen in the Turkic word exclude borrowing from Tocharian.

A last point is that, if TB *mot* 'alcoholic beverage' (§ 2.3.8) underwent Old Steppe Iranian *u*-umlaut (< **madu-*), **peretu* did not, which corresponds to the Ossetic distribution, where we also find *u*-umlaut in *myd* < **madu-* but not in *færæt* < **paratu-*.

In conclusion, previous scholarship correctly interpreted Tocharian A *porat* and Tocharian B *peret* 'axe' as borrowings from an Iranian source. Given the regular sound correspondences between the A and B forms, it is likely that this word was borrowed in the Proto-Tocharian (or pre-PT) period from Old Steppe Iranian. This word is, I argue, ultimately from a BMAC variety. Unlike what has been generally assumed in the past, the Iranian words for 'axe' go back to a common proto-form **paratu-* rather than **paraću-* which is not a viable Proto-Iranian form.

2.2.12 *TB* perne, *TA* paräṃ *'Fortune, Good Fortune'*

The words Tocharian B *perne* and Tocharian A *paräṃ*, usually glossed as 'rank, glory' were first proposed to be loanwords from an Iranian language by Sieg, Siegling & Schulze (1931: 18) and Olaf Hansen (1940: 151–152). Both sources propose the borrowing to be from a Middle Iranian language, cf. *phārra-* in Khotanese and *prn* in Sogdian.

However, there is no reason to suggest that Sogdian *prn* /farn/ or Khotanese *phārra-* would be rendered as TB *perne*: /farn/ would yield TB †*pār(ä)ṃ*, and Khotanese *phārra-* would yield Tocharian A †*pār*, Tocharian B †*pār(r)o* (see Dragoni 2023 *passim*). The correspondence between the Tocharian B and Tocharian A forms is regular: the word-final vowel is lost in Tocharian A, and TB *e* corresponds to TA *a*. Thus, the Proto-Tocharian form can be

reconstructed as **perne*, and it derives straightforwardly from Old Steppe Iranian **farna(h)* by means of borrowing.

As noted above, the meaning of TB *perne*, TA *paräṃ* has been considered to be 'rank, dignity' (as in "dignity/rank of Buddha", etc.) in scholarship. Recent research done by Huard (2024) demonstrated this is not exact. The actual meaning of TB *perne*, as discussed by Huard (2024) seems to be 'fortune, happiness', whether the good fortune that consists in becoming a Buddha, or the happiness that comes from the retribution of actions. The latter renders the Buddhist Sanskrit notion of *phala*, which word *phala* is rendered in Ān Shìgāo's Chinese translation of a text close to the *Peṭakopadesa* by *fú* 福 'happiness, felicity' (Huard 2024). To me, in any case, that *perne* does not render the notion of glory seems reasonable, as the concept of 'glory' is accurately rendered by Tocharian B *ñem kälywe*, TA *ñom klyu*, which are inherited phrases. As to 'rank', which is a meaningless concept outside of the army or monastery, it is rather inappropriate to render the status of the Buddha. If the old meaning were to be maintained, *dignity* would be more appropriate, but the arguments raised by Huard are serious enough to consider the meaning of TB *perne*, TA *paräṃ* as meaning '(good) fortune', with some nuances that I will develop below.

Huard (2024) concludes that the Buddhist meaning of **farnah* was already in use among the earliest Iranian Buddhists. Thus, when TA *puttiśparäṃ* and TB *pañäktäññe perne* render the Sanskrit notion of *bodhi-* 'awakeness ~ awakenness' with the meaning 'quality of awakened beings' (cf. Pinault 2016b: 191), one can understand it as meaning 'fortune of Buddha' = 'good fortune leading to the status of an awakened being'. We could also explain TB *käre-perne*, TA *kär-paräṃ*, which are usually glossed as 'dignity', as a hendiadys 'dignity-fortune'. Iranian Buddhist equivalents can be adapted as well: Buddhist Sogdian *'sk'prn* translated as 'high in rank' (SD: 61) could be rendered as 'of high fortune (= birth?)' and Khotanese *phārra-* 'fortune, high position' (DKS: 261) simply as 'fortune', although naturally this needs further in-depth investigation.

In my view, the meaning of PT **perne* was perhaps closer to the original meaning of Old Iranian **farnah* 'glory' or 'kingly prosperity' (although in fact it is a specific concept which is difficult if not impossible to translate). The meaning of **perne* must have shifted towards 'fortune, happiness' with the established Buddhist meanings it carries and in the Buddhist contexts described by Huard (2024) due to influence from other Iranian Buddhist languages (cf. Pinault 2016b: 191 f.). Nonetheless, as I mentioned earlier, the notion of "farnah" is *difficult* to render, and it is harder to speak of a semantic shift here; if a shift took place, it is a conceptual one rather than a semantic one.

Old Uyghur *kut* can be used in the same contexts as Tocharian B *perne*, Tocharian *paräṃ* (cf. Pinault 2016b: 92). It originally meant 'favour of heaven'

and, later 'good fortune' (Clauson 1972: 594). Possibly, given the results found by Huard (2024), we should no longer reinterpret the Uyghur expressions *arhant kutı* and *burhan kutı* as 'position of an arhat' and 'position of a Buddha' respectively (cf. Bailey 1937: 915, Wilkens 2021: 59), but as 'fortune of an arhat' and 'fortune of a Buddha'. This use seems to follow so closely the Tocharian usage that it might be calqued upon it, more specifically upon the Tocharian A words *arhantiśparäṃ* and *puttiśparäṃ*.

Nonetheless a problem remains with the Tocharian A form, since it has not undergone *p/m*-affection of the *a* as in *porat* 'axe', *opyac* 'memory'. There is no certainty as to why. Professor Georges-Jean Pinault's explanation is that the cluster /rn/ kept this change from happening (p.c.). I would like to propose another solution: it could be a Sogdian loanword. It is not unlikely that Sogdian *prn*, *frn* /farn/ would have been borrowed as Tocharian A *paräṃ*, but I have to admit that one could perhaps also expect †*pāräṃ*. I offer a similar explanation for TA *enāk* (see p. 24), which should have become †*enak*, if directly inherited from Proto-Tocharian **eynake*. For more detail, see the discussion on *puttiśparäṃ* below.

In conclusion, Tocharian B *perne* is from Old Steppe Iranian **farna*(*h*)- 'glory, kingly abundance, *farnah*'. It acquired its meaning 'fortune, happiness' in contact with Iranian Buddhist speakers, and this influenced in its turn the use of Old Uyghur *kut*. The Tocharian A form *paräṃ* is somewhat irregular (one expects †*poräṃ*). It is possible, although historically difficult, that it was borrowed directly from Sogdian *farn*, and would thus not be a cognate of Tocharian B *perne*. This word represents a semantic isogloss with Sogdian and Khotanese, like for the word TB *śāte*, TA *śāt* 'rich' (for which see p. 72–73, pp.158–159 and p. 189). In both cases, this isogloss is not old, but due to common Central Asian Buddhist culture.

2.2.12.1 A Note on TB *pernerñe* and Its Function in the Tocharian Hymn to *pernerñe*

Perhaps calling the hymn found in THT 203 to 205 a "hymn to *pernerñe*" is not one hundred percent accurate. However, the word *pernerñe* occurs in almost every verse, and it is clearly extolled. In the scholarly literature, it is almost systematically translated as 'gloriousness'. Adams translates it as '[awe-inspiring] splendor, glory' (DTB: 426). The word *pernerñe* (TA *parnore*) is the abstract of the adjective *perne*$_u$, obl. sg. *pernent* 'fortunate' (cf. Pinault 2011: 165). It comes from **perne-wer*, like pre-TA **parnar* remade into *parnor* due to the influence of the adjective *parno*, which *parnor* was enlarged with the *-e* collective suffix, hence TA *parnore* (Pinault 2011: 165). On the suffix **-wer* of Tocharian cf. Pinault (op. cit.).

Since the meaning of *perne* was rather 'fortune, good fortune', it is unlikely that *pernerñe* meant 'glory', 'gloriousness'. In my reading of the text, it seems to encompass all the Buddha's qualities such as compassion and selflessness, but also lack of desire, kindness, etc. The meaning 'fortune' also occurs, such as in the passage cited below. A new edition of this text, taking into account all the possible functions of *pernerñe* reflected is needed. In the following lines, I will only comment on a few extracts of this hymn, all taken from THT 204.

THT 204 a2–a3

> *piś cmelaṣṣeṃ onolmeṃts pernerñeś ṣek spelkessu śaulanmaṣṣe pitosa ce pernerñe kraupatai*
> "Always zealous for the *good fortune* of the beings of five births, you gathered this *good fortune* with the price of lives." (adapted from CEToM)

In my view, *good fortune* here refers to the salvation of beings. Through so many births (*jātaka*) the Buddha Śakyamuni has suffered and shown compassion to beings, and so he reached his good fortune. He then tried to save others from the pains, and strove for their own good fortune. In other passages *pernerñe* represents, I believe, other concepts, as shown below. I will replace *glory* in the translation from CEToM by the concept I think is more adapted (in italics). Another way of understanding the first occurrence of *pernerñe* here could perhaps be 'awakening, *bodhi*': for the awakening, for the *bodhi* of the beings of five births, you gathered this good fortune (of Buddha) with the price of lives.

THT 204 a3–a4

> *tañ pernerñe saim yāmoṣ perneñc ka po klautkāre*
> "All who have (ever) taken refuge in your *compassion* have turned *compassionate*." (adapted from CEToM)

It seems clear that *pernerñe* in this context designates compassion, not glory or good fortune.

THT 204 b1

> *snätkūweṣ tañ pernesa indrinta śle-k kektseñe*
> "Your perceptive senses likewise your body are penetrated by *understanding*" (my translation)

This use of *perne* (not *pernerñe*—CEToM suggests this might be for the meter) renders the Sanskrit notion of *abhijñā* which means 'understanding; higher

knowledge' but also, in Buddhist contexts, 'supernatural powers'. Interestingly, Sogdian *prn*, cognate of TB *perne*, is used to translate Chinese *tōng* 通 'penetration, penetrate by intelligence, understand', which is also the standard equivalent of Skt. *abhijñā* (cf. Huard 2024). The word *perne* here does not simply mean 'understanding' but designates 'higher knowledge', 'supernatural knowledge' which corresponds to the Buddhist notion of *abhijñā*.

THT 204 b2–3

pernerñesa wyāksasta tsärśalñenta läkleṣṣeṃ[26]

"With your *awakenness* (*bodhi*) you have turned away the painful torments" (my translation)

It is bodhi, Buddha-like awakenness/awakening that turns away the painful torments, all caused by ignorance and attachment to the senses. There are more examples of *pernerñe* having different meanings, although they always are equivalent to positive notions in Buddhist thought.[27]

From all these examples, it seems to stand out that the notion of *pernerñe* encompasses, at least in this hymn, a great number of functions, basically all the functions connected with the Buddha, his awakenness, his supernatural knowledge, his compassion, his good fortune, and more, as I have only cited a limited number of examples. It is perhaps not impossible to suppose that *perne* also could have a wider set of attributed functions, as representing, besides good fortune, the absoluteness of what was good and to be researched in the Buddhist mindset.

2.2.12.2 A Note on TB *pernesa* 'For the Sake of' and Parthian *frh'h* 'On Account of'

There exists a Parthian word *frh'h*, *frh'* 'for, on account of' (DMMPP: 155), as well as a Bactrian word φαρο 'to' used for personal indirect objects. The Bactrian particle gradually fell out of use in the 7th century (Sims-Williams 2009). These two words were connected by Sims-Williams (2009: 283 f. with references, the connection was originally made by Yutaka Yoshida), who suggests they derive from Old Iranian **frādaϑai* with shortening due to reanalysis of the word as contain-

26 After Ogihara's edition (2019: 13).

27 In a way, this polysemy is similar to that, perhaps less extended, of TB *palsko* 'thought', also used to designate *vijñāna* 'consciousness' in *palsko ekaṣṣe* 'eye consciousness', *pälskonta* 'free will' (cf. Chen 2019: 206, 208). This too probably goes back to an early Tocharian Buddhist convention.

ing a *fra*-prefix (as in Avestan *fradaϑa*, cf. de Vaan 2003: 151f.). It should be noted that Avestan *fradaϑa-* means 'prosperity' and was not grammaticalized.

While Sims-Williams' etymology is certainly attractive, it suffers from a few difficulties, all mentioned by Sims-Williams himself (2009: 284–285). 1. It implies a contraction of *-*δaϑ*- to *-*ϑ*- (> *h*) which is observed in the Sogdian and Khotanese forms of 'to sell'.[28] However, this contraction is not observed in Parthian or Bactrian. 2. This etymology implies a shortening of the initial syllable in this word as in Avestan, which is speculative of course.

Sims-Williams suggests that Bactrian φαρο 'to' underwent the same contraction, but he notes that there is no evidence for a development **fra-* > φαρ- in Bactrian, so this would be "a special case" because it is a monosyllable, or there is an anaptyx **frā-* > **farā-* exactly as in the personal name Φαραλο < **Frāda-*. As Sims-Williams (2009: 283[38]) himself admits, this name, the only reflex of OIr. **frā-*, is not attested at an early date. One can thus not exclude that this anaptyx is later; further the loss of final *-ā* in **farā*, explained by Sims-Williams (2009: 283) as due to its disyllabic constitution and probably unaccented proclitic function, is *ad hoc*.

The Parthian preposition could be more straightforwardly derived from *farrah* 'glory' (< **farnah-*) with an adverbial suffix *-āh* (on which see Henning apud Boyce 1954: 187). Semantically, this would be a shift from 'for the glory of' to 'for the sake of', with a very usual semantic weakening. This could also function for Bactrian φαρο, formally identical to φαρο 'glory' (< φαρρο, attested in coins, cf. Sims-Williams 2007: s.v.).

There is a Tocharian B adverb *pernesa* 'for the sake of', which is technically the perlative singular of *perne* 'fortune'. The function of *pernesa* was studied by Meunier (2015: 344–345). There are many occurrences of *pernesa*, but we can cite the texts below as an example.

THT 28 b8

> *piś cmelṣets wnolmets pernesa*: *ce akālk ṣek ñäṣ(ṣ)ī(tär)*
> "Pour l'amour des êtres des cinq naissances, il formulait toujours ce vœu" (Meunier 2015: 344)

THT 31 b4

> *waike waṣe ṣpä käskor wat wentsi klyin-ne po weṣṣäṃ śwātsintse pernesa*:
> "And if he had to speak a lie and the untruth or gossip, for the sake of food he says anything" (CEToM)

28 And apparently in the Persian reflexes of *frādaϑa-* (cf. Sims-Williams 2009: 280), but this is not clear.

THT 31 b5–b6

[…] *ṣeṃ ywārc ostmeṃ ltweṣ pudñäkteś snaice tallānt ikemeṃ ṣamāni tatākaṣ cai ṣamāññentse maiyyane aklyilñentse ṣ pernesa yark=ālaṃmeṃ klpāṣyeṃ*:

"These ones who [only] went half from the house to the Buddha from a poor, miserable condition had become monks. These achieved, by the power of monkhood and for the sake of [their] scholarship, veneration from elsewhere" (CEToM)

"Ceux-là atteignirent, sous le pouvoir de l'état de moine et par amour pour leur enseignement, la vénération venant des autres" (Meunier 2015: 344, for the part starting with *cai ṣamāññentse*)

In THT 28 b8, the meaning can be "pour l'amour" (for the love) indeed, or perhaps closer to the meaning of *perne* 'for the fortune/happiness': he was making this wish for the fortune or happiness of pentagenic creatures. However, in THT 31 b4, the fool will say anything for the sake of food, not for the fortune of food. This is equally true in THT 31 b5–b6 where the monks achieve veneration for the sake of scholarship, not for the fortune or happiness of scholarship, which is inanimate.

In Tocharian, there are more adverbs with *-sa*, so one could imagine *pernesa* is an independent innovation. However, given the existence of a similar word with a similar function in Bactrian, and in particular in Parthian where it has a suffix, as the Tocharian word does, it is conceivable that *pernesa* in its benefactive function is a calque from one of these or from another Iranian language.

2.2.12.3 A Note on TA *puttiśparäṃ* 'Buddha's Fortune'

Tocharian A *puttiśparäṃ* means 'fortune of Buddha' (previously translated in scholarship as 'dignity of Buddha'). It corresponds to the Tocharian B phrase *pañäktäṃñe perne* (cf. Meunier 2015: 181–182). One can cite, as an illustration, the following passage:

A 18 b4–6

āṣānik ptāñkät käṣṣi nu äntäne mahāprabhā(*se*) *ñoma wäl ṣeṣ waśirr oki prākräṃ pratimyo sne lotkluneyumänt puttiśparṣiṃ śäktälyi ṣñi sant*-(*änaṃ*) *sāryāt*

"Mais le vénérable, le Buddha, le maître, lorsqu'il était le roi nommé Mahāprabhāsa, avec sa détermination aussi dure que le diamant, sans retours dans sa séquence d'incarnations a semé le grain de la dignité de Buddha" (Meunier 2015: 182).

Discussing *puttiśparäṃ*, Pinault (2008: 94–95; 2016b: 191) considers it a compound of *puttiś** and *paräṃ* 'glory; rank'. Pinault views the element *puttiś**, which does not occur elsewhere, nor on its own, as derived from Sogdian *putīšt*, with "simplification prévisible du groupe consonantique final".

Another solution, based on Pinault's proposal, can be found to explain *puttiśparäṃ*: both elements, the first and the second, could have been borrowed together. Thus *puttiśparäṃ* could be viewed, not as a calque, but as a direct loan of the Sogdian expression *putīšt farn* 'good fortune of the Buddhas'.

The sequence *-štf-* could have been simplified, either perhaps in spoken Sogdian or otherwise in Tocharian A, yielding TA *puttiśparäṃ*. This solution could support the hypothesis I mentioned above, that, although Proto-Tocharian **perne* 'fortune, good fortune' was certainly directly borrowed from Old Steppe Iranian, the Tocharian A word, *paräṃ*, was perhaps borrowed from Sogdian. This is only a speculative idea, and Pinault's proposal of *puttiś** as an independent borrowing (although it is only found in this specific word in Tocharian A) is likely more solid, as one would expect an oblique plural *putīštī farn* in Sogdian.

It is an interesting fact that Sogdian *putīšt* was rendered with *-tt-* in Tocharian A. This probably yields some information on the Sogdian sound it renders, although I am uncertain what it should mean phonetically.

A problem remains: why was *putīšt farn* borrowed with *ś* and not with *ṣ*? Since there are not many more examples of secure Sogdian loanwords in Tocharian A,[29] one can simply hypothesize that /š/ in Sogdian was phonetically palatalized due to the preceding *ī*, although this is naturally speculative.

From the TA form *puttiśparäṃ*, perhaps interpreted by speakers as *putt* + *-iś* + *paräṃ*, this construction was extended to other Buddhist concepts, such as *ārāntiśparäṃ* 'good fortune of an *arhat*' corresponding to TB *ar(a)hanteññe perne* (cf. Winter 1971: 218; Wilkens 2021: 59).

In conclusion, Pinault's (2008: 95; 2016b: 191) proposal to derive *puttiś** in TA *puttiśparäṃ* 'dignité (statut) de Buddha' from Sogdian is very plausible. Moreover, I cautiously propose to derive the entire word *puttiśparäṃ* from a Sogdian collocation, directly borrowed in Tocharian A, with simplification of the sequence *-štf-* to *-śp-*, either in Sogdian or in Tocharian A.

29 A possible example is Tocharian A and B *menāk* 'comparison, example' from Sogdian *myn'k* (cf. Peyrot 2015a). If this is confirmed, it would be considered, to my knowledge, as the only loanword into Tocharian B that needs to come from Sogdian, although even this is unsure and it could very well be a Bactrian loanword (see p. 253).

2.2.13 *TB* perpente *'Burden, Load'*

The Tocharian B word *perpente* 'burden, load' was almost always read as *perpette* (cf. DTB: 426). However, a reading *perpente* seems assured in most cases. The meaning 'burden' is assured by the expression THT 554 b1 *piś atseṣṣe perpente täne yatäṃ ṣaläsi* (cf. also THT 1120 a3) "he will be able to throw away the burden of the five aggregates". The metaphor of "throwing away the burden of the five aggregates" corresponds to the Pali formula *panna-bhāra* 'one whose burdens have been laid down' that is, one who has become an arhat (the five aggregates of the Tocharian examples refers to the *five aggregates of clinging* of Buddhism).

I thank Athanaric Huard for having provided me with his new reading of fragment PK NS 216 where line a4 contains, among others, the words *oroce perwendi* 'a great burden' (see now Huard 2022: 268). The spelling ‹nd› and the final *-i* are somewhat mysterious, and I suggest influence of the Sogdian form *prbnty* /parbandē/ (cf. Sims-Williams 2020: 135) on the spelling of the word in PK NS 216. In any case, in the cluster ‹nd› the reading with ‹n› rather than ‹t› is unmistakable, so this form decisively proves that the correct reading elsewhere is *perpente*, not *perpette*. This cannot, however, be a Sogdian loanword, as one would then expect †*parwandi*, †*pärwānt* or the like (pace K.T. Schmidt, apud Sims-Williams 1985: 63).

The etymology of this word is quite transparent: the source must be Old Steppe Iranian **pari-banda-* 'load, burden', itself possibly back-formed from **pari-bandaka-* 'load' → PT **perpente*. The etymology of the Iranian word is *band-* 'bind, attach' prefixed with **pari-* 'around', so 'the thing that is bound around (the ass, the horse)'. Other Iranian cognates of this word are Sogdian *prbnty* 'burden' < **pari-bandaka-* (cf. SD: 283) and Persian *parwanda* 'bundle, roll of tissues; coarse cloth where linens are put; attached document, (gathered) writing relative to a subject, dossier, file, folder', also from **pari-bandaka-*.

An oblique form *perpecce* has often erroneously been read in THT 401 b5 (e.g. in the edition, Sieg and Siegling 1953: 267). Even if it were a correct reading, this could be a writing mistake of the scribe. However, even though the ‹tt› here is indeed somewhat strange, it is clearly ‹tt› or ‹nt›, not *‹cc›, cf. Ogihara (2016: 265, 266; pace Hackstein, Habata & Bross 2021: 83).

To conclude, Tocharian B *perpente* 'burden, load' derives from Old Steppe Iranian **pari-banda-*, of identical meaning.

2.2.14 *TB* mañiye *'Servant'*

Tocharian B *mañiye* 'servant' has been recognized as an Iranian loanword since Sieg, Siegling & Schulze (1931: 11; cf. also Hansen 1940: 149). Isebaert (1980: 204) has correctly classified TB *mañiye* as an Old Iranian borrowing. It derives regu-

larly from Old Steppe Iranian **mānúi̯a-* '(servant) of the house' itself from Proto-Iranian **dmāniu̯a-* 'of the house', see for example Old Persian *māna-* 'house' (MP, NP *mān* 'house, mansion'), Parthian *mān* 'house, mansion', Khotanese *damänā-* and Old Avestan *dəmāna-* 'house'.

TA *māññe* has been taken to be the equivalent of TB *mañiye* but rather corresponds to TB *-māññe* 'belonging to the house, garden', see below.

On the semantics of this word, see p. 189. On the phonetics, see p. 168 f. and 175 f.

2.2.15 *TB* -maññe, *TA* -māññe *'Belonging to the House; Garden'*

Tocharian B *-mañ̃ñe*, borrowed into Tocharian A as *-māññe*, designates a place that is attached to a house, such as a garden (and not necessarily a 'pavilion', pace DToA: 338). It is used in compounded forms such as TB *yärke-mañ̃ñe* 'sanctuary, temple' < 'reverence park', TA *talke-māññe* 'sacrificial park' (for more detail see Pinault 2002: 320 f.). The TB compound *winā-mañ̃ñe* clearly designates a garden or a park:

THT 275 a4

> *kreñc wīnā-mā(ñ̃ñi ¦ ce$_u$ ylai)ñ-ñäktentse ¦ werpiśkesa ma kca me(ṅkicci ¦*
> "les beaux parcs [de Ketumatī] ne sont en rien (inférieurs) au jardin d'Indra" (translation by Athanaric Huard, p.c.)

The meaning should thus be 'place belonging to a house', whether it be a yard, a garden, a pavilion. Pinault (op. cit.) suggests that this word was borrowed from Middle Iranian, which is unlikely, because of the vocalism of the final syllable: Middle Iranian **mānīg* or the like would not have yielded TB *-mañ̃ñe*.

I would like to propose that this word rather derives from Old Steppe Iranian **mān-úi̯a-* 'of the house' (s.v. *mañiye*), cf. Old Persian *māna-* 'house'. I assume that originally the second syllable was stressed, but that it became unstressed in Tocharian because the word was used as a second member in compounds, in which the last syllable of the first compound member is regularly stressed. Because it was unstressed, the *-i-* in *-iye* was lost, i.e. **mañíye* > **mañiye* > **mañye*, and the cluster **-ñy-* then developed to *-ññ-*.[30] The meaning of this word was likely 'belonging to the (noble) house' > '(place) belonging to the house'.

30 Because of the assimilation in this cluster, we cannot see whether the nasal was originally palatal or not in Tocharian, but I assume that it was palatal because of *mañiye* 'servant' (see above).

In conclusion, Tocharian B *-maññe* '-(place) belonging to the house, garden' derives straightforwardly from Old Steppe Iranian **mānúa̯-*, with secondary loss of the stress due to its use as a second member of compounds. It was borrowed into Tocharian A as *-māññe*.

2.2.16 TB melte* *'Pile'*, TA malto *'On Top'*

The basic meaning of Tocharian B *melte* is 'pile' and in some contexts it designates a 'pile of dung' more specifically (cf. DTB: 505). As suggested by Adams (DTB: 505), its closest cognate must be the Tocharian A adverb *malto* 'in first place, first' (DToA: 332). Winter (1965: 205) has suggested that TA *malto* is an adjective derived from an unattested **malt*, in the same way as TA *parno* 'glorious' is derived from *paräṃ* 'glory'.

Winter (1971: 220) separated *malto*, which he connected to Old English *molda* 'head, summit', from TB *melte*, which he connected to Avestan *kamərəδa-* 'head'. Tremblay, however, derives TA *malto* from an Old Iranian **marda-* 'head' and does not cite *melte*.

A problem with Tremblay's explanation is that there is no Iranian word **marda-* with the meaning 'head'. The Old Iranian word **mr̥da-* ~ **marda-* rather meant 'neck' (e.g. Persian *mul* 'neck'), and **kamr̥da-* ~ **kamarda-* means 'head'.[31] As I will argue in a separate article, **ka-* means 'on, together with', and **ka-marda-* initially meant 'on the neck', hence its meaning of 'head'.

From 'neck' to 'top', there is no easy pathway. However, the Sanskrit cognate of **marda-*, namely *mūrdhán-* (borrowed in Tocharian as *murt** 'highness'), has the following meanings: 'forehead, head in general, skull, (fig.) the highest or first part of anything, top, point, summit, etc.' (M-W: 826). From a comparative perspective, we can easily start from a meaning 'top' → 'top of the body'. To the Indo-Aryans, the top of the body would be the head (as it is to us), but to the Iranians, it would be the neck, excluding the head which is not part of the body. As Dr. Athanaric Huard reminds me, the neck of Indian cows is much higher than their head, and it is also possible that this would be the departure point, thus 'neck' as in 'top of the bovine body' → 'top'.

Besides this bovine consideration, the conception of the head as different from the body is not uncommon at all. One *has* a body, but one *is* a head, in a sense. One can, for instance, cite Ronsard (1553: 39) "Voyci deux nuages tous plains / De Mores, qui n'ont point de mains; / Ny de corps: & ont les visages / Semblables à des chats sauuages" ("Behold, two clouds full / Of Moors who have

31 Instead of reading the hitherto unexplained Elamite PN *Muštimarda* as **Muštivarda* as does Tavernier (2007: 486), I suggest to read it as *Mušti-marda* 'neck-fisted' (a compound meaning 'neck-sized fist, or wrist').

no hands / Nor do they have bodies, and their faces / Resemble wild cats"). From this poem it is clear that Ronsard did not understand the face, or the head, as being part of the body, since these Moors had no bodies, yet they had a face. It is possible that this vision of the head as different from the body is linked to the fact that the head is that from which we see, and the body is what we can see, thus excluding the head, but naturally, only a real anthropological study could solve this problem. In any case, Proto-Iranian **marda-* could have designated the 'top', then the 'top of the body', that is, either the neck or the head.

In the present case, we see, once again, a sound change **rd* > **ld* at work in Old Steppe Iranian. This sound change is discussed in greater detail in section 2.7.2.11. The reconstruction of the Old Steppe Iranian form was thus **malda-* 'neck, top of the body; top'.

Here, we see that Old Steppe Iranian had preserved a meaning 'top' of **marda-* which no other known Iranian language has kept. The Indo-European root behind Old Steppe Iranian **marda-* was also inherited in Tocharian B *mrāce** TA *mrāc* 'summit, top of the head' (cf. DTB: 514). TB *murt** 'summit' (the second state of penetration in Buddhist thought) was borrowed from its Sanskrit cognate, *mūrdhan-* (cf. Thomas 1966: 266[15]). Another derivative of the same root was borrowed from Pre-Khotanese: TB *kamartīke* 'ruler', TA *kākmärtik* 'master' etc. (see below). Thus, strikingly, this root was both inherited from Proto-Indo-European and borrowed a number of times: from Old Steppe Iranian, Pre-Khotanese and Sanskrit.

In conclusion, Tocharian B *melte** 'pile', Tocharian A *malto* 'on top' derives from Old Steppe Iranian **malda-* 'top; top of the body' (or 'top of the body, top'), which shows a sound change **-rd-* > **-ld-* seen in other words, such as, for example, *welke** 'stalk (?), petal (?)' < **weltke*.

2.2.16.1 A Note on TB *kamartīke* 'Ruler', TA *kākmart* 'Sovereignty'

The historical derivation of the Tocharian words B *kamartīke* (variant *kamarttike*) 'ruler', *kamarttāññe* 'sovereignty, rulership', Tocharian A *kākmart* 'sovereignty', *kākmärtik* 'ruler, master' (DToA: 101) is very important for both Iranian and Tocharian studies.

Both TB *kamartīke* 'ruler' and *kamarttāññe* 'sovereignty, rulership' have been known for a long time, and their meaning is not debated (see, for instance, DTB: 149). It has equally been noticed since long that these words are likely borrowings from an Iranian language.

Schwartz (1974: 411) proposed that these words and their Tocharian A equivalents are ultimately borrowings from Bactrian **kamarδə* 'head'. The Bactrian word that had to be inferred at the time has since been found in a document dating back to 710 A.D. under the form καμιρδο 'name or title of a god', see

Sims-Williams (1997: 23; 2007: 220). Since Schwartz and with the exception of Dragoni (2023: 73–74), all scholars have considered the Tocharian B words discussed here to be Bactrian loanwords in Tocharian (among others, Isebaert 1980: 198; Sims-Williams 1997: 23; Pinault 2002a: 265 f.; Tremblay 2005: 436; Pinault 2020: 327; Pinault 2020a: 392; DToA: 101). I will discuss some of the points raised by various scholars, in order to further determine the etymology of these words. For this purpose, I will mainly cite Pinault 2002a, Adams (DTB) and Peyrot (2015a).

Pinault (2002a: 263) suggests that both the Tocharian A and B forms go back to an original noun **kamártV* "which would regularly develop to **kamártV* and ultimately, with loss of the final vowel, **kamart*." This would explain why the Tocharian B form is constituted of two *a*-vowels, as can be seen from *kamarttāññe* (/kamartáññe/), while the Bactrian form is καμιρδο, which seems to indicate a schwa as a second vowel, or in any case, a vowel that is not /a/.[32]

To sum up: if the word were from Bactrian, we would in principle expect †*kamärttāññe* for 'rulership' and not *kamarttāññe*, except if the word were very ancient (cf. Pinault 2002a: 263). Thus, the base from which both *kamartīke* and *kamarttāññe* derive is clearly **kamarta* (cf. DTB: 149; Peyrot 2015a). Since this sequence *-art-* cannot correspond to Bactrian -ιρδ-, a derivation from Bactrian is difficult (DTB: 149; Peyrot 2015a).

A more considerable problem is that Bactrian καμιρδο is a late form, while we would expect **καμιρλο for earlier Bactrian (Old Iranian **rd* becomes -ρλ- in Bactrian, which, in late Bactrian, becomes -ρδ- "again"), cf. Sims-Williams (1997: 23) and Peyrot (2015a), yet, for Pinault's idea (2002a) to function, we would need an early Bactrian source. Thus neither the vocalism nor the consonantism fits with a Bactrian etymology.

These considerations are also valid for *kamartīke*, which thus phonologically is /kamartike/, with stress on the *i*, as indicated by the variants with ‹ī›, and as supported by *kamarttāññe*, whose first two vowels are unaccented /a/.

The first element *kāk-* of Tocharian A *kākmart* 'sovereignty' has remained unclear for a long time. Pinault has proposed to explain it as due to contamination with the verb *kāk-* 'to call, invite' in its meaning 'to command' (first in a conference in homage to Bailey at Cambridge in 1999; then in 2002a: 265–266).[33] This seems fairly likely, as the first element **kā-* was not very common in borrowings, and not analyzable within Tocharian either.

---

32 In THT 128 b6, an archaic manuscript, one even finds the form *kamārtaññe*. Pinault (2002a: 263) regards the long vowel in the second syllable as a mistake. However, I do not understand how the scribe could have made the mistake of writing ‹ā› in lieu of ‹ä›.

33 Tremblay (2001: 24[37]) seems to mention Pinault's idea but does not provide a reference to him.

Dragoni (2023: 73f.) suggested that this word was of Proto-Tumshuquese-Khotanese origin, namely, from an accusative singular **kamardu* "with early vocalization of PIr. **r̥* > **ar*." In the first two syllables, Khotanese *-a-* regularly corresponds to Tocharian *a*; the Khotanese final *-u* (of the accusative singular) corresponds regularly to Tocharian *-o*, with a Tocharian oblique *-a* (Dragoni 2023); and Khotanese *-l-* regularly derives from Proto-Iranian **-rd-*. This etymology thus fits perfectly formally, but it is also fitting semantically, since a number of Old Khotanese, Pre-Khotanese or Proto-Tumshuqese-Khotanese loanwords in Tocharian are titles or describe social positions (cf. Dragoni 2023).

The suffix *-īke*, however, cannot be Khotanese, because *-e* is not a normal Khotanese vowel ending in Tocharian. This suffix is nevertheless of Iranian origin (cf. Isebaert 1980). It is found in other loanwords from Khotanese, such as *aṣanīke* 'worthy, worthy one' ← *aṣāṃ* 'worthy' (DTB: 34). I believe that this suffix could be of Bactrian origin, compare Bactrian -ιγο. In Bernard & Chen (2022: 17[41]) as well as in my dissertation, I had suggested that the final vowel *-e* in *sapule* 'pot' came from a final schwa in Pre-Bactrian **sabōlə*, and that the final vowel *-e* in *-īke* also derived from a Pre-Bactrian schwa. Schoubben (2024: 269) has since suggested a different, more convincing derivation. He derives the *-e* in *sapule* from an old nominative singular ending in -ι < **-ah* as in Sogdian and Khotanese. This is supported by a forthcoming work (by Francesca Michetti) with attestations of a Bactrian nominative singular ι-ending. Similarly, Schoubben takes the suffix *-īke* to be from **-īgi*, the nominative singular of **-īga-* (Bactrian -ιγο) < **-i̯ia-ka-*. Conceivably the Tocharian *-īke* suffix was extracted from words such as TB *spaktanīke*, TA *spaktānik* 'servant' ← Pre-Bactrian **spaxtanīgi*, compare Bactrian σπαχνιιο 'relative to service' (cf. Pinault 2020: 392).

In conclusion, Tocharian B *kamartīke*, Tocharian A *kākmart* and their cognates are not of Bactrian origin, unlike what has been suggested in the literature, but they rather derive from Proto-Tumshuqese-Khotanese, as proposed by Dragoni (2023: 73f.). The Tocharian B suffix *-īke* is indeed best derived from Pre-Bactrian.

2.2.17 *TB* maiyya*, arch.* meyyā*, etc. 'Power, Might'*

The Tocharian B word *maiyya* means 'power, might'. The Archaic Tocharian B form of the word is sometimes written as *meyyā* (cf. Peyrot 2008: 58).[34] In fact, the only occurrences of the diphthong ‹ey› in Archaic Tocharian B texts are

34 I was unable to find the Archaic form *meiyyo* noted by Adams between parentheses.

found in the words *eynāke* and *meyyā* (cf. Peyrot 2008: 58), and in *pey-*, an archaic variant of TB *pai-* 'foot': *peynemeṃ* and *peyye*.[35]

For the etymology of *maiyya*, Adams (DTB: 508) suggests that it was originally an abstract meaning 'youthful strength', derived from *$moh_1iu̯o$-, which itself developed to TB *maiwe* 'young; youth'.

In my view, Adams' explanation is not convincing. Firstly, *maiyya* does not refer to youthful strength, but to a great, powerful might, sometimes even of supernatural nature. Secondly, a direct Indo-European transposition of this word, as done by Adams, "[as if] PIE *moh_1iwyeh_a-" does not yield the right outcome. For instance, *$moh_1iwyeha$- would not explain the *-yya* ending of the Tocharian form. Furthermore, it is morphologically problematic from the Indo-European point of view, because it would be a root *moh_1- with an unexplained suffix *-*i*-, another suffix *-*u*-, and yet another suffix *-*ieh_2*. Because of the synchronically exceedingly rare paradigm of TB nom.sg. *maiyya*: obl.sg. *maiyya*, it is rather advisable to investigate whether it can be a loanword.

Van Windekens (1976: 629) has proposed that this word was borrowed from Sanskrit *māyā-*. However, this cannot be correct, as we would then expect †*māy* (cf. DTB: 508). Furthermore, no Sanskrit loanword displays a spelling *ey* in Archaic Tocharian B as in the case of *maiyya*: arch. *meyyā*, a spelling which might imply that the original diphthong was /e+y/ instead of /a+y/.

There exists an Avestan word *maiiā* (< *māiiā-*), which means 'craft, capacity, strength' (but also 'satisfaction, joy'). This meaning is extremely close to the Tocharian meanings of 'power, might'. One can understand the Old Persian name *humāyāsa- (Elamite ‹Ú-mi-ia-áš-šá›) as a bahuvrīhi meaning '(he who has) horses of amazing strength'. The noun *$mā(i)i̯ā$- is also reflected in the Persian name *Humāy* which comes from *$hu\text{-}māi̯ā$- 'having an amazing craft' (cf. Remmer 2006: 139). Outside of Old Persian and Avestan, the word *$māi̯ā$- is also to be found in Sogdian *my'kčyk* /mayākčīk/ 'happy', with regular shortening of *$ā$ before yod (cf. Gershevitch 1961: 17).[36]

The shortening of *$ā$ in front of *yod* is also found in Avestan, Sogdian and Khotanese and other Iranian languages, cf. p. 173 f. It is possible to postulate it for Old Steppe Iranian as well. A reconstructed Old Steppe Iranian *$mai̯(i̯)ā$-

35 The form *peynemeṃ* is found in THT 1859 b5 and the form *peyye* in THT 1314 b7.

36 Provasi (2013: 387[41]) suggests the meaning 'favorable' for Sogdian *my'kčyk*. The semantics behind the Sogdian word remain to be explored: it is unclear to me how 'powerful' ('endowed with *māyā-*') took on the meaning 'happy'. Very speculatively, I wonder if there was a convergence with the meaning of *frn*, which can also, like *māyā-*, describe a power, a magic might (cf. Provasi 2003: 307), and which can also coincide with the notion of happiness.

would perfectly correspond to TB *maiyya*. To support the postulation of this sound law for Old Steppe Iranian, one needs to mention that OSIr. **ā* was also shortened in front of *waw* (s.v. TB *newiya*). This would also explain the Archaic form TB *meyyā*, as one of the only words with the Archaic TB sequence *-ey-* is *eynāke* 'common (vulgar); evil, bad', which is an Old Steppe Iranian loanword. The fact that the word has not been borrowed as ***meya*, and that Classical Tocharian B has *-aiyy-* in this word, and Archaic Tocharian B *-eyy-*, could suggest that the sound change **āi̯ā* > **ai̯ā* went through an intermediary phase **aii̯ā* where the length was transferred from the vowel to the glide (perhaps in front of *waw* as well, with a possible intermediary form **nauu̯ii̯ā-* → *newiya*).

A parallel for the borrowing of OSIr. **ai̯* as PT **ey* is *eynāke*. Instead of taking the latter to derive from PT **en-yənake* (with a Tocharian *en-* prefix) as did Isebaert (1980: 115), it seems more plausible that the Old Steppe Iranian diphthong sounded to the Tocharian ear more similar to **e* (= OSIr. **a*) + *yod* than to the diphthong *ay*. This also seems coherent with the fact that Old Steppe Iranian **au̯* was noted in Tocharian B as *e* + *w* (s.v. *newiya*, TB *kraupe*, Archaic TB *krewpe*).

The variants of this word are: Archaic TB form *meyyā*, and the Classical variant *maiyya*, besides Classical and Late *maiyyo* (cf. Del Tomba 2023: 140 ff. for a careful analysis of the distribution and an in-depth discussion on the morphology of the word). The type of nom. sg. *maiyya*: obl. sg. *maiyya* is aberrant and does not correspond to a noun-class. One would naturally imagine that *maiyya* was turned into *maiyyo* in order to regularize this noun. Del Tomba, nonetheless, suggested the opposite. Del Tomba points out that bahuvrīhi compounds with *maiyya* as a second member are frequent. Such bahuvrīhis have a different declension: nom. sg. *-a*: obl. sg. *-ai*. Del Tomba then logically proposes that *maiyyo* is original, and that *maiyya* is secondary to *maiyyo*, extracted from these compounds (Del Tomba 2023: 141). The form *maiyya*, according to this scenario, is thus secondary. The archaic form *meyyā* only occurs in the oblique and secondary cases, so that it does not contradict this hypothesis (Del Tomba 2023: 141[147]). This scenario seems very likely to me, I would thus like to propose that Old Steppe Iranian **ma(i)i̯ā-* → **meyya* was immediately remade into a noun **meyyo* in Proto-Tocharian.

Semantically, one has to assume a passage of the meaning 'craft, strength' to 'power, might', which is trivial. In any case, since the word **ma(i)i̯ā-* is most probably not inherited, it is hard to have a very distinct idea of its original meaning, and it is also possible that it always meant 'power, might'.

There is yet another possibility for the source of the Tocharian B word: it could be borrowed from the BMAC language directly, especially if the Indo-

Iranian words are themselves borrowed from the BMAC language, and not inherited (Lubotsky 2001: 314 writes that the word looks "fairly IE", and could be inherited, but could also be borrowed from the BMAC language). However, there is no positive indication that this could be a loanword from the BMAC directly. This option is thus much less probable than the Old Steppe Iranian hypothesis, although it definitely cannot be excluded (cf. Del Tomba 2023: 143, who considers my Iranian or BMAC etymology of this word as the most likely).

In conclusion, Tocharian B *maiyya* 'power, might' < PT **meyya* is a direct loanword from Old Steppe Iranian **maɻ(ɻ)ā-*, possibly of identical meaning. The Archaic Tocharian B form *meyyā* confirms, together with Archaic TB *eynāke*, that the Old Steppe Iranian sequence **a* + **ɻ* was borrowed as PT **ey*.

2.2.18 *TB* yentuke* *'India(n)'*

The Tocharian B word *yentuke** occurs so far thrice in the corpus. Two occurrences are the following:

THT 1110 a4–5

tumeṃ karmapāyä weṣeñcatse tonak rekauna yentukäñe pele weṣ(le ||[37]

"thereupon the *karmavācaka* shall speak exactly these words in Indic" (CEToM).

THT 1110 b1–b2

tumeṃ karmavācaketse nano toykk rekauna yeṃtukäñe pele (sāṅ)kāśc(ä) weṣä(le)

"Thereupon again the karmavācaka shall say exactly these words in Indic to (the community)" (CEToM).

As indicated on CEToM, *yentukäñe* stands for **yentukäññe*, a derived adjective based on the noun *yentuke**. The next sentence is in Sanskrit, which implies doubtlessly that *yentukäñe* designates the language of the Indians. Pinault (1989: 21) writes "parallèlement à *kᵤśiñ-pele*, le syntagme de valeur adverbiale *yentukäñe-pele* signifie 'à la manière indienne', 'en indien', i.e. 'en sanskrit'."

As it is an adjective which refers to Sanskrit, it is clear that *yentuke** in this example refers to India. The following example is very fragmentary, but presents us nevertheless with a morphological question:

THT 424 b6

-ka yentukeṃne masa |||

37 Since *reki, rekauna* is of alternant gender, we would perhaps expect **weṣ(lona)*, while the lacuna in the manuscript is too small for ‹lona›. The gerundive of obligation in its predica-

Taking *yentukeṃne* as a locative plural, and, accordingly, *yentukeṃ-* as the oblique plural, one could translate this as "he went to the Yentukes" (as a country) or "he went among the Indians" (DTB: 546). However, perhaps the word had an obl.sg. stem *-eṃ*, in which case *yentukeṃne* could also be the locative singular. If the word had an oblique singular in *-eṃ*, this would mean that the word indicated a human, and the passage would then have to be translated as "he went to the Indian" (if *yentukeṃne* is a locative singular) or as "he went among the Indians" (if *yentukeṃne* is a locative plural).

Schmidt (1985: 764–765) proposed that *yentuke** was of Old Iranian origin, deriving it from Old Iranian **hinduka-*. In my view, this is essentially correct, but his explanation of the initial *ye-* is not satisfactory, as he himself admits. The solution he suggested was that **hi-* was borrowed as **i-*, which became **yä-* at first and then **ye-* under the influence of the following **-u-*.

I believe a more satisfactory solution can now be proposed, because we have another example of an Old Steppe Iranian initial **i-* being rendered as *ye-* in Tocharian B, namely TB *yetse* (TA *yats*) s.v. In view of this second instance, it seems likely that Old Iranian **i-* became **i̯e-* or **i̯a-* (perhaps **i̯i-*) in Old Steppe Iranian. For Old Steppe Iranian, I posit **i̯anduka-* or **i̯enduka-* (perhaps **i̯induka-*?). On the loss of *h-* of Old Steppe Iranian, see section 2.7.2.4 and on the Old Steppe Iranian development of initial **i-* (and **u-*), see section 2.7.2.6.

A very interesting consequence of this is that Old Steppe Iranian speakers must have known about India, and well enough to transmit their name of India to the speakers of Proto-Tocharian.

Nevertheless, it is alternatively possible that a people, an ethnic group, which lost contact with India, also lost usage of the name of the Hindu valley. The fact that Old Steppe Iranians kept it—or acquired it—does not, therefore, necessarily indicate that they used it to designate India and thus that they already had kept knowledge of the region of India. Indeed, for instance, the name of the Yidgha people itself derives from an older form **i̯induka-* (on this, cf. Morgenstierne 1938: 16, 57), although they are, by all accounts, not Indian. It is thus possible that the Old Steppe Iranian name for the 'Indians' came to designate another people like the Yidghas. Because of its resemblance with the name of India, notably in Middle Iranian languages, the Tocharians could then have reinterpereted this word as meaning 'Indian'.

In conclusion, the Tocharian B word *yentuke* 'India(n)' is likely to be an Old Steppe Iranian borrowing. The consequences this may have for the history of Old Steppe Iranian and for Tocharian prehistory are still to be determined.

---

tive use can be in the singular, while taking an accusative object (I thank Athanaric Huard for mentioning this point to me).

2.2.19 *TB* yetse, *TA* yats *'Skin'*

The Tocharian B word *yetse* 'skin' and its Tocharian A correspondent *yats*, reconstructible as Proto-Tocharian **yetse*, have not yet received a satisfactory etymology.

Adams (DTB: 549) suggests that it derives from PIE **h₁ēd-so-*, a lengthened grade derivative of the verbal root **h₁ed-* 'to eat'. This is a mechanical reconstruction, which has to cope with the uncertainty whether PIE **ds* would really become **ts* in Proto-Tocharian. Also, the formation with a lengthened grade coupled with a *-so-* suffix seems designed for this etymology specifically. Adams further supposes that the reconstructed **h₁ēd-so-* would have initially meant 'flesh' and then 'skin'. Adams does not provide any parallel for this semantic development, and I also could not find any. Skin is not an edible part *par excellence*. On the whole, this reasoning is complicated, and requires a large number of intermediary steps.

A more straightforward etymology can be proposed. There is an Old Iranian word **idᶻa-* 'animal skin': Avestan *iza-** in *izaēna-* 'made of leather', Khotanese *häysa-* 'skin, hide', Balochi *hīz* 'leather churn', Yidgha *īze*, *yiǰya* 'goatskin-bag used for carrying sour milk in'. Since we see here the same correspondence between OSIr. **i-* and PT **ye-* as in TB *yentuke** (s.v.), we can assume that this was a regular sound correspondence. This would allow us to derive PT **yetse* from Old Steppe Iranian, since the correspondence of the final vowels and the Tocharian rendering *ts* of OSIr. **dz* < PIr. **ǰ* are securely attested.

Thus, Proto-Iranian **iǰa-* 'skin' would have become Old Steppe Iranian **i̯edᶻa-* or perhaps **i̯idᶻa-* or **i̯adᶻa-*,[38] regularly borrowed as PT **yetse*, and subsequently becoming Tocharian B *yetse*, Tocharian A *yats*. This sound change was perhaps triggered by word-initial stress, but it remains unclear in the detail. As to the semantics, it seems that the Iranian meaning was rather 'hide, leather' and it could have perhaps shifted to the meaning of 'skin' within Tocharian, or this could be a remnant of an earlier Proto-Iranian meaning for this word.

In conclusion, TB *yetse*, TA *yats* 'skin' is regularly borrowed from Old Steppe Iranian **i̯edᶻa-* or **i̯adᶻa-* (perhaps **i̯idᶻa-*?), of similar meaning, perhaps 'hide, animal skin'.

2.2.20 *TB* retke, *TA* rätäk *'Army'*

Both Tocharian B *retke*, Tocharian A *rätäk* 'army' (TA plural *rackiñ*) have been recognized as Iranian loanwords in Tocharian since the earliest times (Schulze

38 For more detail, see section 2.7.2.6.

1932: 212; Hansen 1940: 155). An often cited cognate is Middle Persian *radag* 'row, (army) rank, line'. Although Winter (1971: 217) viewed Tocharian B *retke*, Tocharian A *rätäk* as a Bactrian loanwords, it is clear that this is impossible, as a Bactrian *ραδαγο or *ραδγο *vel sim.* would never yield the *-e-* vocalism that we see reflected in TB *retke*, TA *rätäk*, which regularly go back to PT **retke*.

Because of the vocalism, PT **retke* is rather borrowed from Old Steppe Iranian, and more particularly from a form **rataka-* 'line (of an army)', which underwent syncope of the middle syllable in Tocharian (see section 2.7.2.7). This adds to the Old Steppe Iranian military vocabulary found in Tocharian, which appears to be quite representative of the type of contact Old Steppe Iranians and Proto-Tocharians had.

As to the semantic shift of **rataka-* 'line', one could believe that 'line (of an army)' took the meaning 'army', as is usually assumed. One can also think of a collocation such as **rataka- spadahya* 'line of army' or perhaps a compound **spada-rataka-* 'army line'. Perhaps, in the borrowing process or perhaps in Old Steppe Iranian itself, **rataka-* took the meaning of the whole collocation as a *pars pro toto*, like *babūr* 'steamboat; obs. train' in Tunisian, from *vapeur* in French *bateau à vapeur*, *train à vapeur*. Besides, one can also think of the French expression *bataille rangée*, literally 'ranked battle', designating a battle between two similarly structured armies. In this sort of context, it would not be so surprising that *rangée* 'army line' took the meaning of 'army'. The English expression *rank and file*, designating enlisted troops is a concrete example of how 'rank' (< 'line') can come to designate an army.

In conclusion, Tocharian B *retke* and Tocharian A *rätäk* derive from Old Steppe Iranian **rataka-*, which I tentatively take to have meant 'army'. The Old Steppe Iranian word may have been used synecdochically to designate the army, from a hypothetical earlier construction **rataka spadahya* 'army line', which perhaps itself came to mean 'army' as a *pars pro toto* figure of speech.

2.2.21 *TB* raitwe *and TA* retwe *'Combination, Mixture' and TB* eraitwe *'In Connection with'*

The Tocharian words B *raitwe*, A *retwe* 'combination, mixture', have generally been analyzed as derivatives of TB *rəytta-* 'to join, to connect, to combine' (< PT **rəytwa-*) and TA *ritw-*, so Bailey (1937: 896[1]), Pinault (1988: 183) and Adams (DTB: 588). As Pinault (1988: 183) mentions, the root TB *ritt-*, TA *ritw-* 'to join, to combine' textually corresponds to Sanskrit *yuj-* 'to join, unite, bring together', and he translates the phrase *raitwe preke* as "moyen [et] moment (opportuns)".

The TA word *retwe* has been analyzed as analogically remade from Pre-TA **retu*, either "under the influence of Tocharian B or with the addition of the productive TA abstract suffix *-e*" (Pinault apud Weiss 2015: 191–192[21]). In the fol-

lowing, I will provide a new etymology of Tocharian B *raitwe*, in the framework of which it will be more economical to view TA *retwe* as borrowed from the TB form.

The meaning of the Tocharian A and B forms mentioned above is well established. Often in a medical context, they designate a combination or mixture of ingredients, for instance applied on a specific part of the body. We may also mention the Tocharian B adverb *eraitwe* + comitative 'in combination with, in connection to' as in the following examples.[39]

IOL Toch 244 b2
tumpa eraitwe saṃtke ///
"in combination with this, the medicine …"

PK AS 4A a3
ṣamāññempa eraitwe cimpim palsko yātässi āstreṃ yāmtsi
"In conjunction with monkhood may I be able to tame the spirit [and] to make it pure" (based on CEToM).[40]

PK NS 80.2 a2
/// (*ok*)*ompa nai eraitwe yātalñe cwi orotsts*(*e*)
"Indeed, in connection to the fruit his ability is great" (based on CEToM).

The origin of TB *rəytt-*, TA *rätw-* 'to join, to connect, to combine' (< PT **rəytwa-*) had long been deemed uncertain. Poucha (1955: 260) suggested it was a cognate of Latvian *riedu* (*rist*) 'to arrange' and of Greek ἀριθμός. More recently, Weiss published an article (2015), where he connects TB *rəytt-*, TA *rätw-* to Latin *rīte* 'duly, correctly', *rītus* 'manner, way' (← *rītū* abl. 'according to the customary fashion') and Proto-Iranian **raiϑ-* 'join, mix, combine'. According to Weiss (2015: 190) the Tocharian and Iranian roots go back to PIE *(*H*)*reith*$_2$- 'join, connect', with a present stem in -*ya-* found in Avestan next to a -*ia-* present.

Weiss (2015: 190) is certainly right that for semantic reasons these Latin words cannot easily be connected to PIE **h*$_2$*rei*(*H*)- 'to count' as has been proposed earlier. However, I cannot agree with some of his other suggestions. In the following I will discuss his proposal more in detail, since it is difficult to accept all the points he raised concerning the etymology of these words.

39 Since adverbs do not normally govern cases, one can suppose that *eraitwe* was in the process of becoming a postposition.

40 I wonder if a less literal translation "with regard to monkhood" is not possible too.

As mentioned above, Pinault proposes, apud Weiss (2015: 191–192[21]), to see TA *retwe* as remade from Pre-TA **retu* under the influence of Tocharian B or with the addition of the abstract suffix *-e*. I do not know of any other example of a Pre-TA word having been remade under the influence of Tocharian B. In fact we can either have Pre-TA **retwey* (that is, **retu-* + abstract suffix **-ey*) or a loanword TB *raitwe* → TA *retwe*. Given the large number of TB loanwords in Tocharian A, and given the technical semantics of this word, it would not be surprising if it were simply borrowed from Tocharian B. This would also be more straightforward formally. That TB *-ai-* became TA *-e-* is normal and expected, as *-ai-* in Tocharian A is limited to Sanskrit loanwords (typically TA *śmoññe* 'place' ← TB *śmoññai* obl.sg.).

The existence of the root **raiθ-* in Iranian with the meaning 'to mix, mingle, pervade' is assured (see EDIV: s.v.). However, the morphology of the present stem in *-ya-* found in Avestan is uncertain. Indeed, Iranian has very few present stems in *-ya-* (Kellens 1984: 162–163): **tar(-ya)-* 'to overcome', **har-* 'to guard, observe', **ǰaiH* 'to live', all of which contain **-ya-* already in PIE, and **raiθ-* 'to join, to mingle, to pervade'.[41] Synchronically, **ǰaiH* is not a *-ya-* present (cf. Hoffmann & Forssman 2004: 184), which means we only have **tar(-ya)-* 'to overcome', **har-* 'to guard, observe' and **raiθ-* as Iranian verbs taking a present suffix *-ya-*.[42]

Of all these examples, **raiθ-* stands apart, partly because it contains a diphthong, but mostly because it is the only one without a clear Indo-European origin. The Latin and Tocharian forms connected by Weiss aside, this verbal root is not even found in Indo-Aryan, and would thus appear to be of Iranian date. A root only found in Tocharian and Iranian is generally suspicious in itself, as a borrowing from Iranian into Tocharian would in many cases be the most likely explanation. The Latin forms *rīte* 'duly, correctly', *rītus* 'manner, way' seem difficult to connect to a verb that means 'to mingle, to join, to connect', but their connection is not impossible.

It should be noted that there is also a verb *raēθβa-* in Avestan, with a present in *-ia-* (e.g. *raēθβaiiāṯ*), which Kellens supposes could be a compromise between *iriθiia-* and *raēθβa-* (Kellens 1984: 163). In any case, this shows that at some point **raiθya-* was analyzed as a verbal root, and not as a *-ya-* present, as also happened for **tar(-ya)-* 'to overcome' (cf. Kellens 1984: 163).

41 The present stem of the verb **fiahu* 'to hail' (only attested in Avestan) should be interpreted as a "thematic denominative formation" and not as a stem in *-ya-* (cf. EDIV: 86).

42 In other words, there is no root **ǰaiH-* extended with **-ya-* in Iranian, but the root needs to be reconstructed as **ǰiHu-* for Proto-Indo-Iranian.

Very importantly, there is a noun *raēϑβa-* 'mixture', which corresponds formally and semantically to Tocharian B *raitwe*, Tocharian A *retwe* 'mixture, combination'.[43] In my view, this match is important, as it shows that the word could have been borrowed by Tocharian, either together with the Iranian verb or not.

Since the other *u̯a*-presents in Iranian have Indic and Indo-European cognates, a number of suppositions can be made to explain why they were reanalyzed—or not—as roots enlarged with *-u̯a-*. However, for **raiϑu̯a-*, we have no clear cognates except the Tocharian one to explain its prehistory.

There are a number of possibilities: 1) either Iranian and Tocharian are the only branches to have preserved this verb, and the great similarity of meaning in the verb and noun is due to coincidence or Iranian influence; 2) Iranian is the only branch to have had this verb originally and Tocharian has borrowed it from Iranian; 3) these two verbs are unconnected (so Cheung, EDIV: 310) and the identity of meaning of the derived noun is coincidental; 4) or Tocharian has borrowed the noun and not the verb. It is very unlikely that Tocharian had borrowed the verb and that the noun was made from the verb at a much later period and then became a perfect match with Avestan *raēϑβa-*.

Even if the Tocharian and Iranian forms were cognates, there would be no good Indo-European etymology for both verbs. For Iranian, there is not even a convincing—or known—Indo-Iranian etymology (see below).

An important point to underline is that **tar*(*-u̯a*)- 'to overcome' is only suffixed in *-a̯ia-* although the bare form **taru̯a-* is attested in compounds, cf. Kellens (1984: 163), while **raiϑ-* receives two different suffixes, besides the form **raiϑu̯a-*. This could show that the various treatments **raiϑ-* received as a verbal stem are later than those of **tar*(*-u̯a*)-. This conclusion could be backed by the fact that there is no Indic cognate of PIr. **raiϑ-*.

In fact, I believe that PIr. **raiϑ-* is a properly Iranian formation. I had first believed that, on the basis of the Proto-Indo-European root **leiH-* 'to cling to, to adhere' (Ved. *láyate*, LIV²: 405), an abstract noun was made, probably in Proto-Iranian, but possibly in Proto-Indo-Iranian: **raiϑu̯a-* (with the **-ϑu̯a-* abstract suffix) 'clinging, adhesion (of ingredients)' > 'a mixture, a combination'. This is the noun *raiϑu̯a-* 'mixture', attested in Avestan *raēϑβa-* (Bartholomae, AiW: 1482 'Mengung').

I then renounced this etymology since, as Professor Lubotsky pointed out to me, the existence of the Proto-Indo-Iranian root **raiH-* is uncertain, and the Indic cognates mostly mean 'to dissolve, disappear'. Rather, he suggested

43 Secondary meanings such as 'moyen' discussed by Pinault (see above) may have evolved in a Buddhist context.

that it derives from PIE **loi̯-* 'to pollute' (cf. Milizia 2015), from which Proto-Tocharian **raimo* (s.v. *raimo**) probably derives as well, through borrowing from Old Steppe Iranian. The original meaning would thus be 'to mix' > 'to contaminate', as in Latin *contaminare* 'to defile, corrupt, deteriorate by mingling' ← *contamen* 'contact', Sanskrit *lepa-* 'spot, stain, impurity' ← *lip-* 'to smear, adhere to' etc.

From PIr. **raį̄-* 'to mix' an Iranian noun **raį̄θu̯a-* 'mixture' would then have been made, and the verb **raį̄θu̯a-* 'to mix together, to mix' is clearly denominal. It has been thought that **riθ-i̯a-* 'to die' and **raį̄θu̯a-* 'to mix together, to mix' are etymologically connected; see for instance de Vaan (2003: 216) who assumes one single root *riθ-* 'to mingle; to die'. However, semantically there is little overlap between both meanings and the idea of their cognacy should be abandoned (as in EDIV: 309).

All this explains quite well why the root is completely absent from Indic. If my hypothesis is accepted, then Tocharian either borrowed the noun from Old Iranian, or the verb, or both. In Tocharian, it was in any case possible to derive verbs from borrowed nouns, such as TB *spalk-*, TA *spāltkā-* both 'make an effort' ← PT **speltke* (s.v.) 'zeal' (cf. Peyrot 2013: 837, 837[991]) and **netsu-* (from Old Steppe Iranian **natsu-*) 'dead' from which the verb **nətswa-* was made (s.v. TB *mətstsa-*). Possibly this also happened with **tsəwa-* (s.v. *tsəwa*). This set could have furthered the derivation of the verb *ritwa-* from *raitwe*, since it may have been unclear to the speakers whether the noun or the verb was primary.

There are also parallels for derivation of the noun from the verb: *traiwo* 'mixture' derived from *trəywa-* 'be mixed, get mixed, mingle, shake' could have played a major role, given the semantic closeness to TB *ritt-* / TA *ritw-* 'to join, connect' and TB *raitwe* TA *retwe* 'combination, mixture'. Possibly there even existed a noun *traiwe* with a similar meaning (see DTB: 341). There are other nouns derived from verbs in a similar manner, such as *traike* 'confusion, delusion' ← *trəyka-* 'go astray, be confused'. Nonetheless, there is no trace of a change in that direction for Old Steppe Iranian loanwords in Tocharian, so that it is much less likely, and much more likely for the noun to have been primary and the verb secondary.

Taking TB *raitwe* (probably borrowed in TA *retwe*) as a borrowing from Old Steppe Iranian **raį̄θu̯a-* is thus very tempting. From *raitwe* a verb *ritwa-* could have been made, as a back-formation, on the model of parallel nouns, like *traike* and *trəyka-*. It is also possible, but unlikely, that Tocharian borrowed both the verb (from Old Iranian **raį̄θu̯a-*) and the noun independently (also from Old Iranian **raį̄θu̯a-*). It is also useful to revise the meaning of the preterite participle *rittowo*, at least for some contexts, as in the example I cite below.

PK AS 5C a2

pelaikne stmau prākre perma śayim ṣeme kauṃ su śpālmeṃ: mā no kante pikwala snai pelempa rittowo śaul ārttoymar

"Being installed firmly on the Law, even if I live only one single day, that [would be] an excellent life. But I will not adopt a life lasting a hundred years [if it is] connected with Lawlessness." (Pinault in CEToM)

The last sentence was also translated by Pinault (2020c: 468) as "And for hundred years may I not approve the life, being bound with lawlessness." Given the etymology of **rəytwa-* that I propose, I think one can translate this slightly differently, not as "connected with Lawlessness" but as "blended with Lawlessness" (or "combined with Lawlessness"): I will not adopt a life that contains (any) lawlessness. The difference with 'connected' or 'bound' is slight of course, but 'blended' here is more coherent with the original meaning of **rəytwa-*, namely, 'to combine, to mix'. In the following example, I also suggest a similar translation.

THT 3 b6

[...]*wai*(*me*)*netse śaul totk*= *āttsaik; su ṣp laklempa rittowo; mā no wnolmy aikenträ*:

"The life [of humans] [is] difficult and only short, and it involves suffering, but the beings do not realize [it]." (CEToM)

A more exact translation of the expression *laklempa rittowo*, in my view, is "it is blended with suffering". When milk and water are blended, for instance, it becomes impossible to take away one without taking about the other, they are completely mixed. The original idea must have been the same: life is blended with sufferings, with torments, it is impossible to separate one from the other.

The semantic variations of the verb TB *ritt-* / TA *ritw-* are possibly due to influence from the Indian meanings of the corresponding Sanskrit verb, an influence which seems to be pervading in the texts, as shown by Weiss (2015: 193). The Tocharian verbs also have a meaning 'fitting, suitable' (cf. Weiss 2015: 194). As Weiss (2015) reminds us, in that case they render Sanskrit *yuj-* 'to join' (cf. Sieg & Siegling 1949: 162), which also means 'to be fit or suitable'.

In conclusion, I propose to see TB *raitwe*, TA *retwe* 'combination, mixture', and the connected verbs as borrowings from Old Steppe Iranian. The Tocharian verb and noun have no good Indo-European etymology. While a convincing Indo-European etymology had not yet been found for the Iranian verbal and nominal forms, I propose to see the nominal form as an abstract derivation from a shorter Indo-European root found in Indic but lost in Iranian. Accord-

ing to this hypothesis, the verb is derived secondarily from the noun in Iranian. If this is correct, then there is no reason to assume that the Tocharian forms are inherited from Proto-Indo-European, and they must instead have been borrowed from Iranian.

2.2.22 *TB* raimo*, *TA* rem *'Dust'*

A Tocharian B form *raimaine* has recently been found in THT 1622 (a group of fragments of uneven size probably belonging to the same folio) by Athanaric Huard.[44] It appears in THT 1622.c+b, line a7, in a description of the Tapana Hell.

THT 1622.c+b a7

> /// – *pälkoṣ waraṃśne raimaine* /// *iñcuwasa: kärweñän tarśkaṃ ṣ/pa* ///
> "… dans du sable brûlant et la poussière … sur du fer … des pierres et des tessons …" (Huard 2022: 495).

TB *raimaine* 'dans la poussière', if correctly identified, clearly corresponds to the Tocharian A word *rem* 'dust' (cf. Poucha 1955: 261). These words can hardly be separated from Khotanese *rrīman-* 'faeces, filth, dirt' and Middle Persian *rēm* 'dirt, filth, purulence' (CPD: 71). The etymology of these words is discussed in Milizia (2015). Milizia connects these Iranian cognates, in my view convincingly, to Ancient Greek λοιμός 'plague'. The original idea, according to Milizia (2015: 92, 100 f.) is that of 'bodily rejection; dirt', initially stemming from the human body or corpse (s.v. *raitwe* for more detail on this matter). Since the Proto-Indo-European root has **l-* and not **r-* it is impossible to consider the Tocharian words as cognates by way of inheritance. The Tocharian words TB *raimo** and TA *rem* must thus be, originally, borrowed from another source.

A likely candidate is the Iranian root **raḭ-* referring to polluted substances. If we depart from the root **raḭ-*, it appears that an Iranian source is the most logical choice, since there is no known cognate of this root to be found in Indic. The proto-form would then be PT **reyma* 'dirty substance' ← Old Steppe Iranian **raḭmā-* 'id.'. OSIr. **raḭmā-*, in its turn, would be a feminine derivative of **raḭma-* masc. 'dirt' and probably served as a feminine individualizing noun (on a collective masculine base).

There remains, nonetheless, a problem concerning the inflectional type of this word. We know the oblique singular of this word, but we do not know

44 I wish to thank Athanaric Huard for sending me his notes about various words, among which *raimaine*, and also for drawing my attention to this word, as he himself believed it is of Iranian origin.

its nominative singular yet. There are three possible inflectional classes this word could belong to. First, one could think that it belongs to the type of *aśiya*: nom.sg. *-a* and obl.sg. *-ai*, but this seems very unlikely because this class consists of female referents (Del Tomba 2023: 91). Second, it could belong to the *okso*-type, with a nom.sg. in *-o* and obl.sg. in *-ai*, and third it could belong to the *arṣāklo*-type, also with a nom.sg. in *-o* and an obl.sg. *-ai*. In Tocharian B, words of the *wertsiya*-type have their nominative singular in *-a* and their oblique singular in *-ai*, which theoretically could correspond, but they also contain a palatal element at the end, which is lacking in **raima*.

The Tocharian B word *maiyya*, which is also a borrowing from Old Steppe Iranian, has an oblique singular ending in *-a*, not in *-ai*. Its type is rare, so rare that *maiyya* itself was remade into *maiyyo*: *maiyya* in historical times, corresponding to a much more common paradigm. However, Del Tomba's scenario (Del Tomba 2023: 140 f.) should be taken into account. Having observed that the form *-maiyya* occurs in bahuvrīhi compounds (which have a nom. sg. *-a*: obl. sg. *-ai* inflection), he suggests that the form *maiyya* is extracted from those bahuvrīhis (Del Tomba 2023: 141; see discussion on *maiyya* p. 56 for more detail).

To my knowledge, the oblique singular of *newiya* 'canal' (s.v.) is not known, so that it could have been **newiyai* or **newiya*, but the former is also more likely. These three words, *raima*, *newiya* and *maiyya* are the only relatively certain Old Steppe Iranian long *ā*-stems (*-*ā* < PIE *-eh_2) reflected in Tocharian B, together with *wertsiya* 'assembly'.

Because *wertsiya* and *newiya* were recognized as Iranian loanwords, it is possible that *raima**, also ending in *-a*, was assigned to the *wertsiya*-type. However, *maiyya* contained a *yod*, and was thus more susceptible to belong to this class, also being an Iranian loanword, or, after Del Tomba's discussion, the form *maiyya* is perhaps secondary. It is thus very unlikely that a putative form *raima** was lined up with the *wertsiya*-type only because of its Iranian origin.

A more likely scenario would be that an Archaic Tocharian B form *raima* or a Proto-Tocharian form **reyma* was replaced in Proto-Tocharian or in Classical Tocharian B by *raimo**, like **meyya* was remade into **meyyo* in Proto-Tocharian (following Del Tomba's scenario, s.v. *maiyya*) or alternatively like Archaic Tocharian B *meyyā*, Classical Tocharian B *maiyya*, was later made into *maiyyo* (Peyrot 2008: 99 f.). TB *maiyyo* belongs to the *kantwo*-type (nom.sg. *-o*, obl.sg. *-a*, cf. Del Tomba 2023: 102), which contains a few loanwords (perhaps TB *kāswo*, s.v.; TB *maiyyo*; TB *tāno* s.v.). My suggestion is that *raimo** was part of the *okso*-type (nom.sg. *-o*, obl.sg. *-ai*, cf. Del Tomba 2023: 145 f.) although only the oblique singular *raimai* is known for sure.

In conclusion, the nom.sg. of Tocharian B *raimaine* 'in dust' may at first have been **raima*, but may later have been changed to **raimo*. Together with TA *rem*, it is probably a borrowing from Old Steppe Iranian **raiṃā-* 'dust, dirt' (a feminine derivative of a collective masculine noun), cognate of Khotanese *rrīman-* 'faeces, filth, dirt', Middle Persian *rēm* 'dirt, filth, purulence'.

2.2.23 *TB* weretem* *'Debt; Surety'*

The Tocharian B word *weretem** is not directly attested, but the adjective *weretemaṣe* 'debtor through bankruptcy' (THT 1111 b1) regularly derives from it with the addition of the adjectival suffix *-äṣṣe*. Another possibility, suggested by Prof. Georges-Jean Pinault (p.c.) consists in deriving *weretemaṣe* from *werete** + *-maṣṣe* a late, colloquial suffix ("servant à dériver des adjectifs de relation, surtout avec valeur technique", cf. Pinault 2002: 328). Nonetheless, I believe the first derivation to be more meaningful, because of the ultimate etymology I propose, as I will mention below. Prof. Georges-Jean Pinault (p.c.) also suggests a connection to Tocharian B *peri* 'debt' (probably with the alternation *p* ~ *w* which occurs in some examples), which is semantically very attractive. However, I am not sure how the vocalism could be explained: Tocharian tries to syncopate or dissimilate two or three *e*s in a row, so that we would rather expect †*weritem*, *vel sim.*

Adams (DTB: 663) translates this adjective as "'surety' (?) or, less probably, 'bankrupt person' (?)'". According to Schmidt (1986: 129; 2018: 104) the Chinese version of the text implies a meaning close or identical to 'debtor through bankruptcy': "*weretemaṣe* bedeutet nach seiner Entsprechung in der chinesischen Version etwa "Schuldner durch Bankrott(?)"" (see also Chung 2004: 85, 87).

Adams (DTB: 663–664) convincingly proposes to derive TB *weretem** from Old Iranian **wratam*, related to Avestan *uruuata-* 'promise, contract' and Sanskrit *vratám* 'rule; commitment to the observance of a rule' (cf. Pinault 2017: 363), which go back to Indo-Iranian **u̯rata-* neut. 'duty, promise, contract'.[45]

The meaning of **u̯rata-* is not easy to establish. Schwartz (2002: 61) translates Old Avestan *uruuātā* as 'commitment' and *uruuat-* as '(bound by) solemn

45 For the Indo-European etymology of this Indo-Iranian word, see Pinault (2017). Noting that **u̯rata-* cannot be derived from **u̯erH-* 'to speak solemnly', since we would then expect ***urHata-*, he rather proposes to derive it from **u̯er-* 'to perceive, watch out for' > 'to ward, guard' > 'to observe, respect', and thus **u̯r-etó-* meaning 'that which is to be observed'. According to Katz (2003: 151), Proto-Indo-Iranian or Proto-Iranian **u̯ratám* was borrowed into Proto-Ugric as **wḗrta*, Khanti ("ostjakisch") *vort* 'Nachricht, Botschaft', *wort-χoy* 'bridal suitor' (someone who has promised to marry a specific girl?).

declaration' (2002: 54). For Sanskrit *vratá-*, Pinault (2019: 118) has argued for a meaning 'commitment'. However, Pinault now has a different view (cf. 2016a: 60–61 and especially 2017: 358 f.), and translates the Vedic and Avestan words discussed here with 'rule'.

The meaning of Tocharian B *weretem* is not completely certain (see above). However, based on the meaning 'rule, contract' deducible for PIr. **u̯rata-*, it probably meant 'rule, contract' too, and the derived adjective *weretemaṣ(ṣ)e* could then have meant 'bound by a contract or a rule; bound because of a promise', as is the case with someone in debt, a debtor.

The formal correspondences between Tocharian B *weretem**, which presupposes Proto-Tocharian **weretem*, and the posited Old Steppe Iranian form **u̯rata-* need to be commented. The vowel correspondences for the second and third syllable of the Tocharian word are regular. The final *-m* remarkably retains the final *-m* of the nominative singular of the Old Steppe Iranian neuter (nom. and acc.sg. **-am*). This fact is to be brought together in conjunction with the systematic rendering of Iranian masculine and neuter stems with Proto-Tocharian **-e*, corresponding both to Proto-Iranian **a* and **ah*-stems, since it shows that Tocharian borrowings in *-e* reflect the masculine nom.sg., **-ah* in Proto-Iranian, and the neuter *h*-stem ending, likewise reconstructed as Proto-Iranian **-ah*. It also shows that the Proto-Tocharian rendering **-a* of Iranian feminine **ā*-stems is based on the nominative too, since the **-m* of the accusative would have been preserved in Tocharian.

The first syllable *we-* of the Tocharian word remains to be explained. Perhaps Proto-Iranian **u̯r-* had undergone a change in Old Steppe Iranian. It is tempting to think of the remarkable rendering of Ir. **i-* as *ye-* in Tocharian B *yetse* 'skin' and *yentuke* 'Indian', but the conditioning here is not exactly parallel. This could explain the lack of syncope, as, for instance, the word was not /weretem/, which would have been syncopated to ***wertem*.

In conclusion, Tocharian B *weretem** 'debt', 'surety' is a borrowing from Old Steppe Iranian **u̯rata-*. This word was neuter, and Tocharian preserves the neuter nom.acc.sg. ending *-m* (DTB: 664). The Tocharian reflex *wer-* for Ir. **u̯r-* still needs to be explained.

2.2.24 *TB* waipecce *'Property, Possessions'*

The Tocharian B word *waipecce* means 'property, possessions'. It was mentioned as an Iranian loanword for the first time by Winter (1971: 218), after a suggestion by Warren Cowgill. Its Iranian etymology is generally accepted, see for example Isebaert (1980: 86, 113). Isebaert, however, takes this word as deriving from the accusative "**hvāi-pačyam*" (1980: 118) which is unnecessary and unwarranted as a more straightforward etymology exists, and as the accusative in **-am* would likely be rendered by PT **-em*, as in *weretem* (s.v.).

The identification of this word with Av. *xᵛaēpaiθiia-* 'own' is evident: compare also Old Persian *uvaipašiya-* 'eigen', *uvāipašiya-* 'own possession' (Isebaert 1980: 86), Parthian *wxybyh* (Morano 1982: 39) and Bactrian χοβο/χοβιγο 'one's own, property' (Sims-Williams 1994; Schoubben 2024: 362–365).

As to the meaning, either the (Proto-)Tocharian form underwent a semantic change from 'own' > 'own thing' or it derives from a form similar to OP *uvāipašiya-* 'own', with an identical meaning, thus **hu̯āi̯-paθi̯a*. It appears that both proto-forms would give the same result, *waipecce*, in Classical Tocharian B. Old Steppe Iranian **(h)u̯aɪ̯-paθi̯a-* or **(h)u̯āi̯-paθi̯a* was thus borrowed in Proto-Tocharian as (**waypetʲye* >) PT **waipecye* 'property, possessions (collective)' > TB *waipecce*. On the possible loss of /h/ in Old Steppe Iranian, see section 2.7.2.4. The Tocharian A descendant of this Proto-Tocharian form should have been †*wepac* or †*wepaci*.

While we have Archaic TB *meyyā* and *eynāke* (< OSIr. **ma(i)i̯ā-* and **aɪ̯nāka-*) with Old Steppe Iranian **a+i̯* preserved as Archaic Tocharian B *e* + *yod*, we find no occurrence of †*weypecce*, not even in archaic manuscripts such as THT 128 a2.

A reason for this could be that the Old Steppe Iranian form was indeed **(h)u̯āi̯-paθi̯a-* with **(h)u̯āi̯-* regularly rendered as Proto-Tocharian **a + yod*. However, for *tsain* en *tsaiññe* there is no evidence of an archaic variant †*tseyn* and †*tseyññe* either, and the same explanation cannot be reasonably evoked for these words: it is certainly not warranted to set up **dᶻāɪ̯nu-* and **dᶻāɪ̯n(a)i̯a-* *vel sim.* The *-ey-* variants of *meyyā* and *eynāke* are rather rare and not systematically found in archaic texts, which means that it is also possible that **weypecce* existed, but is not attested, in which case TB *waipecce* might as well simply derive from **hu̯aɪ̯-paθi̯a-*.

2.2.25 *TB* śāte, *TA* śāt *'Rich'*

Tocharian B *śāte* and Tocharian A *śāt* go back to Proto-Tocharian **śate*. The Proto-Tocharian word is a borrowing from Old Steppe Iranian **čyāta-* 'happy'. Hansen (1940: 156) and Tremblay (2005: 448) both consider it a loan from Sogdian, but the consonant *ś-* would then be unexpected, since we would expect Tocharian *ṣ-* for Sogdian *š-* (however, see *puttiśparäṃ*), and the final *-e* of the Tocharian B form would also be unexpected.

Rather, if taken from Old Steppe Iranian, this word has *ś-* for PIr. **čy-*, which can be explained either as a further palatalization of an Old Iranian **č'*, or as a direct borrowing from an Old Steppe Iranian **ś-* (itself from **č'*). Theoretically, Old Steppe Iranian could first have had a stage with a palatalized **ś́i̯-* (OAv. *šiiāta-* 'happy') and then could have lost the *yod* (YAv. *šāta-* 'id.'). This means that, if Tocharian borrowed its *ś-* from Old Iranian, it would have done so at a

later period than if it borrowed a palatalized **č-*, which is supposedly older in Iranian. Another possibility is that an Old Steppe Iranian **čʲi̯-*, with phonetic palatalization of the **č-*, was borrowed into Tocharian as such, and that the development to **ś-* occurred within Tocharian. This problem is discussed in more detail in section 2.7.2.2 of this chapter.

A very important point to note about this word is its semantics: the meaning 'rich' is unexpected for this word, as the original meaning is 'content', and it often means 'happy' in Iranian languages; cf. Avestan *šā̊ti* 'happiness', Middle and New Persian *šād* 'happy', etc.

In my view, the semantic shift from 'happy' to 'rich' is non-trivial, and it only occurs in two other languages: Sogdian *šʾt* /šāt/ 'rich, happy' (SD: 370) and Khotanese *tsāta-* 'rich' (Bailey 1958: 148; DKS: 146). The sister language of Khotanese, Tumshuqese, interestingly also has the word *tsāta-* but with the meaning 'property' (Bailey 1958: 148). In fact, the shift apparently occurred in Tocharian only, in a literary Buddhist context. On same shift in the Sogdian and Khotanese cognates of this word, see p. 189.

In conclusion, Proto-Tocharian **śate* 'happy', which gave Tocharian B *śāte* and Tocharian A *śāt* 'rich', was borrowed from Old Steppe Iranian. The Old Steppe Iranian form from which it was borrowed was either **šʲi̯āta-* or **čʲi̯āta-*, and it meant 'happy, content'. The meaning 'rich' arose in a Buddhist context.

2.2.26 *TB* speltke, *TA* spaltäk *'Zeal, Effort'*

Tocharian B *speltke*, *spel*(*k*)*ke* (on the simplification of *-tk-* to *-kk-* and subsequently to *-k-*, see Peyrot 2020) and TA *spaltäk* mean 'zeal, effort'. From these nouns a verb was derived: TB *spalk-* and TA *spāltkā-* 'make an effort' (cf. Peyrot 2013: 837, 837–838[992]). Because of the different stem patterns of the Tocharian A and B verbs, it seems more likely that this derivation is of post-Proto-Tocharian date, cf. Peyrot (2013: 837–838[992]), and that it occurred independently in the daughter languages.

Tocharian B *speltke* and Tocharian A *spaltäk* were recognized as Iranian loanwords since Winter (1971: 219), who proposed to see them as coming from the unattested Bactrian outcome of "iranisch" **sparδaka-*.

Adams (DTB: 789) is against this proposal of Winter's, and writes that "the phonological connection is not as straightforward as is usually assumed". A similar opinion is also expressed in Pinault (2020: 327). Indeed, a Middle Iranian origin of these words, as has often been assumed, is impossible because of the vowel *e* in both syllables, and because of the sequence *-lt-*, for which no Middle Iranian source seems to qualify. Bactrian, for instance, would have **-rl-* here.

Adams' alternative proposal consists in deriving the verb from PIE **spel-d-ske/o-* "as in Latin *pellō*". According to him, the verb is primary, while the

nouns are deverbal. Adams' etymology is weakened by the fact that there is no other known *s*-mobile form of this root. It would also require an **o*-grade (**spol-d-sk$^{e}/_{o}$-*) which would be aberrant, since verbs in **-sk$^{e}/_{o}$-* regularly have zero-grade in the root.

Pinault (2020: 327 f.) suggests a different etymology: he sees PT **speltke* as deriving from the verb PT **spaltka-* (TB *spalk-*, TA *spāltkā-*) 'make an effort'. This verb is usually considered as denominal (e.g. Peyrot 2013: 837[991]), but Pinault rather views the noun as deverbal (2020a: 327). He sees the *-tk-* sequence in this verb as deriving either from an earlier (Pre-PT) **-t-sk-* < PIE **-d$^{h}h_{1}$-sḱé/ó-* (2020: 328 f.). According to Pinault (op. cit.), the part **spel-* of the verb derives from Proto-Indo-European **spol-*, itself from **sep-* 'to handle (properly), hold' with an *-(e)l-* enlargement, and thus a cognate of Latin *sepeliō* 'to bury' and Vedic *saparyáti* 'to serve ritually, to serve, to honor (a god)'.

Although this is tempting semantically, as one can easily 'serve with zeal', and the notions of 'zeal' and 'service' are often interconnected, Pinault's etymology requires many steps morphologically. This is not impossible, of course, but I would like to propose a solution that I believe to be more straightforward.

If we assume that PIr. **rd* regularly becomes **ld* in Old Steppe Iranian, we cannot only connect *speltke* to Proto-Iranian **spr̥daka-*, but it also explains two other etyma, namely *melte* and *welke** (s.v.). Since there are several cases in which Proto-Iranian vocalic **r̥* seems to have become Old Steppe Iranian **ar* (see p. 180 f.) and short **a* is rendered as Proto-Tocharian **e*, I posit an Old Steppe Iranian **spaldaka-* as the regular outcome of PIr. **spr̥daka-*. This Old Steppe Iranian word **spaldaka-* was borrowed as PT **spelteke*, regularly syncopated to **speltke* (on the syncope, see section 2.7.2.7).

We can assume that the meaning of OSIr. **spaldaka-* was 'zeal, effort' in view of, for instance, Avestan *spərəδa-* 'Eifer', coinciding exactly with the Tocharian meaning. This etymology is therefore straightforward as far as the semantics are concerned. Furthermore, there are parallels for a noun borrowed from Old Steppe Iranian from which a verb was derived in Tocharian (for example, s.v. *raitwe*). This was also the case with Khotanese borrowings, see for instance the verb TB *krasa-* TA *krāṣäyñ-* 'to annoy, vex' or 'to be annoyed' which Dragoni (2023: 114 ff.) analyzes, convincingly in my view, as denominal verbs derivating from TAB (or PT?) *krāso* 'torment'.

In conclusion, while Tocharian B *speltke*, Tocharian A *spaltäk* 'zeal, effort' might theoretically be inherited as suggested by Pinault (2020), it seems that an Iranian origin is also possible, and simpler. However, one has to abandon a Bactrian derivation, as per Winter (1971: 219) and view this word as an Old Steppe Iranian borrowing in Tocharian.

2.2.27 *TB* tsaiñ̈ñe *'Ornament'*

It was Schmidt (1985: 763 f.) who proposed for the first time that Tocharian B *tsaiñ̈ñe* 'ornament' is borrowed from Old Iranian, i.e. Old Steppe Iranian in my terminology. It would thus be a cognate of Avestan *zaiia-* m. 'instrument; weapon', Khotanese *āysän-* 'to adorn, equip'.

In my view, it could be derived from **dᶻaᵢ̯-na-i̯a-* with syncope of the second syllable. It would thus be an adjective 'what belongs to equipment/instrument', possibly designating small ornaments of weapons, or, more straightforwardly 'equipment, adornment' > 'ornament'. Old Steppe Iranian **dᶻaᵢ̯-na-i̯a-* could perhaps indeed yield PT **tseyñ̈ye* which would, in turn, become (Classical) Tocharian B *tsaiñ̈ñe*.

Another solution, suggested by Schmidt (op. cit.), consists in taking **dᶻaᵢ̯a-* as having been borrowed from Iranian as PT **tseye*, to which the adjectival suffix *-ñ̈ñe* was added in Tocharian, yielding regularly *tsaiñ̈ñe*. This would thus originally have meant **'of an ornament, of ornaments' > 'ornament'. In both cases, the formal and semantic details remain not entirely clear, and more work is needed to explain further details of the Tocharian B word *tsaiñ̈ñe* 'ornament'.

In conclusion, Tocharian B *tsaiñ̈ñe* 'ornament' is probably borrowed from Old Steppe Iranian. No etymology from Proto-Indo-European has been put forward to this day, and the element *tsai-* has a very Old Steppe Iranian appearance. There is no immediate morphological comparandum for it in Iranian, although the root it would derive from is well attested among Iranian languages.

2.2.28 *TB* tsain *'Arrow'*

Tocharian B *tsain* 'arrow', plural *tsainwa*, was recognized as being of Iranian origin since Schmidt (1985: 763). In Tocharian texts, it corresponds among others to Sanskrit *iṣu-* 'arrow' (M-W: 168).[46] It is for instance (THT 46 a2; THT 47 a6) used in the expression *swese tsainwaṣṣeṃ* "rain of arrows" (lit. "arrowey rain").

Tocharian B *tsain* probably comes from Proto-Tocharian **tseyn* which in its turn derives from Old Steppe Iranian **dᶻaᵢ̯nu-*, probably meaning 'arrow', cf. Avestan *zaēna-* m. 'weapon', YAv. *zaiia-* 'kind of weapon', Av. *zaēnuš* which means 'baldric'. The form **dᶻaᵢ̯nu-* underwent regular loss of the final *-u*, either in Old Steppe Iranian or, more likely, in Proto-Tocharian, as a trace of its final **-u* is possibly found in the plural *tsainwa* and in the derived adjective *tsainwaṣṣe* see section 2.7.2.6 (cf. also Peyrot 2018a: 270).

46 Tocharian B *tsain-yamäṣeñca* could theoretically correspond to Sanskrit *iṣukāra* 'arrow-maker' as suggested by Adams (DTB: 811), however the manuscript cited by Adams (THT 1318 a2) is too lacunary and the word *tsain* is not even certain.

Tocharian B *prere*, Tocharian A *pär* and Tocharian B *āstär* (borrowed from Sanskrit *astra-*) also designate arrows. These words do not represent different types of arrows: *tsain* and *āstär* correspond to Sanskrit *iṣu-* (see above) and both designate an arrow. It was previously believed, following Couvreur 1964 (so DTB: 64) that TB *āstär* designates Māra's weapon, as opposed to *tsain*, which would be the Buddha's. This is incorrect, both words simply mean 'arrow', whether used by an army, by the Buddha, or by Māra.[47]

To note, Buyaner (2021: 75 f.) has suggested that this word derives from Early Middle Chinese *tsian*^{h} 箭 (Modern Mandarin *jiàn* 'arrow'), cf. Pulleyblank (1991: 148). He writes (loc. cit.) that the Iranian reconstruction **d^{z}aḍn-* does not have a firm basis, but also insists on the fact that the Tocharian word strictly means 'arrow', and not 'weapon'. However, the metathesis he suggests (*ia* > *ai*) is arbitrary, and the Old Steppe Iranian etymology is therefore to be preferred.

In conclusion, Tocharian B *tsain* 'arrow' comes straightforwardly from Old Steppe Iranian **d^{z}aḍnu-*. For a discussion of the semantics, see section 2.7.4.4.

2.3 Old Steppe Iranian Borrowings: Possible Cases

The etymologies presented in this section do function well as such, or seem plausible, but they lack a crucial element, which makes them more hypothetical than the ones presented in the previous section. For instance, no trace of a reflex of the suffix *-iya* of *wertsiya* is found in Iranian; there is no trace in Iranian of the meaning 'or' or 'neither' for **apa* (s.v. *epe*); the meaning of TB *tsetke* is not known; the *-i-* in *ekṣinek* requires a complex explanation; etc. Not all the etymologies here are equally probable: for instance, while it is hardly a matter of doubt that TB *mot* 'alcoholic beverage' is of Iranian origin, and that Old Steppe Iranian is its most likely donor, it cannot be excluded that at some point an Indo-European etymology will be found for *epastye* 'skillful'.

2.3.1 *TB* ākteke, āktike *adj. 'Wonderful, Astonishing; Astonished', Noun 'Wonder'*

The Tocharian B word *ākteke* and its variant *āktike* (also *āktīke*) occur quite often (28 times) in Tocharian texts and its meaning is relatively clear: as an indeclinable adjective it means 'wonderful, incredible', or, sometimes 'astonished, bewondered'. As a noun, it means 'wonder, astonishing thing, surprise'.

47 I thank Professor Pinault for pointing this fact out in his review of my dissertation.

On the details of the semantics of the adjective and the noun, see Pinault (2015: 170–172). TB *ākteke/āktike* is indeclinable as a noun.[48] The form *ākteke* is older than *āktike*, as shown by Peyrot (2008: 171–172). The etymology of *ākteke* has been considered uncertain by Adams (DTB: 40).

Hackstein, Habata & Bross (2019: 24) suggested that *āktike* derives from Parthian *'gtygys'*. However, this is most likely an inadvertent mistake. As Chen (2019: 206) rightly points out, such a Parthian form is not attested and *'gtygys'* is simply the transcription of TB *āktikesa* (or *āktekesa*, perhaps?) in the Manichaean script.

Apart from this, the most recent proposal put forward concerning the etymology of this word is Pinault (2015: 171–172). He writes that the initial accent (‹ā› = /á/) "proves that it is not a single word, but the univerbation of two words *ā* + *kteke*." This argument relies on the fact that Tocharian B stress is usually on the second syllable from the left, but on the last syllable of the first member of compounds. In his opinion, this *ā* is identical to the interjection TA *ā*, *hā*, TB *hā* 'O woe!'. As for "*kteke*", he suggests it is ultimately derived from a particle **kté*. He writes:

> **kté*: < **käté* < CToch. **kwätœ*, an interrogative used as an exclamative and otherwise found in the interrogative adverb TB *mäkte* 'how' < **mœ-kwätœ* [...]. (2015: 172).

Pinault further proposes that *kteke* represents a reduplicated form **kte-kte* dissimilated to *kteke*. Basically, the semantic development which he implies is *'oh! what' > *'what' > 'wonderful, astonishing' etc.

I could not find a parallel in any language for a development from '(oh) what' to 'wonder', and Pinault gives none either, although one could imagine that it could have passed through an intermediary *'surprise' (< *'a (thing about which you say) what!'). Nevertheless, some problems can be found in Pinault's etymology, notably that 'O woe' would hardly make sense in the compound that Pinault proposes.

If the first element of *ākteke* corresponds indeed to *hā* in Tocharian B, perhaps one could expect at least one occurrence spelled as †*hākteke*, given the high number of attestations of this word. There is also a chronological problem: the second element must be of an early date, since **kté* is not found as

48 It does not have a plural form. TB *ākteke* can naturally take secondary case endings, such as the perlative, but does not have a primary declension, and remains identical in the nominative and the oblique.

an analyzable morpheme in attested Tocharian, so that the reduplication and the dissimilation proposed by Pinault must have taken place at least in Pre-TB (if not in Proto-Tocharian or before), while the first element is admittedly "imitative of Skt. *hā*" (Pinault 2015: 171).

I propose another etymology, which accounts for the fixed initial stress as well as for the structure and meaning of this word: *ākteke*, in my opinion, goes back to Old Steppe Iranian **ā-gata-ka-*, a *-ka-* derivate built on the past participle of OIr. **ā-gam-* 'to arrive', **ā-gata-*, cf. Khwarezmian *āγadik* 'what has come' < **āgataka-*; Sogdian *ʾγʾtk* /āγatē/ 'having come; brought' < **āgataka-*. As to the stress, it is on the first syllable, as is usually the case with most secure Old Steppe Iranian loanwords.

I propose the following semantic path to account for the Tocharian B meanings 'wonderful' and 'wonder': 'what has arrived, happened' > *'what has appeared, happened' = '(out of the ordinary) apparition, (unexpected) happening' > 'a wonder, a miracle'. The original meaning is found in words such as Khwarezmian *āγadik* 'what has come' < **āgataka-*, see Benzing (1983: 37; cf. Khwarezmian *ʾγd* f. '(festgesetzte) Zeit' < **āgatā*). For the meaning 'apparition', one can look at Parthian *āγad* 'to appear'; cf. Sanskrit *āgata-* 'newcomer, guest' (M-W: 129). One can think of the French meanings of *apparition*, which can describe the coming of somebody or something, but has a supernatural, magical use, describing, for instance, the sight of a ghost, or that of a divinity or demon.

Another possible argument in favour of the present hypothesis is found in Pashto *aγə́lay* 'handsome, pretty', cf. Morgenstierne (2003: 8) "***ā-gataka-*, with obscure semantics". The semantics could be *'wonderful' (as in Tocharian) > 'handsome, pretty', with a quite usual decrease in semantic emphasis. Cheung (2005: 128) derives it rather from **ā-gu(H)-ta-ka-* 'desirable, agreeable' (Late Khotanese *hagav-* 'to long'), which is possible, of course, although this etymon does not display similar semantics in the rest of Iranian.

It seems to me that the hypothesis that TB *ākteke* 'wonder, wonderful, astonishing' derives from OIr. **ā-gata-ka-* 'wonder' < 'what has appeared' is strengthened by the meaning 'to appear' and 'what has appeared, appeared thing' taken respectively in Parthian and in Khwarezmian by the words **ā-gata-* and **ā-gataka-*.

This word shows initial stress, which is remarkable for trisyllabic words in Tocharian B. As I argue in section 2.7.2.7 of this chapter, this initial stress is regular for Old Steppe Iranian loanwords (with some exceptions, which I think can be explained). The syncope of **ā-gata-ka-* to PT **akteke* is also expected: in early Iranian or BMAC loanwords with three times the same vowel, such as *a_a_a* or *e_e_e*, Proto-Tocharian reduced the syllable that was after or before the stress, cf. section 2.7.2.7.

2.3.2 TB ekṣinek* *'Dove'*

Tocharian B *ekṣinek** 'dove' is a hapax, occurring in the form *ekṣinekaṃñana* 'pertaining to a dove' (DTB: 81). It is viewed, since Schwentner (1956: 238), as an Iranian loanword in Tocharian, corresponding to Ossetic *æxsinæg*, Khotanese *aṣṣänaka-* 'dove', *āṣṣeiṇa-* 'dark blue'. However, the idea that it is a Middle Iranian loanword (DTB: 81) should be abandoned because of the vowel correspondence Iranian *a*: Tocharian *e*, which rather indicates an early borrowing.

The problem with this etymology is that the phonological shape of the Iranian etymon does not fit exactly. The proto-form is **axšai̯na-* 'dark blue' with the **-ai̯na-* suffix of color names (on the etymology of this word, see Garnier 2019: 368). From this, **axšai̯naka-* 'dark blue bird' was derived, like New Persian *kabōtar* 'dove' ← **kapau̯ta-* also 'dark blue'. As we have seen elsewhere, the Old Steppe Iranian diphthong **-ai̯-* remains as such, including in the position in front of **-n-* (s.v. *ainake*).

Since we expect OSIr. **áxšai̯naka-*, **-ai̯-* was probably unaccented in this word, while it was stressed in Old Steppe Iranian **ái̯nāka-*, one could think that unaccented **-ai̯n-* had become **-īn-* in Old Steppe Iranian. However, until confirmation or disproval due to other examples is available, this proposition remains *ad hoc*.

I believe a better solution can be suggested for the phonetics of this word. Kim (2003: 51[20]) proposed the following sound law for Ossetic: Pre-Proto-Ossetic **-ai̯Ci̯a-* > Proto-Ossetic **iyCa*. This could perhaps also have taken place in Old Steppe Iranian, as this word could show. For a more detailed discussion, see section 2.7.2.8.

A last point is that the meaning of the hapax is not assured: its interpretation by Schwentner (loc. cit.) is based on its possible Iranian etymology. The context is *ekṣinekaṃñana misa* "meat of an *ekṣinek**". With Schwentner, we have to admit that it is difficult to see what else it could be.

In conclusion, TB *ekṣinek**, probably meaning 'dove', is likely of Iranian origin. If so, an Old Steppe Iranian etymology is most likely because of the Tocharian *e* corresponding to Iranian *a*. However, the *i*-vowel of the second syllable is not evident to explain. On a possible solution, see section 2.7.2.8.

2.3.3 TB epastye *'Skillful'*, TA opäśśi *'Id.'*

The Tocharian B word *epastye* 'skillful, adept, capable, able' and its Tocharian A cognate *opäśśi* 'skillful' derive from Proto-Tocharian **epəstye*. Various etymological attempts (listed in DTB: 94)[49] suggest a stem **-sth$_2$-* 'to stand' for the

49 To which one ought to add Pinault (2008: 451).

second syllable and a prefix *h_1op- or *h_1opi- for the element *ep-*. It would thus be 'what stands by' or something similar.

This morphological analysis is difficult to accept if one considers the fact that no other Tocharian word contains an Indo-European prefix, except for the negative PT **en-* and the locatives PT **yən-* and **en-*. This would be the only such example, which is uneconomical and unlikely.

Rather, I propose to see this word as a borrowing from Old Steppe Iranian. An Old Steppe Iranian adjective **abí-st(H)-i̯a-* 'relative to excellency' could have easily been made on the basis of the noun **abí-st(H)a-* *'standing over, in front' > *'mastership, excellency', in a parallel way to **abi-staH-ta-* 'standing (lit. stood) in front' > Middle Persian *awistād, ōstād*, New Persian *ustād* 'professor, master' (cf. Horn 1893: 20),[50] and thus referring originally to the position of the teacher. There was no RUKI-effect on **abi-staH-* because it was extracted analogically from unprefixed forms, and thus **abi-staH-* does not come *recto itinere* from Proto-Indo-Iranian.

The adjective would originally have been **abí-st(H)-i̯a-*, with **abí-* carrying fixed accent on the *-i-*, yielding Proto-Tocharian **epə́stiye*. This proto-form **epə́stiye* became Tocharian B *epastye* (< Pre-B **epəstiye*), cf. Peyrot (2021), and Tocharian A *opäśśi* (< **opästyi*).

The meaning of this adjective would have been 'master-like; excellent', and once it was made into a noun, it would have meant 'excellency, mastership', which fits the Tocharian meaning quite well. The form fits perfectly as well. For these reasons, and the fact that it is more difficult to assume a single word going back to a prefixed form in Tocharian, than to view this (cultural) word as a loan, I believe it is indeed another example of a borrowing from Old Steppe Iranian into Tocharian. An obvious drawback to my derivation is that the supposed formation is not so far attested anywhere in Iranian.

In conclusion, Tocharian B *epastye* and Tocharian A *opäśśi* both meaning 'skillful', may be seen as regularly deriving from an otherwise unattested Old Steppe Iranian adjective **abí-st(H)-i̯a-* 'relative to mastership, to excellency', which derivation would fit both the semantics and the phonetics of the Tocharian forms.

2.3.4 *TB* epe *'Or'*

It is agreed upon that the Tocharian A and B word *epe* means 'or' (see for example TEB II: 85). However, the etymology of this word has remained unclear until

50 The form **au̯a-st(H)a-ta-* suggested by Horn (op. cit.) would not have yielded the variant *awistād* (*awestād* according to DMMPP: 76), which variant was not known to Horn.

now (see DTB: 95). The only Indo-European etymology provided for this word by Abaev (1965: 137), who connects it to Ossetic *ævi* 'or, either' and derives both from Proto-Indo-European *$*h_1$ou̯ē*, is not possible since that preform would yield Tocharian B †*eye* (TA †*ay*), as mentioned by Adams (DTB: 95).

I believe that Tocharian B *epe* was borrowed from Old Iranian **apa* 'away from', and more specifically from Old Steppe Iranian **apa* which I suppose had developed a meaning 'neither, nor'. Tocharian A *epe* is a borrowing from Tocharian B. If I am correct, this etymology has important relevance to the interpretation of Iranian–Tocharian contacts. A part of my argumentation is based on my analysis on the difference of function between *epe* and *wat*, which has hitherto not been studied.

2.3.4.1 Examples of Tocharian A and B *epe*

It is important to specify the meaning of Tocharian A and B *epe* in order to establish its etymology. In order to do so, I will give below as many examples of its use and meaning in Tocharian texts as I could find. Indeed, conjunctions and disjunctive particles such as 'or' can have multiple nuances depending on their specific discursive or syntactic function. These nuances can be used in order to suggest a plausible etymology, since usually conjunctions evolve fast in use and meaning, but at various stages, their specific evolution is triggered by their previous function.

Excluded from these examples are incomplete sentences where the first or second element introduced by the disjunction is unknown, and, naturally, very fragmentary manuscripts. In particular I will try to show that *epe* represents originally an exclusive conjunction and not an inclusive one. I will compare it to the use of *wat*, which is also generally translated by 'or', with a few selected examples.

The difference between exclusive and inclusive conjunctions is explained in more detail below (p. 84 f.; cf. also Hurford 1974). An exclusive conjunction, or rather disjunction, expresses the impossibility of the coexistence of two logical elements: *X is a man* ***or*** *a woman* (one cannot be both). An inclusive conjunction, in most basic terms, says that either two things are possible, or only one. I discuss these two concepts in greater detail below.

In the examples below, provided with accepted translations when available, I will endeavour to analyze whether the meaning given by the Tocharian conjunction in bold has an exclusive or inclusive function. All these examples, when clear, are exclusive: for instance in A 6 a2–3, where it is not possible that the mechanical girl is both the sister, and the wife, etc.—only one of these alternatives is possible. In THT 107, the alms can either be given to the speaker, or to someone else, but not both, etc.

Tocharian A

A 6 a2–3

kuss aśśi sās yaṃtrācā(*res mä*)*skatär ṣar ckācar* ***epe*** *śäṃ* ***epe*** *spaktānik* ***epe*** *nṣäkk oki lokit kakmus näṃ*

"Who may she be? Is she the sister, the daughter, **or** the wife, **or** the servant of the mechanic, **or** has she come to him as a guest, just like me?" (Peyrot 2013: 282).

"Qui est donc celle-ci? Est-elle la sœur du mécanicien, **ou bien** sa fille, **ou bien** son épouse, **ou bien** sa servante? **Ou bien** est-elle venue (ici) en hôte exactement comme moi?" (Pinault 2008: 258).

A 342 b2–3

mā te[51] *nātäk caṃ brā*(*maṃ*) *e*(*pe*) *mā ne was entsaträ*

"**ob** der Herr diesen Brahmanen **oder** uns behalten mag" (TEB II: 124).

Tocharian B

Concerning the next example (and the previous one as well), one could think that the conjunction could also be inclusive, in that the lord could possibly keep either the Brahmin or the speakers. But in the context, it is unlikely, and the lord rather has a choice to make concerning who he should keep.

THT 79 a5

(*e*)*pe saswe wess eṅträ epe brāhmaṇeṃ mā ra tsak wes co*(*mp*) ///

"**Whether** the lord keeps us **or** the Brahmin, we are, at any rate, not able, either, to (endure *vel sim.*) that one …" (Melanie Malzahn in CEToM)

"**ob** der Herr diesen Brahmanen **oder** uns behalten mag" (TEB II: 124).

The next example is unfortunately too obscure in order to make a secure judgment, but one could imagine that the *atka*-capacity and the *ṛddhi*-power are two incompatible powers that cannot be used at the same time by the same person. In any case, the *ṛddhi*-power is a strong supernatural power, whereas *atka*, whatever it is, designates here a capacity, rather than a strength.

IOL Toch 178 b7

/// (*po*) *yente kärkāte* (*c*)*w*(*i*) *no tsakṣträ ñke asāṃ · atkane tu cämpamñe* ***epe*** *maiyya räddhiṣṣa* [...][52]

51 Cf. TEB II: 35 for the restoration as (*t*)*e* (CEToM has *ne*).

52 For the restoration of the beginning of the line, I follow Schmidt (1974: 400).

"… he stole the wind and his throne burns; in concentration (?) [is] that power of magical strength" (Adams DTB: 10)
but I think the following translation is better:
"the wind carried all of this (?) away, but now his throne burns. In this *atka*-capacity **or** (with) *ṛddhi*-power …"

As Dr. Athanaric Huard (p.c.) informs me, this text has a parallel in Buddhic literature, namely, Śrāvastī's miracle, where the Buddha defeats heretics through magical powers which they cannot outmatch.

IOL Toch 305 b4–5
kwri no mā yāmtär vaisarp · warkatsa · praśśāṅkaṃ erṣäṃ · epe bhagaṇḍälänta mäskenträ tärrek
"But if one does not do [it], [this] causes erysipelas, hydropsy and *praśśāṅkaṃ*, or [he has] haemorrhoids, is blind […]" (Peyrot 2013: 676).

THT 107 a8–9
*ñi aiścer ce pinwāt **epe** se ññissa śpālmeṃ tākaṃ cwi aiścer*
"[…] do you give those alms to me, **or** do you give it to somebody who is better than me?" (Peyrot 2013: 702) "[…] est-ce à moi que vous donnez cette aumône-ci? **Ou bien**, celui qui d'aventure sera supérieur à moi, est-ce à lui que vous donnerez?" (Pinault 2008: 157)

THT 107 b1
bram-ñikte weñā-meś ṣerśkana: se ñisa śpālmeṃ rṣāke tākaṃ cwi aiścer epe tuwak ñi aiścer
"Le dieu Brahman leur dit: « sœurettes, celui qui d'aventure sera un sage supérieur à moi, est-ce à lui que vous donnerez? **Ou bien**, est-ce à moi que vous donnerez justement ceci?" (Pinault 2008: 157).

THT 107 b8
*ṣerśkana se nomiyeṣṣe bhājaṃ rerinu star-me **epe mā***
"Sisters, is this jewel bowl left by you [to me] **or not**?" (Peyrot 2008: 123)
"Sœurettes, ce bol de joyau, (à moi) est-il laissé par vous **ou non**?" (Pinault 2008: 158).

W 15 a5
*satkenta **epe** puṣne **epe** rohinikene satkenta waltsa*(*na*)
"… remèdes **aussi** en cas de *puṣpa*, **aussi** en cas de tuméfaction à la base de la langue; [sont] les remèdes à combiner …" (Filliozat 1948: 83).

In clearer language, it could be translated as
"… remèdes en cas de *puṣpa* **ou** en cas de tuméfaction à la base de la langue". Here Filliozat uses *aussi* in a quite obsolete meaning (similar to 'or'), because it can be repeated twice in a more beautiful manner than *ou*.

PK AS 3B b3

ṣpel gairikäṣṣe ṣpel te śār kätnālle ***epe*** *no kroścana tonak śār lupṣallona · toṃ prayokänta y*(*amaṣṣällona*) ///
"… pellet, saffron pellet: this is to be strewn over [it], **otherwise** these very same [ingredients] are to be smeared cold over [it]. These treatments have (to be made as)" (CEToM).[53]

2.3.4.2 Discussion on the Origin of 'Or' in Various Languages

In many languages an exclusive conjunction 'either, or' is differentiated from an inclusive conjunction 'or': for instance, in Latin exclusive *aut* is opposed to inclusive *vel*. Among other Indo-European languages, there is for instance Latvian exclusive *vai* and inclusive *jeb*, next to exclusive *vai* … *jeb*.

The difference between an inclusive conjunction and an exclusive one can be seen from a set of examples given in Hurford (1974). "Ivan is an American or a Russian"; "That painting is of a man or a woman" are exclusive, because they naturally exclude each other, it is *either* an American *or* a Russian, *either* a man *or* a woman. On the other hand, "Inmates may smoke or drink" is inclusive, because inmates may do both.

In the cases of the examples given above, we can say that A 6 a2–3, and the following examples are exclusive, whereas PK AS 3B b3 is inclusive. In A 6 a2–3 the mechanical girl can necessarily be either a sister, a daughter, a wife or a servant. Her being one of those things excludes her being any of the others. In A 342 b2–3 and THT 79 a5 the lord can either keep the people who speak, or the brahmins, but not both. In THT 107a 8–9 the alms can only be given to one person: either the speaker, or someone who would be superior to him; likewise the example THT 107 b1. In THT 107 b8 the jewel bowl can be left either to the speaker or to someone else, and *epe mā* is clearly exclusive. In W 15 a5, the remedies can work for either disease, but *epe* retains an exclusive function, in the sense that it is only expected for a diseased person to have either *puṣpa* or tongue tumefaction, in any case even if both can be had at the same

53 Compare Filliozat 1948: 52 "boulette …, à … dans une boulette d'orpiment aussi, donc, ces choses sont bien à appliquer froides, ces moyens …"

time. In PK AS 3B b3 the ingredients can either be smeared cold or perhaps smeared after having been heated? The context seems to be exclusive, but it is unclear.

As mentioned above, many languages differentiate between inclusive "or" and exclusive "or". In many other languages, however, the inclusive "or" should rather be referred to as a neutral "or", as it can take an exclusive function in some cases.

Such is the case of English *or* (neutral, inclusive), opposed to the strictly exclusive *either or ~ either* (exclusive). If I say "here, you can drink either coffee or tea" it clearly indicates that only one of the two options is possible. If I say "here you can drink coffee or tea" it sounds more as if, in this place, both options are available. In English, there are also uses of "or" which can be exclusive (cf. Hurford 1974), such as in the examples cited above (e.g. "Ivan is an American or a Russian"; "That painting is of a man or a woman").

One can cite similar examples in other languages: French neutral, inclusive *ou* vs. exclusive *ou bien* (see Meillet 1921: 164), where *ou* can also be exclusive, and should thus be primarily defined as neutral. One can also think of Contemporary Persian neutral, inclusive *yå* vs. *o yå* lit. exclusive 'and or'. In Tunisian Arabic, the neutral conjunction *wala* 'or' is found alongside exclusive *ou bien*, borrowed from French.

I will describe a functional path taken by a number of disjunctive conjunctions in world languages. I do not pretend that this is the only possible path, or that it is universal, but it is observed in a number of languages, and I propose to reckon Tocharian A and B among them.

(0) A neutral disjunctive morpheme exists. (1) An exclusive conjunction is created, because emphasis on exclusion is felt as pragmatically necessary. It is generally created by (a) the addition of morphemes, (b) the repetition of the neutral morpheme in two places in the sentence, or by (c) grammaticalization processes, and (d) borrowing can occur in cases of strong cultural contact. As Meillet (1921: 169) writes "[i]l n'y a pas d'espèce de mot qui ne puisse livrer des conjonctions." Afterwards, (2) the exclusive conjunction is weakened due to repetition (cf. Meillet 1921: 164 ff.; 169 f.), becomes neutral, and a new exclusive conjunction is created, using one of the processes described above. It becomes weakened in its turn (3) etc.

(a) English *either* is for instance derived from Old English *ægðer*, itself from *a* 'always', *ge-*, a collective prefix, *hwæðer* 'whether'. It originally means 'one of both'. French *ou bien* is constituted of *ou* 'or' + *bien* 'well', as is Dutch *ofwel*.[54]

54 I thank Dr. Benjamin Suchard for reminding me of the Dutch form. He also suggested it is a calque from French, which is very likely according to Dr. Peter-Alexander Kerkhof (p.c.).

Contemporary Persian *yå* derives from Cl. Persian *yā* itself from Middle Persian *ayāb*. The etymology of the latter is disputed, but it is generally admitted that it derives from the addition of three morphemes, **aδa-* 'then' **-u̯ā* 'or' and a third one, which is sometimes derived from **pi-*, an emphatic particle (e.g. Bartholomae 1916: 38[9]; Back 1978: 199).[55] It is probable that this conjunction was formed to be exclusive, opposed to **-u̯ā*, which was neutral. Now that it is neutral, the exclusive conjunction *oyå* has been formed in Contemporary Persian from *-o* 'and' and *yå* 'or'.

(b) In most languages, it is always possible to express exclusivity by repeating the neutral or inclusive 'or' before each proposition. For instance in French "tu prendras ou du thé, ou du whiskey"; "ce livre est ou à toi, ou à moi": there is no doubt here that one has to consume either tea or whiskey, and that this book is either yours or mine. In these examples *ou … ou* corresponds and could be replaced by a single use of *ou bien*. Some languages combine their exclusive conjunction with the neutral one in this type of construction, such as English *either … or*.

(c) Grammaticalization processes in order to create a new exclusive conjunction vary across languages. One of them consists in deriving an exclusive conjunction from a word meaning 'one of both' (as in English *either*, see above). Another grammaticalization process leading to the creation of an exclusive conjunctive particle consists in the use of a particle meaning 'away from' to express disjunction. This is the case of Latin *aut* 'or (exclusive)' which ultimately goes back to Proto-Indo-European **h₂eu̯t-* 'away' (cf. de Vaan 2008: 64). This would also work if my etymology of Cl. Persian *yā* and Middle Persian *ayāb* as containing an element **apa* 'away from' were accepted.

(d) The borrowing of an exclusive conjunction 'or' is not very common, but has been noted to occur in cases of strong cultural contact. One can cite Latvian exclusive *vai*, borrowed from Livonian *või* or dialectal Estonian *vai* opposed to neutral *jeb*, which is inherited (cf. Karulis 1992: s.v.). One can also cite Tunisian Arabic exclusive *ou bien* from French exclusive *ou bien* opposed to inherited and neutral *wala*.

(2) In the second step of this process, the exclusive and inclusive conjunction coexist. Generally the inclusive or neutral conjunction is felt as being "too weak" and loses little by little its importance in the language, before being replaced by the exclusive one, which becomes more neutral. (3) Another exclu-

55 In my view, it is equally possible to suppose **aδa-* 'then' + **au̯a-* 'off, downwards' (or, indeed **-u̯ā*) + **apa* 'away'.

sive conjunction is created, it coexists with the previous exclusive one become neutral, and ultimately replaces it, and so on and so on.

This scenario is entirely schematic, and of course should be nuanced. It is possible for a language to have multiple conjunctions at various levels of semantic development. Besides internal hesitations within the language, dialectal data can add to the variety in the use of conjunctions. I will try to examine the Tocharian data in order to reach a conclusion as to the state of disjunctive particles in Tocharian A and B.

2.3.4.3 The Case of Tocharian

In Tocharian A and B there are three ways to express exclusive or inclusive disjunction: 1. *epe*; 2. TB *wat* TA *pat* and 3. TB *epe wat* TA *pat nu* and possibly TA *epe pat*.[56] Because the three are part of the same system, and it is impossible to analyze a conjunction on its own, especially if one has etymological aims, I will also shortly discuss the function of TB *wat* TA *pat*. Afterwards I will discuss the function of TB *epe wat* and TB *wat no* in a few examples. This will lead to a more elaborate discussion on the disjunctive conjunctions of Tocharian A and B, their prehistory, and naturally the origin of the conjunction *epe* in Tocharian A and B.

TA *pat* and TB *wat*

Tocharian A

A 5 a4–5

täpreṃ ats pälkäṣ mäṃ(tne) tsekeṣi pekeṣi ***pat*** *arämpāt*[57]
"then it looks exactly as if it were a fashioned **or** painted figure." (Peyrot 2013: 263).

A 10 a1

mā nu caṃ täm knā(nmu)neyäṣ wärcetswātsuneyā ṣñi ***pat*** *ālu* ***pat*** *täm ṣurmaṣ pruccamñe ya(ts)i*
"mais, du fait de la déficience de sagesse, pour cette raison, elle ne peut procurer absolument aucun avantage, **ni** pour soi, **ni** pour autrui." (Pinault 2008: 262).

56 TA *epe pat* is only found in an extremely fragmentary context (in A 428) so that its meaning cannot be ascertained.

57 I follow the restoration as presented in Peyrot (2013: 263).

A 59 b3–4

(*tā*)*pärk ṣñi wärpāluneyaṃ* ***pat*** *āssi ālu eluneyaṃ* ***pat*** *nā*(*täk yärm*)
"À présent le seigneur [est] l'autorité, **soit** [qu'il opte] pour sa jouissance, **soit** dans le don [de la jeune femme] aux autres" (Meunier 2015: 56).

A 275 a5

śāt ekro ***pat*** *wrasom tāṣ*
"if this was a rich **or** a poor being" (CEToM).

Tocharian B

IOL Toch 306 b3–4

mā walkeññe päst yā(*mormeṃ*) [...] *yāmormeṃ istak* ***wat*** *päst yāmormeṃ · istak* ***wat*** *prekesa yāmorme*(*ṃ*)
"Having not moved for a long time ... (the tumour (?)/imagined foetus (?) suddenly moves with pain (?))[58] ... having (moved) then suddenly having moved again, suddenly again for a time having moved, (through the accumulation of) *vā*(*ta*)" (Carling 2000: 91).[59]

PK AS 3A a1–2

läksañana misa lykaśke kekarśwa tsatsāpauwa ampoññaṃtse sāṃtke kaṃnte kältsau ṣalype ***wat*** *malkwersa* ***wat*** *pissauṣṣe war* ***wat*** *śār kuṣalle*
"Fish meat finely chopped [and] crushed [is] a remedy against abscess. Oil pressed a hundred [times] **or** [oil] with milk **or** aneth water has to be poured over [it], in case of a gall [boil]" (CEToM).

PK DA M 507.37 and .36a.83–85

saswa ce we[*s*](*s*)*i* (–*Pu*)*ttewante snai paille yāmu ste ce* ***wat*** *wäntare kakāmau ste sū* – (–) *nau*[*ṣ*]*ameṃ papaikau ste*[60]
"Oh lord! The (things that) Puttewante has done lawlessly [to us], **or** the things that have been taken (by him), they have been written above" (adapted from Ching 2010: 215).

---

58 This part corresponds to a restoration done by Carling (2000) on the basis of textual parallels.

59 Cf. Carling (2000: 86) "[...] [it] refers to the movements of the tumour: It does not move for a long time, and thereupon it pulsates like a limbless foetus."

60 I follow the transcription given by Ching (2010: 215), from which I removed the punctuation which she added against the manuscript.

IOL Toch 307 a6–b1

paiykalñesa vācavārg waiyke reki mantanta läññi-ñ (k)oynm(e)ṃ (:) k$_u$cesa ṣañ-añmä karstoymar alyekepi kaloym ***wat*** *mīyäṣṣälyñe*

"May through the writing of the Vācavarga never go out from my mouth untrue speech by which I might destroy myself **or** bring harm upon someone else." (CEToM).

PK AS 2B b5–6

kewiye miśosa malkwersa ***wat*** *nastukārm niryuhaṃ yamaṣle*

"cow urine **or** milk, a nasal injection [or] an enema has to be made." (CEToM).

PK AS 5A a5–6

(wewe)ñor ārtar wat no: ṣe keklyauṣor eñ(c)imar ārth vyajantsa ṣañ yk(nesa)

"May I seize … what (has been) said or what has been praised at the first lecture [lit. hearing] with meaning and literary form (according) to the proper manner. …" (CEToM).

PK AS 6B b1–2

nandeṃ ālts(i)ś pūdñäkte ṣ(amā)neṃ mäntrākk= ālyeṅkä(ṃ): wertsiyaine orotsai ***wat*** *w(e)ña ṣkas toṃ ślokanma*

"In order to restrain Nanda, the Buddha lord told these six stanzas in the same manner to other monks **or** in a large assembly" (CEToM).

Reviewing the examples above, one can see that some are inclusive, while some are exclusive. For instance, the example found in IOL Toch 307 a6–b1 is inclusive: one might both destroy oneself and bring harm upon someone else by untrue speech. PK AS 5A a5–6 is also inclusive, since what has been said is not by nature excluded from what has been praised. The presented texts A 5 a4–5; A 59 b3; A 275 a5; PK AS 2B b5–6; PK AS 3A a1–2 are exclusive disjunctive sentences.

The last example, PK AS 6B b1–2 is neither exclusive nor inclusive, the use of *wat* in it is conjunctive, that is, it is equivalent to an "and". Indeed, the Buddha was repeating the same words to monks and to large assemblies. This conjunctive usage of *wat* in this last example (implying a relative neutrality of the particle) is reminiscent of the one in the example in IOL Toch 306b3–4, where it was even left out of Carling's translation. It is interesting to note that we do not have any examples of a conjunctive use of *epe*, which apparently had a much "stronger" disjunctive function than *wat*.

2.3.4.4 TA *pat nu* and TB *epe wat* and TB *wat no*

To complete the global analysis of disjunctive particles in Tocharian, I believe it is useful to consider the function of the following conjunctions briefly as well: TA *pat nu* and TB *epe wat* and TB *wat no*.

A 69 b5

> *näs wrasaśśi klopaṃ (p)āṣlune yatsi **kupre pat nu** mā yātal*
> "**Oder wenn** (ich) nicht imstande (bin), den Menschen im Leid Schutz zu gewähren, …" (Thomas 1954: 741).

A 226 b6–7

> *ke **pat nu** krī tāṣ ñareyäntwaṣ ke **pat nu** saṃ krī ñäkcī suk näṣ kälpīmār: ke pat nu ākāl ñäkci napeṃṣi ā*
> "Who wants to (be freed?) from the hells, **or** who has this wish, "may I obtain divine happiness!", **or** who has the wish, "… human and divine …"" (Peyrot 2013: 268).

Here Peyrot does not render the first *pat nu* because it is not fitting in English. If one were to translate it completely literally it would be: "**or but** who has the wish from the hells, **or but** who has this wish "may I obtain divine happiness, **or but** who has the wish "… human and divine …"

It is not necessary to examine all examples: while every single use of Tocharian A and B *epe* is exclusive, the other conjunctions studied here can be either exclusive or inclusive. They are thus "neutral". According to the theoretical discussion above, it appears that the meaning 'or' in Tocharian B *wat* and Tocharian A *pat* should be older, since it is weakened, compared to TAB *epe* which has a stronger exclusive disjunctive use as a particle.[61] Tocharian A and B *epe* is thus "newer" in the system, which implies it was either borrowed later, or that it acquired its exclusive function later.

2.3.4.5 The Etymology of TAB *epe* 'Or'

According to ongoing research by Jens Wilkens, Hans Nugteren & Michaël Peyrot, the Old Uyghur word *ap* 'nor' (for which see Clauson 1972: 3) was borrowed from Pre-Tocharian B **epe*. I imagine the meaning of this Pre-Tocharian B **epe* could have been 'nor', as in Old Uyghur.

61 Beekes (2011: 249) explains *wat* as deriving from PIE **-u̯e* 'or' + *-t* (what this *-t* is remains unclear to me). The etymology of *pat* remains obscure.

Indeed, the meaning 'or' can derive straightforwardly from the meaning 'nor', identically to Tunisian Arabic *wala* 'or', which derives from *wa-lā* 'and not' = 'nor'.[62] Logically, 'nor' would have evolved towards an exclusive disjunction first, rather than a neutral one (which it then becomes, for instance in the case of the Tunisian word—the same would be assumed for the Tocharian particle). However, Wilkens (2021: 55) translates Old Uyghur *ap* as 'oder … auch, entweder … auch'. In this case, we can assume a direct borrowing from a possible meaning 'or' of Tocharian A **ap* or Pre-TB **epe*. Nevertheless, in case Clauson's translation is to be preferred, we can use the Uyghur word as a demonstration of the functional trajectory of Proto-Tocharian **epe*. Since TAB *epe* is solely found with an exclusive function in all Tocharian A and B examples reviewed above, it is more likely that this exclusive disjunctive function is rather late, for instance of Pre-Tocharian B (and A) origin, so that the particle did not have the time to be weakened in function. Thus, it would be more fitting to see Old Uyghur *ap* as having at least one meaning 'neither … nor' taken from Tocharian.

The etymology of the Tocharian word I consider straightforward: it was borrowed from the Old Steppe Iranian particle **apa*, which originally meant 'away from', as in Avestan and Old Persian, but took on the meaning 'nor' in a conjunctive use.[63]

A particle meaning 'nor' is always, by definition, exclusive, and when it becomes positive in meaning ('or'), it tends to remain exclusive. The change from the negative to the positive meaning might be due because the negative part of 'nor' is also assumed by the verb, or by another particle (such as PT **ma*).[64] Once this exclusive particle exists, there is also a strong tendency for it to become weakened, and thus become neutral. At that point, usually, a new

62 Moroccan, Algerian Arabic (and even some Tunisian dialects) often have this word under the form *ula*. It is regular (*wa* > *u*, e.g. North African Arabic *u* 'and' < *wa* 'id.'), and it is rather the Tunisian form *wala* which is unexpected. However, in the learned variety of Tunisian, it is not unusual to see Classical Arabic phonological sequences reintroduced, one can think of *dawla* 'state' (vs. Algerian Arabic *dūla*). In the case of *dawla*, it can naturally be viewed as a borrowing as well, but this is not as easy with *wala* (one expects **walē* in this case). It is thus better to see, in my view, the form *wala* as a pseudo-Arabic (in the sense of written/Classical Arabic) secondary form to an original *ula*.

63 For the possible presence of **apa* in the New Persian word for 'or', see fn. 55 p. 86.

64 One can, for instance, suggest that, for Tunisian, sentences like *mē temšīš ġadi wala hnē* "do not go there nor here" were reinterpreted as "do not go there or here", thus yielding the positive equivalent: *emšī ġadi wala hnē* "go there and here (if you want)", although etymologically this meant *"go there and not here".

exclusive disjunction should appear, as per the cycle described above.[65] Here, we are at a stage when *epe* is still strictly exclusive, whereas TB *wat*, TA *pat* have become weakened and neutral. This suggests that the passage from 'nor' to 'or' occurred shortly before the first attested texts, which could fit very well with the Old Uyghur data (if Clauson's translation is to be preferred).

A theory that could explain the absence of a reflex of Pre-TA **ap* 'or' is that it never became a disjunction 'or', but had kept its meaning 'nor': TA *pat* having become neutral, the speakers of A borrowed in turn the exclusive disjunctive particle *epe* from Tocharian B.

In conclusion, Tocharian A and B *epe* 'or' derives from Old Steppe Iranian **apa* 'away from' and more particularly, I argue, from its posited meaning 'nor'. It first took on the meaning 'nor', when the word was borrowed from Tocharian by Old Uyghur speakers. Later, TAB *epe* became an exclusive particle 'or', which is already the case in all our attestations.

2.3.5 *TB* ñyātse ~ ñātse, *TA* ñātse *'Distress, Calamity; Danger'*[66]

There is a set of obviously related words: TB *ñyātse* ~ *ñātse* and TA *ñātse* 'distress, calamity; danger'. Of the two Tocharian B forms *ñātse* and *ñyātse*, *ñyātse* (pl. *ñyatsenta*) is more archaic, while *ñātse* (pl. *ñatsenta*) is later (Peyrot 2008: 63–64). The Tocharian A form *ñātse* was borrowed from the later Tocharian B variant *ñātse*.[67] In one bilingual Tocharian B-Sanskrit text (THT 543 a5) it is translated as *īti* (spelled ‹iti›) 'distress, calamity, plague'.[68]

The etymology of this word is generally considered uncertain (DTB: 291). Hilmarsson (1991: 137–139) proposes to connect it to Proto-Germanic **nīþa-*, "a thematic neuter with the approximate sense 'ill-will, hostility, hate'". He cites cognates such as Goth. *neiþ* 'ill-will, envy', OIcel. *níð* n. 'libel, insult', OE. *níþ* 'enmity, hate, combat', etc., and projects those words back to Proto-Indo-European **níH-to-* ~ **néiH-to-* (Hilmarsson 1991: 138). Based on an observation that "[i]n Tocharian the suffix **-to-* was generally supplanted by **-tḭo-* […]", he also proposes that Proto-Tocharian **ñyātse* (Common Tocharian **ñyātsœ* in his notation) goes back to PIE **niH-tḭo-*.

---

65 On the weakening of conjunctions, see Meillet (1921: 164 ff.).

66 This etymology I have discovered on my own, and I have worked independently on it (as well as on *entse*), and presented it at the Tocharian International Conference in Vienna (October 25th–26th 2019). Afterwards, I have discovered that Isebaert has published the same etymology (Isebaert 2019). My etymological discussion differs a little from his in the morphological and semantic detail.

67 The expected Tocharian A cognate would be †*ñāts*.

68 In the same line *snai ñātse* translates Sanskrit *anīti* 'absence of calamity'.

Although Hilmarsson's etymology could potentially work phonologically, apart from the fact that I expect PIE **niH-* to yield PT **ña-*, not ***ñya-*, I do not find it entirely satisfying semantically. The meanings of 'hatred, enmity' and 'danger, distress', although belonging to a similar semantic field, are not evidently connected. Furthermore, *ñyātse* ~ *ñātse* would be totally isolated in Tocharian A and B, having no verbal root or other related form from which they could derive.

After Isebaert (2019), I rather propose to connect Tocharian B *ñyātse* to Parthian *niyāz* 'need, distress' and Middle and New Persian *niyāz* 'need'.[69] However, these forms cannot be the direct source: with its final *-e*, *ñyātse* should rather have been borrowed from the Old Iranian stage, and it presupposes an Old Steppe Iranian **ni-ādᶻah-* or **ni-ādᶻa-*, depending on the exact etymology: if it goes back to a form closer to Proto-Iranian **andᶻah*, then it should ultimately go back to a proto-form with final **-ah*, whereas if its proto-form is closer to Avestan *āzi*, a reconstruction with **-ah* is not expected.

The problem is that no such formation is attested in Avestan, nor can it in any trivial way be posited for Proto-Iranian as the preform of the Middle Iranian forms just cited. Rather, Avestan has *āzi* 'greed, Greed-Demon' (AiW: 343) attested, a deverbal noun derived from the verb *āz-* 'to long for' (< Proto-Iranian **ā-Haǰ-*). Thus, we would need to assume a thematic derivative **āza-* or **ni-āza-* for Old Steppe Iranian, either from the root **āz-* directly, or on the basis of the **āzi-* attested by Avestan.

The word *āzi* and its cognates are generally believed to be cognates of Skt. *ājí-* 'race, competition, fight' and to be derived from **h_2eǵ-* 'to drive, provoke a motion'. The development of the semantics of *āzi* could thus be conceived of as 'to drive' > 'to strive towards' > 'want' > 'greed', which is naturally a possibility.

As an alternative, I suggest that Tocharian B *ñyātse* and the Middle Iranian forms could go back to a zero-grade **adᶻah* of the root **h_2n̥ǵʰ-* 'to tie up, to restrain'. The long **ā* could then have developed after the prefix **ni-*, as in Eastern Iranian languages **ni-* + *a-* yielded **ni-ā-* (de Vaan 2003: 34). However, this would mean that forms without the preverb **ni-* would have to be analysed as back-formations, since their initial long *ā-* could not have arisen by sound law: Sogdian *ʾz* /āz/ 'greed', Middle Persian and New Persian *āz* 'greed, lust, covetousness'.

As far as the form is concerned, I would also follow Isebaert's explanation (2019: 269 f.), who assumes that an original **ni-anza-* 'nécessité pressante'

69 The Persian form was ultimately borrowed from another Iranian language, as the Persian reflex of PIr. **ni-āǰah* would be †*niyād*.

(derived from *$h_2emǵ^h$-) was influenced by *āza- or *āzi- 'désir, convoitise, empressement' (derived from *$h_2eǵ$- 'conduire, pousser'), yielding a form *ni-āza-. However, he also assumes semantic influence in the same direction, and as I will try to explain below, this is not necessary.

In my view, the semantics of *ni-ādza- can be explained from *$h_2emǵ^h$- as follows. The primary meaning of *$h_2emǵ^h$-, namely 'to tie up, to restrain' is seen in, for example, YAv. *niiāzata* 'tied up',[70] Lat. *angō* 'étreindre, oppresser, serrer (la gorge)', Hittite *ḫamanki* 'binds' (see the LIV^2: 264–265 for more examples, see Isebaert 2019 for a similar discussion). This root is also attested in Germanic, cf. Proto-Germanic *angwu- adj. 'narrow': Gothic *aggwus* adj. 'id.', Dutch *eng* 'scary; narrow' (Kroonen 2013: 28–29). In many branches, derivatives of this root take the meaning 'distress' or, in adjectival derivations 'oppressing, scary, terrible', due to the semantic shift 'tie up' > 'tightened' (> 'narrow'), yielding the notion of a difficulty to breathe, of distress.[71] This can be seen in OHG *angust*, German *Angst* f. 'fear' < Proto-Germanic *angusti-; French *angoisse* 'great anxiety, terrible worry' < Lat. *angustia* 'narrowing'. The same semantic development can be seen in Avestan *ązah-* 'constriction; distress, peril' < *$and^{z}ah$ < PIE *$h_2emǵ^h$-o-.[72]

The German word *Not* can be translated in English by 'need' as well as by 'distress'. It appears that in Germanic languages, too, the notion of 'distress' precedes that of 'need', although they are both intertwined (Kroonen 2013: 385). The same semantic development must have taken place in Iranian languages.

---

70 Ardvi Sura Anahita Yašt, verse 127: [...] *maiδim niiāzata yaϑaca hukərəpta fštāna* [...] "and she tied her middle so that her breasts (are) well-shaped". The form *niiāzata* can also stand for *niiazata with a short *a*, cf. LIV^2: 265 with reference to Hoffmann.

71 Pinault (2019: 394) writes "Ved. *aṁhatí-* fem. [...] refers to 'distress', precisely to the fear caused by the feeling of 'narrowness'." This narrowness, in my opinion, is more precisely narrowness in the throat: difficulty to breathe. Being one of the worst feelings man can endure, difficulty to breathe was seen by most peoples as the worst of all pains. This is perhaps one of the reasons why paradise, in the Zoroastrian tradition, is qualified as *vīspō.xᵛāϑrəm* 'all good-breathing' (although this last word is subject to various scholarly interpretations). The verb *angō* in Latin and many of its close cognates also connote narrowness of the throat (cf. Ernout & Meillet 1967: 33). Another possibility, pointed out to me by Prof. Lubotsky, is that 'narrowness' for nomads meant a lack of pasture, which thus meant death.

72 The notion of need arises from that of distress, in some languages, through the idea of want, of an urgent or vital need, a vital need, for example food for the starving, or medicine for the dying; compare French *être dans le besoin* 'to be in "the" need', semantically almost identical to *être dans la détresse* 'to be in distress' (cf. also Pahl. *niyāzōmand*, NP *niyāzmand* 'needy', Pahl. *niyāzōmandīh* 'poverty', etc.).

Thus, the meaning 'distress' of **ni-ādᶻa-*, preserved by Tocharian and Parthian, must be older than the meaning 'need', found in Middle and New Persian, among others.

The notion of danger, occurring rarely for this word, could be secondary, possibly a Tocharian-internal development. The meaning 'danger' occurs often in Buddhist contexts, in phrases such as "the danger of the *kleśa*s (passions)", where it could perhaps have originally be used in a sense "the distress (or the calamity) relative to the *kleśa*s". All those meanings, although they reflect different semantic developments, are linked within Iranian to the root **Hanǰ-* < **h₂emǵʰ-*.

A formal problem remains in that we have traces of **āzi-*, and none of ***āza-*, and we have no Old Iranian trace of a noun **ni-āza-*: we should rather expect **ni-āzi-*. However, a thematic derivative could have been made in Old Steppe Iranian. We do not have a trace of **ni-āzi-* neither, and it is thus possible that **ni-āza-*, although it must have been secondary, was the prototype of Parthian *niyāz* 'need, distress' Middle and New Persian *niyāz* 'need', etc.

In conclusion, while an Indo-European origin of Tocharian B *ñyātse* cannot be entirely excluded, it appears that an Old Iranian borrowing is a much more plausible explanation. The problem is that no Old Iranian **niādᶻa-* is attested. Therefore, it is advisable to follow Isebaert and assume influence of **āzi-* 'greed' on an original **ni-andᶻah-* 'distress'. However, the meaning of the Old Steppe Iranian **niādᶻa(h)-* that I posit can be derived from PIr. **Hanǰ-* 'to be narrow, to be tight'.

2.3.5.1 A Note on TB *ñyās*, *ñās*, TA *ñās* 'Desire'

Related to the discussion above is the etymology of TB *ñyās*, *ñās* and TA *ñās* 'desire' (e.g. *umpar-ñāsatsum** 'desiring evil' DToA: 66). The form *ñyās* 'desire' in Tocharian B is more archaic, and *ñās* is later (see Peyrot 2008: 63–64). That later form *ñās* was borrowed by Tocharian A, cf. Peyrot (2010: 140). The word is borrowed from a Middle Iranian form *niyāz* 'need': either from Sogdian *niyāz* (SD: 249), Bactrian νιιαζο, Parthian or Middle Persian *niyāz* (DMMPP: 252), or another language.

Here one has to suppose that, in any case, Tocharian B *ñyās* was simplified from an original Middle Iranian *niyāz* → Tocharian B **ñəyās* > *ñyās*.

Since those Middle Iranian forms derive from Old Iranian **ni-ādza(h)-* (from which TB *ñyātse* was also borrowed), the meaning 'desire' of *ñyās* can be seen as being closer to the original meaning **'pressing, vital need' from which the meaning 'need' arose in most Iranian languages.

Malzahn's explanation (2007) that *ñās* is derived from the verb *ñäsk-* 'to desire, to seek' and that *ñyās* is a secondary form (analogically remade from

ñās) does not fit the chronological distribution as established by Peyrot (2008: 63–64; see also DTB: 291). However, one could surmise that the meaning 'desire' of the Tocharian form, instead of 'need' as in most of Iranian, could have been influenced by the meaning of the verb *ñäsk-*.

In conclusion, Tocharian B *ñyās*, *ñās*, borrowed into Tocharian A *ñās*, is itself a borrowing from a Middle Iranian language. Unfortunately, the exact source language of this word is undetermined, as the number of languages from which this word could have come is quite large.

2.3.6 *TB* twere* *'Door'*

Tocharian B *twere** 'door', a masculine noun, has been taken as an inherited formation (cf. DTB: 345; Beekes 2011: 35), but I wish to suggest it is alternatively conceivable that it has been borrowed from a hypothetical Old Steppe Iranian **dṷara-*, also meaning 'door', see for instance Old Persian locative *duvarayā=mai*, *duuara-* 'gate, door', Wanjī *devur* 'door' < **dṷara-* (cf. Lashkarbekov 2008: 69, who derives it from **dṷar-*), Parthian *bar* 'door' (DMMPP: 108, although it could theoretically also derive from **dṷar-*, cf. Avestan *duuar-* 'door'), etc.

The Indo-European **d^{h}ṷoro-* from which TB *twere** should have been inherited would have to be a post-PIE form, because the Proto-Indo-European word was a root noun **d^{h}ṷor-* (cf. Beekes 2011: 35). Reflexes of **d^{h}ṷoro-* are found, but they must all be secondary. For instance, Vedic *dvā́ra-* 'doors', only found after the *Śatapatha Brāhmaṇa*, is a secondary thematization of nom.sg. *dvā́r*, acc.sg. *dvā́ram*. Lat. *forēs* pl. f. 'door, gate; the two leaves of a door'; *forus* 'garden surrounded by walls'; *forum* n. 'market place, public space', and OCS *dvorъ* 'courtyard' designate a place next to doors, and must thus be secondary derivations. It is thus as likely that Tocharian B *twere** is a loanword from Old Steppe Iranian, which would have continued the Iranian form **dṷara-*, a later thematicization of PIr. **dṷar-*.

In conclusion, Tocharian B *twere** 'door' could be inherited from post-PIE **d^{h}ṷoro-* as is commonly believed, or it could have been borrowed from an Old Steppe Iranian **dṷara-*, which has many Iranian cognates, with an identical or almost identical meaning. One argument for it being a loanword is that its preform is post-PIE, although theoretically the thematicization could have occurred in the ancestor of Proto-Tocharian.

2.3.7 *TB* mətstsa-, *TA* nätswā- *'To Starve, to Die of Hunger'*

TB *mətstsa-* and TA *nätswā-* 'to starve' derive from PT **nətswa-* (cf. Peyrot 2013: 790). As noted by Adams (1999: 459; DTB: 493) the initial *m-* in Tocharian B is due to an assimilation of the initial **n-* to the following **-w-*.

Adams (DTB: 493) derives Proto-Tocharian **nətswa-* from "Proto-Indo-European" **n-h₁dˢ-tw-ye/o-*, according to him "a derivative of **h₁ed-* 'eat'". Adams here assumes a sound change **-TT- > -TsT-* which, although it is attested for multiple Indo-European branches, is not known for Tocharian. Also, the supposed formation has no parallels. To my knowledge, apart from this very difficult proposal, the Tocharian A and B verbs did not receive an Indo-European etymology, nor were they explained as loanwords. An argument against the Indo-European inheritance of these verbs is found in the following fact. An Indo-European **n* followed by **u̯* becomes **m* in Proto-Tocharian already: TB *mekwa** (with addition of the plural *-a* suffix), TA *maku* 'nail' < PT **mekwe* < **h₃nogʰu̯o-* (cf. Krause 1971: 10). I do not understand why there was no *p/m*-umlaut in Tocharian A, as the expected form should be ***moku*, but one can imagine that the *u*-vowel blocked the *p/m*-umlaut in Tocharian A. Here, we have to reconstruct Proto-Tocharian **nətswa-* with an *n-*, because of the Tocharian A form, so that it has to be a later sound change.[73]

I believe that TB *mətstsa-* and TA *nätswā-* are derived directly from an unattested Proto-Tocharian noun **netsu*, the **-u* of which is visible in the *-w-* of the verb (PT **nətswa-*). This noun would have either meant 'corpse' or 'demoness of Death', and would have been borrowed either from Old Steppe Iranian **natˢu-* masc. 'corpse' or from **natˢu-* fem. 'corpse; demoness of Death', in its turn from PIr. **naću-* 'corpse', compare Avestan *nasu-* 'corpse; demoness of Death' (from PIE **nek̑u-*). From the Proto-Tocharian noun **netsu-*, a verb was made, **nətswa-*, in the same way that PT **reytwe* 'mixture' yielded a verb **rəytwa-* 'to mix'. Interestingly, TB *eṅkwe* 'man', TA *oṅk* 'id.' derive from the same Indo-European root as OSIr. **natˢu-*, through the semantic development 'mortal' > (**'human being' >) 'man' (cf. DTB: 83).

As to the semantics, there are two different pathways: the first possibility is that **netsu-* was borrowed with religious semantics, as 'the demoness of Death', the verb meaning 'to be overcome by the demoness of Death' > 'to die (as a process)' > 'to die of hunger, to starve', exactly like in Old, Middle and Early New English, where *to starve* indicates dying as a process, as opposed to *to die*, which meant 'to cease to live'.

One can compare also Norwegian (Nynorsk) *starva* 'frieren, dem Tode nahe sein' (cf. Holthausen 1963: 320). In English too, *to starve* could mean 'to freeze

73 It is unclear to me whether the sound change that effected **nətswa-* in Tocharian B is "the same" as in Proto-Indo-European, that never stopped being effective in Tocharian B, or whether it is an identical sound change that occurred a second time in the language.

to death' (Middle English and English) and, transitively, 'to make someone die of hunger'. In this sense, in Tocharian, from a noun 'demoness of Death', a verb meaning 'dem Tode nahe sein' could be made.

The second pathway, which seems a little bit more difficult to me, would make the word go through *'corpse' > *'stiff/thin like a corpse' (perhaps an adjective?) → 'to become thin like a corpse' > 'to starve'. In this case, it is perhaps possible to view the English and Norwegian meanings of 'to freeze to death', and perhaps even 'to starve' as ancient archaic meanings (although they do not appear in Old English, as far as I could find), relative to the original meaning of the verb 'to become stiff' (Kroonen 2013: 477).

In conclusion, PT **nətswa-* is likely the verbal derivative of an unattested Proto-Tocharian noun **netsu*, borrowed from an Old Steppe Iranian noun **natˢu-*, a cognate of Avestan *nasu-* 'dead matter, demoness of Death'. However, it is not entirely sure whether PT **nətswa-* was derived from a noun meaning *'corpse' or *'demoness of Death'.

2.3.8 *TB* mot *'Alcoholic Beverage'*

The TB word *mot* corresponds to Sanskrit *surā-* 'cereal-based alcoholic beverage' (cf. Ogihara 2009: 247). It designates in general various kinds of alcoholic drinks made by fermentation, like Chinese 酒 *jiǔ* (cf. Ching 2010: 387). For a long time, it was believed that TB *mot* was borrowed from Sogdian *mwδ* 'wine' (e.g. Pinault 2003a: 183). However, things are more complicated: the normal Sogdian word for wine is *mδw* /mᵊðu/ (SD: 210).

The form *mwδy* from the Ancient Letter IV, 1.5. is now recognized as meaning 'price' (see Dragoni 2021: 302, with reference to literature). A Christian Sogdian form *mwd[* 'wine' exists, but it is attested in a very late text, and would be too late for Tocharian (cf. Dragoni 2021: 302[31]). In any case, this form *mwd[* likely stands for /muð/, with *u*-umlaut of the schwa, and thus phonetically it is too far off the Tocharian word.

To explain the Tocharian B form *mot*, one can think of three solutions: 1. it is a borrowing from an unattested dialect of Sogdian, which had **mwδ* (< **madu-*) at an earlier stage than late Christian Sogdian *mwd[* 'wine'. This is unlikely, as no other such example exists. 2. It is a borrowing from an unknown Middle Iranian language. This is very speculative. 3. It is a borrowing from Old Steppe Iranian **mod(u)-* < PIr. **madu-* (or **maδu-*) 'mead'.

What could permit us to posit such an umlaut for Old Steppe Iranian is the parallel of Ossetic: in Ossetic, PIr. **madu-* 'mead' became Iron *myd*, Digoron *mud* 'honey'. This *u*-umlaut is documented in Ossetic, with the condition that the *-u* or *-ū* should have been in word-final position and there should be a labial consonant in the word (cf. Cheung 2002: 124 f.). According to Cheung (2002:

125) an intermediary stage of the umlaut was a realization of the *a as [ɔ]. If Old Steppe Iranian had an *[ɔ] in this word, it would have been rendered as *o by Tocharian speakers, who, as far as we know, had no phoneme /ɔ/. Otherwise, we can simply assume that the intermediary form was close enough to Tocharian *o.

Even though this derivation remains speculative since no other case of Old Steppe Iranian u-umlaut is recorded, it is nevertheless clear that Tocharian B *mot* 'alcoholic beverage' must ultimately be related to Proto-Iranian *$madu$- (or *$maδu$-) 'mead'. I do not think that this umlaut could have occurred within Tocharian itself (*$metu$ > *mot) as it has not occurred in other Old Steppe Iranian loanwords (TB *peret* < PT **peretu*, not ***perot*).

In conclusion, the exact origin of TB *mot* 'alcoholic beverage' is unclear for the moment although I believe it is likely to be of Old Steppe Iranian origin. Since we lack clear parallels for the sound change it presupposes, this hypothesis remains speculative. In theory, it could alternatively come from some unknown or unidentified Middle Iranian language.

2.3.9 *TB* wertsiya *'Assembly', TA* wartsi *'Assembly, Crowd'*

The Tocharian B word *wertsiya* 'assembly' (verse variant *wertsya*, cf. Pinault 2006a: 82) and its Tocharian A equivalent *wartsi* 'crowd' have been derived from Proto-Indo-European *$H\mu ord^{h}i̯eh_2$- 'mass, multitude', from *$H\mu erd^{h}$- 'to grow' by Adams (DTB: 665; see also Del Tomba 2023: 178).

In my view, there are several problems with Adams's etymology. First of all, *$H\mu erd^{h}$- means 'to grow (for example, a plant), to make grow (animals, babies); to make strong' and 'to become strong'. These meanings can also be seen in various Uralic loanwords from Indo-Iranian, which have the meanings 'to rear animals', 'to raise children, to give birth'; 'to feed' (cf. Holopainen 2019: 311 f.). Second, *$H\mu erd^{h}$- does not have any assured cognates outside of Indo-Iranian, cf. LIV^2: 228. To assume that, first, this root should have also existed in Tocharian, and second, that it would have evolved to mean 'mass' (which is far from evident) and then 'assembly; crowd' in Proto-Tocharian implies just too many steps for it to be likely.

Lastly, to explain the Proto-Tocharian form **wertsiya* one needs to have both *$H\mu erd^{h}i̯$- (to explain the *ts < *-$d^{h}i̯$-) and, secondarily, *-iya. Thus, for this derivation, a Proto-Indo-European *i would be needed twice.

In my view, a more straightforward etymology can be suggested. I propose that **wertsiya* is a borrowing from Old Steppe Iranian *$\mu ard^{z}i̯\bar{a}$-, a cognate of *$\mu r̥d^{z}\bar{a}na$- 'community', cf. OAv. *vərəzāna-*, YAv. *vərəzāna-* 'community'; OAv. *vərəzāniia-* m. 'member of the community', *varəzāna-* adj. 'belonging to the community' (cf. AiW: 1378 f., 1424 f.).

These words could derive from **u̯r̥dᶻā-*, which presumably meant 'habitation', but also 'village; city'. I suppose that the Avestan and Old Steppe Iranian words were originally substantivized adjectival formations meaning 'those of the village' or 'those of the city' > 'the community'. This could explain why we have here a feminine suffix **-i̯i̯ā-*, denoting appurtenance. As in the Old Steppe Iranian source of the Tocharian B borrowing *newiya* 'canal', this feminine suffix is in origin the substantivization of a feminine adjective.

Unfortunately, the *-i̯i̯ā-* formation that I set up to explain the Tocharian word is not yet attested in Iranian. Alternatively, one could consider that Proto-Tocharian borrowed **u̯r̥dᶻ-* 'village, city' and that the suffix *-iya* was added by the Tocharian speakers themselves, following the pattern outlined by Del Tomba (2023: 178).

If my derivation of Tocharian B *wertsiya* and Tocharian A *wartsi* from Old Steppe Iranian **u̯ardᶻi̯i̯ā-* is correct, the Old Steppe Iranian word would show another instance of **-r̥-* > **-ar-* (cf. p. 180 f.).

A problem for which I have no definitive solution is why the **r* has remained **r* in front of **dᶻ* instead of changing to **l*. Possibly, the change to **l* occurred only before **d*, and not before **dᶻ*, see section 2.7.2.11.

In conclusion, an Iranian source is very likely for Tocharian B *wertsiya* 'assembly, community', Tocharian A *wartsi* 'crowd'. The closest Iranian cognates resemble the Tocharian words very much formally, and, based on their attested cognates, it is likely that their semantics would have been very close to the Tocharian meanings of these words.

2.3.10 *TB* welke*, *TA* waltäk* *'Stamen'*

Tocharian B *welki*, the plural of an unattested noun *welke**, occurs twice. Once in W 11 b5, where it is the only legible word of the line, and once in W 32 b2, which contains a medical recipe. The context as given by Adams (DTB: 665) is:

W 32 b2

> *ke$_u$-pyapy(ai)ntse welki · eñcuwañe keṃtse · te ṣeme yarm*
> "the petals of a dandelion [?] and iron-rust, each the same measure" (Adams DTB: 665).
> "… de fleurs …, de terre ferrugineuse; cela, une mesure […]" (Filliozat 1948: 86).

As noted by Adams (DTB: 665), *welke** (pl. *welki*) designates a part of a flower. He suggests that it can be compared to Vedic *valká-* 'bast, bast fibre' and Avestan *varka-* 'leaf', or Vedic *válśa-* 'shoot, twig'.

In my view, there is a much more straightforward etymology: Old Steppe Iranian **u̯aldaka-* (with PIr. **rd* > OSIr. **ld*, cf. p. 176 f.) '(of the flower', that is, either '(petal) of the flower' or '(stalk) of the flower', but more probably petals, as it is used as an uncountable quantity in the Tocharian B text, and petals were more commonly used as ingredients, and less countable than bare stalks of flowers. On the meaning of the Tocharian word, see below.

The word **u̯aldaka-*, which was probably accented on the first syllable, should regularly have become **weltke* in Proto-Tocharian. The accentuation also would explain why it is the second syllable that was syncopated, and not the third one. The simplification **ltk* > **lk*(*k*) would have occurred in the historical period, after the archaic stage (cf. Peyrot 2020), see for example TB *spelke*, *spelkke* 'zeal' from *speltke*, cf. TA *spaltäk* 'id.'.

This word has a Tocharian A correspondent, as was suggested to me by Dr. Athanaric Huard, although the matter is a little bit more complicated. This has consequences on our understanding of the meaning of the word as well. There is a Tocharian A hapax word *walcisyo* (instrumental plural), which is glossed by DToA as 'braid', and the nominative form reconstructed is *waläc** (DToA: 426). It appears in the Tocharian A version of the *Sūryodgamaprātihārya*, a text that speaks about the multiplication of the Buddha on lotuses. This version is found in A 312 and A 313, and translated by Sieg (1952: 26–33).

A 315 + 316 a4–5

> *cam āsānis mrācaṃ śkaṃ wsāṣinäs kä – – – ·o vaiḍuriṣiṃ oryo nkäñciṃ /// waśir ñemiṣinäs walcisyo wkäṃ vicitra lukśanunt ñemi*(*ṣinās swā*)*ñcenāsyo worpu*(*s*) *– – – (ka)knus oppal pākä*(*r tāk*)
>
> "Et au sommet de ce trône apparut un lotus, pourvu de … d'or …, avec une tige de béryl, … un … d'or …, de *waläk** de joyaux, de diamant, entouré de rayons de joyaux brillant de toutes les couleurs" (Athanaric Huard, p.c.)

A comparison with the descriptions of the wondrous lotus produced during Śrāvastī's miracle shows, according to Huard, that this word *walcisyo* designates the stamina of the lotus—the pollen-producing organ of the plant—, which are mentioned right after the stalk. This story of Śrāvastī's miracle is very similar to the *Mūlasarvāstivāda-vinaya* which describes the same miracle as is described in the TA *Sūryodgamaprātihārya*, although the *Sūryodgamaprātihārya* is more detailed. Sieg & Siegling (1921: 168) have already noted the similarity between the *Sūryodgamaprātihārya* and Śrāvastī's miracle. The stamina are not described in the Sanskrit version included in the *Divyāvadāna* (cf. Rotman 2008: 279), but they are mentioned in the Chinese version of the *Mūlasarvāstivāda-vinaya*:

T1451 XXIV 332b12–13

寶為莖金剛為鬚

"[un lotus] avec une tige de pierres précieuses et des étamines [litt. barbes, *xū* 鬚] de vajra" (Athanaric Huard, p.c.)

It is possible to see *walcisyo* as a scribal mistake for **walckisyo* (with a missing letter). The word **walckisyo* would be the instrumental of **walckiñ*, plural of *waltäk**, in the same way as *rackisyo* is the instrumental of *rackiñ*, plural of *rätäk* (s.v. *retke*, *rätäk*). TB *welke* and TA *waltäk* go back to Proto-Tocharian **weltke* which must be from Old Steppe Iranian **u̯aldaka-*, certainly accented on the first syllable, hence syncope of the second syllable instead of apocope. As to the meaning, it should have been 'stamen', in the plural 'stamina' or what is called in the scientific literature an androecium. It is also likely that W 32 b2 had the same meaning, especially, and we could translate it as "the stamina of a dandelion and iron-rust, each the same measure" (Adams DTB: 665). However, the dandelion when it gets its fruits becomes full of silver-tufted fruits or 'clocks' as they are called, which disperse with the wind—or with children blowing them in each other's face. This is most probably what *welki* is referring too in W 32 b2, and not to an androecium proper. Both dandelion's fruits and iron-rust are light, easily dispersed.

It is not obvious to me whether the meaning 'stamen' was already that of the original Old Steppe Iranian word or not. If it was, then one can speculate that the Old Steppe Iranians were actively artificially pollinating flowers. Otherwise, it is reasonable to think that the Old Steppe Iranian word simply designated some object relating to flowers, 'X of a flower', and that was specified in Proto-Tocharian as referring to stamen, because Tocharians took more interest in the pollinizating aspect of flowers and plants. All in all, this word could have relevance for future archaeological research.

To conclude, TB *welke** and TA *waltäk** 'stamen' may be borrowed from Old Steppe Iranian **u̯aldaka-*, derived from PIr. **u̯arda-*/**u̯r̥da-* 'flower' (on the development of Proto-Iranian vocalic **r̥*, see section 2.7.2.12), with the addition of the suffix *-ka-*.

2.3.11 *TB* tsetke *'A Hundred,* centaine *(?), Century (?)'*

Tocharian B *tsetke* is found in two fragments: IOL Toch 158 and THT 1928. Below I cite the contexts in which it occurs:

IOL Toch 158 a4

/// *kalymi · tsetkesa tucenän kaunän*

"... direction ... · By *tsetke* ... yellow suns ..."

THT 1928 b6

/// ·ṃ – – – *89 tsetke prākre snai* – ///

"... *tsetke* solid without [...]"

Unfortunately, the meaning of the word cannot be established from these fragmentary occurrences. However, the phonological structure of the word is strongly suggestive of Old Steppe Iranian origin: 1) the initial *ts-* need not necessarily, but could well be of Old Steppe Origin; 2) the vowels *e_e* likewise are not necessarily of Old Steppe Origin, but are extremely frequent in words from that source; 3) the final *-tke* is a strong indication of Old Steppe Iranian origin. This latter feature could perhaps have been taken over from a verb if the word was of Proto-Indo-European origin, i.e. *tsətk-* or the like, but in Old Steppe Iranian vocabulary *-tke* is highly frequent. Even though none of these features is absolutely probative, together they make borrowing from Old Steppe Iranian a serious option to consider.

If borrowed from Old Steppe Iranian, *tsetke* should go back to either **dᶻaTaka-* (with *T* representing either **t* or **d*) or **tˢaTaka-*. This easily leads to a possible etymon: **tˢataka-* '*centaine*, around a hundred'. Old Steppe Iranian **tˢataka-* would be a **-ka-* derivative of **tˢata-* 'hundred' (< PIE **(d)ḱmto-*), see for instance Middle Persian *sadag* '*centaine* (around a hundred); century'; New Persian *sada* 'century' (and, of course, MP *hazārag* 'millenium' ← *hazār* 'thousand', cf. CPD: 43). The syncope of the second vowel of **tˢataka-* is regular (cf. section 2.7.2.7).

Although I stress that the meaning of *tsetke* cannot be established independently at this point, it seems to me that "a hundred" is compatible with the attestion in IOL Toch 158. Since there is no English equivalent to French *centaine*, which is to a hundred what a *dozen* (← Fr. *douzaine*) is to 'twelve', I kept the French word '*centaine*' to translate *tsetke*. For this reason I also think it would be easier to translate IOL Toch 158 a4 to French:

"des soleils jaunes par centaines" "yellow suns by the hundred"

The perlative of *tsetkesa* also corresponds to Fr. *par* in *par centaines* (and also to English *by* in *by the hundred*). A full translation for THT 1928 b6 is not feasible. As Dr. Dragoni informed me (p.c.) this could be "a variant of a frequent Buddhist stock phrase referring to the innumerable Buddha suns, which are often hundreds or thousands in number" (cf. Skjærvø 2004: I, 68–69). An issue with this explanation is that this expression of the "many Buddha suns" is typically Mahāyāna, which is a branch of Buddhism. The Khotanese were Mahāyāna, but the Tocharians were Sarvāstivādin, which is another branch of Buddhism,

where this phrase does not occur. However, as Athanaric Huard told me, the type of texts to which this Tocharian text belongs (meditation texts) "multiplies mental objects" and mentions elements in extraordinary numbers, filling the universe with them, so that the meaning 'centaines' is likely in this context.

As a conclusion, the meaning of Tocharian B *tsetke*, which appears in IOL Toch 158 a4 and THT 1928 b6, cannot be firmly established. As far as its phonological structure is concerned, the word could well be of Old Steppe Iranian origin. A possible source would then be Old Steppe Iranian **t^{s}ataka-*, a **-ka-* derivative of **t^{s}ata-* 'hundred', the reflex of which is found in all Iranian languages, for instance Avestan *satəm*, Middle and New Persian *sad* 'a hundred', and this could be supported by a possible analysis of one occurrence as a frequent Buddhist literary image.[74]

2.3.12 TB tsere *'A Measure of Liquid Volume of Half a* lwāke *(?)' or 'Cup (?)'*

The Tocharian B word *tsere* designates "a kind of vessel or [...] even a measure of capacity equal to 0.5 *lwāke*" (cf. Ching 2011: 68[21]). It is worth, in size or volume, half a *lwāke* 'jar'.

As a jar could very well fill only two big cups, TB *tsere* might have designated a cup, and could come from Old Steppe Iranian **t^{s}ara-* 'head' (Avestan *sara-*, Persian *sar*, etc.). An argument in favor of this etymology could be the practice of drinking from dead enemies' skulls as famously reported about the Scythians by Herodotus.

Perhaps, less cruelly, one can also think of a metaphorical designation. This is a known process, where the parts of the jug are identified with parts of the human body. For instance, one can think of Dutch *kop* 'cup; head', where the meaning 'cup' is original, cf. also Latin *testa* 'jug, cup' > French *tête* 'head'.[75] With suffixes, one can also think of English *handle*, Persian *dasta* 'handle' < **dastaka-* built on **dasta-* 'hand'.[76]

---

74 Since Middle Persian *sadag* < **θataka-* < **t^{s}ata-ka-* meant 'century (a hundred years)', one can also imagine that the Old Steppe Iranian word had this meaning as well.

75 I would even like to suggest that the *ʔ* in Arabic *kaʔs* 'cup, glass', which is not etymological, is due to the influence of Ar. *raʔs* 'head', where the *ʔ* is inherited (compare Hebrew *rōš* 'head' < **rās* < **raʔs*).

76 On this theme, one can evoke Omar Khayyām's moving quatrain:

Īn kūza čō man āšiq=i zār=ē būda-st
Dar band=i sar=i zulf=i nigār=ē būda-st
Īn dasta ki bar gardan=i ō mē=bīn-ī
Dast=ē-st ki bar gardan=i yār=ē būda-st

"This jug was a desperate lover, like me, / (Like me,) it was in chains, tied to the hair of an idol / This handle that thou seest on its neck / It is an arm, which was around a lover's

Evidently, a 'head' may be called a 'cup' because of their resemblance. The Old Steppe Iranians would have processed in the opposite direction, and called a 'cup' a head. Unfortunately, I could not find an example among Iranian languages of *t^s*ara-* (or a derivative therefrom) taking on the meaning 'cup'.

An advantage of this etymology is that it works perfectly formally, and there are some parallels, although not exact ones. PT **tsere* could very easily derive from Old Steppe Iranian **t^s*ara-*. Two disadvantages of this etymology lie in the fact that (1) the Tocharian meaning is not certain and (2), as far as I know, there are no examples of the specific shift from 'head' to 'cup' among Iranian languages.

2.4 Old Steppe Iranian Borrowings: Difficult Cases

The words discussed in this section are of possible Old Steppe Iranian origin. However, they present serious difficulties, e.g. clear Iranian cognates are lacking, the form is both problematic and its meaning unclear, and so on. In the case of TA *kump* 'cauldron', for instance, where an Old Steppe Iranian etymology is possible, an Indian origin is possible as well. The meaning of TB *tseriteke* is not clear, but its *-i-* is also unexpected, so that an etymology will always be doubtful until at least its meaning is assured. The Old Steppe Iranian origin of TB *ecce*, TA *aci* 'hither, from' I find tempting, but there remain many problems in the details of this etymology.

2.4.1 *TB* ecce, *TA* aci *'Hither, From'*

Tocharian B *ecce* means 'hither, from' and Tocharian A *aci* 'beginning with … up to; hither; from' (e.g. TA *tmäṣṣ aci* 'from there', *tanäṣṣ aci* 'beginning with all that', *mkältoräṣṣ aci* 'from childhood', cf. DToA: 5). Both forms regularly go back to PT **ecye*. These words remain without clear etymology (cf. DTB: 83). The meaning of these words was discussed in detail by Winter (1984). He concluded (1984: 122) that they are synonymous with TB *śka* 'hither'.

Dunkel (2014: 262) has proposed an Indo-European etymology for PT **ecye*: an Indo-European particle **$h_1$oti* 'away from' → neuter adjective **otḭom* > PT **ecce*. This etymology seems implausible to me because **tḭ* would have yielded

neck." As we all become earth and soil, this soil is used by potters to make jugs, which, once part of the human body, now mimic the body, as the *tsere* of the Tocharians perhaps reminded them of the human head. In them we drink wine, and tomorrow, we will become the soil from which new jugs are made.

PT **ts* rather than **cy* or **cc*, and also because **$h_1$oti* is only supported by two very doubtful Balto-Slavic forms (see Dunkel 2014: 262).[77]

I am tempted to connect TB *ecce*, TA *aci* to Proto-Iranian **hačā* 'from', cf. Old Persian and Avestan *hacā*, and to see in it a borrowing from Old Steppe Iranian. However, **ecye* cannot have been borrowed from **hačā* directly, as this would have yielded PT ***eca*. Therefore, in order for this etymology to work, I have to posit a derived adjective **hačā-i̯a-*. As no such adjectival formation is attested, and no close parallel for this formation can be found either, I classify this etymology as difficult.

If a derived adjective **hačā-i̯a-* existed, it could have become **(h)ačai̯a-* through shortening of **ā* in front of *yod* in Old Steppe Iranian according to the rule discussed in section 2.7.2.9. This hypothetical Old Steppe Iranian **(h)ačai̯a-* would probably have become Proto-Tocharian **ecye* according to regular syncope rules (cf. Winter 1994).

In the posited Old Steppe Iranian **(h)ačai̯a-*, the suffix **-(i)i̯a* would be adjectival, and then the neutral adjective would have become an adverb again, according to the well-known circle adverb → derived adjective → (instrumental function or form of the adjective) > adverb. A parallel can for instance be found in Latin: *super*(*i*) adv. 'above' → adj. (**superinos* >) *supernus* > adv. *superne*.

A point which could be useful to mention in connection to this etymology is that usually two variants of **hačā* 'from' are assumed to have existed in Old Persian: **háčā* and **hačā́*, the first yielding New Persian *az*, and the second yielding New Persian *zi*, both meaning 'from'. I wonder, however, if New Persian *zi* cannot be derived from **hačíi̯a-*. Formally, this is not impossible, but it has not, to my knowledge, been proposed before. Since there are no other examples of **-á* > *-i*, the alternative derivation from **hačíi̯a-* may even have to be preferred. A form **hačii̯a-* would be parallel to **hačā-i̯a-*, but it would have to be derived from **hač-* rather than **hačā*. Since no such base **hač-* is attested, my suggestion has to remain hypothetical for the time being.

Notably on the basis of other words which obviously underwent syncope, I would like to suggest that **(h)ačai̯a-* was stressed on the first syllable, so **(h)áčai̯a-*, as it would otherwise have become **ecey*, similarly to how **kenek* possibly derives from **keneke* (cf. § 3.2.4). There is a tendency, or even a regularity, in Old Steppe Iranian having first syllable stress in words where stress can be observed, with few exceptions linked to prefixal and suffixal stress; so

77 A connection with PIE **$h_1$éti* is also impossible: this would have yielded PT **yəc*, like PIE **h_1éḱu̯o-* > PT **yəkwe* > TA *yuk*, TB *yakwe* 'horse'.

this should not be taken as having any consequences on the stress of this word in Proto-Iranian in general. For more detail, see discussion in § 2.7.2.7.

In conclusion, I have proposed to consider Tocharian B *ecce*, Tocharian A *aci* 'hither, from', which both go back to Proto-Tocharian **ecye*, as ultimately borrowed from an Old Steppe Iranian adverb, itself derived from an adjective, **(h)ačaḭa-* which could also have meant 'hither' and ultimately goes back to the preposition **hačā* 'from' suffixed with the adjectival suffix **-ḭa-*. However, this etymology remains difficult since **hačăḭa-* is attested nowhere and its derivational pattern has no parallels.

2.4.2 TB eprete *'Resolute, Steadfast'*

The Tocharian B word *eprete* 'resolute, steadfast' is considered by Adams (following earlier literature, see DTB: 95 with references) as being constituted of the prefix *e(n)-* and **prete-* 'decision, resolution', "the unattested TchB counterpart of TchA *pratim* (the latter borrowed in B as *pratiṃ*)" (DTB: 95; similar view in Pinault 2008: 334).

Adams' interpretation (already in 1999: 90) has been accepted in the scientific literature (e.g. LIV^2: 493). This form **prete-* has been connected to a root **pret-* 'erkennen', with the two cognates Gothic *fraþjan* and Lithuanian *prantù*, both meaning 'to understand', adduced to the reconstruction by the LIV^2: 493.

This etymology is not without problems. As far as we know, there is no clear example of a borrowing from Tocharian A into Tocharian B, while the other direction is much more common. Because of its semantics, a word such as TB *pratiṃ* 'decision, resolve, conversion' belongs to a rather intellectual or literary stratum, making it even more unlikely to have been borrowed from Tocharian A.

Under a different entry, Adams himself (DTB: 442) suggests that TA *pratim* was borrowed from TB *pratiṃ*, and not the other way around, contradicting himself on this point. He proposes that TB *pratiṃ* is ultimately derived from "the Prakrit descendant of Sanskrit *pratijña-* 'agreement, promise, decision, assertion' ". Ji (1998: 291) also provides a plausible etymology in Sanskrit *pratimā-* 'decision, likeness, symbol'. No matter the etymology of Tocharian A *pratim*, Tocharian B *pratiṃ*, these two cannot be formally compared to Tocharian B *eprete*, despite the semantic closeness of all those words.

Apart from this, there is no Tocharian B word †*prete* (*vel sim.*) from which *eprete* would have been made, nor is there a Tocharian A cognate of this word (†*prat*?).

In my view, it is much more likely that TB *eprete* regularly derives from an Old Steppe Iranian word **abi-ratu-* 'on (= according to) the decision, the judgment'.

On the meaning 'decision, judgment' of *ratu-*, see for instance Bartholomae with Old Avestan *ratu-* 'judicium, Richterspruch' (AiW: 1502).

If **abi-ratu-* existed and was borrowed, it would most likely have been rendered as PT **epəret*, with regular loss of the *-u* (s.v. TB *peret*, TA *porat* for instance); and reduction of **abi-* to **epə-* (s.v. *epastye* for another such example). This reconstructed form **epəret* would then have been reduced, possibly already in Proto-Tocharian, to **epret*. In Tocharian, a stem vowel *-e* would have been added secondarily to make adjectival inflection possible.

If the word was **ep(ä)rete* in Proto-Tocharian, when the syncope took place (see for instance section 2.7.2.7), the word would have been reduced to either ***epärte* or ***epräte*, because **é_e_e* or **e_á_e_e* was systematically reduced to **e_o_e* in Proto-Tocharian. This suggests that the adjectival *-e* was added after the Proto-Tocharian stage.

In conclusion, Tocharian B *eprete* 'resolute, steadfast', which does not have a convincing etymology so far, may derive from Old Steppe Iranian **abi-ratu-* 'according to the resolution, the judgment', which one can further understand as 'following one's resolution, steadfast'.

2.4.3 *TA* kump *'Cauldron'*

The Tocharian A word *kump* is often translated as 'pot' (e.g. Carling 2009: 150; Meunier 2015: 280; DToA: 141). The context of most attestations is fragmentary. However, in one well-preserved text, which I give below, the meaning is clearly 'cauldron' rather than 'pot'.

A 341 a7

> *lyalypurṣi want worta-ṃ kumpaṃ ṣū[t]i śpāl pakt-äṃ ymār tmaṃ kapśañi: ṣñi lyalypäṃntu(yo) ///*
>
> "The wind of deeds threw her in the caldron. She quickly boiled her body head first [in it]. (By) her deeds … […]" (adapted from CEToM).[78]

78 Carling and Pinault (DToA: 494) have a new interpretation of *ṣū[t]i śpāl* in the passage: they read it as *ṣūci-śpāl* 'needle-headed', a kind of hungry ghost (preta). The original manuscript is lost and it is impossible to be certain of the reading. According to the DToA, *ṣūti* is a hypersanskritization of Skt. *sūcī-* 'needle or any sharp-pointed instrument' and the compound would correspond to Sanskrit *sūcīmukha-* "literally 'needle-headed'". They disagree with Sieg's tentative interpretation as 'kopfüber' (Sieg, Siegling & Schulz 1931: 247; Sieg 1952: 40). Further, they state that this translation has no support in the narrative, and cite Chavannes' Chinese parallel (see p. 109 of the present book) as an argument. However, I disagree with these statements for a number of reasons. The Sanskrit compound *sūcīmukha-* means 'needle-mouth' and refers to the fact that the preta's mouth is too small for them to eat, this is not the same as 'needle-head'; the hypersanskritization is unwar-

It is, I believe, unthinkable that an entire woman's body would fit in a pot. The translation 'pot' is based on the Sanskrit word *kumbha-* 'jar, pitcher, water pot, ewer, small water-jar' (M-W: 293). However, in this precise context, it should be a cauldron, in which an entire woman's body could fit. This is confirmed by the Chinese parallel (Chavannes 1911: 251), which has a word translated by Chavannes as *marmite* (cauldron, big cooking pot) and *chaudière* (boiler, furnace). In particular, we read in the Chinese text that: "une marmite à trois pieds apparut; un feu de charbon la faisait bouillonner; cette femme enleva ses vêtements, les mit de côté et entra dans la chaudière; sa peau et sa chair furent entièrement cuites; il ne resta plus que de petits morceaux d'os; mais alors un vent frais vint à souffler; elle put sortir de la marmite et revenir à la vie; elle mit ses vêtements et dévora sa chair cuite. Quand elle l'eut dévorée, elle partit." (Chavannes 1911: 251).

Clearly, the object that is mentioned in both the Chinese and Tocharian versions is very different from the water-jar or pitcher that is denoted by Sanskrit *kumbha-*. It is bigger, and used for cooking, or at least for boiling water.

This text is an extract of the *Koṭikarṇa-Avadāna*, which is also part of the Sanskrit *Divyāvadāna*. Sieg (1952: 37) writes that the Chinese version is closer to the Tocharian one than both are to the Sanskrit text. In the Sanskrit version, it is only said that "whatever food she eats turns into her own flesh." (cf. Rotman 2008: 74).

Another indication on the meaning of TA *kump* is the distributive *kumpa-kump* (e.g. A 318 a1; YQ III.2 b4), translated as 'pot by pot' by Adams (2017: 1384) but more correctly 'in crowds, in droves' (cf. Carling 2009: 151, DToA: 141), as pot by pot does not fit the relevant contexts, where it often occurs together with *kropa-krop* 'group by group' (cf. Carling 2009: 174 "crowd by crowd"). However, the meaning 'crowd by crowd' or 'group by group' cannot be easily derived from 'cauldron-per-cauldron', since cauldrons rarely come in groups.

---

ranted, especially in view of the absence of an identical original compound. Further, Dr. Athanaric Huard made a suggestion to me, which I reproduce here:

> le passage est inspiré d'une description d'une torture très commune dans les descriptions d'enfers bouddhiques où les damnés sont jetés la tête la première dans des chaudrons (*lohakumbhī*): *adhosiraṃ gahetvā tattāya lohakumbhiyā pakkhipanti*. Il n'est pas surprenant qu'il y ait des échanges entre les tortures des enfers et des preta, notamment pour celles liées à la nourriture. Pour la confusion entre ces deux catégories, voir par exemple Zin 2014: 284.

This is not a case-settling argument, however, in view of all these elements, I believe there is not enough evidence for a reading *ṣūci-śpāl* with the meaning 'needle-headed' for A 341 a7.

In my view the original meaning was rather 'cauldronful by cauldronful', which exists with a distributive use also in other languages (cf. Ugaritic *dd dd*, Del Olmo Lete & Sanmartín 2003: 265), or perhaps better, in order to give the meaning of various people together, 'quantity (of what is contained in the *kump*) by quantity'. Strangely enough, *kumbha-* in Sanskrit also designates a rather large quantity: a measure of grain, which is "equal to twenty Droṇas, a little more than three bushels and three gallons; some make it two Droṇas [...]" (M-W: 293). Even only two droṇas would be equal to more than twenty kilos. There is a real discrepancy between *kumbha-* as a receptacle, which is rather small, and as a measure of quantity, which is very large.

To bring some nuance to these facts, it should be noted that in some Prakrits *kumbhī-* fem. designates a 'large round pot', or even, in some modern languages, a pool or a bucket (cf. CDIAL: 170). In that sense, it is not impossible that Tocharian A *kump* 'cauldron' comes from a Middle Indic language.

In Iranian, reflexes of **xumba-* designate either a pot (Av. *xumba-*), a pool (Balochi *kumb*), or a jar (Persian *xum*(*b*), cf. Horn 1898–1901: 59). Kumzari, a Persic language, has a word *xumba* (< **xumbaka-*) 'clay storage jar'. Interestingly Wakhi has a word *xəmbák*, which designates a very big bin or chest to store grain (cf. Steblin-Kamenskij 1999: 411). It thus seems that the discrepancy in sizes is seen among Iranian languages as well.

In these languages, the *xumba-* is either a jar to keep water or other liquids, as in Sanskrit, or it is a big container that contains grain (it is likely that Balochi *kumb* initially designated a grain-storage pool). The two meanings found in Sanskrit thus may be connected by comparison with Iranian languages. I cannot easily understand the two distant meanings: 'jar, pot' and 'big container of grain', but both meanings, that is, small jar and big container, both being storage tools, might have coexisted originally in Indo-Iranian.

A scenario that I can propose here is that Proto-Tocharian borrowed **kumpe* from Old Steppe Iranian **xumba-*, which had the meaning 'big receptacle', possibly used as a kitchenware, and which gave the meaning 'cauldron' to TA *kump*. Because the meaning 'cauldron' is absent in Indo-Aryan, it is somewhat conceivable, in my view, that Tocharian A *kump* 'cauldron' ultimately goes back to an Old Steppe Iranian form.

It is striking that there is no Tocharian B †*kump*(*e*) word or the like. To translate Sanskrit *kumbha-*, TB uses the word *lwāke* (possibly inherited, see Garnier & François 2020: 51–52), and the semantic equivalent of TA *kump* 'cauldron' seems to be TB *aise*, which Pinault (2008: 127) plausibly derives from the Proto-Indo-European root **h*$_1$*eiˀ-* meaning 'to be hot'.

However, it is also possible that TA *kump* was borrowed from a Middle Iranian language with the meaning 'cauldron'. This would make the absence of

the word in Tocharian B less problematic, especially given that the Tocharian B equivalent might be inherited. Of course, we cannot exclude that there was a technical difference between **kumpe* and **aise*, which would have disappeared due to lexical generalization at a later stage.

To conclude, I propose that Tocharian A *kump* 'cauldron' could derive from Old Steppe Iranian **xumba-*, which should also have meant 'big receptacle', and has many cognates among Iranian languages. In my view, the meaning 'big receptacle' can be reconstructed for Proto-Iranian, and it could easily have evolved to mean 'cauldron' through the meaning *'big receptacle (for cooking)'. However, a Middle Indic origin for TA *kump* cannot be entirely excluded, especially in view of the parallel cited by Huard (fn. 78 p. 108) where the Indic word used is *lohakumbhī*.

2.4.4 TB kercci *'Palace'*

The Tocharian B word *kercci* (variants *kerci* and *kerc(c)ī*, oblique *kercïyeṃ**) means 'palace'. Its etymology is difficult. I will present the data, then discuss previous etymologies (2.4.4.1), detail a specific French etymology which I believe has had an impact on the previous etymologies (2.4.4.2), then explain my own etymology and problems relative to it (2.4.4.3).

TB *kercci* is a *plurale tantum*, whose nominative plural ending *-i*, written ‹i› or ‹ī›, comes from **-iyi*, cf. Peyrot (2021: 458). It occurs in many texts. In the past, two of its occurrences have been interpreted as the plural of the word *kertte* 'sword', but Del Tomba has convincingly shown that for those two occurrences the meaning 'palace' is preferable (s.v. *kertte*; cf. Del Tomba 2020; Del Tomba 2023: 378–379). Tocharian B *kercci* 'palace' was borrowed into Old Uyghur as *karšı* '(royal) palace' and, from there, into Mongolian (cf. Clauson 1972: 664).

Tocharian A does not have an equivalent of TB *kercci* but uses a periphrastic figure *lāñci waṣt*, literally 'royal house', to designate a palace (cf. Meunier 2015: 84). Because of this, I believe the current meaning of *kercci* to be a secondary, post-Proto-Tocharian development. It is rather unnatural for a language that has a perfectly fitting word to discard it in favour of a more complicated periphrasis. In the following pages, I will argue that the word *kercci*, a *plurale tantum*, originally meant 'servants, people of the house', before meaning 'palace', and that it is likely an Iranian loanword. It is possible that the unattested Tocharian A cognate of Tocharian B *kercci* existed (perhaps under the form **karcañ* or **karccañ*)[79] took on another meaning, and no longer corre-

79 According to Michaël Peyrot (p.c.), the plural (PT) **kercciyeñ* would have become (Pre-A) **karccyañ* > TA **karcañ* or **karccañ*, the geminate of which would be due to the **-cy-*

sponded to 'palace', but this seems less likely and less economical than the option that Tocharian B *kercci* took on the meaning 'palace' secondarily.[80]

2.4.4.1 Previous Etymologies

Previous etymologies of Tocharian B *kercci* 'palace' can be classified in two categories: inherited from Proto-Indo-European or borrowed from Iranian.

Of the first kind, Meillet (in Hoernle 1916: 379) and Lidén (1916: 21–22) proposed to connect TB *kercci* to Gothic *gards* 'house', Old Norse *garðr* 'fence, hedge, court', Old English *ġeard* 'enclosed space, garden, dwelling; land, region' (Modern English *yard*), OCS *gradŭ* 'enclosure, city', Sanskrit *gṛ́há-* 'house, habitation, home', etc. Adams (DTB: 210–211) hesitates between cognacy to this group of words and to another group of words: either "PIE **gʰordʰi̯o-*" or what he calls a "putative PIE **gʰortiyo-*" (DTB: 210), which would in its turn be related to Greek χόρτος 'enclosed place, feeding place', Latin *hortus* m. 'garden', Latin *cohors* 'yard, enclosure', and possibly to English *garden*.

These two reconstructions, namely **gʰordʰi̯o-* and **gʰorto-*, are connected, although hesitantly, by Beekes (2010: 1645). The formal discrepancy between all those forms indicates that there is a real problem in their reconstruction.[81] De Vaan (2008: 291) provides a solution: for him the Latin noun reflecting **-to-* (and its Indo-European cognates) "might be regarded as a verbal adjective to a root **gʰr-* 'to enclose'". He also reconstructs two PIE forms: **gʰr-to-* and **gʰordʰo-* which might have contaminated each other in the daughter languages or in PIE itself (2008: 291).[82]

On the borrowed side, Isebaert (1980: 88, 114) derives Tocharian B *kercci* from Old Iranian **gardi̯a-* 'servant, courtesan', corresponding to Sanskrit *gṛ́hya-* 'servant', etymologically '(the one) of the house'. In the plural, this would mean 'entourage, (royal) court'.[83]

---

cluster, similarly to *opäśśi* 'expert' which derives from **epəstəye* > TB *epastye* 'idem', s.v. TB *epastye*, TA *opäśśi*.

80 The Tocharian A word *wimāṃ*, Tocharian B *wimāṃ**, from Buddhist Sanskrit *vimāna-*, designates a specific divine type of palace (cf. DTB: 656).

81 Cf. LIV²: 197, where the verbal root is reconstructed as **gʰerdʰ-* 'umschließen, umgürten'.

82 The forms going back to **gʰordʰo-* could ultimately reflect a univerbated PIE compound **gʰr̥-dʰh₁-ó* 'enclosed'. There is a parallel for this formation in Hittite *u̯arpa dai-ⁱ* 'to enclose' (cf. Kloekhorst 2008: 966).

83 Joki (1973: 269) suggests that Tocharian B *kercci* derives from the Old Scythian ancestor of the Ossetic word Iron *kært* 'Hof, Bauernhof', adding that the ultimate etymology of the Ossetic word is uncertain. Normally **rd* and **rt* yield Ossetic *rd* (cf. Cheung 2002: 29) and only **ϑr* yields Ossetic *rt* (cf. Cheung 2002: 38). There are some cases of word-final devoicing in some Ossetic words, including the variant *art* of *ard* 'oath' (cf. Cheung 2002: 38). However, the initial *k-* in *kært* cannot derive from an ancient **g-* (see Cheung 2002: 22). In

Tremblay (2005: 426–427) apparently did not understand Isebaert's etymology, since he writes that Joki and Isebaert agree that TB *kercci* derives from Old Iranian **gr̥da-* 'house'. He himself proposes two solutions: the first is a borrowing from Old Iranian **gr̥da-* > Proto-Tocharian **karta* (his notation) → adjective **karciya-*, made within Proto-Tocharian. That adjective "eventually ousted its related substantive". Tremblay's second solution is that PT **kerciye-* "was borrowed from a *vr̥ddhi*ed collective **gardiia-* (instead of the expected **ǰardiia-*)." (2005: 427). Despite all this, Tremblay (2005: 427) believes that Adams' etymological proposal from either **gʰordʰiio-* or **gʰortiio-* is also possible.

As regards semantics, the etymological proposals mentioned here can be divided in two groups:

1. Adams (and others) propose an adjectival formation based on Indo-European reconstructed forms for 'yard', assuming a development 'yard' > *'lordly yard' > *'belonging to lordly yard' > *'belonging to lordly yards' > 'palace'. This is perhaps Joki's interpretation as well (1973: 269), as he suggested a link between the Tocharian word and the Ossetic one meaning 'garden; enclosure, fence'. The original meaning of the Tocharian would thus be 'enclosure', evolving into 'yard', etc.
2. The second type, which is that of Isebaert, consists of the following semantic pathway: 'of the house' > *'servant' > pl. *'servants' > *'entourage, court' > 'palace'.

My own explanation follows a similar pathway to Isebaert's. In my view, the first series of proposals, concerning an Indo-European etymology for the Tocharian word, rely on a specific semantic evolution, from 'yard' to 'royal court' which is based on a Romance (specifically French) parallel. I wish to show that this parallel is less self-evident than it initially appears, and I thus need to detail the French etymology that, in my opinion, implicitly motivated the etymology of TB *kercci* from **gʰordʰiio-* and related.

2.4.4.2 French *Cour* 'Yard; Court' and Its Bearing on the Etymology

The first type of proposals were presumably influenced by the semantic development seen in French *cour* 'yard; royal court'; secondarily 'palace (as a building)' (when it designates the royal court, *cour* is usually written as *Cour*),[84]

Joki's opinion (op. cit.), Old Persian **garda-* 'Hausgesinde' (< **gr̥da-* 'Haus') is not a cognate of this Ossetic word. In the absence of a clear origin of the Ossetic word this corresponds to explaining *obscurum per obscurius*.

84 The Old French word *cort* was borrowed in English as *court*. The English judiciary meaning also found in *court* derives naturally from the fact that judiciary processes originally occurred in the royal court. The modern French orthography *Cour* (instead of *Court*)

Spanish *corte* 'idem', Italian *corte* 'idem', German *Hof* 'idem', Dutch *hof* 'yard, royal court'. Then, by metonymy, at least in French, *cour* can designate the palace itself.

Incidentally, *cour* derives from the accusative *cōrtem* of Vulgar Latin *cōrs*, *cōrtis*, itself from Classical Latin *cohors*, *cohortis* and is thus directly connected to the PIE form **g^{h}ortos* mentioned above. This connection has perhaps further prompted the etymological proposals discussed here.

French *cour*, Old French *cort*, *curt*, initially designated an enclosed space, with a general meaning "terrain découvert, souvent entouré de murs ou de bâtiments, devant ou derrière l'habitation principale" (von Wartburg FEW II: 849). Later, in medieval times (only attested since the 12th century, von Wartburg FEW II: 850),[85] the lord's enclosure (stronghold) offered protection against raids to the 'lower people' (peasants) in the *basse-court* and to 'higher people' in the *haute-court* or *cour d'honneur* (in Spanish *corte noble*). In Modern French *basse-court* came to designate the animals associated with the lowest caste (poultry, rabbits, pigs). Because justice was held in the *haute-court*, the terms *cour de justice* and *haute cour de justice* are still used in Modern French.

By metonymy, this term came to designate the members of this *haute-cour* and the castle or palace where the lord (and later the king) resided. It is rather clear that the identical meanings found in Germanic languages for words that originally meant 'yard, enclosure' are due to French influence. Such an influence on Germanic semantics is often attested, and seems quite likely.[86] If indeed, Low and High German acquired this meaning for *hof* under French influence (and most other Germanic languages from German, cf. Hellquist 1922: 244 for Swedish),[87] as did Russian,[88] then the frequency of this semantic change is even reduced.

---

derives from an etymological confusion with Lat. *curia*, French *curie* 'Roman senate; the assembly thereof' (see Ménage 1694: 227–228, for references, arguments, and a discussion).

85 If this semantic shift already occurred in Vulgar Latin, one cannot eliminate the possibility of the influence of Greek αὐλή 'open court; courtyard'; post-classical Greek 'house of a lord; royal entourage' (Bailly 2000: 309). This is however not certain, given the relatively late attestation of this meaning in French and the other Romance languages (12th century in French, even later in Italian and Spanish). A further difficulty of this etymology is that Lat. *aula* only designates the lordly court, so that the confusion with 'courtyard' cannot have happened, except among Hellenists. It should further be added that the meaning 'house of a lord' of the Greek word was first used to designate Persian palaces, which were very different from Greek ones, and contained large courtyards and gardens (I thank Romain Garnier for informing me of this fact).

86 Cf. *"Hof", in: Pfeifer & al.* (*online, accessed on the 15th of September 2020*).

87 I thank Professor Martin Kümmel for providing me with this information.

88 As Professor Lubotsky (p.c.) informs me, French influence is generally accepted to explain the fact that Russian *dvor* 'yard; court' shares those two meanings.

The semantic evolution of this word from 'yard, enclosure' to 'royal court' (and then to palace) took place in a very specific context: medieval Western Europe, in a world of strongholds, lords, and courtesans, with rather rigid social divisions reflected in 'low yard' and 'high yard'; so the semantic trait 'lordly, royal' could emerge. It is hard to believe such a specific evolution could have happened in the Tocharian world, given its cultural, geographical and historical distance with the Latin-speaking medieval world.

This casts much doubt on the inherited etymologies of TB *kercci*. Instead, like Isebaert (1980) and Tremblay (2005) I believe this word to be an Iranian loanword in Tocharian, as I will explain below.

2.4.4.3 TB *kercci* from OSIr. **gr̥di̯a-*

Having disposed of the inherited etymologies, I adhere to Isebaert's proposal, with some new considerations and arguments presented here. We start with Old Iranian **gr̥di̯a-* 'servant' lit. 'that (one) of the house',[89] attested for example in Elamite ‹kur-ti-e-bar-šá› ← **gr̥di̯a-bŗza-* 'the exalted servant' (Tavernier 2007: 191–192) and Elamite ‹kur-ti-ia-ma› < **gr̥di̯a-u̯ant-* 'having servants' (Tavernier 2007: 192); also Babylonian *gardapata*, *gardapatu*, Elamite ‹kur-da-bat-ti-iš› 'majordomus, steward, house-holder' ← Old Persian **gr̥da-pati-* 'house master' (Tavernier 2007: 424).

The pathway 'of the house' > 'servant' is relatively straightforward, often attested; in Old Iranian itself, we have *(*d*)*māna* 'mansion, house' → **māni̯a* 'of the house' > 'servant', borrowed with this latter meaning in Tocharian (s.v. *mañiye*).

The Iranian noun from which the adjective **gr̥di̯a-* 'servant' derives is probably **gr̥da-* '(noble, rich) house', which had a variant **garda-* in Old Iranian. These variants are due to generalizations of the zero-grade and of the *a*-grade, respectively.[90] The form **gr̥da-* is seen, among others, in Avestan *gərəδa-* '(daevic) cave, burrow of obnoxious creatures' and in the Armenian loanword *gerdastan* 'household, body of servants and captives' (cf. Olsen 1999: 333 and 333[290]). From **garda-* derives for example Middle Persian *gāl* coll. 'the gang, the villeins labouring on the estates of the kings, the satraps, the magnates, etc.; in war their military crew' (Nyberg 1974: 80).

89 This is exactly parallel to French *domestique* 'servant', also originally a substantivized adjective *domestique* 'of the house', and also to Tocharian *mañiye* (s.v.), from an Old Steppe Iranian word that originally meant 'of/belonging to the house'.

90 Either of two related nouns with different accentuation: **gr̥dá-* ~ **gárda-* (cf. Hoffmann 1992: 840, 854[8] for a discussion of such variations in Iranian). I thank Nicholas Sims-Williams for providing me with this information and the reference.

As one can see in these examples, **garda-/gṛda-* already had from Old Iranian times, alongside the basic meaning of '(important) house', the metonymic meaning 'workers and people of that house, household'. It is more likely that **gṛdi̯a-* ~ **gardi̯a-* 'servants, people of the house' is derived from the basic meaning.

The notion of 'court people' or rather, of 'people attached to the court' appears in the Elamite form *kurtaš* discussed in detail in Henkelman (2018). He calls *kurtaš* "dependent workers" (2018: 224). As can be seen from the texts he analyzes, these *kurtaš* were foreigners, coming from Bactriana, Lycia, or even Sogdiana. They could be men, women or boys. This implies that the *kurtaš* were, so to say, there to stay (Henkelman 2018: 235). A possible difference between the status of **gṛdi̯a-* and that of **māni̯a-* could have been that the former were permanently included into a class or a group, as possibly were the *kurtaš* in the Achaemenid Empire (cf. Henkelman 2018: 239).

As can be seen from the semantic range of the Middle Persian, Elamite, Babylonian, Armenian, the type of house and the type of servants described by **gṛda-* etc. is of a high standing. Not every house had butlers, house-holders, bodies of servants (and captives!), etc. Clearly, the type of house and households we are talking of are those of noble people and of kings. A royal meaning is even found in one of the meanings of Middle Persian *gāl*: 'the villeins labouring on the estates of the kings'. In the Middle Persian form and in many others, as can be seen above, there is a plural or collective meaning 'group of servants'.

Thus the semantic trait 'lordly, royal' is already attached to the Iranian word, and does not need to be acquired in Tocharian, which is a strong point for Isebaert's etymology. Logically, it would have been a singular noun in Old Iranian still, and only made into a *plurale tantum* within Tocharian. The shift from 'court, servants' to 'palace', which may have been connected with its becoming a *plurale tantum*, would thus have been made within Tocharian.

If this word is indeed borrowed from Iranian, it needs to have been borrowed from Old Steppe Iranian because of its vocalism, i.e. TB *e* for OIr. **a*.

An Old Steppe Iranian **gṛdi̯a-* 'servant' would have been borrowed in Proto-Tocharian as **kercye* or **kercce* 'servant' → nom.pl. **ker(c)ciyi* obl.pl. **kerciyeṃ* 'body of servants, entourage',[91] which became our *plurale tantum* TB *kercci* 'palace', obl. *kercciyeṃ** (this last form proves that the nominative was originally **kercciyi*).

The geminate *-cc-* can be explained in three ways: (a) it is due to the preceding *r* (cf. *kertte* ← OSIr. **karta-*); (b) it is due to the simplification of the cluster

91 Compare TB *epastye* 'skillful' → nom. pl. m. *epasti* < **epástiyi*, for which see Peyrot (2021).

**ciyi* > **cyi* > *cci* in the plural, with influence of the spelling of the nominative on that of the oblique—however, **-iyi* most probably became *-i*, not *-yi* (cf. Peyrot 2021); (c) it is a regular Tocharian reflex, or correspondent, of Old Steppe Iranian **-Ti̯V-*, as one can perhaps see in TB *waipecce* ← Old Steppe Iranian **(h)u̯ai-paϑi̯a-*.

We cannot ignore an important problem: I have suggested elsewhere (s.v. *melte*; *speltke*; *welke*) that in Old Steppe Iranian, **rd* had become **ld*, while **rt* had become **rd*. Thus we would expect †*kelcci*. If we reject this sound law, the *-lt-* in *melte*, *speltke* and *welke* needs to be explained differently only in order to explain *kercci*.

I propose three solutions to account for the *r* of *kercci*: first of all it is possible that **-rdʲ-* (that is, **r* + palatalized **d*) had not, in Old Steppe Iranian and at the time of borrowing, become **ld*. This is of course *ad hoc* in the absence of any other example, but a phonetic explanation is available:

If we consider that OSIr. */d/ was realized as a dental, then **rd* > **ld* can be explained as an assimilation (/l/ being generally more dental than /r/ which is generally apical, cf. Ladefoged & Maddieson (1996: 182 f., 215 f.) for a detailed discussion).[92] In a cluster **-rdʲ-* the **d* would be palatalized and thus no longer dental.[93] Perhaps it was already closer to **[dᶻ] than to **[d̪]. Following this, **-D(i)yV-* (where *D* notes any dental sound) would be rendered as Proto-Tocharian **-c(i)yV-*, as seen in *waipecce* 'good, belonging' < **hu̯ai-paϑi̯a-*.[94]

A better solution would be that the PIr. sequence **-di̯V-* was already palatalized in Old Steppe Iranian (on this see the section 2.7.2.1). This solution differs in that it would suggest that this palatalization would be older than the sound law **rd* > **ld* of Old Steppe Iranian. This also implies that **rʃ* would have been borrowed into Tocharian as **rcc* and only **rd* in a non-palatal context had become **ld* in Old Steppe Iranian.

An even better solution, in my view, consists in viewing the sound change of **rd* to **ld* as more specifically **[rð] to **ld*, then explaining the palatalization of **rdi̯* becomes easier. For more detail on this solution, see section 2.7.2.11.

In conclusion, the inherited etymologies lack semantic ground for the incorporation of the 'lordly, royal' trait in the etymon, while Isebaert's etymology

92 This explains why, in some Iranian languages, both **rd* (< PIr. **rd*) and **rz* (< PIr. **rʃ*) became *l*.

93 I also believe that Tocharian *l* was alveolar or dental, because it was in opposition with palatalized *l* noted ‹ly›, and in this type of opposition the two elements need to be maximally differentiated in the place of realization. However, in Polish and Russian, the non-palatal *l* is velarized, which is also a possibility for Tocharian.

94 Palatalization in front of Old Iranian **-i̯a-* is also seen in Tocharian B *mañiye* ← OIr. **māni̯a*.

accounts for it. The meaning 'palace' is most probably a Tocharian B innovation from 'royal household', as Tocharian A uses a periphrasis instead. Thus, on a semantic basis, I favour an Old Steppe Iranian etymology for this word, although it requires some phonetic explanations in order to explain TB *-rcc-* from OSIr. **-rdi̯-* (instead of, for instance, **-lcc-*). In general it can be said that if this phonetic problem could be solved, this etymology would yield a precious amount of information concerning Tocharian society and culture, as well as Tocharian prehistoric architecture.

2.4.5 *TB* leṅke *'Valley, Cleft'*

The Tocharian B word *leṅke* means 'valley' or 'cleft' (DTB: 608). Its meaning is discussed by Bailey (1967: 304). Bailey considers the word *rrätu* 'cleft' in the story of *Mahākāśyapa* entering into trance in a mountain, and finds that it corresponds to Tocharian THT 12 a3 "where the phrase is 12a3 *śłyīye leṅke* mountain cleft". The exact citation is:

Book of Zambasta 20.281

> *hamata śśandā rrätu ya–nde ggarū sarbite käḍä mästä*
> *Myāño tṛṃkhānu samāhāña mahākālśavi āste*
> "The earth itself will split apart and a very large mountain will rise up. Amid its peaks, Mahākāśyapa will be sitting in meditation." (Emmerick 1968: 331).

Although the meaning must correspond, the equation is not perfect. No clear Indo-European etymology has been found for this Tocharian word.

As observed by Abaev (1973: 29) and Adams (DTB: 608) Tocharian B *leṅke* is very similar to Ossetic Iron *lænk* Digoron *læncæ* 'ravine' and possibly a cognate thereof. Due to the vocalism, the word should have been borrowed into Proto-Tocharian from Old Steppe Iranian **lanka-* or **langa-* or **lanak/ga-*. If this is correct, it is very important for two reasons, 1. this is the only example of the phoneme **l* in Old Steppe Iranian and 2. this word is otherwise specific to Ossetic only, thus this would constitute one more argument to suggest a strong kinship between Old Steppe Iranian and Alanic. In order to better understand the origin of this Old Steppe Iranian word, it is necessary to discuss the etymology of the Ossetic word. Abaev (1973: 29–30) connects it to some Slavic words, and he also suggests some place names as possible cognates: Sanskrit *Lankāpura*, Khotanese *Laṃggä-kamtha* 'Lanka city'. However, as these are toponyms designating the capital city of Ceylon, it is not possible to do much with the latter words. Viredaz (2017: 36[14]) views the Ossetic forms as borrowings from Slavic.

The Slavic root is probably **lę̀ga* f. 'depression', Russian *ljága* 'swampy place, swamp, depression', also in Baltic Lithuanian *léngė* (arch.) f. 'depression, small meadow between two hills', etc. (Derksen 2008: 276), and **lôgъ* (Derksen 2008: 288): Old Church Slavonic *lǫga* genitive sg. 'wood', Russian *lug* 'meadow', Czech *luh* 'damp depression overgrown with shrubs and trees', Old Czech *luh* 'forest, woods, grove, wooded meadow'. Here also, Baltic is represented with Old Prussian *Langa* 'name of a brook', *Langodis* 'name of a swamp'. According to Derksen, these words should also be connected to **lǭkà* f. 'low-lying meadow, water-meadow' (Derksen 2008: 288 f.) and he notes that the variation between roots with and without a nasal could imply a substratum origin, while evoking the influence of **lenk-* 'bend'. The root **lenk-* 'bend' is limited to Balto-Slavic (Derksen 2008: 277), so it is thinkable that it is of foreign origin as well. If the root **lenk-* was attested beyond Balto-Slavic, one could also imagine that it could be a plausible etymon for the Tocharian form, especially in the meaning 'valley'.

It is obvious that semantically as well as formally, the Balto-Slavic forms are very attractively similar to the Ossetic and Old Steppe Iranian forms. Nonetheless, it seems historically and geographically difficult to accept a Slavic loanword in Old Steppe Iranian, or even in Alanic (the ancestor of Ossetic) at the time of Old Steppe Iranian–Tocharian contacts. Old Steppe Iranian was situated far too much to the East for such contacts to have taken place, and no other such example can be mentioned.

An ancient cluster **-nk-* becomes *-ng-* in Ossetic, merging with PIr. **-ng-* (Cheung 2002: 31). The cluster *-nk-* derives from syncopated **-nVk-*, as in Iron *fynk*, Digoron *finkæ* 'foam' < **faṇ̥nakā-* (Cheung 2002: 17). One could thus posit a proto-form **lanaka-*, regularly yielding Iron *lænk*. The Digoron form *læncæ* needs to derive from **lanakṇ̥iā-* or **lanakṇ̥ia-* > **lænc* → *læncæ* with paragoge due to the rarity of *-ænc* in Digoron (similar to other examples mentioned by Cheung 2002: 55 f.). As per syncope rules discussed p. 169, an etymon **lanaka-* (probably with initial accent) could be reconstructed for Old Steppe Iranian as well, and it would regularly yield PT **lenke*, TB *leṅke*.

A Slavic origin of the Ossetic word is also not obvious, especially because it derives from **lanaka-*, and the Slavic words themselves seem to have a complex, unclear history. I thus wish to suggest two other hypotheses: 1. both (Balto-)Slavic on the one hand and Ossetic and Old Steppe Iranian on the other borrowed this word from another source language; 2. the similarity between the Ossetic and Old Steppe Iranian word and the Slavic words is due to coincidence. In any case, given the general proximity between Ossetic and Old Steppe Iranian (see § 2.7.2 and § 2.7.3), it seems likely that these two words are, indeed, connected. A last point which is unclear, is whether **lanaka- vel sim.* was bor-

rowed independently in Ossetic and Old Steppe Iranian, or only once in the common ancestor of both languages. Interestingly, this example shows that **l-* existed in Old Steppe Iranian, perhaps only in words of foreign origin, and that **l* in Ossetic existed even before secondary *l* < **r* in the vicinity of **i*, a development of which there is no trace in Old Steppe Iranian.

In summary, I propose that Tocharian B *leṅke* 'valley, cleft' derives from Old Steppe Iranian **lanaka-*, which I reconstruct on the basis of Ossetic Iron *lænk* Digoron *læncæ* 'ravine'. However, as the Ossetic word has no clear etymology, this proposal remains not entirely certain.

2.4.6 *TA* wankā- *'To Joke, Chatter'*

Tocharian A *wankā-* 'to joke, chatter' (Peyrot 2013: 813; DToA: 429), deverbal noun TA *waṅke* 'chattering' (DToA: 422; cf. Peyrot 2012: 212, with references), has no clear etymology. Isebaert (1980: 90–91) has proposed to derive it from Iranian. Indeed, as he notes, there is an Iranian form **u̯ānk(a)-* (or **u̯āng(a)-*) 'sound, cry' (Bailey writes "*vank-* (or *vang-*) 'make sounds'" DKS: 373) which can be reconstructed.

Khwarezmian has *w'nk* 'Ruf' (Benzing 1983: 635); Middle Persian has *wāng* 'voice, call, cry, noise' (DMMPP: 335), borrowed into Armenian *vang* 'Laut, Ton, Silbe' (cf. Hübschmann 1897: 243); New Persian *bāng* 'loud shout, outcry', Gilaki *väng* 'cries, lamentation' and Balochi *gwānk* 'cry' (Korn 2005: 99); Zaza *vang*, *van*, Bakhtiari *bang* 'shout, shriek, hail', and the list can be much longer.[95]

This root in Iranian is strictly nominal, and I was unable to find any primary verbal form relating to it. Bailey (DKS, s.v.) saw it in a few Khotanese verbs, but better etymologies have been found: *vaj-* 'to dispute' has been tentatively derived from **au̯a-Hi̯au̯d-* (Emmerick apud Degener 1989: 101);[96] *pyūṃj-* 'to deny' is convincingly explained by Federico Dragoni (p.c.) as deriving from the root **u̯a(n)č-* 'to stagger, shake' ('to shake away' > 'to deny'). As to Balochi *gwanǰ-* 'to cry', it is convincingly explained by Korn (2005: 137; already in essence in Gershevitch 1971: 284) as a back-formation based on the past stem of *gwānk ǰan-* 'to cry' (lit. 'to strike a cry'), *gwānk ǰat*, reinterpreted as a simplex *gwanǰat*.

It has been proposed that this **u̯ānk(a)-* derives from **u̯ač-* 'to speak', with a "nasal increment", or a nasal infix (cf. Gershevitch 1971: 279 f.). Alternatively, it has been suggested that it derives from Old Iranian **u̯āna-ka-*, a cognate of the Vedic *vāṇá-* 'voice, music' (cf. Hasandoust 2014: 398 f., with references).

95 For the Bakhtiari shortening of **ā* before *-ng*, cf. Vahman & Asatrian (1987: 71).

96 One often finds the spelling *vaṃj-* but, according to Emmerick's etymology, it might be better interpreted as *vaj-*, with unetymological *anusvāra*, common in Late Khotanese, and the spelling *vaj-* is more precise (I thank Federico Dragoni for informing me of this).

The first hypothesis is a bit difficult because of the *-k-* consonant, the second one because we would expect **u̯ānaka-* as a proto-form (yielding, for instance, Balochi ***gwānak*, etc.), and also because a retroflex *ṇ* in Sanskrit does not regularly correspond to **n* in Iranian.

In both cases we could imagine the influence of the mimophone "*bang*" (English *bang*, French *pan*). The first part of the word could be connected to Vedic *vāṇá-* while the element *-ng* could be due to the influence of the mimophone. A slight problem is that the mimophone usually starts with a *b-* or *p-*, because these sounds imitate an initial explosion (that of the original sound). Perhaps an even better solution would be to see it as a cross-form between the root **u̯ak-*/**u̯ač-* 'to speak' and this same mimophone *bang*. This hypothesis also helps explain why this root is strictly nominal in Iranian.

This discussion serves to explain the appearance of this purely Iranian nominal root, which is unusual from an Indo-European point of view (there is nothing in Indo-European that could yield a **-ng-* or **-nk-* sequence in Proto-Iranian).[97] If the Tocharian verb were also derived from the mimophone "*bang*", it would rather be something such as †*pańkā- vel sim*. Most probably the Tocharian word thus derives from Iranian, as already suggested by Isebaert (1980: 90–91). The semantic pathway is very straightforward 'noise' > *'rumor, unimportant speech' > 'chatter, chitchat, chat'. It is a quite commonly attested semantic change: for instance French *bruit* 'noise; gossip'. A parallel semantic change is probably behind the Sogdian form *wnxrš* 'rumour, news'.[98]

The question that now concerns Tocharian is: what is the exact source, and was the verb borrowed first, or the noun? One might think it is a Pre-Bactrian or Bactrian borrowing, *οαγγο → **wańk*. From this noun a verb *wankā-* would have been made. The Bactrian form itself is unattested, but it is very likely to have existed, as this root is (almost) pervasive throughout Iranian.

There is a possible problem, since the first vowel would be expected to have been long, and thus probably yield †*wāńke*. The verb can derive from either a long or a short vowel in the first syllable, but no Bactrian word was recorded as being borrowed as a verb in Tocharian to this day. If it is not a borrowing from an unattested Tocharian B word, the Tocharian A form rather constitutes an abstract derived from the verb (cf. Peyrot 2012: 211 f.).

97 Schwartz (1969: 446) saw such a root reflected in Sogdian *wnγr* /wanxr/, comparing also OInd. *vaṅkú-* (as 'noisy') and mentioning Tocharian A *wańke*. Prof. Schwartz now tells me that the OInd. word should be removed from consideration since it is semantically uncertain, and since the root is limited to Iranian, he endorses an Iranian origin for the Tocharian A word.

98 I thank Prof. Nicholas Sims-Williams for informing me of this example.

Another possibility is that the Tocharian A words are of an older date, and derive from Old Steppe Iranian. In this case, it would be a Proto-Tocharian borrowing **wanke* 'gossip' ← Old Steppe Iranian **u̯ānka-*, or **u̯ānaka-* (or **u̯ānga-*) 'noise; gossip' although this latter etymon would normally yield TA **wāṅk*. A verb would then easily have been derived therefrom, according to the same model as the other Tocharian verbs deriving from Old Steppe Iranian nouns (for instance s.v. *raitwe*). Then, to explain the vocalism of the noun TA *waṅke*, it would have been an inner Tocharian A creation, a secondary abstract in *-e*, as suggested by Peyrot (2012: 211 f.). Although it is difficult to ascertain this, it is in my view a more likely solution. According to this hypothesis, Tocharian B lost this lexeme, and replaced it with *käskor** at least for the noun.[99] Here again, one can evoke Professor Schwartz's solution of a proto-form **u̯ank-* (with only secondary lengthening), to explain the vocalism of the Tocharian A form, since Old Steppe Iranian **u̯ank-* would regularly yield Tocharian A *waṅk-*. However, this would imply that the word was borrowed directly as a verb, for which we have no parallel.[100]

To sum up, a possible trajectory for the Tocharian A word *wankā-* 'to chat, gossip', without going through a hypothetical Tocharian B loanword, is as follows: Old Steppe Iranian **u̯ān(a)ka-* 'noise; gossip' would have been borrowed as a Proto-Tocharian noun **wanke* 'gossip', from which a verb PT **wanka-* (prs. *wonko-*) was made, and later a Tocharian A deverbal noun *waṅke* was derived from the verb.

In conclusion, this discussion involves multiple assumptions and hypotheses, since we cannot be entirely sure of the etymology of Tocharian A *wankā-* 'to chat, gossip'. However, I consider an Iranian origin of this verb very likely, as no other origin can easily be put forward and we have a suitable Iranian source-word.

---

99 Alternatively, it is possible that TB **wanka-* belonged to a much lower sociolect and was not represented in the texts, or perhaps even that it is simply not yet attested in our extant corpus.

100 Professor Martin Schwartz has suggested another solution to me, which I cite verbatim: "I suggest that an Iranian root **u̯ank-* 'to make a sound', reflected in Sogdian /wanxr/ 'voice' < **u̯ank-ra-*, gave the noun **u̯ankV-* in Old Steppe Iranian, which is reflected in Tocharian A *waṅke*, etc. The early date of this event left the short vocalism of the first syllable (cf. the etymologically unrelated stem Avestan *x^{v}anaŧ-* 'making a noise' in compounds) unaffected by the later development whereby **hu̯ānaia-* furnished the verb stem for 'to call' etc. (> Manichaean Middle Persian *xwān-*, Khotanese Saka *hvāñ-*, Khwarezmian *m|ʾfxwʾny-*, Bactrian χοαν-, χοην- and Ossetic *xon-*).

I suspect this affection occurred via phrasal collocation of **hu̯ānaia-* with **u̯angV-*, giving **u̯āngV-* reflecting in Middle Persian *wʾng*, Khwarezmian *wʾnk*, Balochi *gwānk*, etc. It seems that **hu̯ana-* survived in Parthian *xwn-*, Sogdian *xwn-* and Bactrian χοαν-."

2.4.7 TA nātäk *'Lord', fem.* nāśi *'Lady'*

The Tocharian A word *nātäk* 'lord' (nom.pl. *nācki*), corresponding semantically to Tocharian B *saswe*, and its feminine counterpart *nāśi* 'lady' have previously been seen as cognates of Greek ἄναξ, -κτος, 'lord, master' (cf. Mycenaean *wa-na-ka*, Beotian Fάναξ, etc) and ἄνασσα, -ης 'lady, queen' (Winter 1970: 53; Adams 2017a:1376).[101] This connection is rightly rejected by Del Tomba (2023: 96), who notes that the initial **u̯-* seen in the Greek forms should not have been lost in Tocharian, and that Greek has *-kt-* while Tocharian shows *-tk-*. The etymology of the Greek word is difficult (cf. Beekes 2010: 98).

Del Tomba concludes that "TchA *nātäk* cannot be derived from any internal source" and alternatively proposes it to be a borrowing from Skt. *nāthá-* m. 'protector, possessor, lord' (2023: 96 f.). As Del Tomba (loc. cit.) notes, this noun is frequently attested in epithetic compounds such as *nāka-nātha*, *nāka-nāthaka-* 'sky-lord' (epithet of Indra),[102] *loka-nātha* 'saviour of the world (epithet of the Buddha)', etc.

Although very attractive from the semantic perspective, Del Tomba's Sanskrit etymology of TA *nātäk* 'lord' is difficult on the formal level. From Skt. *nāthaka-*, one would rather expect TA †*nātak* (type Sanskrit *kacchapa-* 'turtle' → TA *kāccap*, cf. Schwartz 1974: 406). Another problem, in my opinion, is the archaic nature of the morphology of Tocharian A *nātäk*: the plural *nācki* and the feminine *nāśi* 'lady'. The form *nātäk* seems to reflect a Proto-Tocharian syncope (PT **nateke* > **natke*, of the type of OIr. **rataka-* > PT **retke*, TB *retke*, TA *ratäk* 'army'), and it cannot be inherited (cf. Del Tomba 2023: 96). Those two facts brought together strongly suggest a very early date for the borrowing of this word, namely, that it is a Proto-Tocharian borrowing, despite the fact that no Tocharian B cognate has yet been found for this word.

The vowel *ä* in *nātäk* could hardly be due to an early Tocharian A sound change. For older loanwords and inherited words, there existed a rule of vowel weakening: in the second syllable, Pre-TA **ā* was reduced to **a* if the first syllable contained a vowel **a* or **ā* or a diphthong (cf. Kim 2007, with references). However, this rule is not found in Sanskrit loanwords (e.g. *āsāṃ* 'throne') nor in late Iranian loanwords. The spelling *kātäk* 'householder' is sometimes found in the literature as a counterpart of TB *kattāke* 'id.' (e.g. Pinault 2008: 234), but the form *kātäk* is not found in any text with the meaning 'householder', and as

101 To note, Winter (1970: 53) suggests that both the Greek and the Tocharian words are of (shared) non-Indo-European origin.

102 I wonder if the translation 'supporter of the sky' (as in, the one who keeps the sky from falling) would not be more appropriate.

the counterpart of TB *kattāke* one should rather set up *kātak** (cf. DToA: 103).[103] There is thus no clear example of a secondary *ä* < *a* in the (TA) second syllable in any trisyllabic Indic loanword.

No borrowings from Sanskrit in Tocharian that predate the separation of Tocharian A and B are known, that is, Sanskrit borrowings for which we have a regular correspondence between A and B, as we find for Old Iranian words. Although in this case there is no known Tocharian B correspondent, the Tocharian A word is still much more archaic in its derivation than any Sanskrit loanword in Tocharian. I cannot accept Del Tomba's assumption (2023: 97) that the word was borrowed into Proto-Tocharian from an intermediary language, although tempting, suffers from the same difficulty concerning the second vowel of the word, as admitted by Del Tomba himself (2023: 97). Such a solution would also suggest a rather complicated inflection and derivation for a word taken from a Middle Indic or Middle Iranian language.

Because it would explain the problems mentioned above better, in particular the syncope or the vowel-weakening, I prefer to see this word as an Old Steppe Iranian loanword in Proto-Tocharian. The major obstacle to this etymology lies in the fact that there is no word **nāʔaka-* attested in Old Iranian. This is, however, only a superficial problem, as I hope to demonstrate.

First, the basic meaning of Ved. *nāthá-* is 'refuge, help', cf. for example *anāthá-* n. 'Schutzlosigkeit'. This word has no clear etymology: a root **h₃neh₂-* 'to help', suffixed with **-tHo-* thus yielding the meaning 'helping' cannot yield the correct meaning, since no root with such a meaning is attested, cf. Frisk (1960: 395 f.; LIV²: 302; *pace* EWAia²: 33 f. and Beekes 2010:1083 f.).

It has been suggested by Kroonen (2013: 388) that Vedic *nāthá-* is connected to Proto-Germanic **nēþō* f. 'mercy, safety'. He thus projects *nāthá-* 'refuge, help' back as PIE **nēt-h₂-o-*, and further connects the Indic and Germanic forms to a Proto-Indo-European root **net-* 'to protect, favor', cf. Proto-Germanic **neþan-* s.v. 'to grant safety (?)'. In any case, Kroonen's reconstruction (**nēt-h₂-o-*) is preferable to the earlier postulation of a root **h₃neh₂-* 'to help', for the reason evoked above (simply that such a meaning is not found, or not primary in the Greek forms such as ὀνίνημι, primarily meaning 'to be of use').

---

103 The form *kātäk* found in Tocharian A texts can rather be interpreted as the 3sg preterite active of *kātkā-* 'to arise'. The DToA (p. 103) writes, concerning *kātak**, that it is borrowed from Middle Iranian, "cf. Khot. *ggāṭhaa*, itself borrowed from M[iddle] I[ndian], cf. Gāndhārī Pkt. **ghahaṭha*, from Skt. *gṛhastha*, see Pinault (1995: 21–24)." Dragoni (2023: 72) considers the immediate origin of *kattāke* and *kātak** unclear, and a direct derivation from Khotanese difficult (the *-e* ending of the TB form suggests a late borrowing, the reflected *-ak-* suffix suggests a prehistoric loanword).

In Indic, the meaning 'saviour', 'lord', 'protector' might have been back-formed, or semantically reinterpreted from the epithetic compounds mentioned above. This well-known process can be described as follows: *loka-nātha-* initially meant 'refuge of the world', but since it was attributed to an individual or a god, it was taken to mean 'protector of the world', and *nātha-* was thus reinterpreted as a word meaning 'protector'. Another possibility is that a neuter **nātha-* 'protection' was derived in **nāthaka-* 'protector' and the meaning 'lord' of *nātha-* m. is secondary, as I explain below.

Once reinterpreted as a noun 'protector, lord' or 'saviour', *nātha-* could easily be enlarged with *-ka-*, as shown in the examples above (for more examples, see Del Tomba 2023: 96). For the Iranian part, it would be easy to simply suggest that **nāϑa-* 'protection' was enlarged with the denominal suffix **-ka-* and thus took on the meaning of 'protector, lord'. The semantic shift from protector to lord is not difficult, especially in ancient times when lords were, first and foremost, protectors, or supposed to be.

The Indo-Iranian noun **nāϑa-* is not attested in Old or Middle Iranian, but seems to subsist in at least one word: New Persian *panāh* 'protection, refuge' < **pad-nāh* < **pati-nāϑa-* (Hübschmann 1895: 43).[104] This etymology of Hübschmann's has long been neglected, with many preferring to see this word as a continuation of the Persic cognate of Avestan *nas-* 'hingelangen zu' (so Nyberg 1974: 150). In my view, this second explanation fits neither the form nor the meaning of New Persian *panāh*: indeed, 'to go somewhere, to return somewhere' seems rather far removed from the notion of 'protection'. Rather, Hübschmann's etymology should be retained as I will explain below.

I would also like to adduce another Persian word which, in my opinion, goes back to OIr. **nāϑa-*. It is MP *nihān*, NP *nihān* ~ *nahān* noun, 'secret, hiding' adj. 'secret, hidden', *panhān* < **pad-nahān* adj. 'secret, hidden'. This word is usually explained as deriving from **ni-δāna-* 'set down' (from **dā-* 'to set') (so Monchi-Zadeh 1990: 106[132], 127; Hasandoust 2014: 2814). This cannot be correct, as *-δ-* does not become *-h-* but *-y-* in Middle and New Persian. Furthermore, PIIr. **ni-dʰāna-* seemingly means 'treasure' (lit. 'what has been deposited'), as one can deduce from the comparison between Sanskrit *nidhāna-* 'treasure' (M-W: 548) and Middle Persian *niyān*, also 'treasure' (DMMPP: 252).

In my view, *nihān* derives from **nāϑa-* 'protection, refuge' suffixed with *-ān*. In the present case, this would not be the verbal suffix *-ān* (deriving from the athematic present middle participle **-āna-*, cf. Bartholomae 1895–1904: 109 f.

104 Hübschmann (loc. cit.) was still unsure whether **patn-* could yield *pan-*, but I believe this is now firmly established (cf. e.g. *panhān* < **pad-nahān*, discussed here as well).

and Cheung 2015: 273) but the "other" suffix *-ān* (synchronically different: while it originally goes back to the middle participle in **-āna-*, it is mostly nominal and often non-analyzable already in the Middle Persian period), of the type of Middle Persian *wiyābān* 'astray, wandering' (cf. Cheung 2015: 273), and Parthian *žafrān* 'depth' (from *žafr* 'deep'). With the addition of this suffix, the word took a progressive, continuous meaning, so that we can propose the following semantic shift: *'protected' > *'covered' > 'hidden, secret'.[105] As to the form, we must first posit reduction of the **ā* of the first syllable, which is common before **h* (see Pahlavi ‹šh› *šah* 'king', cf. also Lazard 1963: 182), but could perhaps also be explained as a dissimilation (**nāhān → nahān*).

The *-i-* vocalism of Middle Persian *nihān* is, I contend, due to the influence of the verb *nihuft-/nihumb-* 'to wear, to hide'. Both words have strongly influenced each other. As is known, *nihuft-* derives from **ni-gufta-*, but **-g-* does not regularly become *-h-*. In the present case, it has been proposed that the *-h-* of *nihuft-/nihumb-* is due to the contamination of *nihān* (cf. Hasandoust 2014: 2814). I thus also propose that the *-i-* vowel of *nihān* is due to contamination with *nihuft-/nihumb-* (and the *-a-* in NP *nahān* is secondary, cf. Pisowicz 1985: 15). There probably was also interaction between the meanings of both *nahān* and *nihuft-*.[106]

Now that I have given my arguments for the existence of an Iranian word **nāϑa-* 'protection, refuge',[107] it is not difficult to propose an Old Iranian **nāϑaka-* 'protector, lord', along with **nāϑa-* 'protector', either back-formed from the same type of compound, as in Sanskrit, or rather from **nāϑaka-* 'protector, lord' itself. The semantic shift from 'protection, refuge' to 'protector, lord' is facilitated by the very common phraseology of the type "you are (like) a refuge", French *tu es mon refuge*, or *tu es ma protection*, usually said in a roman-

105 It is not even necessary to posit a verbal form **nāϑ-* 'to protect', since the suffix *-ān* could simply be added to adjectival or nominal forms, once most of the inherited words containing this suffix were no longer perceived as verbal or rather deverbal forms.

106 One could also think that **nāhān* directly became *nihān* under the influence of *nihuft-*. In that case the *-a-* vowel of *nahān* is simply due to assimilation of the second *-ā-*, as occurred in Iranian Persian in other words.

107 Federico Dragoni has suggested a few words to me whose semantics are attractively close to that of the concerned root. These are Khot. *ānatu* used together with the verb *yan-* ('to do') in the collocation *ānatu yan-* 'to take care of, protect', and with *häm-* in the meaning 'to be taken care of, to be protected' (cf. also DKS: 18); Christian Sogdian *'ntqy'* 'modesty', and some other cognates. Although the forms look strikingly similar, with a semantic derivation that is far from absurd ('protecting' > *'hiding' > 'modest'), I do not know why they present a *-t-* and do not reflect a **-ϑ-*. Perhaps one can think of a PIE **nēteh$_2$-* 'protection' → **neth$_2$-o-* 'relative to protection, protecting, refuge' to explain PIIr. **nāϑa-* and **ā-nat-* (< PIE **h$_1$ed-net-*) 'conferring protection' to explain PIr. **ānata-*.

tic context, but not far removed from a literal use (that is, when said to the lord, who literally is meant to be the protection of his people). The Old Steppe Iranian word **nāϑaka-* 'protector, lord' was borrowed in Proto-Tocharian as **natke*, which regularly yields Tocharian A *nātäk* (and should have yielded TB †*nātke*).

The derivation of TA *nāśi* 'lady' is somewhat more complicated. As Del Tomba (2023: 97) notes, palatalization of the cluster *-tk-* yields *-ck-* and not *-ś-*. This is obvious from the palatalized nom. plural of *nātäk*, *nācki*. Del Tomba's suggestion (2023: 97) of a Middle Indic continuant of **nāthyā-* 'the one related to the lord' with a similar result as TA *āmāś* 'minister' ultimately from Skt. *amātya-* is tempting, nonetheless *ad hoc* as Del Tomba admits (2023: 97). In my view, it is clear that *nāśi* is related to *nātäk* with the addition of the **-ya-* suffix of the *aśiya*-type, although the matter is intricate.

The first proposal is difficult to accept because the syncope of **nateke* happened very early: this type of syncope (see section 2.7.2.7) occurred either in Old Steppe Iranian or in Proto-Tocharian already. In that sense, I am unsure whether we can even reconstruct a **nateke* stage. Most of all, I do not see how there could have been both *k* and *ś* in the sequence *-təkśʸ-*. In any case, there is no parallel for a simplification or palatalization of such a sequence to *-ś-*. The second proposal seems less complicated. Of course, the palatalization **natkya* > **naśśi* is unprecedented and *ad hoc*, but it is the best solution at hand, it seems.

In conclusion, I have tried to show the existence of Proto-Iranian **nāϑa-* 'protection, refuge', a cognate of Vedic *nāthá-* 'idem'. I have also argued that TA *nātäk* 'lord' ultimately derives from Old Steppe Iranian **nāϑaka-* or **nātaka-* (depending on whether there existed a phoneme /ϑ/ in Old Steppe Iranian, see section 2.7.2.1) meaning 'lord' or 'protector'. As to the exact form of TA *nāśi* 'lady', there is no entirely convincing explanation, although Del Tomba (2023: 97) has advanced two hypotheses, one of which seems more convincing. In any case, the problems of the derivation of TA *nāśi* 'lady' exclude a borrowing from Sanskrit.

2.4.8 *TB* tˢəw- *'Attach Oneself to, Stick to', TA* tˢäw- *'To Put Together'*

Before discussing the etymology of Tocharian B *tˢəw-* 'add' and Tocharian A *tˢäw-* 'to put together', I wish to present and discuss the data at hand. Peyrot (2013: 846) sets up the following verbs:

– TB *tˢəwa-* 'attach oneself to, stick to' TA *tˢäwā-* 'fit (intransitive, active); obey (middle)'
– TB *tˢəw-* causative 'add (transitive)' TA *tˢäw-* causative 'put together'.

Peyrot reconstructs for both languages a Proto-Tocharian form **tˢəwa-*. Adams (DTB: 808), giving the Tocharian B form as *tsu-*, translates it as 'cohere, adhere;

embrace, contain' (active); 'adhere, stick, cling, attach oneself' (middle) and causative 'make cohere, add to'; cf. also Pinault (2016b: 183f.) who adds other meanings such as a causative meaning 'prepare, compose a work' parallel to the Sanskrit causative *yojayati* of *yujyate*. Adams further writes that the etymology of this verb is uncertain. He mentions a connection proposed by Seržant (2007) of this verb with Greek δύναμαι 'to be able' and Gothic *taujan* 'to make'.

Greek δύναμαι means 'to have the capacity of doing something potentially' (like 'I can swim'). Thus, the meaning of the Greek verb is far off that of the Tocharian one. In the Addenda and Corrigenda to the LIV^2 (available online)[108] Kümmel posits a PIE root **dehꜧu̯-* 'zusammenfügen, ordnen' with only the Tocharian verbal forms.

Apart from the semantic weakness of these connections, we would in Tocharian expect **śew-* from **dehꜧu̯-* (> **tsēw* > *tsyew-* > **śew-*), or perhaps **śaw-*. This is not a major problem, since the initial could have been generalized from a different ablaut grade, for instance from the zero grade. The lack of cognates is more problematic for the Tocharian etymology.

In my view, an Iranian etymology is also possible. If we start from the causative meaning 'to add to, to mix', and see the meaning 'adhere, stick to' as a back-formed decausative, we have a good Iranian candidate: Proto-Iranian **ʃau̯-* 'to pour a libation',[109] reflected in Avestan *zaoϑra-* 'libation', Buddhist Sogdian *zwt'k* and Christian Sogdian *zwty* 'beer, liquor' (according to Gauthiot 1913: 102 it could mean 'wine'), Yazgholami *zaw-*/*zod* 'to tread down, compress, squeeze', etc. (cf. EDIV: 471f.).

It is the meaning 'compress, squeeze' which I find quite strikingly close to the Tocharian meaning of 'put together, make adhere, fit'.[110] In my view, Buddhist Sogdian *zwt'k* 'wine' and Christian Sogdian *zwty* 'alcoholic liquor' (cf. Sims-Williams 2014: 102) originally comes from **zuta-ka-* 'what has been squeezed, trodden down', and refers to the process of squeezing fruit for the fabrication of liquor, or of treading down the grapes to make wine.[111]

From 'to squeeze fruit' to 'to make adhere, add', the semantic pathway is not evident. However, in a number of languages the root **ʃau̯-* gave a word for

108 https://www.academia.edu/402269/Addenda_und_Corrigenda_zu_LIV_ (accessed on the 20th of May 20, 2021).

109 PT **kəw-* 'to pour' is the inherited Tocharian cognate of this Iranian root.

110 The word for 'pus' is also related to this verb 'to pour' in Ormuri *zū*, Waziri *zawa* 'pus, matter' (cf. Morgenstierne 1932) and in Khotanese *ysūa-* gen. *ysūna*, from *ysun-* 'to pour, strain', cf. DKS: 353. Pus needs to be squeezed and compressed to be extracted.

111 Henning (1946: 720) wonders if there is a connection of the Sogdian word to Greek ζῦτος, ζῦθος, an Egyptian type of beer (Lat. *zythum*), which, he writes, has an unclear etymology. He proposes that both the Greek and the Sogdian words have a common origin, "possibly in some Scythian language".

'broth': Khotanese *ysūma-* 'broth', Waziri *zēmna* 'soup' (DKS: 353). Interestingly, Tocharian B *sumo* 'libation, oblation' was probably borrowed from Khotanese *ysūma-* 'broth' (cf. Dragoni 2023: 196–197). Broth is constituted of hot water to which bones, meat or vegetables are added, or put together. If, again, we depart from the meaning 'to put together, to add' that the Tocharian A and B have in their causative, there is a much stronger semantic connection to be made between the Iranian forms and the Tocharian.

If this is correct, although the data at hand seems a bit unsure, and the semantic details still need to be made clear, it is possible that the Proto-Tocharian verb **tˢəwa-* 'to add, to put together (?)' derives from an Old Steppe Iranian verb **dᶻu-* 'to add to a liquid, to compress together'. The semantic divergence from the original meaning of 'to pour (a libation)' can perhaps be explained through a preverb which was later removed, for instance **ni-dᶻu-* 'to pour down (as a recipe)'. It is perhaps more likely, as no direct verbal borrowing from Old Steppe Iranian has been found out yet, that there was an Old Steppe Iranian noun **dᶻu-ta-* or **dᶻu-ka-* 'mixture, broth' from which the verb **tˢəwa-* would have been extracted within Tocharian, or perhaps, which would have influenced the Tocharian and the Old Steppe Iranian meaning. It is also possible that the verb is inherited from Indo-European, according to the model I suggested above.

To conclude, the Proto-Tocharian verb **tˢəwa-* and its Tocharian A and B descendants TB *tˢəwa-* 'attach oneself to, stick to' TA *tˢäwā-* 'fit (intransitive, active); obey (middle)' could be inherited, deriving from a zero-grade **dh₁u-* (in my view from **deh₁-* 'to bind' with a present stem in *-u-*) becoming **duh₁-* with regular metathesis and resulting in Proto-Tocharian **tˢəwa-*. It could also likely be derived from an Old Steppe Iranian form **dᶻu-* ultimately going back to Proto-Iranian **ǰau̯-*.

2.4.9 *TB* tseriteke *'Young Grass (?), Lizard (?), Turtle Shell (?)'*

As pointed out to me by Federico Dragoni, Tocharian B *tseriteke* has a very Old Steppe Iranian appearance, despite some phonetic problems (see p. 163 f.). Its meaning is unclear. Below I cite one of the sentences in which it appears:

THT 324 a3

(*me*)*nāktse ṣamāne: tseriteke menākäccepi* ///

"a monk comparable with … to the comparable […]" (Ogihara 2009: 406).

Based on this context, the meaning of *tseriteke* cannot be established. However, if one takes this word to be of Old Steppe Iranian origin, it is tempting to see in it the rendering of **dᶻaritaka-*, a derivative of **dᶻarita-* 'green'. Ossetic *zældæ*,

the descendant of Old Iranian **dʒarita-*, means 'young grass, grass' (cf. Cheung 2002: 253–254). If this meaning is very provisionally accepted for Tocharian B *tseriteke*, one can perhaps restore it elsewhere:

THT 3279 a6
/// – *r kaccāp tseri*(*tekene*) ///

The Tocharian B word *kaccāp* meaning 'turtle, tortoise', one could venture to think that a tortoise would naturally be in grass. Then perhaps one could read the first occurrence as meaning "a monk comparable to young grass", meaning "a young monk". Indeed, the metaphor of green grass, used to designate young, inexperienced people is quite common cross-linguistically; cf. French *un vert*, Dutch *groen als gras zijn* 'to be green as grass', *een groentje* 'a novice, uninitiated'. Of course, in the lack of a clear context or of bilingual evidence, this remains provisional. One could perhaps think as well of a turtle or a tortoise in a shell, as turtle and tortoise shells are also green. The translation is based completely on the assumed etymology and cannot be confirmed by textual evidence. Furthermore, it remains problematic that Old Steppe Iranian **i* would be reflected as Tocharian B *i* in this position (see section 2.7.2.6).

There is another occurrence of *tseriteke*, which I cite below:

PK AS 15D b4
(*we*)*śeñaṃ su mā kca ṣäṃṣāte tserītekets wiyälyñe ramt* ///
"Il ne tint pas compte des voix [des démons] … comme un effroi de *tseriteke*" (Athanaric Huard, p.c.)

As Athanaric Huard (p.c.) tells me, since we do not know if *tserītekets wiyälyñe* refers to the previous part of the sentence or not, this example is not very helpful. Nonetheless, from the context it seems unlikely that it refers to 'grass', as 'grass' can hardly be afraid. It could possibly refer to a green or yellow animal, such as a lizard.[112] In fact, lizard could function also for THT 324 a3, which probably is from a chapter on diseases, where skin diseases are discussed. It could thus designate ichthyosis, a disease which makes one's skin look like that of a lizard, scaly, dry, and thickened.

112 Athanaric Huard (p.c.) informs me that this word could correspond to Tocharian A *tsuri*, found in a A437 b6 together with other animal names. However, the vocalism of the two forms is difficult to reconcile.

In conclusion, *tseriteke* could be a word of Old Steppe Iranian origin meaning '(young) grass', 'lizard', or perhaps 'turtle shell', depending on how we interpret the scarce evidence we have at hand. The meaning '(young) grass' would fit with the Ossetic meaning of *zældæ* 'young grass, grass'. Whatever its meaning, the appearance of TB *tseriteke* makes it very likely that it is an Old Steppe Iranian word. If the meaning 'grass' is wrong, another meaning such as 'green' or 'yellow (thing)' could also be possible for this word.

2.5 Old Steppe Iranian Borrowings: Rejected Cases

In this section, I will discuss words which have been proposed to be or could perhaps be thought to be of Old Steppe Iranian origin, but should in my view be derived from another source or cannot at present be plausibly attributed to Old Steppe Iranian.

2.5.1 *TB* atiya*, *TA* āti *'Grass; Straw'*

Tocharian B *atiya** can mean either 'grass' or 'straw' (in THT 194 b1 a fire of *atiya** is mentioned, which can hardly imply green grass). In a bilingual text, THT 530 b3, *atiya** corresponds to Sanskrit *tṛṇa-*. The word *tṛṇa-* itself has a wide range of meanings, such as 'grass; herb; any graminous plant; straw' (M-W: 453) and this wide range of meanings could very well be reflected in Tocharian texts, with calqued meanings as is not unusual. The precise meaning of Tocharian A *āti* is 'grass' as well, as in A 92 b2 *ātyāsā lmo naṣt* "you have taken a seat on the grasses" (DToA: 33).

Hilmarsson (1996: 51, with references) viewed these two words as cognates of Latin *ador* 'spelt, emmer wheat' and as deriving from PIE *h_2ed- 'to be dry', enlarged within Tocharian with the suffix -(*i*)*ya*. However, the type of grass referred to by these Tocharian words is not necessarily dry, and an Iranian derivation of this word might perhaps be considered.

Tocharian **atiya* could in theory derive from an Old Steppe Iranian form *$*ādi̯i̯ā$-. However, evidence within Iranian for such a form is extremely slim, as it is based on Scythian αδιγόρ 'locust'. This word has been analysed as consisting of *$*ādī$- 'grass' and *$*x^{w}ara$- 'eater', a locust being a "grass-eater" (cf. DTB: 9). However, this comparison remains extremely weak in view of the uncertain analysis of this "Scythian" word and its apparently very limited distribution within Iranian.

One might alternatively think that *$*ādī$- is related to PIr. *$*ādu$- 'cereal', on which a lot has been written (see for instance Rossi 2010), but little consensus has been reached. PIr. *$*ādu$- is based mainly on Sogdian *$^{\nu}$δwkh* 'grain,

cereal'.[113] It would take too long to enter on a discussion of these words here (cf. most recently Kölligan 2020: 227 f.), but it is unclear how a stem **ādī-* could have been formed to **ādu-*. One could perhaps imagine that an Old Steppe Iranian **ādu-* was borrowed into Proto-Tocharian as **at*, which was later suffixed with **-iya*. One can wonder why there was no assibilation, although *epastye* could provide a parallel. This etymological option, too, remains too hypothetical to be acceptable.

In conclusion, the etymology of Tocharian B *atiya**, Tocharian A *āti* remains difficult. A derivation from an Old Steppe Iranian form **ādi̯ā-* would work formally, but such a word cannot be safely set up on the sole basis of Sogdian *"δwkh* 'grain, cereal' and Scythian αδι* in αδιγόρ 'locust'.

2.5.2 *TB, TA* āp *'Water; River; Flood'*

Tocharian A and B *āp* mean 'water' but also 'river' and 'flood', thus referring to water in movement rather than stagnant water. Theoretically, these Tocharian A and B words could either be inherited from Proto-Indo-European **h₂ep-* or borrowed from Old Steppe Iranian **āp-* 'water' (cf. Avestan *ap-*, nom.sg. *āfš* 'water', Sogdian *"p*, Persian *āb*, etc.). However, it would be remarkable if such an essential word were a borrowing. At the same time, if it is a borrowing, the most likely source would be Old Steppe Iranian, as no other language was the source of the borrowing of such fundamental vocabulary in Tocharian.

The plural of the Tocharian B form could be telling. In IOL Toch 23 a5, we could read either *āpäṃ* or *āpäṃ*(*ta*), since the fragment is torn off afterwards:[114]

IOL Toch 23 a5
y(*ku*)*wermeṃ orotstsana āpäṃ* ///
"Having gone, the great rivers …" (based on CEToM).

If the correct reading is *āpäṃ*, this would be a good argument for inheritance rather than borrowing, as this type of plural is rare, unproductive and archaic, but if it is *āpäṃta*, then it could be either inherited or borrowed, as *-nta* plurals were common and productive at the Proto-Tocharian period (examples of the *-nta* plural found in Old Steppe Iranian loanwords are TB *waipeccenta* 'possessions', *pakenta* 'portions, shares').

113 The Scythian form αδι-γόρ has, to my knowledge, never been brought in the discussion of these Iranian words.

114 There is perhaps another example, IOL Toch 74 b1, which Adams (DTB: 47) and others have read as (*śt*)*w*(*āra*) *a*(*päṃ*) "four rivers". However, Peyrot (CEToM) reads it as (*śt*)*w*(*āra*) *k · t · –* ///, and this reading is accepted by Hartmann (2013: 448).

There are three arguments against a reading *āpäṃta* in IOL Toch 23 a5: first, we would expect the stress to go to the right, and we would thus expect †*apaṃta*; second, the spelling with ‹ṃ› is unexpected for this ending, which is usually spelled ‹nta›; third, the word is feminine, and feminine nouns normally do not form *-nta* plurals. In sum, a reading *āpäṃ* is more likely, and this in turn points to inheritance.

The semantics are not in favor of an Iranian borrowing either: in Iranian, **āp-* means 'water', and does not necessarily refer to rivers. Further, it is such an essential element of the vocabulary that it seems difficult to accept the borrowing hypothesis. If it meant 'river' in Iranian, a loanword hypothesis would be more plausible, as examples are more common, for instance English *river* ← Old French *riviere*.

In conclusion, although Tocharian A and B *āp* 'water; river; flood' could theoretically derive from an Old Steppe Iranian form, it could formally very well derive from Proto-Indo-European **h*$_2$*ep-*, from which the Iranian word for 'water' also derives. In the absence of evidence suggesting a borrowed origin, and with a plural that is apparently *āpäṃ*, inheritance is much more likely, and this option should be preferred.

2.5.3 *TB* **ām** *'Silence'*

It has often been proposed that Tocharian B *ām* 'silence'; as an adverb 'quietly, still', derives from Iranian. Tremblay (2005: 438), for instance, suggests that it is a direct borrowing from Sogdian *ʾrmyh*. However, this Sogdian word means 'deserted, isolated' (SD: 57) or 'solitary' (DKS: 244), which semantically does not fit. In fact, Tremblay probably took this word to mean 'still' in the phrase *ʾrmyh nyδ* "to sit *ʾrmyh*", but this should rather be translated as 's'asseoir à l'écart' (cf. Benveniste 1946: 66 "elle s'assit à l'écart").[115]

Isebaert (1980: 47) non-committally ("[g]eheel vrijblijvend") suggests to derive the Tocharian B word from another Iranian etymon: a Middle Iranian **āh*(*a*)*m* 'sitting' < **āh-*(*a*)*ma-*, from the root **āh-* 'to sit' (Cheung EDIV: 153–154), semantically parallel to Latin *sēdāre* 'to calm down, to restrain (cf. Eng. *to sedate*)' from *sedeō* 'to sit, be seated'. Naturally, 'to sit' and 'silence' can be connected to one another thanks to the formulation "to sit in silence" which often occurs in Tocharian as well. However, in my view, this derivation is difficult as no Middle Iranian **āh*(*a*)*m* is attested; indeed, there is no trace at all of an Iranian **āh-*(*a*)*ma-*.

115 Note Gauthiot's outdated translation (1912: 485) "elle s'assit soumise".

On the other hand, **arma-* 'quiet' as extracted from Avestan *ārmaiti-* 'name of an Amesha Spenta' and Vedic *arámati-* 'piety, devotion' would theoretically fit semantically. However, it does not fit phonetically: Proto-(Indo-)Iranian **aráma-* is to be reconstructed to account for **arámati-*, which is to be metrically restored for Avestan *ārmaiti-* 'Proper, Devoted thought' (cf. Hoffmann & Forssman 2004: 36). The reconstruction of a form **arma-* coupled with the meaning 'quiet' is also anything but assured.[116] Nevertheless, an Iranian adverb **armaį* must have existed. It is found in the Avestan phrase *armaē šad-* 'sitting still, ruhig stehend', in the Avestan adverb *airime* 'still' (cf. Narten 1968: 247), perhaps in Sogdian *ʾrmyh* 'à l'écart' and in Bactrian αρμαυ- / αρμαστο 'to be present, be (temporarily) resident, stay, remain', which Sims-Williams (2007: 194) explains as deriving from **arma-* 'still' and **āh/āsta-* 'to sit, remain'. This derivation is further supported by the Avestan collocation *airime.aŋhaδō* 'sitting still' (Yašt 13:73, cf. Narten 1968: 247).

There is no trace of an Old Iranian noun **arma-* 'stillness' or 'silence', or 'quiet', but if such a noun were borrowed in Tocharian B, it should have yielded ***erme*. Perhaps one could suggest an *ad hoc* simplification of the cluster **rm* to *m* (in Iranian), but then it would have become ***āme* in Tocharian (if the vowel was lengthened in Old Steppe Iranian) or ***eme* (if it was not lengthened). From a Middle Iranian language one could perhaps indeed expect *ām*, again with *ad hoc* simplification of the cluster, but such a noun seems to be lacking from all Middle Iranian languages. Moreover, no meaning close to 'silence' is found in the Middle Iranian hypothetical cognate Sogdian *ʾrmyh* 'à l'écart'. For completeness' sake I want to add that the hypothesis that TB *ām* 'silence' would be a cognate of Sogdian *ʾwrʾm* 'peace, quietude' (Sims-Williams 2020: 36), Persian *ārām* 'peaceful, quiet', Persian *ārāmiš* 'quietude', etc. (not found in the scientific literature) is also to be excluded, because these words are too different from the Tocharian form. There is no way to expect a reduction of **ārām* to **ām*, whether in Old or in Middle Iranian.

In conclusion, there is no clear Iranian source for Tocharian B *ām* 'silence', adv. 'quietly' (*pace* Tremblay 2005: 438): none that would fit for the meaning and none that would fit for the form. An Iranian source is theoretically not entirely excluded, but no suitable source form has yet been identified. Therefore, it is advisable to follow Adams (DTB: 47) for the moment, and to consider this word to be of unknown etymology.

116 Pinault (2016: 123 f.) proposes that there existed a PIIr. adjective **Hára-* 'fitting, proper, right' and a stem **Haráma-*, enlarged in *-ti-* (Avestan *ārmaiti-*, Vedic *arámati-* 'Devotion').

2.5.4 TB kāswo *'Skin Disease, in Particular Leprosy and Alphos'*

Adams (DTB: 165) translates Tocharian B *kāswo* as 'eruption, inflammation of the skin'. It corresponds to Buddhist Sanskrit *kuṣṭha-* which designates all kinds of skin diseases, among which 'leprosy' (see e.g. M-W: 297–298; Emmerick 1970: 79; on *kuṣṭha-* in Pali, see Huard 2022: 196–197). For a number of reasons, 'leprosy' seems a fitting translation of *kāswo* (cf. Isebaert's translation as *melaatsheid*, that is, leprosy, 1980: 196). For a discussion of the meaning of Tocharian B *kāswo* see below.

This word does not have an evident Indo-European etymology. Isebaert (1980: 197), following an idea by Van Windekens (1977), proposes to derive *kāswo* from PIE *$*g^{\text{u}}$os-u̯ā/ō(n)*, a derivative from the root *(s)g^{u}es- 'to go out, to burn out' (cf. LIV^2: 541–542).

Van Windekens' main argument (1977: 141) is based on the parallel with Vedic *jásu-* 'Erschöpfung, Verschmachten' and especially *jásvan-* 'elend', which "offre une structure morphologique qui se superpose nettement à celle de tokh. B *kāswo*: ce substantif féminin s'explique excellemment à partir d'i.-e. *(z)g^{u}ōs-u̯ō(n) [...]". Naturally, *jásvan-* needs to derive rather from an *e*-grade *g^{u}es-u̯ō(n) because of the *j-*, and the supposed *ō*-grade in the root has no morphological justification. Further, the semantic shift proposed by Van Windekens relies on an intermediary meaning *'disease' > *'leprosy', which relies on Dutch *melaats* 'leper' ← French *malade*.

From a methodological point of view, I think that loanwords should preferably not be used as parallels to support language-internal semantic changes. Nevertheless, 'sick' > 'leper' is an attested semantic change, in cultures where leprosy was "the sickness *par excellence*" (or "the grave sickness *par excellence*"). An example is found in Middle Cornish *claf* 'sick' > 'sick, leprous' and, in the plural, *cleyvon* 'lepers'.[117] In this way, I slightly disagree with Adams (DTB: 165) who writes about Van Windekens' proposal that "the semantics are anything but compelling", but he is not entirely mistaken, since the Sanskrit word does not mean 'sick'. It is principally on morphophonological grounds that Van Windekens' proposal needs to be discarded, the decisive argument being that his Indo-European derivation does not work.

Winter (1962: 113[10]) proposed to link TB *kāswo* to TA *kāsu* 'good thing', saying it seemed "inadvisable to separate [them]". Obviously, the semantic connection is difficult to see. For any semantic relationship to exist, there would perhaps need to be a certain cultural setting, in which diseases are particularly avoided

117 Cf. Williams (2011). Probably through **claves bras* 'big disease' which designates leprosy in Modern Cornish (cf. George 2020). I thank Pierre Faure for bringing this latter fact to my attention.

subjects, and in which leprosy is one of the very worst diseases. One could then perhaps think that a word meaning 'good' was euphemistically taken to designate a specific (in fact terribly bad) disease.

There are no indications that there was such a cultural setting in the Tocharian environment, nor are there any clues for another way to explain this shift. Another argument against this etymology would be that, for this word meaning 'good', the original meaning would have been completely abandoned in Tocharian B, whereas its secondary meaning 'leprosy' would have never arisen (or not be attested) in Tocharian A. This makes the argument, already weak, even less probable. Kim (2019: 186[29]) also rejects Winter's suggestion.

Hilmarsson (1996: 107) connected this word to Proto-Germanic **haswa-* 'grey, white' (on which Kroonen 2013: 213–214). This will be discussed more in detail below, in connection with Del Tomba's renewed etymology.

Del Tomba (2023: 130 ff.) has proposed a semantically attractive connection of Tocharian B *kāswo* to the Proto-Indo-European root **ḱes-*/**ḱseu̯-* 'to comb, scratch', based on the idea that the skin of lepers often shows lesions which are similar to scratches and scabs, and look like they can be scratched off easily. He nevertheless notes that the vocalism of the Tocharian form is problematic.

Assuming that *kāswo* designates not leprosy in general, but "white (tuberculoid) leprosy" (cf. also Schmidt 1986; 68–70; 2018: 74), Del Tomba further makes a series of assumptions. Before addressing these, I need to mention a number of facts. White leprosy, or rather alphos or vitiligo, is a disease with which patients lose the pigmentation of (parts) of their skin. It differs in fact from tuberculoid leprosy, which is a less dangerous and contagious form of leprosy, with which patients usually have pinkish patches appearing on their skin. It is unclear to me whether white leprosy in the texts cited by Del Tomba (2020: 123 and 124) refers to alphos or to tuberculoid leprosy.

It is clear that whenever *kāswo* corresponds to Sanskrit *kuṣṭha-*, it definitely designates leprosy (either tuberculoid, borderline or lepromatous). This meaning of *kuṣṭha-* as referring to the most dangerous, limb-decaying forms of leprosy is confirmed by early Indian medical texts (Rastogi & Rastogi 1984).

As Del Tomba (2023: 131) establishes, Tocharian B *kāswo* translates Sanskrit *kilāsa-*, and thus designates rather alphos, or skin depigmentation (see Del Tomba 2023: 131–132 with references). Thus, in general, it seems safer to assume that the primary meaning of *kāswo* was 'skin disease', but a skin disease of a rather grave nature, not a simple rash.

Because of the correspondence of *kāswo* with *kilāsa-*, Del Tomba (2023: 130 f.) rehabilitates Hilmarsson's etymology. He proposes to etymologize *kāswo* as from PIE **ḱh₂s-u̯o-* 'having whiteness' → **ḱh₂s-u̯e-h₂* 'mass of whiteness' > PTch

**kaswå* > TchB *kāswo* 'white leprosy; skin disease'.[118] In my opinion, this etymology is not impossible but there is another argument to be taken in account. Since TB *kāswo* does not seem to refer to one disease in particular, but rather to a series of diseases, all related to the notion of '(serious) skin disease', there are two main possibilities for its derivation: 1. either its initial meaning was much less specific or 2. it was borrowed and its exact meaning was unstable for some time.

If the first hypothesis is to be preferred, then I suggest a slightly different evolution, while accepting the main points of Del Tomba's etymology. From **ḱh₁-s-u̯o-* 'being grey or pale' → **ḱh₁-s-u̯e-h₂* 'greyness or paleness' one could imagine a result TB *kāswo* 'skin disease which makes one's skin dark (as in the lesions of leprosy) or pale (as in alphos or lepromatous leprosy; or perhaps also pink as in tuberculoid leprosy)'. A parallel for another, perhaps better solution, is found in a semantic path to be observed in Champenois dialects. It is the word *havé* adj. (cf. Tarbé 1851: 74), variant *havi* 'desséché, contracté' (cf. French *hâve* 'pale and thin because of a disease or hunger'). Those forms are ultimately borrowings from the Germanic word **haswa-* discussed above. The semantic shift is *'grey or pale' > *'sicklish' > 'weak, made thin' > 'dried up'. The skin of lepers appears to be particularly dry and ribbed. The meaning of TB *kāswo* would thus add one supplementary step from 'dried up', attested for Champenois, to 'leper'. This supplementary step I believe, is very plausible given the appearance of lepers. Whether the first hypothesis (which is more similar to Del Tomba's) or the second should be preferred depends on how one sees the disease or interprets the Tocharian word.[119]

It is also possible that the word *kāswo* is ultimately a loanword. In this case, the source is probably to be found in Iranian. Lidén (1916) and others after him (e.g. Oettinger 1983: 330; Tremblay 2005: 441) have suggested an Iranian source. The point of departure is found in Avestan *kasuuiš* designating 'someone having a specific disease'.

The etymology of Avestan *kasuuiš* is discussed in Kellens (1974: 367–368; cf. also Bartholomae 1906: 158). After rightly rejecting a number of etymologies he proposes to read it as a *bahuvrīhi *kasu-viš-* 'qui a un petit poison, des pustules'. This proposal seems very unlikely to me: in what way are pustules small poisons? However, my main argument against this etymology is that *viš-* in Avestan and in Iranian in general designates liquid poison, a poison that springs out.[120]

118 There is no reason to reconstruct this root with **h₂*; usually it is reconstructed as **ḱeh₁-* (cf. Lubotsky 1989: 57).

119 I thank Philippe Hattat for providing me with the Champenois examples and references.

120 As in the Hom Yasht XI:11, when the dragon Sravara ejects a yellowish poison; or as in the

One could perhaps venture to say that **kasu-viš-* designates pus, and thus the pustule containing the pus by metonymy, but pus is not a poison. Naturally pus should not be drunk: but it is not poisonous (except that it might perhaps look like the yellowish poison of dragons?).

Rather than accepting Kellens' etymology, I believe Avestan *kasuuiš* to be an *-iš* adjective of a noun **kasū-*, cognate of Sanskrit *kacchū-* 'scabies', and probably meaning the same thing. This would fit very well the context of both attestations (cf. Bartholomae 1906: 158 for more details).[121] Lepers are already cited in Vendidad 11:29 and Yašt 5:92 (*paēsō*). The list of people who are forbidden to approach the libation in Yašt 5:92 is based on physical appearance (cf. Bartholomae 1906: 158). It is possible that the translations by Pahlavi ‹kynyk› and Persian کینه کش *kīna-kaš* 'vengeful' (cf. fn. 121 below) are based on a confusion with NP *kana* کَنه 'scab-causing mite; leech'.[122]

It is not difficult to imagine that an original Pahlavi *𐭪𐭭𐭠𐭪𐭥 **kanagīg* 'mangy'[123] was corrected in 𐭪𐭩𐭭𐭩𐭥 *kēnīg* 'vengeful' and translated as such in Persian. The reason could be that the Middle Persian word **kanag*, from which NP *kana*, itself rare, must derive, is not otherwise attested as far as I know, and the copyists perhaps did not know it. This "correction" also happened with the Avestan word *kapasti-* 'name of a disease or a poisonous plant (?)' also translated by NP *kīna* 'vengeance' (cf. Shapira 1998: 220[93]). Positing the existence of the predecessor of the New Persian word *kana* in Middle Persian further solves another problem: if the form (*)𐭪𐭭𐭩𐭲𐭥𐭥 *kēnxūn* as Justi read it (cf. fn. 121) indeed

---

Zand-i Wahman Yasn IV:2 when the invaders of Iran, those of the race of Xešm, will "piss poison (*wīš*)".

121 The Pahlavi translation of *kasuuiš* is a real problem. Bartholomae (1906: 158) already notes Justi's reading *kēnxūn* 'angry-blooded' (transcribed *kīnḫūn*) *𐭪𐭭𐭲𐭥𐭥 which he could not find in the manuscripts, which have 𐭪𐭭𐭥𐭥 and 𐭪𐭭𐭲𐭥 or 𐭪𐭭𐭲𐭩𐭥𐭥.

I have found the form 𐭪𐭭𐭪𐭲𐭥𐭥 in L4 and ‹wdzwytk› in 4711 B1, where it is translated (in Persian) as رنجیده *ranǰīda* 'hurt' (maybe based on a reading *‹r(n)dčytk›). I am unsure of the reading of those words, and they could be read in many various ways. It is true that F10 has ‹kynyk› 'vengeful' rendered as کینه کش *kīna-kaš* 'idem' in the Persian translation. In general it seems that this word (*kasuuiš*) was not correctly understood by the commentators (or at least, some later ones). Perhaps the translation 'vengeful' was influenced by a folk-etymology based on *kasu-* 'small' (as in 'petty'), as was proposed by Justi (1905: 95).

122 Perhaps etymologically related to *kand-*/*kan-* 'to dig' (so Skjærvø 1994: 277) as in 'the digging one' because of the hole-like spots it leaves in the skin, if the scabs remain untreated, or, if the meaning 'leech' is primary, perhaps because of the animal's "blood-digging" habits? Perhaps also it is somehow related to Waxi *kanek* 'mosquito' (cf. Morgenstierne 1938: 526 for the Waxi word; cf. Hasandoust 2014: 2281 for the Waxi connection to the Persian word).

123 Mange is a category of diseases that includes scabies. People sick with scabies are usually called *mangy*.

exists, then one could easily see it as a deformation of **kanxūn* 'blood-digger' (of which *kanag* is perhaps originally a back-formation), rather than **kēnxʷāh* 'rachsüchtig' as Justi (1905: 95) proposed. That *kasuuiš* was seemingly less and less clear to the commentators is also evident from the various translations and unclear paraphrases in various manuscripts (cf. fn. 121 on the previous page).

If my explanation of the deformation of an original Pahlavi translation as **kanagīg* 'scabby' is to be accepted, then it can be proposed that *kasuuiš* indeed meant 'scabby' and the original stem **kasū-* 'scabies' would thus be a perfect cognate of Sanskrit *kacchū-* 'scabies'. In that case, they would both need to go back to **kasćū-*. Since this word does not have any Indo-European cognate, it could perhaps be of BMAC origin.[124]

If PIIr. **kasćū-* is indeed a borrowing from the BMAC language, one could perhaps imagine that TB *kāswo* was borrowed from the same source. However, this is difficult, since whatever sequence of sounds gave rise to PIIr. **-sć-* would not likely be rendered with *-s-* in Tocharian. For instance, what we reconstruct as BMAC **ć* on the basis of Indo-Iranian, is rendered as PT **ś* (PIIr. **ćaru̯a-*: PT **śer(ə)we*). Furthermore, an accented vowel **a* has not been found in other BMAC loanwords in Tocharian (see chapter 3). The word is thus unlikely to be a direct BMAC loanword in Tocharian.

Alternatively, one could also imagine that TB *kāswo* was borrowed from Old Steppe Iranian. A Proto-Indo-Iranian form **kasćū-* should become **kasū-* in Old Iranian (cf. Lubotsky 2001a). However, an Old Steppe Iranian form **kasū-* would be borrowed as †*kes* or perhaps †*kos* < **kesu* in Tocharian. Therefore, an Old Steppe Iranian origin can be excluded for the Tocharian B word *kāswo*.

Yet another option would be to consider a Bactrian origin for this word. A Pre-Bactrian form **kaswā* (< feminine OIr. **kasu-ā-*) would have given Pre-Tocharian B **kas(u)wa*, readapted as *kāswo*, like Pre-Bactrian **kōškā* → TB *koṣko* (see Bernard & Chen 2022) or rather **kōžgā* → TB *koṣko*,[125] possibly on the model of *maiyya* '(supernatural) power' remade into *maiyyo*. From the Iranian point of view, this would be an *-ā-* abstract on an adjectival stem **kasu-* 'mangy'. Reconstructing such an adjectival stem is nevertheless gratuitous, and there is

124 Mordvin *kośkilda* / *kośkelda* 'scabies' might perhaps be connected to this Indo-Iranian etymon by means of borrowing.

125 Schoubben (2024: 59[129]), commenting on Bernard & Chen (2022) rightly notes that there is no reason to ascribe this devoicing of *-žg-* to the Bactrian language, and that it should simply be attributed to Tocharian. Obviously, he is right, and this mistake (entirely mine) should be corrected in future publications if need be.

no evident Middle Iranian candidate with such a formation in this word. A Pre-Bactrian origin is thus also unlikely for TB *kāswo*.

Dragoni (2023: 84–85) has proposed to see in *kāswo* a borrowing from Pre-Khotanese **kasūwa-*, more specifically from its accusative singular **kasūwu*. As to the exact meaning of this Pre-Khotanese word, it might have designated skin eruptions or inflammations, like Late Khotanese *kasaa-* (cf. Dragoni 2023: 85). Del Tomba (2023: 132) engages with the etymology of TB *kāswo* provided by Dragoni and myself. Although he considers it possible, he views the isolation of **kasū-* within Iranian as well as the absence of an attested source form for *kāswo* as urging caution. In principle I agree, which is also why I have no issue with accepting a possible Indo-European etymology (namely, Del Tomba's) for this word. However, one should note that in this specific field, it is sometimes necessary to posit source forms with marginal attestations as it often remains, after all, lacunary. For this same reason, Holopainen (2019: 93–95) finds fault with the Iranian source to which Finnish *iho* 'skin', North Saami *assa* 'back of skin or furs' etc. go back, while cognates with the specific meaning skin are more numerous than usually thought, including a borrowing in Tocharian (s.v. *yetse*).

The semantic shift from 'scabby' to 'leprous' is not uncommon, because leprous skin is very often scabby. However, 'scabies' on one hand and 'leprosy' on the other, as well as 'leper' and 'mangy' are often carefully differentiated, as far as I could find. This is not trivial, as leprosy and scabies, especially in some variants of each disease, present similar alterations of the skin. This is notably true of the many scabs and reddish spots found on the skin of sufferers of both diseases. Because of their never healing lesions, lepers are also prone to mange.

The fact that these two diseases, scabies and leprosy, are not easily confused must be due to the fact that their consequences are very different. Both scabies and leprosy are contagious (although scabies is much more contagious), but it is leprosy that has terrified the imagination of European, Caucasian, Indian and Chinese peoples. It is leprosy with which the highest social stigmata are attached: the falling of the limbs and the neurological damage it causes are irreversible. Therefore, the semantic change between 'scabies' and 'leprosy' could in my view not have happened language-internally. Nevertheless, it is not impossible to imagine a setting where mange was uncommon, such as Central Asia, and upon borrowing such a word from Iranian tribes, Tocharians used it to refer to skin disease in general, but to the most fierce leprosy in particular. This would explain the various meanings that can be attributed to it in the various texts, where it corresponds to both Sanskrit *kuṣṭha-* and *kilāsa-*. Perhaps also the Iranian donor language had a more general meaning for this word, such as 'skin disease'. Cf. Sinhalese *kas* 'itch, skin disease' < OInd. *kacchū-* 'scabies' (cf. CDIAL: 203).

To conclude, although the inherited Indo-European etymology from the root *$\acute{k}h_1$-*s*- 'to be grey, white' is possible to explain the Tocharian B word *kāswo* 'skin disease, in particular leprosy and alphos', I propose that the Iranian etymology which was already proposed in the past might function as well. I suggest an emendation of the Pahlavi translation of the Avestan word *kasuuiš* on the basis of a New Persian word which designates the animal that causes scabies. With this emendation, the Avestan word *kasuuiš* may be taken to mean 'mangy'. If Tocharian B word *kāswo* was borrowed from Iranian, it should for phonetic reasons have been borrowed from Pre-Khotanese, because of the regular correspondence between Pre-Khotanese **a*: Tocharian *a* and Pre-Khotanese *-a*, acc.sg. *-u* regularly corresponding to Tocharian B *-o*, as proposed by Dragoni (2023).

2.5.5 *TA* kāwṣtā- *'Strike, Kill'*

The Tocharian A verb *kāwṣtā* 'to strike' (DToA: 156), with a derived form *koṣtlune* 'beating', does not have an identified Tocharian B equivalent. The root vocalism *-āw-* is needed to account for the present participle *kākoṣtu* (Peyrot 2013: 729). It can mean 'to beat an object', as in the example below.

A 218 a3
(*bārānas*)*y*(*a*)*c yäm koṣtam oṅkraci kumpäc*
"I am going to Benares and I will beat the drum of immortality" (DToA: 156)

Or in this other example.

A 359 a18
kanti koṣtluneyo prā(*mañi*) (*talke*) ///
"by the beating of the Brahmanic sacrifice gong" (my translation)

The verb *kāwṣtā* can also have a more somber meaning, as seen in the sentence below.

A 56 b3
mā =ryu praṣtaṃ kāreyo kākoṣtu nkatär
"before long he will perish, killed by the sword" (Peyrot 2013: 224).

The adjective *koṣtluneṣi* means 'pertaining to killing' (DToA: 156), as in *koṣtluneṣiṃ watku* 'order of killing' (A 76 b6, cf. DToA: 156).

Because of its form and meaning, one could be tempted to connect this word to an Old Steppe Iranian hypothetical **kaꞎš-*, whose meaning would be similar,

as it would derive from Proto-Iranian **kau̯š-* 'to fight, struggle; kill' (cf. EDIV: 251). This etymology, although tempting at first, should be rejected, for reasons I explain below.

First, the form that would yield TA *kāwṣtā*, perhaps a past participle 'stricken, killed' needs to contain a diphthong in the Iranian source itself. Such a diphthong is posited by Cheung for the Iranian root (cf. EDIV: 251), and it would seem to be reflected in Middle Persian Pahlavi *kwhš-* 'to strive, struggle, endeavour' and *kwš-* in Manichaean Middle Persian and Parthian, although this could be a *mater lectionis* for *kuš-* (DMMPP: 215). However, apart from Middle Persian and Parthian, not a single Iranian form is found with a long vowel, that is, the reflex of a diphthong **-au̯-*, all forms go back to a short **u*, including Avestan *kušaiti* 'kills' from *kuš-* 'to fight, struggle' and Khwarezmian *m|kš-* 'to butt (with the horns)'. In my view, the Middle Persian form ‹kwhš› is simply due to a contamination from the verb ‹twhš› /tuxš-/ which has almost the same meaning ('to strive, struggle with') and which has an inherited *-xš-*.

However, even if the Middle Persian and the Parthian forms (*kwš-*) were not *matres lectionis* but represented */kōš/, nothing shows that it is not a secondary full grade remade after the model of many verbs that have *uCt* in the past stem and *ōC* in the present stem. The reason why all the Iranian forms derive from *kuš-*, which is a zero grade is certainly that, semantically, 'killed' or 'stricken' is a much more frequent utterance than 'kill' in the present tense, and the verb was already built on the participle stem. There is thus no Iranian **kau̯šta-* on which a Tocharian form could have been derived: only **kušta-*.

This also means that the traditional derivation of the Iranian root from PIE **keh$_2$u̯-* 'schlagen, spalten' (cf. LIV^2: 345) is erroneous. Indeed: **kh$_2$ú-s-to-* would have yielded ***xušta-* regularly (e.g. Kümmel 2018: 165 f.), and **kuh$_2$-s-tó-* would have yielded ***kūšta-*. We are thus bound to reconstruct a root **keu-s-* or **keus-* 'to fight, kill', the details of which remain to be elucidated. The Tocharian A verb *kāwṣtā-* is not a borrowing from Old Steppe Iranian: nothing warrants such a reconstruction for Old Steppe Iranian, and, besides, the morphology of the Tocharian verb is archaic. While this is not an argument in itself (e.g. § 2.2.8), it stands together with all the other elements supporting an inherited origin, or at least a non-Iranian origin, for this verb. The meaning 'kill' of *kāwṣtā-* is perhaps influenced by TA *kāw-* 'to kill'. It is perhaps the case that TA *kāwṣtā-* is a descendant of the root **keu*(-)*s-*, and thus a cognate of the Iranian forms, but the data remain ambiguous.

In conclusion, Tocharian A *kāwṣtā-* 'strike', which can also mean 'kill', is not a borrowing from Old Steppe Iranian, but most probably an inherited word.

2.5.6 *TB* tāno *'Seed, Grain'*

It is generally assumed that Tocharian B *tāno* 'grain', obl.sg. *tāna* is inherited from Proto-Indo-European (cf. DTB: 303 with references). That, however, seems difficult in regard of the fact that the only cognates of this word, in my view, are Indo-Iranian: Vedic *dhānā́-* f.pl. 'geröstete Getreidekörner' (cf. EWAia[1]: 787), Avestan *dānā** in *dānō.karš* adj. '(Getreide)körner (ver)schleppend' (cf. AiW: 734), Khotanese *dānā-* f. 'grain, corn', Sogdian *δān* 'seed', Middle Persian *dān* and *dānag* 'seed, grain' (CPD: 24), New Persian *dāna* 'grain, unit',[126] Xorāsāni *deng* 'corn' (Monchi-Zadeh 1990: 50). Bailey also noted some assimilated forms (going back to **nānā* and **nānā-ka-*): Yidgha *nānoγō*, Pashto *nīne* 'parched grain', *nānga* 'blackberry' (DKS: 156).

The connection with Lithuanian *dúona* 'bread' and Latvian *duõna* 'slice of bread, heel of a loaf' (cf. EWAia[1]: 787) seems rather weak to me. We have a group of close-knit words meaning 'grain' in Tocharian and Iranian, and this meaning is not even found in the Indic cognate. On the other hand, the Baltic words designate bread, and there is a long way from the whole grain to bread. Even more so: the Latvian form also designates the crust of a loaf of bread, which Peyrot (2018: 258) connects with an unaccented word *duona*, which has various meanings relative to an edge (see details in Peyrot, loc. cit.). He suggests an evolution 'edge' > 'edge, heel of a loaf' > 'slice of bread' (> 'bread' in Lithuanian) (2018: 259), which I find more convincing than an evolution, to my knowledge unparalleled, from 'grain' to 'bread' (even with multiple intermediary steps).

Even if the Baltic words were cognates of the Indo-Iranian ones, which is far from evident, it seems difficult to believe in a Baltic–Indo-Iranian–Tocharian isogloss, where Tocharian shares the exact same meaning as Iranian, and not Indic: in no modern Indo-Aryan word does the descendant of this word mean 'grain', cf. CDIAL: 387. All Middle and New Indo-Aryan languages show a different meaning for this word, such as 'parched grain' and 'growing rice' (cf. CDIAL: 387), which implies that the Proto-Indo-Aryan meaning of **dhānā́-* was different from simply 'grain, cereal', a meaning which can on the other hand safely be reconstructed for Proto-Iranian.

Logically, then, Tocharian B *tāno* should be an Iranian loanword in Tocharian (cf. already Duchesne-Guillemin 1941: 180; later Schmidt *apud* EWAia[1]: 787; Peyrot 2018: 258 and Del Tomba 2023: 128, 150[169]). A problem with this hypothesis, which otherwise seems rather evident, is, as Peyrot mentions (2018: 258–259), that the class to which *tāno* belongs, namely the *kantwo*-type, is rather

126 A meaning 'seed' can also be seen in the New Persian compound *wan-dāna* 'seed of the *wan*-tree' (for which see Maggi 2003: 123[37]).

small and comprises no loanwords. Nevertheless, Del Tomba recently (2023: 144) adduced a strong argument in favor of this hypothesis: Tocharian B *mālo* 'alcohol', obl.sg. *māla*, which doubtlessly was borrowed from an Iranian language, namely Pre-Bactrian, and also belongs to this same class. I have also suggested two more members of this noun class to be loanwords from Iranian:

a) Tocharian B *maiyya* (s.v. *maiyya*) 'power, might'.
b) Tocharian B *kāswo* 'skin disease; in particular leprosy and alphos', also has its oblique singular in *-a* and belongs to the same class as *tāno* and *mālo*. I have argued (s.v. *kāswo*) that this is also likely to be an Iranian loanword.

In case my argumentation is to be followed, there are at least four loanwords in this class: *maiyyo*, *tāno*, *kāswo* and *mālo*. Dragoni (2023) has identified even more loanwords in this category, such as TB *kātso* A *kāts* 'stomach, belly, womb' and TB *tvāṅkaro* 'ginger'.

If Tocharian B *tāno* is borrowed from Iranian, the exact source must be determined. The main clue here is the ending and class appurtenance of the Tocharian word. I have shown that Old Steppe Iranian final **-ā* is in Tocharian rendered in the same way as in Old Steppe Iranian **ā* in other positions: Proto-Tocharian **a* > Tocharian B *a*. This correspondence is seen in TB *newiya*. In the case of Archaic TB *meyyā*, it can perhaps be speculated that a new form *maiyyo* was made, to follow a more productive model than a nom.sg. *-a*: obl.sg. *-a* declension, although this is no longer likely, see doubts expressed by Del Tomba (2023: 140 ff.), see § 2.2.17. However, I have also suggested that TB *raimo** 'dust, dirt' originally comes from **reyma* (Old Steppe Iranian **raiṃā-* 'dirt') but was transferred to the productive *okso*-type (nom.sg. *-o*: obl.sg. *-ai*).

Thus, if an Old Steppe Iranian word **dānā-* f. 'grain, seed' was originally borrowed as PT **tana*, it could at any period of time have been remade into a noun *tāno*, of the *kantwo*-type.[127]

On the other hand, an Old or Pre-Khotanese (or Proto-Khotano-Tumshuqese) form **dānā-* could also have been borrowed as PT **tano*. Even a Pre-Khotanese or Old Khotanese form **dāna-* m. could in theory have yielded Tocharian B *tāno* (cf. Peyrot, Dragoni & Bernard 2022: 408 for a discussion of this form).

There is yet another argument that can be used to determine the source of borrowing: there exists a diminutive form *tanākko*. The suffix *-kko*, as Dragoni

127 However, one would have to exclude a source form OSIr. **dāna-*, which could only have led to TB †*tāne*. The protoform **dāna-* needs to be reconstructed for the Middle and New Persian forms: only **dānaka-* or **dānakā-* can be the source of Middle and New Persian *dān*(*ag*).

(2023: 107 f.) demonstrated, is likely to be of Pre-Khotanese or Proto-Tumshuqese-Khotanese origin (from **kka-* < **-ka-ka-*). Thus, the presence of this diminutive suffix renders more likely the hypothesis that the base word *tāno* derives from some stage of Khotanese as well.

Finally, it is striking that there is a high number of botanical terms among Khotanese borrowings in Tocharian, where *tāno* would fit well, while there are not that many from Old Steppe Iranian, the only plausible candidate being TB *welke** 'stalk' or 'petal' (s.v.).

Apart from these two sources, there can hardly have been another Iranian donor language, both for historical and morphological reasons. The word-final vowel warrants an earlier stage, which excludes Sogdian, Parthian and historical Bactrian. Pre-Bactrian is also impossible here: it would have given TB †*lāno* in the same way it gave Tocharian B *mālo* < **madu-*, and, as I argue together with Ruixuan Chen, in the same way that Pre-Bactrian **kōžgā-* was rendered as TB *koṣko* (Bernard & Chen 2022). It could also hardly have been a smaller, undetermined Middle Iranian language: how could a word like 'seed, grain' have been borrowed from a smaller, distant language of little political importance?

In conclusion, Tocharian B *tāno* 'seed, grain' can hardly be inherited, as non-Iranian cognates with this precise meaning are lacking. It is very likely an Iranian borrowing, as has been proposed multiple times before. More specifically, it can be either an Old Steppe Iranian loanword, which implies that it was first borrowed as a noun of the *maiyya*-type (nom.sg. *-a*: obl.sg. *-a*), and later shifted to a different declension class, the *kantwo*-type, or it was borrowed from Pre-Khotanese, Proto-Khotano-Tumshuqese, or Old Khotanese (cf. Dragoni 2023: 122). The first hypothesis is unlikely. In favour of the Khotanese hypothesis is the existence of the diminutive *tanākko*, whose suffix *-kko* ist best derived from some stage of Khotanese.

2.5.7 *TB* waipte *'Separately, Apart'*

Tocharian B *waipte* 'separately, apart' is generally considered to be of Iranian origin (e.g. DTB: 668), after Schmidt (1985: 760 f.). Schmidt's etymology consists in deriving *waipte* from **waiptay* ← Old Iranian **hwai-patayai* '(dat.) self-master', "für sich".

From the semantic point of view this derivation is not fully convincing. It is, of course, difficult to exclude that **hwai-patayai* 'für sich' could eventually come to mean 'for itself' = '(each thing) for itself', but the assumed semantic development is far from compelling.

Formally, the problems with Schmidt's etymology seem insuperable. According to the correspondences established, "**hwai-patayai*" would yield †*waiptyai*

or †*waipcai vel sim*. There seems to be no way to derive the final *-e* of TB *waipte* from the sequences *-ay-* or *-ai* of **hwai-patayai*, nor from a combination of the two.

To conclude, Schmidt's etymology of Tocharian B *waipte* 'separately, apart' is impossible formally. For the moment, this adverb remains without a clear etymology.

2.5.8 *TB* waimene *'Difficult'*

The etymology of the Tocharian B word *waimene* 'difficult' is arduous. As recalled by Adams (DTB: 668), Van Windekens (1976: 643) suggested to view it as a borrowing from Sanskrit **vaimanas-* 'tired, discouraged', a derivative of *vimanas-* 'perplexed, discouraged'. As Adams writes, "the putative Sanskrit morphology would be unparalleled". The semantic change from 'perplexed, discouraged' to 'tired' and 'difficult' is not evident either.

Isebaert (1980: 184) derives the word from Proto-Indo-European and sees in it a lexicalized locative form of an unattested **dṷoi-mo-* (derived from 'two'), but, as Adams writes (DTB: 668) there is no other example of a frozen locative becoming an adjective in Tocharian. Further, the semantics are not evident either.

Pinault (2015: 170) proposes to derive this word from **wai-me-ne* 'woe on us': **wai* 'woe' + *-me*, the plural suffixed pronoun + *-ne* locative ending. He explains the functional change as being perhaps the result of a reinterpretation of the interjective phrase as a predicate in deontic sentences, and gives a possible example. However, in Tocharian pronoun suffixes cannot be attached to interjections, so that this explanation is syntactically problematic.

Indeed, although it is tempting to see in this word a trace of the Proto-Indo-European interjection **ṷoi* 'woe', I believe a real problem lies in the fact that woe and difficulty are concepts that are quite far apart. Woe relates to the curse of destiny on man, and to the sufferings that ensue. Difficulty relates more to the obstacles man has to overcome. Apart from that, it is not that easy, I believe, for an interjection to become frozen, to acquire a locative suffix, and then become an adjective.

Schmidt's explanation (1985: 762) from Old Iranian **hṷai-manah-*, with the semantic development 'self-willed' to 'difficult' ("wer oder was seinen eigenen Sinn hat", d.h. "eigensinnig", "schwierig" [...]") is also unlikely. Indeed, in French, in English, in German, and in many other languages, someone who is self-willed is called 'difficult' (*don't be difficult*, French *ne fais pas le difficile*, etc.). However, in this case the development would be precisely the other way around. That something difficult would be deemed 'self-willed' seems to require a very specific phraseology in a specific language, of which we have no trace in Iranian, and which seems thus unwarranted.

In conclusion, although the Tocharian B word *waimene* 'difficult' does have an Old Steppe Iranian appearance, its etymology remains unclear for the moment, and it could theoretically be either borrowed or inherited.

2.6 Old Steppe Iranian Calques in Tocharian

In the previous sections of the chapter, I have focused solely on loanwords from Iranian. Such a type of language contact as the one sketched by the present book does not limit itself to the borrowing process: often, calques are also involved. While Tocharian is replete with calques from Sanskrit (even the originally Old Steppe Iranian words TB *śāte*, TA *śāt* 'rich' were calqued, as to its meaning on Sanskrit *bhoga-*, see § 2.7.4.2), these are usually literary calques, calques of compounds and expressions, and the like. I am proposing here to discuss three alledged Old Steppe Iranian calques in Tocharian, where the calquing process concerns the semantics of originally inherited Tocharian words—or, in any case, Tocharian words from a non-Iranian source. More calques could be found, both from Iranian (TB *onolme*, TA *wrasom* 'being' < 'breathing', compare Khotanese and Tumshuqese forms going back to **uzāna-bara-*, 'breath-bearing', Parthian *gyānβar*)[128] and perhaps more specifically from Old Steppe Iranian, however, this topic requires even further research. It seems likely that some Tocharian inherited words were semantically influenced by either similar-sounding or functionally similar Old Steppe Iranian words. It should also be mentioned that calques, especially calques from a reconstructed language such as Old Steppe Iranian, represent a much weaker evidence, as the same semantic evolutions could have occurred in the target language on its own, and as the semantics of the source language are not directly known. I therefore present them as an interesting research direction, but not as a hard fact demonstrative of Old Steppe Iranian ↔ Tocharian contacts.

2.6.1 *TB* ñäkte, *TA* ñkät *'God, Lord'*

TB *ñäkte*, TA *ñkät* 'god, lord' is definitely an inherited Tocharian word. It comes from **neḱ-th$_2$-o-* 'qui surmonte la mort' (cf. Pinault 2008: 36, with references to previous etymologies). It is not the meaning 'lord' or 'god' of TB *ñäkte*, TA *ñkät* which, I contend, might be calqued on Old Steppe Iranian, but rather the reason why it is used in compounds.[129]

128 This example was found by Prof. Michaël Peyrot (p.c.).

129 This idea was first shortly discussed in Bernard (2022).

Both Tocharian B and A adduce, without any major semantic change, the word *-ñäkte* and *-ñkät*, respectively, to the word for 'moon' (TB *meñ-ñäkte*, TA *mañ-ñkät*), 'earth' (TB *keṃ-ñäkte*, TA *tkaṃ-ñkät*) and 'sun' (TB *kauṃ-ñäkte*, TA *koṃ-ñkät*). Such a construction was calqued in ancient Turkic *kün-tängri* 'sun', *ay-tängri* 'moon', *yir-tängri* 'earth' (cf. Pinault 1998: 358; 2008: 32 f.). Pinault considers this phenomenon to be a trace of "la divinisation des forces naturelles dans la religion pré-bouddhique des Tokhariens […]" (2008: 32; already Pinault 1996:20; 1998). While I agree that this is necessarily the case, it is also reminiscent of the Iranian usage of calling the Sun 'bright-sun': Av. *hvarə.xšaēta-* (Bernard 2022: 28).

The element *xšaēta-* also means 'lord' in Avestan and, more generally, in Iranian. As I argue (Bernard 2022), **xšaiṯa-* originally meant 'bright'. Due to a contamination with the root **xšaiṯ-* 'to rule', it also could mean 'lord, sovereign', possibly through the reinterpretation of a noun **xšaiṯar-*: nom. sg. *xšaitā* 'ruler' (Bernard 2022: 27).

Iranian languages have only kept the meaning 'ruler' or 'bright' of this word, except sometimes in some fixed constructions (Bernard 2022: 26), but the compound **hwar-xšaiṯa-* is widespread (e.g. Persian *xuršēd* 'sun'). It is thus conceivable that a reflex of **hwar-xšaiṯa-* 'sun' existed in Old Steppe Iranian with the meaning 'sun', and that, in Old Steppe Iranian, **xšaiṯa-* only meant 'lord', as in Sogdian *xšyδ* /xšēδ/ 'lord'. Tocharian speakers would have naturally reinterpreted **xwar-xšaiṯa-* as meaning 'lord-sun' and would have calqued their compound **kaun-ñəkte* thereon. This only implies that the meaning 'lord' was already present in the Proto-Tocharian ancestor of *ñäkte* and *ñkät*. 'Moon-lord' and 'earth-lord' would then have been made after 'sun-lord'.

Naturally, this type of compound could not have been created if Tocharians did not divinize these natural beings. This is not a difficult process, as the Tocharian compounds themselves were calqued in Old Uyghur (see above). The source of this compound should be Old Steppe Iranian because of its Proto-Tocharian antiquity and because there is no trace of it in Khotano-Tumshuqese as far as I know. The Tocharian A word *ārkiśoṣi* means 'the world' (DToA: 44), literally 'the bright, white world/people' (*ārki* means 'white'). Pinault (1994: 366; 2008: 234) explains it as underlying a compound *ārki* + *śoṣi* "espace brillant et ensemble des êtres vivants", the first part of which refers to the world enlightened by the sun.[130]

130 Pan (2021) views it as a calque from Sanskrit *sabhāloka-*, with reinterpretation of *sabhā-* as containing the element *bhā-* 'bright'.

One could be tempted to see it as an adaptation of **gaiϑā-xšaiṭa* (*vel sim.*) in Old Iranian, with the meaning 'bright world'. However, such a compound is not attested, as far as I know, in any Iranian language, and besides, as I mentioned above, Iranian languages preserve the word **xšaiṭa-* with only one meaning, either 'lord' or 'bright', so that the idea of *ārkiśoṣi* as a calque of such an expression would contradict the argumentation above. Pinault's explanation is to be preferred. However, it is impossible not to think of Zarathushtra Spitama's words, recorded in the Uštavaitī Gāϑā, which I quote below.

Y. 45:1
nū īm vīspā ciϑrē zī mazdåŋhōdūm
"now conceive indeed (*zī*) of this entire [existence] as bright (*ciϑrē*)".

One could imagine an ancient Iranian expression *ciϑra ahu*, *ciϑra gaēϑā* or the like, meaning 'bright life', 'bright world' on which Tocharian A *ārkiśoṣi* would have been calqued. Until more evidence is found, this remains eminently speculative. It would further be more than unexpected to see an Old Iranian calque being only reflected in Tocharian A and not in Tocharian B as well.

In conclusion, I suggest that the Proto-Tocharian predecessor of TB *kauṃñäkte*, TA *koṃ-ñkät* 'sun-god/lord' was produced due to a calque of the Old Steppe Iranian reflex of **hwar-xšaiṭa-* 'bright sun' ~ 'lord sun'.

2.6.2 *TB* yok, *TA* yok *'Hair, Body Hair, Color'*

Tocharian B *yok* means both 'hair, body hair' and 'color' (DTB: 550; DToA: 373) as in *āssa yok* 'head hair', *oppal-yok* 'lotus-colored'. TAB *yok* does not have a convincing etymology (DTB: 550–551). The semantics of TAB *yok* correspond to the meaning of OIr. **gaṷna-* 'color, hair': Av. *gaona-* 'color, hair'; Pashto *γúna* f. 'color, complexion; hair (of body)' (Cheung 2010: 119); Pahlavi *gōn* 'color, complexion' (CPD: 37); Sogdian *γwnč* 'color', *γwn'k βwn'kh* 'hair pore' (SD: 175), etc.

The basic idea is certainly 'animal coat' > 'animal color'. It is not a trivial semantic development, but it not a very surprising one neither. It thus seems not out of place to suggest that the meaning of PT **yok* was extended to include 'color' or 'hair'. This naturally implies that the word TAB *yok* goes back to Proto-Tocharian, although it says nothing about its ultimate etymology. If one views this development as a calque, the directionality of the semantic extension is not entirely secure: while '(animal) hair' > 'color' is the obvious original shift, a word meaning 'color' could also mean 'hair' in a calque situation. We could perhaps connect this calque to the borrowing of the Old Steppe Iranian word for 'skin': TB *yetse*, TA *yats* (s.v.), for which one can imagine that it originally designated animal skin, or hide.

A Proto-Tumshuqese-Khotanese origin is also conceivable for this calque. A Middle Iranian source is less economical, as it would imply two separate calquing processes, one in TA and one in TB except, of course, if we take TA *yok* to be a borrowing from TB, which is not impossible, but relies on no positive evidence.

In conclusion, I suggest that the meaning 'hair, color' of TAB *yok* comes from a calquing of Old Steppe Iranian **gau̯na- vel sim*. This semantic distribution is observed in many Iranian languages, so it could also be a later calque from another Iranian language.

2.6.3 *TB* śaiṣṣe *'World', TA* śoṣi *'People'*

TB *śaiṣṣe* 'world; people' (DTB: 696) and TA *śoṣi* 'people' (DToA: 477) go back to PT **śaẃəṣye*, ultimately from the verb **śaw-* 'to live', TB *śay-*, *śaw-*, TA *śāw-*. Textually the Tocharian B word corresponds to both Sanskrit *jagat-* 'living world, living beings, world' and *loka-* 'world', which, in Tocharian A, is better rendered with *ārkiśoṣi* (see above). There is definite semantic permeability between 'world' and 'people', however, in my view, the shift is more commonly from 'world' to 'people' as in French *tout le monde*, *le monde entier* 'everybody', lit. 'all the world, the entire world', Classical Persian *ʿālam* 'the world' = 'people' (e.g. *ʿālam gūyand*, عالم گویند "the world says").[131] I thus want to suggest that the original meaning of PT **śaẃəṣye* was 'world' and that it took the meaning 'people' only secondarily, in the individual languages. In Tocharian A, the compound *ārkiśoṣi* was created in order to compensate for the semantic shift of *śoṣi* (see the etymological remarks concerning *ārkiśoṣi* in § 2.6.1).

The meaning 'world' of PT **śaẃəṣye*, related to the root 'to live' is reminiscent of the Iranian word for 'world': **gai̯ϑā-*, from the same root **gᵘih₃-* as Tocharian **śaw-*, as in Avestan *gaēϑā-* 'world, living beings' (perhaps originally 'livestock'), *gaēϑā- astuuaiti-* 'material world' (lit. 'bony world'). It is thus possible to view PT **śaẃəṣye* 'world' as a calque from Old Steppe Iranian. The calque process cannot be later than Proto-Tocharian, but it can also not be due to Khotanese-Tumshuqese contact, as **gai̯ϑā-* in that branch meant 'help, assistance, property' (cf. Dragoni 2023: 87 f.). The remaining candidate is thus Old Steppe Iranian.

In conclusion, I propose that the Proto-Tocharian word **śaẃəṣye* 'world' was derived from the root PT **śaw-* 'to live' as a calque of Old Steppe Iranian **gai̯ϑā-*,

131 Sanskrit *jagat-* has undergone a different pathway: it designates originally the moving world, either moving out of Brahmā's mind, or as the collective of all moving beings (including celestial beings).

synchronically understood as deriving from the verb 'to live' (**ǰaiH-*) and probably designating both the living world and the world in general, as in Avestan.

2.7 Discussion of the Features of Old Steppe Iranian

I hope that the multiple arguments developed throughout this chapter have been consistent enough to convince the reader that there was an Old Iranian language, of a quite archaic nature, that came in contact with both Tocharian languages at a very early stage or, more likely so, with Proto-Tocharian. The loanwords from this layer are coherent in nature, and the sound correspondences I have been able to establish are regular. The lexical categories these loanwords belong to are also limited and give a coherent idea of the type of contact that existed: military terminology; various objects; animals and animal products; grammatical words; abstract notions. A logical conclusion from this fact is that this layer actually only represents one single language, the so-called Old Steppe Iranian language.

This layer of Old Steppe Iranian loanwords in Tocharian is homogeneous, and almost all—if not all—of these loanwords were borrowed according to well-established borrowing patterns, implying a high degree of fluency or knowledge of this language among Tocharian speakers.

Although a few words played an important role in the delimitation of this language, one in particular played a decisive role. It was previously believed that an Old Persian word had spread as far as Tocharian, or to the Old Iranian language in contact with Tocharian, namely TB *peret*, TA *porat* 'axe'. By determining that the predecessor of TB *peret*, TA *porat* must have been regularly Old Iranian **paratu-* and was not an indirect borrowing from Old Persian as previously assumed, I hope to have demonstrated that no Old Iranian loanword in Tocharian was demonstrably borrowed from a Persic language.

The question that ensues is of course the following: to which branch of the Iranian tree did this language belong? To which known (i.e. written) Iranian language was this language closer (or closest)?

Answering this question is not easy. The Iranian branch is by far the most geographically extended and, in terms of the number of languages, the richest of all the Indo-European branches. Out of an unknown number of branchings,[132] only two Old Iranian languages are better attested, Avestan and Old Persian, and a third one, Old Scythian, is attested in an extremely fragmen-

132 They are unknown for two reasons: 1. Branches may have existed of which we have no

tary and mostly indirect fashion. Until now, no other Old Iranian language has been attested, although we know they must have existed, since very few of the many existing Iranian languages descend directly from the three mentioned above.[133] Attested Middle Iranian languages are more numerous. We can count eight of them: Middle Persian (in fact two main varieties: Pahlavi and Manichean Middle Persian), Parthian, Bactrian, Sogdian, Khotanese and Tumshuqese, Khwarezmian, and Alanic. A ninth one, the Kushan language, was recently deciphered (Bonmann, Halfmann, Korobzow & Bobomulloev 2023).

After the Middle Iranian period, the number of attested New Iranian languages is much greater, and too uncertain to be enumerated here. We are thus presented with the figure of a funnel, where—traditionally—all the modern Iranian languages are supposed to descend from a group to which one of these older languages, both Middle and Old, belonged. It has been demonstrated recently that this model is impossible (Korn 2016; 2019). At least several Middle and New Iranian languages go back to unattested Old Iranian languages, which were not necessarily in a closer relationship with other known branches. Such is the case, for instance, of Balochi (see Korn 2005).

Based on this state of affairs, it cannot be stressed enough how crucial the discovery of the Old Iranian layer of borrowings (dating back to Isebaert 1980, Schmidt 1985, Tremblay 2005, detailed and made explicit in Peyrot 2015a; 2018a) is for Iranian studies. Not only do we have one more Old Iranian language to count on for drawing the phylogenetic tree of Iranian languages, one that is not hypothesized for the sake of it, but attested through another language, but we also have the confirmation and rejection of a number of predictions made by Iranists over the past two hundred years (see below).

A very important step in the present study consisted of excluding a group of words which were considered Iranian, or Old Iranian, and prevented scholarship from establishing straightforward facts (for example: *witsako* 'root', s.v.). I also had to review a number of etymologies, and revise various facts concerning these etymologies, for example by proposing a post Proto-Iranian **paratu-* and not **paraću-* for the word for 'axe'. As a result, I have gathered the features of the Old Iranian language in question, as found in Tocharian A and B, in a consistent and systematic comprehensive manner.

---

trace or knowledge, and 2. Even for the attested languages, there is no consensus as to how many inner-Iranian branches have to be posited.

133 More precisely, Avestan has no known (direct) descendant, and an undocumented dialect of Old Persian gave birth to Middle Persian, another to New Persian (through its unattested direct predecessor, cf. Korn 2021), and possibly to one or two other dialects. Little is known about Old Scythian (or Sarmatian), for a summary on our knowledge thereof, see Novák 2013: 9 f., but it seems to be the ancestor of the Scythian or Alanic languages, which constitute two closely related groups of languages.

An important feature (see Schmidt 1985; Peyrot 2018a) is that PIr. **ć* in that language is reflected by **t^{s}*, and that PIr. **ǰ* (PIIr. **ʃ* and **ʃ^{h}*) is reflected by **d^{z}*, both rendered as *ts* in Tocharian. The realization *[ts] and *[dz] of the PIE palato-velars at a very early stage of Iranian has been assumed by Iranists for a long time, but Old Steppe Iranian offers concrete evidence for this realization.

Before discussing the features of Old Steppe Iranian in connection to the identification of that language, it should be mentioned that there will always be two difficulties that will make this identification arduous. The first is the difficulty in deciding whether the relevant sound changes occurred within Old Iranian or in the borrowing process to Tocharian. The second is that this Old Iranian language is evidently old and archaic, and that consequently some of the innovations found in supposedly related or even descendant languages might not be found in it.

In the following, I will first present a table of Old Steppe Iranian–Tocharian vowel correspondences, based on the word-studies of the present chapter (2.7.1). I will then discuss all the relevant phonetic innovations of that Old Iranian language, that is, those that differentiate it from Proto-Iranian in a clear manner. Those innovations are divided between weaker and stronger ones, from two points of view: on the one side we have clear innovations, seen in multiple Tocharian words, such as PIr. **rd* > OIr. **ld* and on the other side we have possible innovations, such as the *u*-umlaut that can perhaps be seen in *mot* 'wine' (from **madu-*), for which I was not able to find any other example or counter-example. In section 2.7.3, I will present some lexical isoglosses, although with the caution that, not all Iranian languages being as well documented as each other, there is an inevitable bias in terms of lexical isoglosses. In section 2.7.4, I will discuss relevant semantic changes. These will often be more certain than the phonetic changes in the sense that their being Iranian or Tocharian developments is easier to establish.

Both in section 2.7.2 and 2.7.4, when discussing those changes, I will try to systematically compare them with a number of Iranian branches or languages. This will inevitably be impaired by a few difficulties, the greatest of which being the fact that some languages are much better studied than others. Ideally one should compare the Old Steppe Iranian language with every Iranian language, but this would lead to a much longer and not necessarily more useful discussion, as many Iranian languages are both attested too late and situated too far away to yield any significant result.

The languages I will compare this Old Iranian language with are: Old Persian (and the Persic group in general), Sogdian, Yidgha (or Yidgha-Munǰī when necessary and possible), Balochi, Bactrian, Khotanese (and Tumshuqese), Khwarezmian, Pashto, and, very importantly, Ossetic, since Ossetic is a direct descendant of a Scythian steppe language. I will also compare it when possible with

Wanjī, an extinct modern Iranian language spoken up to the end of the 19th century in the Wanǰ valley of Tajikistan (in the center south part of the country). The data we have on Wanjī is scarce, so comparison with this language will be very limited. It will be shown that Wanjī is not phylogenetically close to Old Steppe Iranian.

The goal of this comparison is to permit us to ascertain whether an ancestorship status can be given to Old Steppe Iranian for any specific known Iranian language, on one hand, and what type of linguistic proximity Old Steppe Iranian has with other Iranian languages—notably, to which branch of the Iranian tree does Old Steppe Iranian belong. The results of this analysis are still subject to caution for the time being. To have a better idea of the answer, one has to combine data from this section and from the two next sections, which discuss lexical isoglosses of Old Steppe Iranian and other Iranian languages (2.7.3) and the semantic innovations of Old Steppe Iranian compared to Proto-Iranian reconstructible meanings of words (section 2.7.4.).

2.7.1 *Old Steppe Iranian–Tocharian Vowel Correspondences*

This table is not exhaustive, in the sense that it does not include exceptions, or problematic and secondary sound changes. It is thus not to be used as a reference for these sound correspondences, but rather as a succinct convenient summary. All the relevant details and problems are discussed in the following pages. As questions relative to consonants are too intricate for a simple table, I refer the reader to the following pages for the discussion of relevant questions.

TABLE 1 Sound substitutions between Old Steppe Iranian and Tocharian loanwords

Old Steppe Iranian reconstruction	**Proto-Tocharian rendering**	**Tocharian B reflex**	**Tocharian A reflex**
**a*	**e*	*e*	*a*
**ā*	**a*	*a*	*ā*
**aɪ̯*	**ey*	*ey*, *ai*	*e*
**aʊ̯*	**ew*	*ew*, *au*	*o*
**ɪ* (or **ə*)	**ə*	*ä*	*ä*
**ɪ́*	**ə́*	*a*	*ä*
**-ɪ́ɪ̯V-*	**-əy-*	*-iy-*	*-i-*
**-ɪɪ̯n-*	**-in-*	*-in-*	n.a.
**-u-*	**-u-*	*-u-*	*-u-*
**-u*	**-u* or **-∅*	∅	∅
**ŕ̥/ár*	**er*	*er*	*ar*

2.7.2 *Phonetic Changes of Old Steppe Iranian*

In order to make inferences about the phonetics and phonology of Old Steppe Iranian, we have to differentiate sound changes from sound correspondences. For instance, Old Iranian **a* was rendered as Proto-Tocharian **e*, and Old Iranian **ā* was rendered as Proto-Tocharian **a*. This different rendering strongly suggests that there was an important phonetic distinction in timbre between Old Steppe Iranian **a* and **ā*, on top of the difference in length.

One option is to assume that Old Steppe Iranian **a* was more front [æ] and **ā* was more back [ɑː]. Indeed, Tocharian *a* was perhaps also rather back, i.e. [ɑ] or [ɒ] (cf. Pinault 2008: 421). These values would fit the phonetics of these two phonemes in Persian very well,[134] and could perhaps reflect the values of Proto-Iranian without further changes, or with only slight changes.

Another option is that originally Proto-Iranian **a* was closer to schwa ([ə]), as can be found in the literature (e.g. Holopainen 2019: 30 for Proto-Indo-Iranian). In case Proto-Iranian **a* was indeed phonetically a schwa, I would like to suggest that Old Steppe Iranian underwent a chain shift, where **a* became more front, and perhaps **i* and **u* more central.

2.7.2.1 The Treatment of **-θi̯-* and **-di̯-* and the Phonetic Value of **-θ-* in Old Steppe Iranian

An interesting problem concerning Tocharian adaptations of Iranian sounds is the correspondence between Proto-Iranian **-θi̯-*, **-di̯-* and PT **-cc-* seen in TB *waipecce* 'possession' and TB *kercci* 'palace', and possibly in TB *iścem** 'clay'. Here we have the Proto-Iranian reconstructed forms **hu̯ai-paθi̯a-*, **gr̥di̯a-*, **ištu̯am* and the Tocharian borrowings from the Old Iranian language: *waipecce*, *kercci*, *iścem*, but we cannot reconstruct the Old Iranian forms easily on this basis. We have three choices here: α. either the Old Steppe Iranian forms were **hu̯ai-paθi̯a-*, **gardi̯a-* and **ištu̯am*, and Tocharian palatalized them to *waipecce*, **kercci*(*ye*) and **iṣcem* or β. they had become **-čč-* and **-j̈j̈-* already in the Old Iranian language, or γ. in the sequences **-θi̯-* and **-di̯-*, the initial consonant was a palatalized sound, with an outcome **tʲy* and **dʲy*, which were further palatalized in Tocharian.

There are arguments in favour of all three options, and one has to propose a coherent solution before proceeding to the comparison with any documented Iranian language.

---

134 A great variety of languages use ‹e› to transcribe [æ] and ‹a› to transcribe [ɑ] or [ɒ], e.g. most foreign notations of Persian *a* ([æ]) ‹e› and *ā* ([ɒ]) ‹a›. There exists also a tendency, among languages, for [æ] to go to [e], and for [ɑ] or [ɒ] to go to [a], as in, for example, Xurāsān varieties of Persian (Monchi-Zadeh 1990: 1).

α. The first solution is the simplest, but it relies on a supposed Tocharian change *tyV* > *-cc-* that did not occur as far as we know (cf. Pinault 2008: 551). As is known, the PIE cluster *-tḭ-* became **t^s* in Proto-Tocharian, while later *-ty-* sequences were borrowed as such (e.g. *pratyaikapañäktäññe* 'pertaining to a buddha who reserves his enlightenment to himself' DTB: 443). That this supposed Tocharian sound change *tyV* > *-cc-* would have occurred only in Old Steppe Iranian borrowings seems very unlikely, but not completely impossible: it would have been a punctual sound change, in order to render a sequence in Tocharian that was impossible at the time. This is still difficult, because this replacement of **-ty-* to *-cc-* would have only occurred once and for a short time. One can now add one more example, not from Old Steppe Iranian, but nonetheless from an early borrowing: TB *cowo** 'theft', which Federico Dragoni (2023: 118 f.) interprets as deriving from Pre-Khotanese **dyūa-* 'theft', through its accusative singular **dyūwu*. If this etymology is correct, this particular sound substitution may have taken place at a very early period, preceding later Middle Iranian and Indic loanwords.

β. The second solution, namely **(-ϑḭ-) > *-tḭ-*, **-dḭ- > -čč-* and **-ǰǰ-* within Iranian, has the advantage of explaining the Tocharian forms in a very simple way. It would imply that the gemination was preserved in Tocharian. An argument in favor of this option is that we find *kercci* rather than **kelcci*, whereas this Old Iranian language had undergone **rd* > **ld*, at least in some contexts. In this scenario, **-dḭ-* would have become **-ǰǰ-* before **rd* had become **ld*, thus yielding a cluster **-rǰ-*. A difficulty with this scenario lies in the fact that no known Iranian language—to my knowledge—presents such a change.

γ. The third hypothesis, namely assuming that **ϑḭ-* and **-dḭ* had become **tʲy* and **dʲy* in Old Steppe Iranian, presents no particular drawbacks, in my view. In its favor: it necessitates no sound change from the Iranian side other than **-ϑ- > *-t-* which is needed for the previous hypothesis as well (see below), since an interdental fricative would need to become a stop before in order to become a postalveolar affricate. It has two major advantages: it explains the palatalization and provides a coherent system for the Iranian side. The posited Old Steppe Iranian **tʲy* and **dʲy* could also have been rendered at first as **cy* in Proto-Tocharian, assimilated to **cc* in Pre-Tocharian B, and turning to **ci* in Pre-Tocharian A.

If scenario α. were to be chosen, then nothing could be said about the nature of **-ϑ-*, as both **-ϑḭ-* and **-tḭ-* would yield *-cc-* in Tocharian. Again, Proto-Tocharian would have borrowed **-ϑḭ-*, **-tḭ-* and **-dḭ-* as such and palatalized them into **-cy-* "itself".

If scenario α. is discarded, no matter whether scenario β. or γ. is chosen, a sound change of *-ϑ- to *-t-* would seem to have happened first, at least before *yod*. Since I was unable to find a parallel for a conditional defricativization of *-ϑ-* in front of *yod* only, it is possible that *-ϑ- > -t-* was a general shift in Old Steppe Iranian. However, I admit that there is no other evidence for this shift than the palatalization under discussion here.

If a general shift *-ϑ- > -t-* occurred, Old Steppe Iranian is closer to Ossetic, which had *-ϑ- > -t-* (Cheung 2002: 21), but farther away from Old Persian, Yidgha-Munǰī, Khotanese, Khwarezmian, Sogdian and Bactrian, for all of which *-ϑ- rather than *-t-* needs to be reconstructed.[135] Balochi also underwent a change *-ϑ- > -t-* (Korn 2005: 81), but it would be illusory to compare it here, as the sound change in Balochi is obviously secondary (since *-ϑr- became -ss-*, for example, see Korn 2005: 89), and is part of a single phenomenon of occlusion of fricatives, which is posterior to the time period we are speaking of (cf. Korn 2005: 323 f.). Pashto is more difficult to evaluate. In intervocalic position *ϑ does indeed merge with *-t- (> -l-*), as in *plən*, Waziri *plan* 'broad, wide' from **paϑana-*, but most probably this change went through an intermediate *δ, not through *t* (cf. already Geiger 1901: 209). The only other word from our list which may go back to a form with *-ϑ- is Tocharian A *nātäk* 'lord' (← OSIr. **nātaka-* or **nāϑaka-* 'protector').

If scenario β. is favored, then the proximity with Ossetic is even more prominent, as Ossetic is the only Iranian language known to have undergone *ϑi̯ > *-č- > -ts-* (transcribed as ‹c› in the transcription based on Cyrillic), cf. Iron *syfc*, Digor *sufcæ* 'shoulder' < **sufϑi̯a-* (Cheung 2002: 34); Iron *fælmæcyn*, Digor *fælmæcun*, 'to get tired', from **pari-maϑi̯a-* (Cheung 2002: 185). There is no trace of gemination in Ossetic here, but there is no clear indication against it having been there at an earlier stage either.

The words TB *epiyac*, TA *opyāc* 'memory' could also belong here, given their final *-c*. These words correspond to PIr. **abi-i̯āta-* but the final *-c* finds no easy explanation. It has been usually admitted that the proto-form from which Tocharian derives was **abi-i̯āti-* but there is no evidence for an *i*-stem for this word. More importantly even, there is no proof that *-ti* became *-c* in Old Steppe Iranian, or in Tocharian. If my derivation of TB *epastye* 'skillful' < PT **epəstiye* is accepted, it would rather show that Old Steppe Iranian *-ti-* was not palatalized in Tocharian.

135 Khwarezmian even seems to preserve PIr. *ϑ in some positions.

2.7.2.2 Old Iranian **čį́* or **šį́* → Proto-Tocharian **ś*?

The Proto-Tocharian word **śate*, TB *śāte* TA *śāt* 'rich', was borrowed from a word that goes back to Proto-Iranian **čį́āta-* 'happy'. To explain this we can assume that the palatalized phoneme **č* + *yod* or **š* + *yod* was rendered as **ś-* or as **tś-* (with secondary "simplification" **tś-* > **ś-*) in Proto-Tocharian, where word-initial gemination was not possible. This is more economical than suggesting a sound change within the Old Iranian language to a palatal sound,[136] which corresponded to the palatal phoneme of Proto-Tocharian. The main question is to determine whether the Tocharian **ś-* derives from a palatalized **š*, as we have it in Avestan and many other Iranian languages, or from a palatalized **č*.

Proto-Iranian **čį́āta-* has an Indo-European etymology: it derives from **$k^{u}ieh_1$-* 'to be quiet, calm' (Lat. *quietus* 'calm'), and the Proto-Iranian root can be set up as **čį́aH-* (cf. EDIV: s.v.). However, from the Iranian point of view, the Common Iranian form is often reconstructed as **šį́āta-* (e.g. in Korn 2005). This is erroneous: the Common Iranian form is also **čį́āta-*, as we can conclude among others from the Ossetic form *cadæg* 'quiet, slow' (EDIV: 38).

First, a look at Khotanese and Tumshuqese may be useful. The Khotanese and Tumshuqese languages, which constitute, together with Wakhi, their own branch of the Iranian languages family, have the cognate form *tsāta-* 'rich', which also goes back to **čį́āta-* (cf. Bailey 1961: 54). Since the Khotanese-Wakhi-Tumshuqese branch is the first to have branched out, our notion of Common Iranian excludes these languages. Thus, the form *tsāta-* < **čį́āta-* does not directly help us for the reconstruction of the Common Iranian descendant of Proto-Iranian **čį́āta-*, but it shows with certainty that the Proto-Iranian form was not **šį́āta-* but indeed **čį́āta-*.

Ossetic, however, is included in Common Iranian, and the form *cadæg* 'quiet, slow' mentioned above thus shows that there was no Common Iranian sound change **čį́-* > **šį́-*. Furthermore, the Khwarezmian form *hncy-* 'to rest, relax', Sogdian *ʾnčʾy* 'to stop, to cease, to rest' and the Ossetic forms Iron *æncajyn*/*æncad* Digor *æncajun*/*æncad* 'to rest', all deriving from **ham-čį́ā-*, show that **čį́* did not become **šį́* in word-internal position in the ancestors of these languages (so Bailey 1961: 54).

If one accepts that PT **cy* or **cc* derives from Proto-Iranian **ϑį́*, **dį́* by means of further palatalization in Tocharian of an already palatalized consonant in Old Steppe Iranian (see above), then I also believe it is more economical

136 Avestan *ś́-* originally derives from a palatalized **š* (**šʲ*), in the sequence of sound changes **čį́-* > **šį́* > OAv. *ś́y-* (‹ś́ii-›) > YAv. *ś́-*, and Old Persian *š*(*i*)*y-* simply shows **čį́-* > **šį́*, so that an intermediary step with a palatal **ś* is not needed for the reconstruction of the word in Avestan.

to assume that a palatalized sequence *$\check{c}^{j}_{\underline{i}}$ [t͡ɕj] would be borrowed as *$ś$- in Proto-Tocharian. With the caveat that the evidence is circumstantial, I thus assume that the phonetic shape of the word was *$\check{c}^{j}_{\underline{i}}āta$- in Old Steppe Iranian.

A form derived from Proto-Iranian *$\check{c}^{j}_{\underline{i}}āta$- was also borrowed into Proto-Permic *$šud$ 'luck', as well as into Khanti and Mansi (Katz 2003: 159; Holopainen 2019: 265 f.). In every case, the forms suggest an original *$š$- and no trace of a *yod*, which must have disappeared quite early in the source language. Nevertheless, as Holopainen (2019: 266) writes, "[i]n theory, Mansi *š* could also reflect earlier *č*, so the word can also be borrowed from an Iranian variety which preserved the affricate *č*." I thus wonder if Old Steppe Iranian or a closely related language could have been the source of the Mansi borrowing. The difference in meaning would not be an obstacle to this etymology, as the Tocharian meaning is due to a later evolution (see section 2.7.4.2).

It seems difficult to establish any specific isogloss here, both from the Iranian and from the Tocharian side. It is possible to speculate on a form with a palatal *$śāta$- in the Old Steppe Iranian language, but, although it fits the Avestan data, this would not be warranted by any other Iranian data, and would serve no other purpose than to explain the Tocharian form. An important point to note is that there is no parallel for *cy- > *$ś$- in Tocharian. This seems to suggest that, indeed, the Old Steppe Iranian word showed some degree of palatalization of its initial sound or sequence of sounds.

2.7.2.3 Old Steppe Iranian *aha > *$ā$

A very important feature of Old Steppe Iranian is the phonological change of *aha (possibly *[æhæ]) > *$ā$ (possibly *[ɑː] or *[ɒː]). This change is only seen in one Old Steppe Iranian word in Tocharian, TB *ainake* (Archaic TB *eynāke*), TA *enāk* 'evil, bad', but it is quite clear from it. Indeed, Old Iranian *a was borrowed as Proto-Tocharian *e, and Old Iranian *$ā$ was borrowed as Proto-Tocharian *a. The two Old Iranian phonemes *a and *$ā$ were different, not only quantitatively but also qualitatively according to the data at hand. There is thus very little place for the possibility of a sequence of two Old Iranian vowels *aha to be borrowed as *a in Tocharian, as it would imply the backing *within* Tocharian of two front vowels, to render a length that would be, in Tocharian, non-phonemic. In other words, it is not plausible that *$e.e$ would have become *a in Tocharian, cf. TB *pernent*, the obl.sg.m. of TB $perne_u$ 'glorious, worthy (one)', which derives from *$perne.ent$ < *$pernewent$ (and is not †$pernant$).

Once this fact is established, two main consequences can be proposed for our Old Iranian:

1. Despite the fact that those two vowels **a* and **ā* were certainly phonetically articulated differently (as they were rendered quite differently in Tocharian), it is possible that they were phonologically the short and long counterpart of each other.
2. It is possible, if not likely, that this Old Iranian language had no phoneme *h* in intervocalic position, at least clearly not in this specific position (between two *a*s).

Other Iranian languages have undergone the same change (**aha* > **ā*) at a relatively early stage. Sogdian is one of them, e.g. *sāk* 'number' < **sahaka-* cf. Gharib, SD: s.v., but so is Ossetic, e.g. *wat* 'room' < **u̯ahaϑa-*, cf. Cheung 2002: 24, and Khwarezmian too (*s'k* 'hare' < **sahaka-* cf. Benzing 1983: 567; *rāk* n.f. 'vein' < **rahakā-*, cf. MacKenzie 1990: 121), Yidgha-Munǰī (*sīγ* 'hare' < **sahaka-*, cf. Morgenstierne 1938: 57) and Pashto (*wéra* f. 'fear' < **wāryā* < **au̯aharį̄ā-*, cf. Cheung 2011: 192). Khotanese, Tumshuqese, Parthian, Balochi, Old Persian, Bactrian, all preserve the sequence **aha* (e.g. MP *rahag* 'vein',[137] Parthian *rhq* 'vein', Bactrian πιδοοαυανο 'request' < **pati-u̯ahana-*). The Pahlavi form ‹s'k› (either *sāk* or *sāg*) 'number' is considered of obscure origin by Ciancaglini (2015: 299, 299[12]). In my view, it should be connected to Sogdian *sāk* 'number' (see above) and, in view of, for instance, *rahag*, I believe it is perhaps a borrowing from another language, perhaps Sogdian.

2.7.2.4 Loss of PIr. **h-* in Old Steppe Iranian

Apart from the contraction of **aha* to **ā*, there is another fact which suggests that **h-* was dropped in Old Steppe Iranian: the Tocharian B name of India, *yentuke**, which shows the same initial vowel as TB *yetse*, TA *yats* 'skin', ultimately from PIr. **iǰa-* (Av. *izaēna-* 'made of skin'). There are Tocharian loanwords with *i-* in initial position: see for example TB *iṣcäke* 'clay'. Even if word-initial *i-* were impossible to pronounce for Proto-Tocharian speakers, why would they not have rendered it with *yə-*, which existed in the language (cf. TB verb *i-* 'to go' /yə/), rather than with *ye-*?

It is more plausible that initial *i-* in that Iranian language was rendered as **i̯a-*, **i̯i-* or a variant thereof, such as **i̯e-*, adapted in Proto-Tocharian as **ye-*. And, again, it is much more economical to suggest that the initial *h-* was dropped in that Iranian language, like in intervocalic position, because **hi-* in **hinduka-* has the same representation in TB *yentuke* as **i* in **idza-* 'skin' in TB

137 Note, however, MP *afsānag*, NP *afsāna* 'fable' of which one of the possible etymologies derives it from **abi-sahanaka-* (see Gershevitch 1971: 274; cf. Hasandoust 2014: 242 f.), or from **upa-sahanaka-*, if it is not borrowed from Parthian. Compared to, for instance, NP

yetse, TA *yats*. I assume that only after the loss of **h-* was initial *i-* (both primary and secondary) turned to **i̯a-* or similar.

For the loss of *h-*, one can cite Sogdian, Pashto, Ossetic, Khwarezmian and Yidgha-Munǰī, against Bactrian, Balochi, Avestan, Old Persian, Khotanese and Tumshuqese. Here we need to address two questions of relative chronology. First, the loss of *h-* in Sogdian and Khwarezmian is relatively old, since it precedes the apheresis of *a-* in both languages (**haC-*, **aC-* > *C-*).

Another argument suggests that **h* was lost in initial position: the apparent reconstruction of Old Steppe Iranian **u̯ai-* (see below). Indeed, in Sogdian, Ossetic and Khwarezmian, the loss of *h-* is later than the sound change **hŭ-* and **hu̯-* > *xu-*, *xw-*. Compare, for instance, Buddhist Sogdian *ʾγwšh* 'sister-in-law' < **hu̯asru-* (cf. Gershevitch 1961: 51), Ossetic Iron $x_{o}y$, Digor *xu* 'pig, boar' < **hū-* (cf. Cheung 2002: 16, 25), or Khwarezmian *xubisk* ‹xbsk› 'eigen' < **hu̯ai-paϑi̯a-ka-* (cf. Benzing 1983: s.v.). Thus 1. **hŭ-* and **hu̯-* > **xw-* and then 2. **h-* > **∅-*. As a comparison point, one can observe that both Khotanese and Tumshuqese, which do not lose **h-* at all did not undergo such a sound change: Khot. *hĭvia-*, Tumshuqese *havya*, *hävyä* 'one's own'. This is also the case for Northern Talyshi, a geographically very distant Iranian language, which has *hān-* 'read, sing' (past stem *hānd*) and *hārd* 'ate', obviously from **hu̯ān-* and **hu̯ar-*.

The importance of this observation is due to the fact that, in Old Steppe Iranian, there was no apheresis nor was there any sound change **hu̯-* > **xu̯-*, as is seen in Tocharian B *waipecce* (from **xu̯aipaϑi̯a-* one expects TB ***kwaipecce*, cf. TA *kump* 'bowl' ← Old Steppe Iranian **xumba-*, unless this is an Indic loanword).

This seems to imply that, in this Old Iranian language, *h-* was dropped first before any type of vowel, unlike what happened in Sogdian, Ossetic and Khwarezmian. This lowers the chances that this change occurred as part of a single areal phenomenon, and brings to zero the chances of this change having occurred as part of a common innovation with these languages. Naturally, it cannot be excluded that **hu-* and **hu̯-* were still pronounced as such in Old Steppe Iranian, and that the **h-* is simply not reflected in Tocharian, although this is not congruent with the other data presented here and above.

In Pashto, **hu̯-* became **xw-* (e.g. **hu̯ásru̯i̯ā-* 'mother-in-law' > Pashto *xwā́xe*), but **hu-* seems to have become **u-* (**hufta-* 'asleep' > Eastern Pashto *udə́*, Waziri *wewd* 'id.'). In any case, the change **hu̯-* > **xw-* of Pashto excludes it here as well. Because this change also occurred in Yidgha-Munǰī (Morgenstierne 1938: 68), these two should be excluded as well.

---

rag 'vein' < MP *rahag*, one might consider that the *-ā-* as a reflex of **-aha-* in *afsānag* is not regular.

The Persic group represents a difficulty here: we have, on the one hand, Middle and New Persian which show very clearly **hu̯-* > **xw-* (*xwad* 'self', for example) and on the other hand, there is documented Old Persian, which did not undergo this sound change: rather, in Old Persian **hu̯-* > **huw-* > *uw-* (cf. Hübschmann 1895: 217 f.; Korn 2021: 100 f.).[138]

To note, Proto-Iranian final **-ah* became **-i* in Ossetic (cf. Cheung 2002: 56 f.; Kim 2003: 57 f.), which very likely indicates that **-h* was lost later in Ossetic. In Old Steppe Iranian, there is no trace of such a change, with, for instance, PT **perne* and **entse* pointing to Old Steppe Iranian **a(h)* as a reflex of Proto-Iranian **-ah*. On the basis of the other elements here, it is possible to speculate that the Old Steppe Iranian ending was simply **-a*, without final **-h*, as /h/ had disappeared in all other positions, as far as we can tell.

2.7.2.5 Old Steppe Iranian Reflexes of PIIr. **ʃ* and **ć*

A useful feature to look at is the result of PIIr. **ʃ* and **ć* in Old Steppe Iranian, that is, the Proto-Indo-European palatovelars. Both phonemes are rendered as **ts* in Proto-Tocharian, and as *ts* in Tocharian A and B, in the words TB *etswe* 'mule', *ñyātse*, TA *ñātse* 'danger', TB *entse*, TA *eṃts* 'envy, greed', TB *yetse*, TA *yats* 'skin', TB *tsain* 'arrow', TB *tsaiññe* 'ornament', and, if my etymological proposals are accepted, also in TB *wertsiya* 'community, assembly', TB *tˢəw-*, TA *tˢäw-* 'to adhere, to fit, to put together', TB *mətstsa-* and TA *nätswā-* 'to die of hunger, to starve', TB *tsetke* 'centaine (?)', TB *tsere* 'a measure of liquid volume of half a *lwāke* (?), cup (?)' and TB *tseriteke* 'young grass (?), green animal (?), turtle shell (?)'.

The reconstruction with **tˢ* and **dᶻ* was established by scholars of Iranian since a few decades (e.g. Klingenschmitt 1975; Mayrhofer 1983; 1989: 6 f.). Ever since it was noted that the earliest Iranian loanwords in Tocharian show *ts* for PIIr. **ʃ* and **ć* (Schmidt 1985), we actually have under our eyes the reflexes of these Old Iranian sounds, which were until then only reconstructed.

It has been postulated for multiple reasons that Proto-Iranian **ć* and **ʃ* did not become **tˢ* and **dᶻ* in all positions. The question is too thorny to be discussed here in its entirety. However, the Khotanese-Tumshuqese-Wakhi branch preserved a palatal pronunciation of **ć* and **ʃ* in front of **u̯*, that is, this branch never had **tsw* (Sims-Williams 1998: 136; 2017: 264; Peyrot 2018a).[139] As Peyrot

138 In her article, Korn argues that the loss of **h-* in documented Old Persian is one of the features that exclude the possibility that it is the direct parent of Middle and New Persian.

139 We cannot, at this point, be sure that Khotanese-Tumshuqese *s*, *z* derive from an interme-

(2018a) demonstrates, the Tocharian B word *etswe* 'mule' derives from Proto-Iranian **aćm̯a-* 'horse' (< PIE **h₁ék̑m̯o-*) through a borrowing from Old Iranian **atˢm̯a-*.

As Peyrot (2018a: 271f.) argues, the Khotanese-Tumshuqese-Wakhi branch is excluded for Old Steppe Iranian because of the outcome **-tsm̯-* in Old Steppe Iranian of Proto-Iranian **-ćm̯-*. The Khotanese word for 'horse' is *aśśa-*, which goes back directly to **aćm̯a-* with preservation of the palatal and assimilation of the glide.

Some scholars do not regard Khot. *-śś-* as the regular development of **-ćm̯-*, but rather the outcome of **-tsm̯-*, for instance Kümmel (2007: 234) and Novák (2013: 121–122[186]).[140] However, new considerations about TB *eñcuwo*, TA *añcu** 'iron', which was borrowed from a preform of Khotanese *hīśśana-* 'iron' (s.v. *eñcuwo*; cf. also Peyrot, Dragoni & Bernard 2022), show that the sequence **-ńćm̯-* in Pre-Khotanese or Proto-Khotanese-Tumshuqese became a sequence of nasal + palatal, which must have been a palatal sound as **-ć-* in **-ńćm̯-* did not merge with the result of PIr. **-č-*. This discovery thus invalidates the hypotheses of Kümmel, Novák, and confirms the exclusion of the Khotanese-Tumshuqese-Wakhi branch as a close relative of the Old Steppe Iranian language, as already argued in Peyrot (2018a).

Thus, the Old Steppe Iranian language cannot be closely related to Khotanese, Tumshuqese and Wakhi, but is phylogenetically closer to the rest of Iranian languages.

It is still unclear how Old Persian, and the Persic group in general,[141] which show **s* for **ćm̯*, arrived at that result (cf. for instance Kümmel 2007: 234). In any case, the Persic group did not preserve **ćm̯* as such and cannot be excluded from the comparanda on the basis of this feature.

2.7.2.6 Proto-Iranian **i* and **u* in Old Steppe Iranian

A thorny question concerning Old Steppe Iranian is that of the treatment of the short vowels **i* and **u*. As is known, PIE short **i* and **u* were changed to schwa in Proto-Tocharian (see for instance Ringe 1996: 133f.). New **i* and **u* phonemes arose from PT **əy* and **əw*. However, Tocharian speakers seem to have had continuous problems with the pronunciation of /i/ and /u/, as can be seen in, for instance, TB *pañäkte* 'Buddha' with *pa-* from **pät* 'Buddha', itself a

---

diary stage **tˢ*, **dᶻ* (respectively) and not directly from a palatal phoneme as in **ʃ* > **ź* > *z* instead of **ʃ* > **dᶻ* > *z*, for instance.

140 Including a suggestion by Reiner Lipp.

141 That is, all Iranian languages that show a number of common innovations, such as **ϑ* and **z* for **ć* and **ʃ* in a number of positions, and **s* in pre-labial context.

borrowing from **bud* or **but* (perhaps from Sogdian /but/) compared with the poetic variant *pūdñäkte*, thus with "learned" *pūd-*. One can also look at the variation between *kuñcit* and *kwäñcit* 'sesame' (cf. Bernard 2020: 52 f.), and *sakw* 'luck' ← Skt. *sukha-* 'happiness'.

In our list of Old Iranian loanwords in Tocharian, there are a number of words which had a short **i* or **u*, in several positions. In initial position there are TB *yetse*, TA *yats* 'skin' going back to a form borrowed from the Old Steppe Iranian reflex of Common Iranian **idᶻa-* in Proto-Iranian, and TB *yentuke* 'India' which goes back to an Old Iranian form **induka-* < **hinduka-* 'India' (for the loss of the initial /h/, see above). There was thus a sound change **i-* to **i̯V-* (**i̯a-* or **i̯i*(*/e*)-?) in Old Steppe Iranian.

The change of **i-* to **i̯V-* is not entirely uncommon in Iranian, but it is not found in Khotanese, Pashto, Ossetic, nor in Old Persian or Bactrian. There is definitely an alternation, in writing, between *i-* and *yV-* in Sogdian and possibly in Khwarezmian, for instance in the Sogdian variants *ʾync ~ ynch* 'woman' (SD: s.v.), and in the Khwarezmian variants *ʾyδʾβrc ~ yδʾβrc* 'Zwischenräume habend' (Benzing 1983: s.v.).

However, Yidgha and Munǰī do show traces of an ancient development of **i* to **y-* in initial position (Morgenstierne 1938: 95), including after the loss of pre-vocalic initial **h*, as in Munǰī *yumar* 'to count' < **hišmar-* (vs. Yidgha *imar-*); *yimsāl* 'this year' < **ima-sarda-* (perhaps a loanword from Persian?), cf. Morgenstierne (1938: 95).

The Yidgha word *Idγ* 'Yidgha' itself shows a secondary **yV-* > *i-* (and perhaps *imar-* as well), cf. the name of the language in both Yidgha and Munǰī, *Yedγă* (cf. Morgenstierne 1930: 443), itself from **yenduka-*. The Yidghas are called *Inǰigān* in both Khowar and Persian (cf. Morgenstierne 1930: 443).

In this sense, I consider the Yidgha-Munǰī group closer to this evolution of Old Iranian. Possibly this evolution can be connected to the constant first-syllable stress of this Old Iranian language. If **i* had a tendency to be reduced to schwa, and the first-syllable was accented, then initial *i-* would probably change to another sound, but possibly not to initial schwa. We can speculate that word-initial schwa was not allowed in Old Steppe Iranian, as in many languages, including Tocharian. We would, in a parallel fashion, expect **u̯a-* or **u̯e-* for initial **u-*. One can think, for instance, of TB *wertsiya* 'community, assembly', if ultimately going back to the zero-grade form **u̯rdᶻi̯i̯ā-* (*-i̯ya* might also have been added in Tocharian itself to a base **u̯r̥dᶻ-*). I have also suggested that the *we-* in TB *weretem** 'promise, contract' might possibly go back to an initial **u-* or **u̯-* (see p. 70 f.).

A Proto-Iranian **i* has been reduced to naught in the two following cases: TB *perpente* 'burden, load' from **pari-banda-* 'attached around' (but with a much

more specific meaning attested already in Iranian), and, if my etymology is accepted, in *eprete* 'resolute, steadfast' from **abi-ratu-* 'according to the resolution'.

It has become *yod* in *ñyātse* 'danger' from **ni-ādᶻah-* 'distress', *-māññe* 'pavilion' from **mānii̯a-* 'of the house'.

It may have been reduced to schwa, as shown by **abi-st(H)-i̯a-* → TB *epastye*, if that derivation is accepted.

It has possibly been maintained as **i* only in front of *yod*. However, since the sequence **-əyV-* would also yield *-i̯y-* in Tocharian (or get reduced to *-y-*, s.v. *epastye*), these examples are not probative. The examples are: TB *epi̯yac* (TA *opyāc*) 'memory' ← **abi-i̯āt* ... 'idem', *mañi̯ye* 'servant' (← **mānii̯a-* '(servant) of the house') *newi̯ya* 'canal' ← **nau̯ii̯ā-* 'idem', and TB *wertsi̯ya* 'community, assembly' ← **u̯rdᶻii̯ā-*. The word *ekṣineke** is no counterexample, as the *-i-* in it goes back to **-ii̯-* (cf. section 2.7.2.8), which was simplified either in Old Steppe Iranian itself, or in (Proto-)Tocharian.

There are no traces of word-final **i* in our loanwords.

The TB word *tseriteke*, whatever its meaning (I argue, speculatively, for 'young grass', 'green animal' or 'turtle shell', s.v.) does have an Iranian appearance. It would correspond, if the etymology I developed together with Federico Dragoni is correct, to Ossetic *zældæ* 'young turf', with *-ka-* suffix. I have no explanation for the preservation of the **i* here. As a rule, we must assume that Old Steppe Iranian **-i-* has become schwa in Proto-Tocharian.

An early **u* has been reduced to naught in word-final position: TB *peret*, TA *parat* 'axe' from Old Iranian **paratu-* 'idem' and TB *tsain* 'arrow' from Old Steppe Iranian **dᶻainu-* (the **-u-* is still visible in the plural TB *tsainwa*, see below). Possibly it is also lost in *mot* 'wine', if from Old Steppe Iranian **madu* 'mead, wine' and if the original **-u* ending was not lost in the umlaut process. Compare further perhaps TB *eprete* 'steadfast' (< **epret* + adjectival *-e*) from Old Iranian **abi-ratu* 'according to the resolution' (> 'steadfast'). For the possible preservation of the final **-u* when the words were borrowed in Tocharian, see below.

The phoneme **u* has been maintained in contact with consonantal clusters in Tocharian B *yentuke* 'India' (← OSIr. **i̯anduka-* or **i̯enduka-*) and perhaps in TA *kump* 'bowl' (from OIr. **xumba-*), if not from Sanskrit or Middle Indic. This preservation of **-u-* is unexpected. In the case of *wertsi̯ya* 'community, assembly', it is conceivable that the initial *we-* derives from **u-*. In that case initial **u-* had become **u̯e-* or **u̯a-* in that Old Iranian language, **urʃii̯ā-* or **u̯rʃii̯ā-* > **u̯ardᶻii̯a-*.

We do not know whether the reduction of **i* and **u* to schwa or the loss thereof is Tocharian-internal or already Old Iranian, but there are strong ele-

ments in favor of the loss having occurred within Tocharian. One can depart from the plural of *tsain* 'weapon' (← **dᶻainu-*), which is *tsainwa* and seems to indicate that Tocharians borrowed **tsainu*, made a plural **tsainwa*, morphologically originally |tsainu-a|, and only then lost the **-u* (cf. Peyrot 2018a: 270).[142] One can also add the adjective *tsainwaṣṣe* as a further argument. A similar argument based on the verb **nətswa-* 'to starve' can be adduced: indeed, this verb implies a Proto-Tocharian noun **netsu-* 'corpse'. In both cases the preservation of the **-u-* is due to a following vowel. These elements would suggest that the loss of *-u* occurred within Proto-Tocharian. However, this does not preclude that the reduction of short **i* and **u*, at least in other positions, is an Iranian phenomenon.

It is an important fact to note that, if Dragoni (2023: 223ff.) is right in suggesting that the final *u*-ending of Pre-Khotanese accusative (**-əm* < **-am*) is the source of the *-o* stem of the numerous Khotanese, Pre-Khotanese etc. borrowings in Tocharian, it is remarkable that Tocharian did not make a single Old Steppe Iranian *u*-stem into an *o*-stem. Tocharian *o*-stems, in the system I described in this chapter, are both rare and originally *a*-stems (OSIr. *ā*-stems) made into more common *o*-stem types, although this is disputable for *maiyyo*, *maiyya*, see discussion with references to Del Tomba 2023, however it still was an *ā*-stem in Old Steppe Iranian, and not an *u*-stem.

It would have been possible, theoretically, for Proto-Tocharian speakers to borrow a noun **pereto* instead of **peret(u)*. Why this did not happen, while it happened later for Khotanese and Pre-Khotanese loanwords is a mystery to me. One could imagine that this is due to the Old Steppe Iranian reduction of **-u*; however, however, how can one then explain the preserved **-u* in *tsainwa* (cf. fn. 142 below)?

As I wrote above, the fact that in Old Steppe Iranian accented word-initial short **i* had turned into something else (s.v. TB *yentuke*, TB *yetse* TA *yats*) is possibly connected to a general reduction of **i*. Reductions of **i* and **u* are not observed in the generally phonologically more conservative Old Persian and Balochi groups.[143] Neither does Ossetic reduce short **i* and **u*: on

142 To explain the plural *tsainwa*, one has to take into account that, *a*-plurals being rare in Tocharian, it is also conceivable that Tocharian speakers chose to make a *-wa* plural, which is more productive, from *tsain*. This would nonetheless not explain the adjective *tsainwaṣṣe*.

143 This is, of course, a generalization. There are positions where **i* and **u* were reduced, of old, in Balochi and also in languages related to Old Persian. This is nevertheless not a general tendency, but only relates to specific sound laws.

the contrary, they merged with *$\bar{i}$ and *$\bar{u}$ (Cheung 2002: 15–16).[144] Reductions of **i* and **u* are, however, common to multiple languages spoken to the East.

In Sogdian, for instance, PIr. **i*, **u* and **a*, when they are unaccented, have been reduced to schwa (Yoshida 2009: 285; Novák 2013: 124), namely Gershevitch's "indistinct vowel" (Gershevitch 1961: 15–16). In Khwarezmian as well, there is a "reduction of unstressed short vowels" (MacKenzie 1990: 94). This phenomenon is also common to Yidgha and Munǰī for **i* and **u* (cf. Morgenstierne 1938: 95–96), but this should be rather late, in particular since it does not apply in the same way to both languages. To some degree **i* and **u* were reduced in Bactrian as well, but not **a* (cf. Gholami 2014: 58 f.). This also occurred within historical Bactrian: archaic Bactrian νιϸαλμο /nišalmə/ 'seat' became later νοϸαλμο /nəšalmə/.

I believe that the reduction that took place in Pashto and, more particularly, the conditions for the preservation of **u* are of particular interest for our topic.

In Wanjī, an extinct Modern Iranian language of Tajikistan, of which we only have very scarce documentation, it seems that both short **i* and **u* merged into a single phoneme, which was reflected as either *i* or *ə*, but was probably /ə/. After this reduction, long *$\bar{i}$ and long *$\bar{u}$ became /i/ and /u/ (cf. Novák 2013: 36), as there was no longer a length distinction in that language. Theoretically it could fit with the Tocharian data to some extent, but there is no way to know when any of these changes occurred.

In contrast, in Pashto, *$\bar{i}$ and *$\bar{u}$ were shortened, and thus merged with **i* and **u* (Cheung 2011: 199). The "new" vowel **u* was lost in unstressed open syllable, as in *bar* 'top, summit' < **upára-* (in detail in Cheung 2011: 199, also for the following). It became **ə́* in stressed open syllables, for example in *γanə́m*, Wanetsi *γandə́m* 'wheat' < **gantúma-*. So did **u* < **-am*, as in *zə* 'I' (< **azám* < **ajám*).

However, in a closed syllable, **u* was preserved: *búr* 'having lost a son, childless' < **apúϑra-*, *gúta* f., Wanetsi *nəgút* 'finger' < **angúštā-*, etc. This situation can be somewhat compared to the Tocharian one, where TA *kump* 'bowl' ← **xumba-* (if from OSIr.) and *yentuke* 'India' ← **i̯anduka-* show preservation of **u* either in a closed syllable or after a closed syllable. The same rules also apply to ancient **i* (< **i* and *$\bar{i}$), see Cheung (2011: 199 f.).

In view of this, one could now say that **i* was preserved in front of *yod* (**i̯a-* preserved as a group), for instance in TB *mañiye* 'servant', but the fact remains that schwa in Tocharian automatically became *i* in front of *yod*, which means

144 In Ossetic, **i̯a-* becomes **i-* (cf. Cheung 2002: 18).

we cannot be sure of the preservation of Old Steppe Iranian **i* in front of *yod*. As to the Tocharian B word *epastye* < **abí-st*(*H*)-*i̯a*-, it shows a reduced **i* (of **abi*-) in a closed syllable. This **i* carried the stress, and it is thus normal that it did not disappear giving a form †*epstye*, or, if the reduction occurred in Old Steppe Iranian, ***abst*(*H*)*i̯a*-. As stated earlier, the preservation of the *-i-* in *tseriteke* is definitely problematic. Due to syncope rules, one might expect rather †*tseritke* where the *-i-* would be preserved in a closed syllable, perhaps, but this is in any case not what we have.[145]

It should be underlined that, although **a* was reduced in Sogdian and possibly in Khwarezmian, in unstressed positions, neither in Yidgha-Munǰī nor in Pashto was the treatment of **a* identical to that of **i* and **u*.

It seems likely that the general tendency in Iranian languages spoken to the East to reduce their vowels, at least **i* and **u*, was an areal phenomenon. Possibly, more specific changes occurred in parts of that area, and possibly, these changes occurred at different points of time.

I would like to suggest here that the reduction of **i* and **u* which we find in loanwords in Tocharian, even in later ones, was also due to the influence of this areal phenomenon. For the discussion of the features of the Old Iranian language, it is relevant to note that we cannot know for sure whether every point discussed here pertains to Old Iranian or to Tocharian. It could be a Tocharian development, although *yentuke* is definitely a problem. However, it does seem to me that the Old Iranian language itself underwent some reductions of **i* and **u* in some contexts, including when stressed, as seen in the shift of initial **i* and **u* to **i̯V*- and **u̯V*-.

2.7.2.7 The Old Steppe Iranian Accent as Reflected in Tocharian Loanwords

I have suggested at multiple points throughout the volume that Old Steppe Iranian had fixed first-syllable stress, with morphological exceptions. It is now the place to examine the evidence in its totality. The stress can be seen in the following secure loanwords:

145 An *ad hoc* solution to account for the preservation of **i* and **u* in *tseriteke* and *yentuke* is by supposing that the words where pronounced with voiced intervocalic **d* (**dzaridaga*) and **g* (**i̯anduga*). Then, we should suppose that the *-u-* and *-i-* in contact with a **d* or **g* were lengthened, as they were at a very late stage of Old Persian or a very early stage of Middle Persian, as per Korn's law (Korn 2009). Since an ancient **rd* had become **ld* in Old Steppe Iranian, it is probable that **rt* had then become **rd*, implying that voicing of ancient unvoiced stops in this language may possibly have occurred. Since there is no other reason to assume this solution apart from the fact that it occurred in Middle Persian, I view it as a very doubtful one.

Tocharian B: *śāte*, *ākteke*, *ainake*, *epiyac*, *pāke*, *ñyātse*, *mañiye*, *-māññe*

The stress in *epiyac* and *mañiye* can be seen from the fact that the *-i-* was preserved, and the second syllable must therefore have been stressed. As far as *epiyac* is concerned, this fits the facts because **abí-* probably carried a fixed accent on the *-i-*, as it did in Indic. Unstressed **-iy̯a-* secondarily becomes *-ya-* in Tocharian, as in *epastye* and *-māññe* (for more on this, see below).

Due to syncope rules, where in a series of three identical vowels a non-accented vowel was reduced, creating a consonant cluster (cf. Winter 1994), the stress can be deduced in TB: *ākteke* ← **ā́gataka-*, *retke* ← **rátaka-*, *speltke* ← **spáldaka-*, *welke* ←**u̯áldaka-* and TA: *nātäk* < **natke* ← **nā́taka-* or **nā́ϑaka-* (although in the latter instance the first vowel was long).[146] One can also add TB *leṅke* ← **lánaka-*, where the original structure can be reconstructed from the Ossetic (Iron) cognate, which needs to come from an etymon **lanVka-*.

Where the stress can be deduced from syncope rules, it is always on the first syllable. In the word *kenek* 'cotton cloth', which needs to be a BMAC word because there is no suitable Iranian etymon, the stress was probably on the middle syllable, thus **kenéke* > *kenek*, with apocope rather than syncope, due to the middle syllable stress. If the word had been of Old Steppe Iranian origin, we would have had **kéneke* > TB †*keṅke*. Where the stress can be seen, it is constantly on the first syllable, with two exceptions: 1. it is on the second syllable of the prefix **abi-*, 2. it is fixed on the *-i-* or the *-ə-* in the adjectival suffix **-iy̯a-* (known from comparative Indo-Iranian data to have been accented **-íy̯a-*). In *-māññe* 'pavilion', it shifted from the **-íy̯a-* suffix to the preceding syllable (showing thus that unaccented *ñíye* > *ññe* within Tocharian). This shift is not surprising, since *-māññe*, the attested spelling, thus /máññe/, is a second member of compounds, and it is the last syllable of the first member of compounds that carries the accent in Tocharian B. As to 1., the prefix **abi-* was originally stressed on the second syllable, which fits with the rest of the Indo-Iranian data, for instance with Vedic *abhí*, and explains why we have TB *epiyac* and TB *epastye* (and not †*epyac* and †*epstye* or rather †*epästiye*).

Another word, of a more complex formation, also seems to have had initial stress, it is Tocharian B *ecce* 'hither, from', cf. TA *aci* 'id.', both come from PT **ecye* which I proposed to view as a borrowing from Old Steppe Iranian **(*h*)ačaiy̯a-* (§ 2.4.1), a form itself derived from the Proto-Iranian adverb **hačā* 'from', cf. Avestan *hacaṯ* 'from' in the ablative, Old Persian *hačā* 'from', Middle and New

146 Tocharian B *perpente* 'burden, load' does not show the expected syncope: this may be because a regular reduction would have led to a form **perpnte*, with a cluster *-rpnt-* which was certainly too difficult to pronounce.

Persian *az* 'from', etc. Were **(h)ačaḍa-* accented on the second syllable, the second vowel would have remained, and we would have expected another vowel to have dropped, perhaps even the initial one, although this is unprecedented. It seems likely that this word as well was thus accented on the first syllable, thus **(h)áčaḍa-*.

We thus have the portrait of a language with fixed lexical stress, always on the first syllable, but where the stress could shift due to already stressed suffixes and prefixes. Resembling this portrait is Proto-Ossetic, which also had fixed initial stress (Cheung 2002: 123), as discussed below.

Possibly, **i-* in *iścem* 'clay' (< OSIr. **ištiám*) did not become **ḍa* (or **ḍe*) as expected specifically because it was not stressed, as it was originally a neuter noun ending in *-ám* (or by influence of *iṣcäke*, an earlier borrowing with a similar meaning). In any case, the reflex of stressed initial **i-* can be seen in *yentuke* (< **(h)índuka-*) and *yetse* (< **ídᶻa-*), see section 2.7.2.4. This hypothesis has not been tested to the full yet due to the scarcity of data, and it remains to be confirmed or disproved.

Sogdian stress is not archaic, and is mostly the result of inner-Sogdian innovation (the so-called "rhythmic law"), which was not even shared by its closest sister-language, Yaghnobi (cf. Novák 2013: 80). It was initially free, and had to be learned for each word or word-class, but then a new system emerged, where stress was fixed on the first long vowel or diphthong, and otherwise on the last syllable of the word (Sims-Williams 1989: 181 f.; Novák 2013: 77 f.).

Pashto, Old Persian and Balochi all have their own stress rules. In general, Pashto accentuation is conservative, but it also has rules concerning heavy and light syllables (Cheung 2010).

As Dr. Niels Schoubben informs me, Bactrian stress seems to have been on the first heavy syllable from the last one, as can be seen from syncope rules. Fixed initial accent is excluded for Bactrian: cf. σαζαμανο 'as soon as' < **čiḍátǰamā́na-* (etymologically 'whenever time'). However, Bactrian had initial accent in light-syllable words. In the prefix **abi-*, the accent was even on the first syllable, unlike in Old Steppe Iranian: αβδδινο 'custom' < **ábi-daḍana-*. Note however αβιβταδο 'master' < **abi-štāta-* (cognate of TB *epastye* § 2.3.3), where the accent could be on the second syllable as well, because no syncope took place (although it is also conceivable that it was accented on the first syllable, but syncope did not occur because a cluster **-ββτ-* would have been too complex).

Morgenstierne (1938: 103) has shown that a great number of Yidgha-Munǰī words go back to preforms with initial stress, similar to Old Steppe Iranian. However, he posits a different position of the stress for some words, like *dram* 'inside' < **antaráhmi*, and the situation remains to be clarified.

It thus seems that only Proto-Ossetic has been reconstructed with fixed initial accent. A reason for this fixed initial stress could perhaps be the Uralic tribes to the north, which also had initial stress, or some other substratum/superstratum. Possibly, initial stress is a marked feature of "Steppe Iranian".

My conclusions concerning fixed word-initial accent with exceptions due to prefixes or suffixes seem to contradict Kuryłowicz's conclusions that the accent was fixed on the penultimate syllable in all of Old Iranian (cf. Kuryłowicz 1964; 1975). He was basing himself on Old Persian and Avestan spellings. However, his conclusions were not accepted, as it has been shown that, for instance, Avestan preserved an old stress system (cf. Beekes 1988). One of his main arguments is that the phonological opposition between *a* and *ā* has been neutralized in Old Persian and in Avestan (cf. 1975: 500 f.). With Old Steppe Iranian, we have now the example of a third Old Iranian language, and this language clearly preserves the opposition *a*: *ā* of Proto-Iranian (e.g., *newiya* vs. *perne*, *perpente*), but does not have a fixed stress on the penultimate syllable.

2.7.2.8 Old Steppe Iranian and the Name of the 'Dove'

Since we know the stress to have been on the first syllable, we can propose that unaccented **aḭ* became **ī* in front of **n* in order to explain TB *ekṣinek** (possibly *ekṣineke**) 'dove' from OIr. **axšaḭnaka-* (a *-ka-* derivative of PIr. **axšaḭna-* 'dark blue'). I posit the additional condition of unaccented **aḭ* because we also have *ainake* 'evil' (and not †*inake*) and *tsain* 'arrow', which have both preserved **aḭ* before **n*, possibly due to the fact that **aḭ* was accented in these words. In Ossetic, **aḭ* became **ī* in front of **n* (through **ē*?) before **aḭ* became **ē* in other positions (Cheung 2002: 17). A notable example is *æxsin* (Digor) 'dark blue', a cognate of **axšaḭnaka-* 'dove'. Nevertheless, in Ossetic this rule was not regulated by the position of the stress. Although it is hard to reject the possibility of unaccented **-aḭ-* becoming **-i-* in front of **-n-* in Old Steppe Iranian, it is also hard to accept it on the basis of one single example, and I thus indicate it with much caution.

A different solution to the problem can be suggested. It differs from the posited sound change of **aḭ* to **ī* in front of **n* mentioned above. Kim (2003: 51[20]) proposed a sound law Pre-Proto-Ossetic **-aḭnḭa-* > Proto-Ossetic **ijna*. The examples he presents for it are: **abi-šaḭϑnḭā* > POss. **æfsijnæ* > Digoron *æfsijnæ*, Iron *æfsin* 'landlady', **xšaḭϑnḭā-* > POss. **(æ)xsijnæ* > Digoron *æxsijnæ*, Iron *æxsin* 'lady, mistress', **axšaḭnḭaka-* > POss. **æxsijnæg* > Digoron *æxsijnæg*, Iron *æxsinæg* 'wild dove, pigeon' and **kaḭnḭā-* > POss. **kijnæ* > Digoron *kijnæg* 'reward, revenge'. I recently discussed this sound law, which I have named Kim's law, in Bernard (2022a: 32 f.).

Kim's law is a very interesting phenomenon, especially since it could also explain the Tocharian B form, if the sound change had occurred in Old Steppe Iranian as well. From the Ossetic point of view this sound law would explain why a so-called long *ī*, contrasting with *i*, occurs in Ossetic only in front of **n* (cf. Kim 2003: 51[20]). Furthermore, as Kim (op. cit.) points out, Digoron [ī] is phonemically /ij/, which supports Kim's suggested sound law.[147]

We are here facing a sound law that operated in both Ossetic and Old Steppe Iranian, namely **aiṇiṇV* > **iiṇV* (cf. Bernard 2022a: 34). Technically, one can suppose that first a vocalic palatalization took place: **aiṇiṇV* > **iniṇV*, followed by a metathesis, **iniṇV* > **iiṇV* or, differently, **aiṇiṇV* > **iiniṇV* > **iiṇV*. Another possibility, which I suggested in Bernard (2022a: 34) is that "there was first palatalization of **n*, subsequent depalatalization, and palatalization of the preceding vowel", by which I meant a process such as **ayny* > *añy* > **iñ* > **iyn*. The fact that *i*-epenthesis in Ossetic is also found after **n* but not, say, after **m* (cf. Cheung 2002: 107) could be an argument in favor of a connection of these phenomena to a palatalization of the *n*. No matter the phonetic detail of the intermediary phases, it would be uneconomical to assume two identical sound laws in both Ossetic and Old Steppe Iranian, given how uncommon this sound law is. Thus, it is an extremely important piece of evidence for the prehistory of these languages, as we have to assume that this sound law happened in the ancestor language of those two languages (since Old Steppe Iranian is not the ancestor of Ossetic).

This sound law, namely Pre-Proto-Ossetic **-aiṇia-* > Proto-Ossetic **iyna*, should not be confused (despite what I mistakenly claimed concerning *-œrijnœ* 'ell' in Bernard 2022a: 33) with the *i*-epenthesis after **niṇ* of Ossetic (Cheung 2002: 103—not mentioned in Bernard 2022a), although I admit historically situating the *i*-epenthesis of Ossetic is not an easy question, and both the epenthesis and the sound law discussed above yield similar (but not identical) results. The example of Digoron *mojnœ* (Iron *moj*) 'husband' < **māniia-* (Cheung 2002: 205–206) implies that the sound change **ān* > **ōn* is older than the epenthesis. In view of Old Steppe Iranian **māníia-*, and not **mīn(íi)a- vel sim.*, one can

147 In Bernard (2019), I proposed that New Persian *yak* 'one' went back to Middle Persian *yak* < **īyak*, itself ultimately from **aiiaiaka-*, analogical form based on either **θraiah-* 'three', or on **diaia-* 'double' with the original meaning of **aiiaiaka-* being **'unique', cf. Balochi *ēyōk* 'alone' < **aiiaia-* + suffix *-ōk*, cf. Bernard (2019: 52). To explain the initial **i-* of MP **iyak*, I had some difficulty, and resorted to a dissimilation of the two **-ai-*, as suggested to me by Johnny Cheung, which I still believe is possible. However, if the Ossetic sound law proposed by Kim is correct, we could see here a parallel sound law operating: **aiiaiaka-* > **aiiiaka-* > **īyak*. This idea is further developed in Bernard (2022a).

safely assume that no such epenthesis had taken place in Old Steppe Iranian at the time it was in contact with Tocharian. I would thus rather take the *i*-epenthesis before **n* as a secondary step in Ossetic, although it is perhaps part of the same larger phenomenon.[148]

The **aⱬiⱬV* > **iⱬnV* sound law provides information of various sorts: it must have occurred in the ancestor language of Old Steppe Iranian and Ossetic; before the monophthongization of Proto-Iranian diphthongs (since Old Steppe Iranian had undergone this change before monophthongization); and possibly, I assume, before the shift of both languages to word-initial stress. Indeed, this sound law should rather have operated on a stressed syllable, as is more likely typologically, thus **axšáⱬiⱬaka-* > **áxšiⱬnaka* (→ PT **ekṣineke*). Since, as far as I could find, no other Iranian language shares this exact sound law, we can safely assume that Old Steppe Iranian and Ossetic were closer to each other than to any other (documented) Iranian language, as other data also suggests (see table 2.7.2.13 for a summary).

2.7.2.9 Shortening of **ā* Before **-u̯ā-* and **-(i̯)i̯ā-* in Old Steppe Iranian
Another notable innovation of Old Steppe Iranian seems to be the shortening of **ā* before **-u̯-* and **-i̯-*. There are three examples: TB *newiya* 'canal', TB *maiyya*, Archaic TB *meyyā* 'power, might, strength', and TB *ecce* 'hither, from'.

Tocharian B *maiyya* has been borrowed from OSIr. **maii̯ā-*, ultimately going back to Proto-Indo-Iranian **māi̯ā-* 'magic power', a BMAC loanword (cf. Lubotsky 2001: 314). Originally the word was perhaps accented on the last syllable, as in Vedic *māyā́-*.

PK AS 17B a2
cwī maiyyane klye(ma)ne a(kā)lkäntaṃts ta(r)n(ene)
"in his power standing on top of the wishes" (translation from CEToM).

Naturally, a Tocharian B word cannot be accented on the final syllable (see p. 9), so the accentuation of **maii̯ā-* on the first syllable is mostly based on the other Old Steppe Iranian examples showing first syllable accentuation in

148 What could have happened in Old Steppe Iranian, though, is that the sound law **aⱬiⱬV* > **iⱬnV* could also have concerned the **au̯* diphthong, thus **au̯ⱬiⱬV-* > **uⱬi̯n(i̯)V-* as in Ossetic Iron *sin*, Digoron *sujnæ* 'hip' < **sraⱬiⱬā-* (cf. Cheung 2002: 104). This would only apply if **au̯ⱬiⱬV-* > **uⱬi̯n(i̯)V-* is part of the same sound law as **aⱬiⱬV* > **iⱬnV*, thus concerning both diphthongs. As we still lack any Old Steppe Iranian examples, it is impossible to prove and I only mention it for completeness' sake.

unsuffixed, unprefixed words. The Archaic TB form *meyyā* confirms the antiquity of the (Old Steppe Iranian) shortening of long **ā* in this word.[149]

Tocharian B *ecce* 'hither, from', if my etymology thereof is correct, also displays such a vowel shortening. It is suggested (§ 2.4.1) that its source form, Old Steppe Iranian **(h)ačaia-* derives from PIr. **hačā-ia-*, with shortening in front of *yod*. However, this is a disputably less secure example than the two mentioned above.

Avestan might show a similar, if not identical sound law: cf. for the two first words Avestan *nauuāza-* 'Schiffer', Young Avestan *maiiā-* 'witchcraft power' (vs. OAv. *māiiā-*). In Avestan this is the effect of the sound change **-āu̯ă-* > **-au̯ă-* and **-āiă-* > **-aiă-*, cf. Hoffmann & Forssman (2004: 58). A similar sound law also operated in Sogdian and in Khotanese (cf. Sogdian *nw"z* 'sailor', but also *n'wzy* 'id.'); cf. for example Gershevitch (1961: 17). It is to be noted that de Vaan (2003: 118 ff.) rejects this sound law for Avestan, and rather considers case by case explanations for words which show it, such as *nauuāza-* which he explains as a dissimilation **ā_ā* > *a_ā* (2003: 124).

Szemerényi (1951: 159) proposed that this Avestan sound law is correlated to stress, that is, the shortening only applied when the syllable was in pretonic position. I am not able to judge the pertinence of this observation for Avestan (strongly doubted by de Vaan 2003: 118), but this would work if the Proto-Iranian stress was the same as in Sanskrit (cf. Vedic *mayá-*), and remained there long enough for the shortening to occur for the word.

There is no Ossetic example or counter-example of the sound change **āu̯ă-* > **-au̯ă-* (see Cheung 2002: 17), but **-āi-* > **-ai-* is recorded (cf. Cheung 2002: 15). It is difficult to know whether either happened in Yidgha-Munǰī, as many shortenings of **ā* occurred in Yidgha-Munǰī, and the historical phonology of this branch is not sufficiently known. It did not occur in Old Persian, nor did it in Balochi, and as to Pashto, **ā* is only preserved as such in unstressed closed syllables (Cheung 2011: 178), so that it there is no way to know if this law operated here.

Since the sound law as known from Avestan, Sogdian and Khotanese is **-āu̯ă-* to *-au̯ă-*, and here we have **-āu̯íi-* > **-au̯íi-*, we have to assume that 1. the sound law applied to **nāu̯ā-* > **nau̯ā-*, only after which we had **nau̯íia-* (analogically of **nau̯ā-* and possibly before the fixed initial stress pattern was

149 An important point to note is that the *-aiyy-* sequence in Classical and Late Tocharian corresponds to *-ey-* in Archaic Tocharian B, the only other constant example thereof being Archaic TB *eynake* 'evil, bad', corresponding to Classical TB *ainake*. This suggests that the Old Steppe Iranian diphthong **ai* was not borrowed as a diphthong **ay* in Proto-Tocharian.

introduced), or 2. that the sound change was extended to or only concerned the sequence *-*āu̯úi̯*(*a*)-. If the Old Steppe Iranian shortening is shared with other branches, then we can imagine that it was shared at an earlier level, and thus defines a greater branch. However, it is difficult to believe it was, as the detail of this shortening varies among the different languages mentioned.

In any case, this sound change or variants of it occurred among Iranian languages spoken to the east, and it certainly concerned Old Steppe Iranian. Since it is not a trivial change, it is more logical to suppose that it occurred as part of a continuum, or of an areal feature, rather than assuming parallel developments. In that case, it is easier to explain the fact that the sound law is not identical in all these languages.

2.7.2.10 Simplification of the Cluster *_dm-_ in Old Steppe Iranian

Another possible sound change, only seen in two words, is the apparent simplification of Proto-Iranian *_dm-_ to *_m-_ in TB _-māññe_ '(place) belonging to the house, garden, outbuilding'; TB _mañiye_ (from the same Old Steppe Iranian root, with different accentuation, see section 2.7.2.7). These words go back to the PIE root *_dem-_ 'house' (Latin _domus_, Greek δόμος), and can be reconstructed for Proto-Iranian as *_dmāna-_ (OAv. _dəmāna-_). The initial cluster *_dm-_ was reduced to _m-_ in several Iranian languages: Old Persian _māna-_ 'house' (MP, NP _mān_ 'house, mansion');[150] Parthian _mān_ 'house, mansion'. It was changed to _nm-_ in Young Avestan _nmāna_, from Old Avestan _dəmāna-_ (/dmāna/).

There are three possibilities to explain the initial _m-_ in the Tocharian words. The most straightforward consists in suggesting a simplification *_dm-_ > _m-_ in the Old Steppe Iranian source, yielding *_mānii̯a-_. The second one consists in suggesting a sound change *_dm-_ > *_nmānii̯a-_ → PT *_nmānii̯a-_ which could have been simplified to *_māñiye_ in Pre-B. Both are theoretically possible, and unfortunately we do not have a Tocharian A cognate to be sure of it, but possibly the first hypothesis should be favoured, as it is much more economical. Furthermore, only Young Avestan is known to have undergone _dm-_ > _nm-_ in this word, to my knowledge.

150 It has been claimed since long that *_dm-_ changed to _b-_ (through *_db-_) in Middle Persian, hence _bānūg_ 'lady' < *_bān_ + _-ūg_; cf. also _bānbišn_ 'queen' < *_dmāna-paϑnī-_ (see Benveniste 1966: 29), see also Persian _bām_, _bān_, Pahlavi _bān_ 'Dach, Haus' (cf. Horn 1898–1901: 60). If this is correct, which is far from certain (see doubts expressed already in Hübschmann 1895: 25 and in Back 1978: 200) this must be in another dialect or variety than the one which simplified *_dm-_ > *_m-_ (as in Middle and New Persian _mān_ 'house').

The third possibility is a simplification of **dm-* to **m-* in Tocharian itself. However, this simplification seems completely unwarranted: initial clusters are not uncommon in Tocharian, and we have a word *tmāne* 'ten thousand' which could theoretically go back to an initial **tm-*, although of course its antiquity is not assured, and it can thus not be used as a certain example or counterexample.

Considering the possible simplification of **dm-* > **m-* in Old Steppe Iranian, it should not be compared to other languages which underwent the same simplification, as it is a rather trivial one, which could have occurred independently in many languages. For instance, it occurred in Ossetic and Old Persian as well as in Pashto (Morgenstierne 1927: 44) and others. Rather, we should compare the possible Old Steppe Iranian simplification with the languages that did not undergo it, to exclude common branching. Thus, Sogdian, which has *δm'(')n*, and Avestan are to be excluded here; Khwarezmian is ambiguous but seems to point towards a preservation of **dm-* as *δm-*;[151] Khotanese is also excluded, as it has *damänā-* 'house' (DKS: 152, written *damäna-*). To my knowledge the word is not attested in Yidgha-Munǰī; nor could I find it in Balochi.

In Bactrian I could not find a cognate of this word, however, at least word-internally, **-dm-* was preserved: archaic νιϸαλμο, ναϸαλμο, νιϸλμο 'seat (of the gods), throne, dwelling' < **ni-šadman-* (cf. Sims-Williams 2007: 240).

One can also exclude Zazaki, which has *bān*, certainly also from **dm-* (possibly through **db-*).

Whether TB *mañiye* derives from OSIr. **nmānίa-* or **mānίa-*, it is unlikely to derive from **dmānίa-*, which I believe would have yielded †*tmañiye* in Tocharian B. It is thus a relevant exclusionary isogloss, which leaves Khotanese, Sogdian, Avestan and Zazaki out of the possible languages that would share a direct (immediate) ancestor with Old Steppe Iranian.

2.7.2.11 Old Steppe Iranian Sound Change **rd* > **ld*

The study of Old Steppe Iranian has led me to conclude that, possibly apart from the palatalized context **rdʲi̯*, and **rdᶻ* (§ 2.7.2.1), the regular outcome

151 In Khwarezmian two forms with *-δm-* are attested, and in both cases it occurs word-internally: one example is *γrδm'n* 'Paradise'. However, here the word is very likely a religious borrowing (← Av. *garō-dəmāna-*) and it could also have been preserved in word-internal position only (cf. Middle Persian *garōdmān* 'Paradise', New Persian *garuϑmān bihišt 'garuthmān* Paradise'). The other form is (')*δm's-*/*δm'δẙk* 'dick werden' < **ā-dmasa-* (Benzing 1983: 28).

of Proto-Iranian *-*rd*- seems to have been Old Steppe Iranian *-*ld*-. This was already suggested by Tremblay (2005: 424) who explains it through a very contrived chain of changes: "**rd* becoming something like *ḍ*, *ẟ* or *ḷ*, whereas **rt* was retained or did not go further than **rṭ* or **rḍ*".

There is no need for such a complex scenario: Proto-Iranian **rd* in my view had rather directly become **ld* in Old Steppe Iranian. There are many arguments one could evoke against Tremblay's reasoning, but simply from the phonetic point of view *ḍ*, *ẟ* or *ḷ* to *ld* is very unlikely, although theoretically a substitution of *ḷ* by *ld* cannot be totally excluded. Instead, one can view the change from **rd* to **ld* as an assimilation of the place of articulation. If, for instance, **r* was an alveolar flap or trill, and **l* a dental liquid, then this sound change would suggest that the Old Steppe Iranian **d* was a dental plosive. Such a change is quite frequent cross-linguistically, including among Iranian languages, e.g. Middle and New Persian *l* < **rd*, with compensatory lengthening, as in *sāl* 'year' < PIr. **sarda*- 'idem'.

The Old Steppe Iranian sound change **rd* > **ld* is seen in the following words: TB *speltke* from OSIr. **spardaka*- or **spṛdaka*- 'zeal', TB *melte*, from OSIr. **mṛda*- 'top'; TB *welke* from OSIr. **u̯ardaka*- or **u̯ṛdaka*- 'stamen'.

The TB word *kercci*, possibly from OSIr. **gardi̯a*-, poses an obvious problem here, because it can imply that **rd* > **ld* was only a phonetic, not phonologized development, that is, */rd/ had an allophone *[ld] in some contexts. If TB *kercci* is to be removed from the list of Iranian etymologies, or if my explanation of its sequence *-rcc*- (s.v.) is to be accepted, then the sound change **rd* > **ld* of Old Steppe Iranian is without exception.

One should however note that TB *wertsiya*, TA *wartsi* 'assembly', which I derived from OSIr. **u̯ardᶻ(i̯i̯ā)*-, shows that this sound change did not occur in front of the phoneme **dᶻ*. This could either imply that the sound change was affecting */r/ + */d/ but not */r/ + */dᶻ/ and thus contradict my previous statement concerning *kercci*. A possible solution is **dᶻ* was not pronounced at the same point of articulation as **d*. For one, **z* cannot be pronounced at the dental point of articulation, and a simple pronunciation exercise will prove without difficulty that /dᶻ/ needs to be pronounced at least in the apex.[152] A third possibility is that the sound change occurred when **dᶻ* (and **tˢ*) were still the palatal

152 A dentalized laminal alveolar variant of the /z/ phoneme exists (noted as ‹z̪› in the IPA), however it is only found cross-linguistically in word-initial prevocalic or in intervocalic position. As Professor Lubotsky points out to me (p.c.), **-d*- might also be apical, as it is usually dental only when opposed to, for instance, retroflexes.

sounds **ǰ* and **ć* of Proto-Iranian when the sound law **rd* > **ld* took place. An obvious difficulty with this hypothesis is that it requires the sound change to have taken place very early, at a stage when (most) Iranian languages had not diverged from each other.

Here I wish to suggest another solution, which could enlighten us on the prehistory of both TB *wertsiya* and *kercci*. If **rd*—at least in pre-vocalic position—was actually realized as [rð], because **d* would have been systematically realized [ð] according to this hypothesis, then we could very easily understand why **rdᶻ* is not reflected as **rcc* in Tocharian: **dᶻ* was not realized as [ðz], which is a difficult cluster. Furthermore, one could surmise that **-di̯-* was not realized as [ðj] but either as [dj] or as [θj], in both cases **gardi̯a* would be palatalized to PT **kercye*. This would be either because **d* was only realized as **[ð] in pre-vocal position, or because of a sound change **[ðj] > **[θj]. There is meager evidence for the latter among Iranian languages, so this solution is rather speculative, but one can nonetheless cite punctual examples (that is, not occurring due to a specific sound law), such as Shughni *yēϑ* 'nest' < **āhadi̯a-*.[153]

The sound change **rd* > **ld*, sometimes accompanied by **rt* > **rd*, occurred in many Iranian languages, sometimes as part of shared innovations, and in other cases as part of an areal phenomenon.

Here again, only languages that did not partake in this innovation can be said to be not descending from Old Steppe Iranian. These are: Old Persian (however, Middle and New Persian share this innovation); Sogdian (*wrδ* 'rose', cf. Gershevitch 1961: 44); Khwarezmian (*wrδc* 'Blüten', cf. Benzing 1983: 651); Pashto (*zṛə* 'heart' < **ǰṛdai̯a-*, cf. Cheung 2011: 187); Ossetic (cf. Cheung 2002: 29); Parthian (*wār* 'flower', cf. DMMPP: 335).

The situation in Pashto and Ossetic is slightly more complex. Since it might help us shed some light on what happened in Old Steppe Iranian, I wish to enter it in more detail. In Ossetic, **rd* and **rt* merged as *rd* (cf. Cheung 2002: 29), like in Pashto where **rt* and **rd* also merged to **rd*, which later became *ṛ* (cf. Cheung 2013: 622–623). It seems slightly counter-intuitive to suppose that these two clusters simply merged, without any intermediary steps, although this is what seems to be the case at a first look.

For the sake of the argument, one can suppose that **rd* > **ld*, and **rt* > **rd*, both in Ossetic and Pashto: then, **ld* merges again with *rd* later in Ossetic, simply because the phoneme /l/ was rare in preconsonantal position, if not non-existent.[154] Once **rd* shifted to **ld*, a chain shift may have occurred, leading

153 I thank Professor Martin Kümmel for this example.

154 Another theoretical possibility is that the "new" *l* (< **ṛi̯*) was phonetically different from

to **rt* becoming **rd*. In Pashto it is possible that either the same phenomenon as the one proposed above for Ossetic took place, or that both **ḷ* (< **rd*) and *ṛ* (< **rt*) coexisted, before merging at a later stage. This situation would perhaps be similar to that of Yidgha-Munǰī, where **rt* > *ṛ*, but **rd* > **ḷ* > *l* (cf. Morgenstierne 1938: 81).

Thus, **rd* could have indeed become **ld* in some earlier stages of Pashto and Ossetic, before it merged with **rd* < **rt*. It should be stressed that this remains hypothetical, but could theoretically explain the merger of **rt* and **rd* in Ossetic.

It is of note that Bactrian ϱλ (< **rd*) reverted to ϱδ in later stages of the language. This is not identical to Pashto or Ossetic, of course, as it is the λ < **d* which changed, here, but it could be seen as a somewhat comparable sound shift.

Wanjī apparently underwent the sound change **rd* > *l*, possibly through **ld* as can perhaps be seen from *zil* 'heart' < **ǰ̌rdaḍa-* (cf. Lashkarbekov 2008: 83). However, in this language **d* (and **d* < **t*) becomes *l* in many positions, including word-finally: OIr. **pāda-* 'foot' > *pal* 'leg' and OIr. **gātu-* 'moment' > **gād* > *amγal* (< **ima-gātu-*), *γal* 'now', so one could also imagine that **zǰ̌d-* became **zid* and then, regularly, *zil*. See p. 215n20 for a possible example of **-ti-* > *-ϑ-* in Wanjī.

Of particular interest concerning the question of the fate of **rd* in Iranian languages is the fact that in the Persic group, for instance in Middle and New Persian, both **rd* from PIr. **rd* (< PIE **rd*, **rdʰ*, **ld*, **ldʰ*) and **rd* from PIr. **rǰ* (< PIE **rǵ(ʰ)*, **lǵ(ʰ)*) have become *l*. This would naturally only have happened after the sound law **dᶻ* > *d* that defines Proto-Persic. In the second case, we have OSIr. **rts* rather than **lts*: TB *wertsiya* 'community, assembly' ← OSIr. **u̯ardᶻiḍā* < PIr. **uǰ̌ḍā*. Changes of **rd* to **ld* or *l* are thus very old among Iranian languages, and have occurred at multiple stages, in multiple languages, including languages geographically separated from each other. The assumption of a continuum or areal phenomenon cannot justify every change from **rd* to **ld* among Iranian languages, and we have thus found the earliest example of that change in Old Steppe Iranian.

It is noteworthy that Khotanese seems to be the only known Middle Iranian language in the neighbourhood of Tocharian to have without a doubt undergone the sound change of **rd* > *l*. Wanjī, which is not in the immediate neighbourhood of Tocharian, could also have undergone this change, but it is not possible to determine this with certainty.

---

the **l* in **ld*, for instance more retroflex (as it was perhaps originally palatalized), and the existence of these two different *l*s "pushed back" the **l* in **ld* to *rd*.

2.7.2.12 The Old Steppe Iranian Treatment of Vocalic *$\mathring{r}$

To determine the reflex of *$\mathring{r}$ in Old Steppe Iranian, we need to turn to the Tocharian words TB *melte*, TA *malt** 'pile, summit', TB *spelkke* (< **spel̥tke*), TA *spaltäk* 'zeal', TB *welke* 'petal of a flower', TB *kercci* 'palace', TB *wertsiya*, TA *wartsi* 'community' (< Common Iranian **u̯r̥dzi̯ā-*).

All these words could represent either *$\mathring{r}$ or **ar*, since both are attested among Iranian languages, but in each case *$\mathring{r}$ is original. As examples I can cite **kam̥rda-* 'head', related to **mr̥da-* 'neck', where one has Bactrian καμιρδο 'head; chief-god' < **kam̥rda-*, but Persian also shows a reflex of **kamarda-* in *kamāl* 'face (vulgar, derogatory)'. One also finds Middle and New Persian *gul* 'rose, flower' < **u̯r̥da-*, as opposed to Avestan *varəδa-* 'rose' (AiW: 1369) < **u̯arda-*. The same **u̯arda-* was also borrowed in Arabic *ward* 'rose' (secondary collective from *wardah* reinterpreted as a singulative).[155]

Having considered this, it appears very unlikely to me that Old Steppe Iranian would have kept only *-ar-* variants for all four words cited above, since these reflect original zero-grades, so that *-ar-* is secondary in any case. This could be theoretically possible for one, two or maybe three, but the likelihood of that is very low. Thus, I consider **-ar-* to be the Old Steppe Iranian reflex of PIr. **-r̥-*, at least before **-d-* and **-dz-*. It is likely that **-ar-* was the phonological development of **-r̥-* in all four words.

In order to explain forms going back to **-ar-* instead of **-r̥-* in Iranian languages, it has been suggested by scholars that accented vocalic **-r̥-* yielded **-ar-*. For instance, Cantera (2017: 489) explains Pahlavi *kamāl* from **ka-mr̥da-*. The same principle could also be invoked here: all the examples we have at hand (*kercci*; *melte*; *spelkke*; *welke**; *wertsiya*) contain a vocalic **-r̥-* in the first syllable, which we have reasons to believe was always accented in Old Steppe Iranian, with only few exceptions, (see p. 168 f.).

Based on the evidence at hand, there are two possibilities that we can consider for Old Steppe Iranian: (1) either **-r̥-* became **-ar-* in every position, or (2) **-r̥-* only became **-ar-* under the effect of stress and had other reflexes or realizations in other positions. An argument in favor of the first hypothesis, I believe, is found in the fact that the cluster **rd*, even when going back to **ŕ̥*, had become **-ld-* in Old Steppe Iranian. This might show that the sequence was phonologically not /ŕ̥+d/ but /ar+d/, or in any case phonetically [ərd] and not [ərəd] or [rəd]. If this is correct, one can observe that such is not

155 Pace Eilers (1962: 207), who writes that "*ward* „Rose(n)": vgl. av. *varəδa-*; np. *gul* (und *vil* „Liebchen") gehen auf ein **wr̥da* zurück". Neither the Avestan nor the Arabic forms are likely to come from the zero-grade **u̯r̥da-*.

the case of Old Persian (see Korn 2021 with discussion of the literature); Bactrian (e.g. καμιρδο 'head; chief-god' < **kamr̥da-*); Pashto (cf. Cheung 2011: 187); Khotanese (cf. Emmerick 1989: 211 f.); Yidgha-Munǰī (cf. Morgenstierne 1938: 97 f.).[156]

Khwarezmian seems more ambiguous and more research is needed to determine the result of vocalic *r* in Khwarezmian. Sogdian and its sister language Yaghnobi also have a different reflex of **r̥* than Old Steppe Iranian: in both languages, **r̥* becomes variously *ər*, *ir*, ur, etc. depending on the context (cf. Novák 2013: 95; Gershevitch 1961: 19 f.).

Wanjī has three different reflexes of vocalic *r*: **r̥* either becomes *-i-* or *ir* as in **kr̥naṷ-* > *kin-* 'to do' (compare Persian *kun-*) and *kirmγaz* 'wormy', with *kirm-* 'worm' < **kr̥mi-* (cf. Lashkarbekov 2008: 67), and it also becomes *ai* in front of a nasal (cf. Novák 2013: 36). For the latter change, it is the same reflex as that of **ar* and **a*, **ā* in front of a nasal, as in Wanjī *main* 'apple' < **amarna-*, *γain* 'woman' < **gan*(*ā*)- < **gnā-*, *kain* 'source, spring' < **kāna-* (cf. Lashkarbekov 2008: 71). This common reflex for all four sounds or sequences is somehow disturbing, as it would suggest a merger of **r̥* and **ar* in front of a nasal (> **arN* > **aN*). This does not square well with the fact that **r̥* is otherwise reduced to *i* or to *ir* (possibly /ər/). Perhaps—but one has to recall this is a dead language of which we have no texts and only a limited amount of words—**r̥* became /ər/ in most positions and /ar/ in front of nasal consonants, or perhaps /ər/ became /ar/ in front of nasal consonants in Wanjī. It is also possible to imagine that the reflex of Proto-Iranian **r̥* was /a/ before nasal consonants in Pre-Wanjī, before it became *-i-* and *-ir-* in the other positions. This, however, would not explain why **ar* also became **a* in front of nasals, and it is thus a less preferred hypothesis, in my view. In any case, the reflexes of **r̥* as *-i-* and *-ir-* exclude Wanjī as a possible descendant of Old Steppe Iranian, as one has to reconstruct a vocalic *r̥* phoneme for Pre-Wanjī.

The only language of our list which has a consistent reflex of PIr. **-r̥-* as *-ar-* is Ossetic (cf. Cheung 2002: 24). The change of **-r̥-* to *-ar-* could thus be a common development of Pre-Proto-Ossetic and of the ancestor language of Old Steppe Iranian. However, this change might perhaps have occurred independently. Nevertheless, the very fact that it occurred so early in Old Steppe Iranian, when it did not in other Iranian languages spoken in Asia that we know of, is non-trivial.

156 Khotanese has multiple outcomes of vocalic *r* depending on the position, as listed in Emmerick (1989: 211–212).

As Korn (2016: 410–411) notes "as soon as **r̥* yielded *ir*, *ur* or *ar*, this sequence is indistinguishable from old sequences of vowel plus *r* […]. The development is thus a "point of no return" in dividing Middle Persian plus Parthian from other WIr. languages […]." The same is true about Old Steppe Iranian: PIr. **r̥* > OSIr. **ar* is indistinguishable from PIr. **ar* > OSIr. **ar*.

However, if Old Steppe Iranian underwent the same sound law as—possibly—, other languages, such as Middle and New Persian (according to Cantera and others), namely that accented *-*r̥*- had become *-*ar*-, the isogloss question becomes more obscure. It is unfortunate that we do not have an Old Steppe Iranian word with **r̥* in unaccented position, for instance a reflex of **ka-mr̥da-* (or another word with vocalic -*r̥*- in the second syllable), since it is the syllable **ka-* that would be the one carrying the accent. This would have been helpful in order to prove or disprove Cantera's suggestion (see above).

In conclusion, although it is clear that all examples going back to vocalic *-*r̥*- in Proto-Iranian are reflected with Old Steppe Iranian *-*ar*-, which suggests a regular development from PIr. *-*r̥*- to OSIr. *-*ar*-, all examples also go back to an *-*r̥*- that should have been accented in Old Steppe Iranian, making it difficult to determine if the sound change was triggered by the stress or not.

2.7.2.13 Table of Phonetic Innovations and Isoglosses

Below is a table of all the phonetic innovations of Old Steppe Iranian discussed here, compared to the other Iranian languages I have considered, in order to sum up and give an overview of the types of isoglosses found.

Based on this table, we can say that Old Steppe Iranian agrees with Ossetic at least nine times;[157] with Sogdian five times; with Yidgha-Munǰī also five times; with Old and Middle Persian four times; with Pashto also four times; with Avestan thrice; with Khwarezmian thrice; with Khotanese-Tumshuqese twice; with Bactrian once or twice; with Balochi once. Naturally this should be taken together with lexical shared innovations, and a number of other elements in order to be significant, so that one cannot say that Old Steppe Iranian is closer to Old Persian than to Avestan or Khwarezmian for instance. Nonetheless, the amount of times Old Steppe Iranian agrees with Ossetic, including for non-trivial innovations, and including for changes which no other listed Iranian language shares (such as the development of vocalic **r̥* to *ar*) is remarkable, and suggests a real genetic proximity between both languages, as has been suggested in the past (e.g. cautiously, Pinault 2002a: 245). Further, the connection

157 Since in Ossetic **i̯a-* becomes *i-* (Cheung 2002: 18), there is no way to know if word-initial **i-* had also become **i̯a-* in Pre-Proto-Ossetic or not.

TABLE 2 List of phonetic isoglosses of Old Steppe Iranian compared to a selected amount of Iranian languages

Phonetic innovations of OSIr.	Avestan	OP | MP	Sogdian	Khwar.	Khotanese-Tumshuqese	Bactrian	Yidgha-Munǰī	Pashto	Balochi	Ossetic
Palatalization of *$\vartheta_{\underline{i}}$ and *$d_{\underline{i}}$	No	No	No	No	No	No	No	No	No	Yes
*$a_{\underline{i}}n_{\underline{i}}a$- > *-$i_{\underline{i}}na$-	No	No	No	No	No	No	No	No	No	Yes
Preservation of *$\check{c}_{\underline{i}}$- or palatalization of *$\check{s}_{\underline{i}}$-	Yes (pal.)	Yes (preservation)	No	No	Yes (preservation)	?	?	?	No	Yes (preservation in post-consonantal pos.)
Contraction of *$a(h)a$ > *$\bar{a}$	No	No	Yes	Yes	No	No	Yes	Yes	No	Yes
Loss of *h	No	No	Yes	Yes	No	No	Yes	Yes	No	Yes
Possible loss of *h in front of *u	No	Yes|No	No	No	No	No	No	No	No	No
PIIr.*$ću$ reflected as *tsw	Yes	Possibly	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Initial *i > *$i_{\underline{u}}a$- or *$i_{\underline{u}}e$; *u > *$u_{\underline{u}}a$- or *$u_{\underline{u}}e$-	No	No	Occasional palatalization of *i-	No	No	Tendency to have *i- > *ya, yi*.	No	No	No	Unclear
Reduction of short *i and *u in word-internal position	No	No	Yes	Unclear	No	Yes	Yes	Yes	No	No
Fixed word-initial accent	No	No	No	No	No	No	Tendency	No	No	Yes
Shortening $\bar{a}_{\underline{u}}V$- > *-$a_{\underline{u}}V$- and *-$\bar{a}_{\underline{i}}\breve{a}$- > *-$a_{\underline{i}}\breve{a}$-	Yes	No	Yes	No	Yes	Unclear	Unclear	?	No	Yes
Simplification *dm- > *m-	No	Yes	No	Unlikely	No	Unlikely	Unclear	Unclear	Unknown	Yes
Change from *rd to *ld	No	No|Yes	No	No	No	No	Yes	No (not clearly)	No	No (not clearly)
Vocalic *$\underline{r}$ > ar	No	No	No	Unclear	No	No	No	No	No	Yes

between the Tocharian B word *leṅke* (s.v.), corresponding to Ossetic Iron *lænk* Digoron *læncæ*, if true, should count as another argument towards genetic proximity between Old Steppe Iranian and Ossetic.

2.7.3 *Lexical Isoglosses of Old Steppe Iranian*

To have a complete overview of the features of Old Steppe Iranian, one must not stop at the study of its phonological innovations and archaisms, but it is also necessary to look at the lexical isoglosses it shares with other Iranian languages and the semantic shifts, developments and archaisms of this language. I should like to underline a few reasons which might explain the formation of both the lexical and the semantic isoglosses.

First, most languages surrounding Old Steppe Iranian might have disappeared without leaving any trace. Indeed, Old Steppe Iranian itself would have left no trace if it had not influenced Tocharian. This could explain why no closely related variants are known. Second, Persic is by far the best attested Iranian subgroup, since besides Old Persian, we have Middle and New Persian, which have enormous corpora, and have been lexically studied both by native and foreign lexicographers since quite a long time. Other Persic languages such as Kumzari and Bakhtiari also provide us useful information whenever data is lacking from the three languages mentioned above. It is thus not entirely a coincidence that this group is where many of our isoglosses are found.

An ancient, prehistorical connection between Old Steppe Iranian speakers and Proto-Persic (the ancestor language of Old Persian and all other Persic languages, such as New Persian and Bakhtiari, for instance) could be imagined, by means of the Caspian sea. If one goes to the North East of the Caspian, one encounters the endless steppes that lead, ultimately, to South Siberia and Xinjiang. This hypothesis is not very likely, however, notably because of the geographical distance between Old Persian and these very steppes.

Avestan, in its turn, was more probably spoken on the plains of modern-day Afghanistan, or in the region which is globally situated in the South of the steppes, and I would be more inclined to believe that the lexical isoglosses one sees here are due to chance, in the sense that there is a higher probability that an archaic rare Iranian word is reflected in Avestan, which is a well documented archaic Iranian language, rather than in another language that has undergone more lexical replacement. Another factor here is naturally the extensive study that has been done on Avestan vocabulary since at least two hundred and fifty years in the West (and much longer in the East).

In this section 2.7.3, I will analyze and discuss a number of isoglosses that include Old Steppe Iranian and a minority of Iranian languages. It will be an

overview, and for details I refer to each specific word in the individual discussion above (sections 2.2, 2.3 and 2.4).

2.7.3.1 Old Steppe Iranian–Persic

Old Steppe Iranian and Persic share at least two specific lexical isoglosses:

1. Old Steppe Iranian **rataka-* 'army' (TB *retke*, TA *rätäk* 'id.'), with its Persic correspondent **rataka-* 'row, (army) line', represented by Middle Persian *radag* 'row, (army) rank, line', New Persian 'line, row'. A cognate of this word that is formally farther away is found in Ossetic Iron *rad*, Digoron *radæ* 'row' < **rātā-* (cf. Abaev 1973: 338). The military semantics of this word must have arisen in Old Steppe Iranian, possibly due to a *pars pro toto* formula, since it is not particularly salient in the Middle Persian and New Persian words.
2. Old Steppe Iranian **nāϑaka-* / *nātaka-* 'protector, lord' (Tocharian A *nātäk* 'lord') has Indo-Aryan cognates but very few Iranian cognates. All those Iranian cognates are exclusively found in the Persic branch, namely New Persian *panāh* 'protection' (< **pad-nāh* < **pati-nāϑa-*) and Middle, New Persian *nihān* 'secret, hiding'. The root **nāϑa-* 'protection' as such seems to be exclusively shared by Old Steppe Iranian and Persic.

2.7.3.2 Old Steppe Iranian–Balochi–Persic

Old Steppe Iranian shares one isogloss together with Balochi and Persic.

Old Steppe Iranian **graṷpa-* or **graṷfa-* 'group' (TB *kraupe*, TA *krop* 'id.') which is found in a very limited number of languages: Balochi *grōp*, Middle Persian *grōh*, New Persian *gurōh* 'id.' Dawāni *gorufa* 'ball of fiber or rope' represents a derivative of this word, (with the **-ak* suffix) which has a parallel in New Persian *gurōha* 'globe, bowl, or any spherical figure; ball for a cannon; a kind of sweetmeat; an assembly, multitude of people'.

Dawāni belongs to the Persic group (cf. Mahamedi 1994). It thus seems that this word is only found in the Persic branch, Balochi and Old Steppe Iranian. It is possible that the Balochi word was itself borrowed from an unattested Old Persian **graṷpa-*, making this an isogloss between Old Steppe Iranian and Persic. The other direction of borrowing, namely, from Balochi to Old Persian, is unlikely.

2.7.3.3 Old Steppe Iranian–Avestan

Old Steppe Iranian shares at least two exclusive isoglosses with Avestan. These can naturally be due to chance, as Avestan preserves a number of ancient lexical items and verbs that have been lost in the rest of Iranian.

1. Old Steppe Iranian **raiϑṷa-* 'combination, mixture' (TB *raitwe* and TA *retwe* 'combination, mixture' *passim*), Avestan *raēϑβa-* 'to mix'.

2. Old Steppe Iranian **andza(h)* 'greed' (TB *entse*, TA *eṃts*), Avestan *ązah* 'constriction; distress, peril'. While many cognates of the root (PIr. **Hanǰ-*) are found in Iranian languages, I was unable to find any other exact formal cognates.

2.7.3.4 Old Steppe Iranian–Avestan–Persic–Parthian

Old Steppe Iranian **dᶻaᵢ̯nu-* 'arrow' (TB *tsain* 'id.'), Avestan *zaēnuš* 'baldric', Avestan *zaēna-* m. 'weapon'. Here we observe an interesting semantic evolution in Old Steppe Iranian, for more detail see section 2.7.4.5.

Cognates are also found in Parthian and Manichean Middle Persian *zēn* 'weapon, sword, armor, arms' (DMMPP: 386). The Parthian (or, perhaps, the Middle Persian) word was borrowed into Armenian *zēn* 'weapon', *-zēn* (as second member of compounds) 'weapon, armor'.

2.7.3.5 Old Steppe Iranian–Avestan–Persic–Sogdian

Further, Old Steppe Iranian shares one isogloss with Avestan, Sogdian and Persic.

Old Steppe Iranian **maiį̯ā-* 'power, strength', Avestan *maiiā-* 'strength, craft', Old Persian **māį̯ā-* 'craft, strength', Sogdian *my'kčyk* /mayākčīk/ 'happy'. The word probably existed in other branches as well, but can only be found in Avestan, Sogdian, and, residually, in the Persic branch (through names attested in Middle and New Persian as well as Elamite).

2.7.3.6 Old Steppe Iranian–Avestan–Khotanese–Ossetic–Balochi–Yidgha

See the discussion of *yetse* p. 61 for more detail on this possible (but complicated) isogloss.

2.7.3.7 Old Steppe Iranian–"Eastern" languages–Scythian and Ossetic–Khotanese–Khwarezmian

This is the opposite of the Old Steppe Iranian–Persic isoglosses mentioned above, since the Old Steppe Iranian word for 'axe', **paratu-*, is only shared with Ossetic *færcæt*, Scythian Φαραδο[υ] (inscription of a person's name or nickname, see also fn. 23 p. 41), Khwarezmian *pdyk*, Khotanese *paḍa-* (for etymological details, see section 2.2.11). Furthermore we have another (probably) Iranian language from which Old Turkic borrowed its word for 'axe': *baltū* < *baltō* 'axe', but that language remains unknown. Dragoni (2023: 159) proposed to view *baltō* as a borrowing from the Old Khotanese acc. sg. *paḍu*.

In Avestan and Sogdian another word for 'axe' is used: Avestan *taša-*, Sogdian *taš*, and in the west, the word **tapara-* and descendants of it are found. The word **paratu-* is thus quite limited in distribution. Naturally it could have

existed in Proto-Iranian and been replaced elsewhere, but it seems more likely that there was no Proto-Iranian word for 'axe', or perhaps there were different types of axes in Proto-Iranian times, each of them having a different name, and these types were generalized among different groups. This question needs to be answered through an archaeological lens, which goes beyond the scope of the present study. This isogloss is nevertheless both relevant and useful, in that it places Old Steppe Iranian in a group of Iranian languages which shared the same word for 'axe', although they were not (all) closely related.

2.7.3.8 Old Steppe Iranian–Alanic

Old Steppe Iranian and Alanic (namely, the branch to which belongs Ossetic and various closely related Iranian languages) share two notable isoglosses:

1. The Old Steppe Iranian meaning 'sword' of **karta-* (TB *kertte* 'sword') is only to be found in Alanic, the linguistic branch of Ossetic and other closely related languages (cf. pp. 27–30). In all other Iranian languages, the bare, unsuffixed word **karta-* strictly means 'knife' (New Persian *kārd*, etc.) so that this convergence of meaning is of particular interest. This isogloss is of rather important consequence if correct. It would suggest that the ancestor of Old Steppe Iranian and that of Ossetic, either areally or genetically, developed the meaning 'sword' or, in any case 'cutting war weapon' for **karta-*. Given the amount of proximity between Ossetic and Old Steppe Iranian (see previous section), the fact that these two languages would have separately developed this meaning for this word separately seems rather improbable.
2. The Old Steppe Iranian word **lanka-* 'cleft, ravine' (s.v. TB *leṅke*) corresponds to Ossetic Iron *lænk* Digoron *læncæ* 'ravine'. The word has further no Iranian cognates, and possibly no Indo-European etymology, making the possibility of a shared retention from Proto-Iranian thinner.

2.7.3.9 Old Steppe Iranian–Sogdian–Persic

The Old Steppe Iranian word **paribanda-* 'load, burden' (TB *perpente* 'burden, load') corresponds exactly, apart from the lack of a reflex of the suffix *-aka-*, to Sogdian *prbnty* 'idem' and Persian *parwanda* 'bundle of clothes, etc.' (s.v. TB *perpente*). Although it is likely that any Iranian language could have created such a word at the Old Iranian stage, or perhaps even later, it is still noticeable that only Sogdian and (Old) Persian did, to our knowledge.

There is a small chance, I believe, that the Persian word *parwanda* is borrowed from Sogdian, or from a hypothetical Bactrian cognate. Indeed, the expected Persian form is †*parbanda*. However, the shift *rb* > *rw* is dialectally attested in Persian, so that the word could also be dialectal, which fits well with its material semantics, but not with its meaning 'file, case'.

2.7.3.10 Summary

Old Steppe Iranian is thus found to share exclusive isoglosses with most groups of Iranian languages, including Persic. Interestingly, no Old Steppe Iranian–Khotanese-Tumshuqese(-Wakhi) isogloss has yet been found. Since, on the one hand, some Iranian languages (such as Persic ones) have been much more documented than others, and on the other hand, Old Steppe Iranian is an archaic language, these isoglosses should naturally not be all taken at face value. An important isogloss is the semantic isogloss for **karta-* shared by Old Steppe Iranian and Ossetic, as well as the isogloss OSIr. **lanka-* : Ossetic Iron *lænk* Digoron *læncæ* 'ravine'. If we bring these together with the observations of sound changes in the previous section, it seems to strengthen the idea of kinship between Old Steppe Iranian and Ossetic; the Iranian branch from which Old Steppe Iranian and Ossetic descend could tentatively be named "Scytho-Steppic".[158] All in all, these isoglosses do represent important departure points for further research on the phylogenetic identity of Old Steppe Iranian, and perhaps for research concerning its geography and the migratory history of its speakers. This data should be coupled to data on semantic innovations of Old Steppe Iranian (see next section).

2.7.4 *Semantic Innovations of Old Steppe Iranian*

I believe it could be useful to discuss a few semantic developments found only in Old Steppe Iranian, besides the lexical isoglosses mentioned in the previous section.

In the same way that it was sometimes difficult to determine whether a specific sound change had occurred within Old Steppe Iranian or in Tocharian, or in the process of borrowing, it is sometimes difficult to determine whether a semantic shift occurred within Tocharian or already in Old Steppe Iranian, although it is generally easier than for sound changes as, naturally, less examples are needed to propose a semantic change than to establish a sound law.

2.7.4.1 Old Steppe Iranian **and^{z}a*(*h*) 'Envy, Greed'

While multiple Iranian languages have a word *āz*, *āzi* with the meaning 'greed' (cf. p. 14 f.), the Proto-Iranian form **Hanǰah-* has only one other known cognate: Avestan *ązah-*, with a very different meaning: 'constriction; distress, peril'. I wonder if Buddhist influence could be the source of the semantic change, that is, after the Proto-Tocharian period, and quite some time after the word was borrowed.

158 This name is the translation of the French term *scytho-steppique*, spontaneously coined by my friend Maxime Canin when discussing this topic.

2.7.4.2 Old Steppe Iranian **čʲāta-* 'Happy, Content'

Together with Federico Dragoni I was able to establish that the meaning 'rich' of TB *śāte*, TA *śāt* is probably due to Buddhist influence. Here, by "Buddhist influence", I do not mean the influence of Buddhism as a religion and ideology (as the word for 'rich' is not particularly important in Buddhist phraseology), but that of Buddhist literary culture in general, which seems to have influenced meanings and functions of words beyond purely religious lexicon. Indeed, the meaning 'rich' is found in Khotanese *tsāta-* as well and in Buddhist Sogdian *šʾt* /šāt/ exclusively, not in Christian or Manichean Sogdian. Also, in Tumshuqese, the sister language of Khotanese, and also a Buddhist language, *tsāta-* means 'possession' (possibly this is a backformed noun, derived from an unattested adjective **tsāta-* 'rich'). No other Iranian language attests a meaning 'rich' for any cognate of these words.

This apparently purely Buddhist semantic shift has presumably occurred in a literary context, due to the influence of Buddhist Sanskrit *bhoga-* and derivatives, which ambiguously meant 'enjoyment, happiness' and 'possession, property, wealth' (cf. e.g. M-W: 767), which *śāte*, *tsāta-* and *šʾt* often translate. One can take as an example *bhoja-* 'bestowing enjoyment, generous', and the adjective *bhogavat-*, which means both 'happy' and 'wealthy, opulent'. A concrete example of this can be seen from the *Suvarṇabhāsottamasūtra* (cf. Skjærvø 2004: I, 111), where Khotanese *tsāta-* (§ 6.2.64) translates Sanskrit *bhoga-*.

Since it is impossible that Old Steppe Iranian speakers were Buddhists, it is more than probable that the Old Steppe Iranian language still had the meaning 'happy, content' for **čʲāta-*, and that the semantic shift occurred within Tocharian, alongside Khotanese, Tumshuqese and Buddhist Sogdian, due to a calque of Sanskrit *bhoga-* 'happy, wealthy, opulent'.

2.7.4.3 Old Steppe Iranian **mānίa-* 'Servant'

The Old Steppe Iranian word **mānίa-* 'servant' (Tocharian B *mañiye* 'id.'), initially a derived adjective 'of the house', is perfectly identical to the French word *domestique* 'servant' (< Lat. *domesticus* 'of the house') in its semantic derivation. Typologically, one can also compare Armenian *ałaxin* 'female servant' derived from *ałx* which originally meant **'house' (cf. Martirosyan 2010: 24 f.).

We find an identical semantic shift in Old Persian *māniya-* (translated as 'household slave(s)' by Kent (1953: 202). However, an interesting, very different semantic shift is found in Ossetic Iron *moj*, Digoron *mojnæ* 'husband' < **mānίa-* as convincingly argued by Cheung (2002: 205–206). From 'the one of the house' one can easily derive 'husband' and 'servant', but 'servant' is

hardly derived from 'husband', and vice-versa. It thus appears that the meaning 'servant' must be the result of a separate innovation in Old Steppe Iranian, and the meaning 'husband' in Ossetic also, both branching out early enough. To sum up: the original meaning, in Proto-Iranian, was simply 'of the house', and the shift to 'husband' and 'servant' must be independent from each other.

Of course, it can be argued that, if the meaning 'of the house' was retained long enough, it would have been possible to remake a new word meaning 'servant' or 'husband' at any point in time. However, given the strong lexical preference that must be given to one or the other meaning, this hypothesis seems quite unlikely to me. The specialization must have occurred early enough.

2.7.4.4 Old Steppe Iranian **dᶻainu-* 'Arrow'

In Tocharian B *tsain* means 'arrow'. It derives from Old Steppe Iranian **dᶻainu-*. In Avestan there exists a word *zaēna-* m. 'weapon', also YAv. *zaiia-* 'kind of weapon', cf. also Vedic *hetí-* 'missile, weapon', and also Avestan *zaēnuš* which means 'baldric' (cf. Lubotsky 2021: 228). The original meaning of the root was 'to hit' (cf. Lubotsky, op. cit.). The meaning 'arrow' could have arisen within Old Steppe Iranian.

In Parthian and Manichean Middle Persian *zēn*, also going back to **dᶻainu-*, means 'weapon, sword, armor, arms' (DMMPP: 386). The meaning 'arrow' is thus a specific specialization of this word. Possibly, Old Steppe Iranians had remarkable arrows, which they, or the Tocharians, designated as the weapon *par excellence*. Old Steppe Iranians can perhaps be associated with a horse-riding group, for archaeological reasons (see p. 246). It is possible, if we combine this with the semantic shift 'weapon' > 'arrow', to suggest that a sizable part of these horse-riders were mounted archers. Perhaps, as many soldiers of the antiquity, they carried both arrows and a sword (*kertte*), which permitted them to attack both from afar and from near. At least two mounted warriors were recognized among the tombs of the Yanghai cemetery (cf. Wertmann, Yibulayinmu, Wagner & al. 2023: 14), and it is likely, in my opinion, that these were Old Steppe Iranian warriors, or that these were related to the Old Steppe Iranians.

2.7.4.5 Old Steppe Iranian **apa* 'Neither … Nor'

On both typological and historical grounds I inferred that the intermediary meaning between the Proto-Iranian meaning **'away from' of **apa* (documented in Old Iranian languages) and the Tocharian meaning 'or' of *epe* should have been **'neither' (with a secondary change 'neither' > 'or', compare North

African Arabic *ula* ~ *wala* 'or' < *wa-lā* lit. 'and-not'). We can add Old Uyghur *ap* 'nor', *ap* … *ap* 'neither … nor'; 'oder … auch, entweder … auch' (cf. Clauson 1972: 3; Wilkens 2021: 55) to the comparison, which was probably borrowed from some early, prehistorical stage of Tocharian.[159]

If my argument is accepted, the Old Steppe Iranian meaning may have been 'neither … nor' rather than 'or' at the time of the borrowing. Such a meaning cannot be found in any other Iranian language. It should thus be a specific semantic development of Old Steppe Iranian.

159 Rather than directly from Old Steppe Iranian, as we have no other trace of such a contact (cf. ongoing research by Hans Nugteren, Michaël Peyrot and Jens Wilkens).

CHAPTER 3

BMAC Words in Tocharian (a Selective Survey)

3.1 Introduction

The study in the previous chapter of Old Steppe Iranian loanwords in Tocharian led to both the inclusion and the exclusion of a number of Tocharian words from the list of Old Steppe Iranian borrowings. Some of them cannot be of Indic or of Iranian origin, but they are not inherited either, or at least, an Indo-European etymology has never been found for them, such as TB *witsako* 'root'. Besides, a number of Tocharian words have been analyzed as of "substratal" origin, and this origin has usually been ascribed to the same substratum as that of Indo-Iranian, also tentatively called the language of the BMAC (for Bactriana-Margiana Archaeological Complex), nowadays also called Oxus Civilization (see e.g. Francfort 2009). Carbondating dates the civilization of the BMAC to 2250–1700–1500 BCE (cf. Parpola 2022: 26 with references).

The pioneering study of Tocharian loanwords of BMAC origin is that of Pinault (2006). In that paper, Pinault extends Lubotsky's (2001) substratal study on the Indo-Iranian vocabulary to some elements of the Tocharian lexicon. As a conclusion, he deduces that the BMAC language had definitely enough influence on Tocharian speakers to provide them with a number of important words. Some of these elements are already present in Pinault (2002).

A number of points need to be expounded before delving further into this issue: 1. it is unsure whether this so-called substratum language was indeed a substratum; 2. even if it were a substratum for Indo-Iranian, it is not certain at all that it was a substratum for Proto-Tocharian. Lubotsky (2001) himself writes that he is unsure whether this language was a substratum or rather an adstrate. Since its contribution to Indo-Iranian was apparently mostly lexical, it is more likely that it was an adstrate or a superstratum (cf. e.g. Kümmel 2020: 255); 3. for practical purposes I will call it "BMAC language", in the case of Tocharian contact, and "Indo-Iranian substratum" to refer to the variety in contact with Indo-Iranian in particular. In the discussion at the end of this chapter, I will investigate whether we can tell if the BMAC language that influenced Proto-Tocharian was the same as the one that influenced Indo-Iranian, or whether they were for instance sister languages.

In the present chapter, I will discuss a selected number of Tocharian words that appear to have substratal features, or have been claimed to be BMAC words.

© KONINKLIJKE BRILL BV, LEIDEN, 2025 | DOI:10.1163/9789004732537_004
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

In general, an Iranian origin has been ascribed to them, although it can, in my opinion, no longer be maintained, in light of the phonetic correspondences discussed in the previous chapter. This study is selective in the sense that I was not able to accomplish a full survey of all potential BMAC words in the Tocharian corpora, and, unlike for Old Steppe Iranian, I did not discuss every possible BMAC loanword in Tocharian mentioned in scholarship, because it was not the initial purpose of my research. For instance, I have not discussed the proposed BMAC loanword **āni-* 'hip' (cf. Pinault 2003; 2006) because its etymology is rather complex. In general, there are not many proposed BMAC loanwords, and this topic deserves, in my view, greater consideration.

Some common features can be recognized for most of those words: 1. apart from TB *etre* TA *atär* 'hero', they all designate realia: plants, animals, food, or construction material. 2. Most of them are trisyllabic, and, when visible, the stress is constantly on the second syllable. This reminds us of the feature of the BMAC loanwords as described by Lubotsky (2001: 303), namely: an unusual syllable structure, mostly trisyllabic nouns with a long vowel in the middle syllable. I will discuss these features in greater detail in the discussion at the end of the chapter.

3.2 Analysis of Potential BMAC Loanwords in Tocharian

3.2.1 *TB* **iṣcäke** *'Clay', TB* **iścem** *'Clay'*

The Tocharian B words *iṣcäke* 'clay' or 'a specific type of clay' and *iścem* 'clay' are related, both formally and semantically. Their etymology, however, is complex and intricate. I will discuss various etymological proposals concerning these words, and try to put forward my own. In 3.2.1.1, I will discuss Pinault's and Adams' proposal. In 3.2.1.2, I will discuss cognates of these words in other languages, which will permit me to propose another solution for *iṣcäke*. In 3.2.1.3, I will discuss the origin in Tocharian of *iścem*, which I believe to be related but not identical to that of *iṣcäke*, before concluding (§3.2.1.4). Finally, as an appendix, I add a philological study of the Avestan word *zəmōištuua* and related forms in 3.2.1.5, which will be useful to the present discussion.

3.2.1.1 On the Meaning and Etymology of TB *iṣcäke* 'Clay'

The word *iṣcäke* occurs only once, in the form *iṣ̱cq̱ake*, in a bilingual St.-Petersburg manuscript of the Petrovsky collection. The line where it appears reads *tokharika*: k_u*cañ̈ñe iṣ̱cake*. The sequence ‹ṣ̱ca› stands for ‹ṣ̱cq̱›, i.e. *ṣcä* (cf. Peyrot 2008: 179). Since the text in which *iṣcäke* occurs is archaic, it is impossible to determine whether the schwa was accented or not.

This line has been much discussed, but it seems that a definitive breakthrough was made by Pinault (2002), according to whom *iṣcaṃke* designates a type of clay, and translates *tokharika*, a pseudo-Sanskrit form corresponding to Sanskrit *tūbarika*, a word designating a sort of clay. As Adams (DTB: 191–192) suggests, a meaning 'a sort of' is probably to be posited for k_u*cañ*. Thus we could translate k_u*cañ iṣcaṃke* as "a type of brick" or "a type of clay". Theoretically, "a type of clay" could metonymically also designate a specific sort of brick based on the stuff of which it is made, but the translation as 'clay' is simpler and more straightforward.

Pinault (2002: 325–335 and 2006: 171) was the first to connect TB *iṣcäke* etymologically to TB *iścem* 'clay', sometimes translated as 'clay brick' (on which see below). Pinault suggested that there existed a noun **iśc*, a cognate of unattested TA **iśäc*, borrowed in Old Uyghur *išič* (*ešič* 'Kessel, Topf', cf. Wilkens 2021: 264). As a paradigmatic analogy with TB *āśce* 'head': obl.sg. *āśc* and other nouns having the same pattern, a nominative **iśce* was formed. This noun **iśc* would derive from a BMAC word **išti-* with **-ti* > **-cä* as in words from PIE (cf. Pinault 2002: 330).

The form **iśce* would have been enlarged with the suffix *-maṣṣe* "servant à dériver des adjectifs de relation, surtout avec valeur technique" (2002: 328), to form *iścemaṣṣe* 'earthenware; (thing) made of clay', from which *iścem* 'clay brick' would have been extracted. However, it was later demonstrated by Peyrot (2008: 94) that the suffix *-maṣṣe* is both late and colloquial, while *iścem* appears in classical texts. It is thus unlikely that *iścemaṣṣe* derives from *iśce**, rather than from *iścem*.

The form *iṣcäke* itself would have, according to Pinault (2002: 331), derived from this same **iśce*, through the addition of a suffix *-ke* of Iranian origin. Pinault interprets this suffix as possibly indicating "une sorte de", and thus glosses *iṣcäke* as "une sorte de terre" or "une sorte d'argile" (2002: 331). Nonetheless, this suffix only occurs in Iranian, Indic or BMAC loanwords, where it never has the meaning 'a sort of', as one can see in the examples cited by Pinault: TB *kattāke*, TA *kātäk* 'maître de maison' (actually *kātak** see p. 123-124, 124n103); *ṛṣāke*, TA *riṣak* 'sage, ascète', either from Indic or from Indic through an Iranian intermediary; *ṣecake* (s.v.) 'lion' from BMAC, TB *ainake*, TA *enāk* 'bas, vil' (s.v.) from Old Steppe Iranian. Pinault suggests this meaning 'a sort of' because it is a meaning this *-ka* suffix can bear in Indic, but there is no need for that, since the meaning 'a sort of' is already provided by k_u*cañ* (cf. DTB: 191–192).

There is only one other example of a borrowed noun following a pattern similar to that of *āśce* in Tocharian, a Sanskrit loanword, namely *akāśe*, obl. *akāś* 'sky, ether' ← Skt. *ākāśa-*, and there is no positive evidence for BMAC **-ti-* > PT **-cə-*, nor is there negative evidence against it, I have to admit. Although

Pinault's interpretation is very enlightening on many points, I believe a more straightforward scenario can be presented to account for the etymology of both *iṣcäke* and *iścem*. Before presenting my own theory, I wish to discuss another etymology of *iṣcäke* and related words, namely, Adams'.

Adams (DTB: 72) wrongly writes that Pinault takes *iṣcäke* to "represent an earlier Iranian **iṣtyaka-*." Adams further (DTB: 73) takes *iścem* to derive from an eastern Iranian language form "*ištyám*". In his opinion these Iranian words derive from **h*$_2$*eis-* 'fire (clay)' → **h*$_2$*isti-*.

There are some problems with this proposal, the most important being the semantic aspect of it. Indeed, the meaning of **h*$_2$*eis-*[1] (cf. Av. *aiiaŋha-* 'cauldron' < PIIr. **aiasa-* 'cauldron' < *'metal(-ware)') does not mean 'fire (clay)', but refers to the process of heating metals, cf. YAv. *aiiah-* n. 'metal', Vedic *áyas-* n. 'metal, Nutzmetall' (EWAia[1]: 104). This is a totally different meaning from that of 'brick'. Admittedly, it is conceivable that words for brick refer to the process of brick-baking. However, in most Indic and Iranian languages, and in the Burushaski cognates discussed below, these words refer to sun-dried bricks, i.e. simple mud-bricks, which involve a completely different preparation process than metal. Both the objects (clay pots, clay bricks vs. metal-ware, metal pots) and the preparation processes (burning, branding vs. sun-drying) are fundamentally different. I therefore do not believe that this etymology is possible on semantic grounds.

Rather than looking for a complex, unclear Indo-European origin for *iṣcäke* as Adams does, I believe, like Pinault (2002), that TB *iṣcäke* can be explained as a borrowing from the BMAC language, although my solution differs from his in the detail. I also explain *iścem* directly from Old Steppe Iranian.

3.2.1.2 Cognates and Further Etymology of TB *iṣcäke* 'Clay'

Cognates of the Tocharian B word *iṣcäke* are found in Indo-Aryan and Iranian languages as well as in Burushaski. I believe it is useful to cite them here, before discussing the etymology of TB *iṣcäke* in more detail.

Some of the Indic cognates are: Vedic *íṣṭakā-* f. 'brick', Sanskrit *iṣṭikā-* f. 'id.', Buddhist Sanskrit *iṣṭā-* (BHSD: 115: "[p]erhaps loss of suffixal ka [...]"), Prakrit *iṭṭhakā-* 'tile, brick', Nepali *ĩṭ*, Assamese *iṭā*, Hindi *ĩṭh*, Gujarati *ĩṭi* f. 'brick'. Although it could be inherited, Balochi *īt* 'brick' could also be a borrowing from an Indic language (Korn 2005: 137).

1 The **h*$_2$ here is not assured by means of reconstruction, but based on the hypothetical connection with **h*$_2$*eid*^{h}*-* (cf. LIV^2: 229).

Some of the Iranian cognates of these words are: Old Persian *išti-* 'sun-dried brick' (Kent 1953:175), YAv. *ištiia-* n. 'Ziegel, Backstein' (AIW: 378), YAv. *zəmōištuua-* (AiW:1691) 'Lehmziegel' (= Eng. 'adobe'), Khwarezmian *štyc*, < *(*i*)*šti-c*(*y*) pl. *'štyc* (Benzing 1983:99, 601), MP *xišt* (CPD: 94), NP *xišt*. The latter was borrowed into Pashto as *xaxta* f. Geiger (cited by EVP: 98) was the first to propose it as a borrowing from Persian. Cheung (2013: 618–619) considers the borrowing to be quite old.

Burushaski, a language isolate, has the forms *diṣcík* (sg. and pl.) and in the Yasini dialect *giṣṭék*, *giṣtík*, *kiśtíki*, *diṣtík* 'sonnengetrockneter Ziegel', cf. Shina (a Dardic language) *diṣtík* (Berger 1998: 121). Since neither *d*(*i*)- nor *g*(*i*)- are nominal prefixes in Burushaski, these forms cannot be immediately explained.

In any case, the Indo-Iranian forms go back to a form with the cluster *-*št*- while the Tocharian forms go back to a form with the cluster *-*šč*-. This might suggest that the Indo-Iranian languages borrowed this word independently of Tocharian. Perhaps Indo-Iranian adapted an original cluster *-*šč*- as *-*št*-, or perhaps this variation was found in different BMAC dialects. According to Pinault (2002: 330), the Burushaski forms also presuppose an early cluster *-*šč*-, but in other instances where *ṣç* and *ṣṭ* interchange, the latter seems to be original (Berger 1998: 25).

In passing, I would like to exclude another etymology for Tocharian B *iṣcäke*. Because of the Yidgha sound change *šť > šč through *šť^{y} (Morgenstierne 1938: 73), I had wondered if this word was borrowed from Pre-Yidgha into (Proto-)Tocharian. In that case, it was possible to consider such a Pre-Yidgha or Proto-Yidgha-Munǰī word as an Old Iranian loanword into Tocharian.

However, this hypothesis is weakened by two major points: if *iṣcäke* goes back to a Proto-Yidgha descendant of an earlier **ištika-*, then the **-i-* of the Old Iranian form was reduced to schwa while the word final *-a* was preserved, which is chronologically difficult, if not impossible, because the sporadic reduction of unstressed short vowels to schwa (notably of *i* and *u*, cf. Morgenstierne 1938: 95) is a much later phenomenon in Yidgha than the loss of word-final stem short vowels. This is especially true of the *-aka-* suffix, which was reduced to *-ë* ~ *-ëy* in Yidgha (Morgenstierne 1938: 114), although it was retained longer after *-u* and consonant (Morgenstierne, op. cit.). We can thus expect TB †*iṣcike* or †*iṣcäk*, but a derivation of *iṣcäke* from Pre-Yidgha seems to be a chronological problem in itself.

The second difficulty with a Yidgha origin for the Tocharian B form *iṣcäke* 'brick' is that the words for 'brick' in Yidgha (and Munǰī) are *uštu* (← Khowar) and *xišt* (← Persian), thus later borrowings. This does not imply that an inherited word for 'brick' could not have existed in Yidgha, but since it was replaced, it was possibly not a prominent technological feature of the Pre-Yidgha people.

Both these arguments make the hypothesis of a borrowing from Yidgha or Pre-Yidgha for this word very unlikely. An independent borrowing from a BMAC language remains the most likely option for TB *iṣcäke*.

The *-äke* ending of the Tocharian word cannot derive from the Iranian *-aka-* suffix. Indeed, the latter should be reflected as †*-eke* if from Old Steppe Iranian and there would not be a final *-e* if from Middle Iranian (for Khotanese, see Dragoni 2023). I propose that this *-äke* element goes back to the **-ka-* suffix of BMAC, discussed in Lubotsky (2001) and seen for example in the Indo-Iranian words **atka-* 'cloak', **stuka-* 'tuft of hair'. It is likely that the word reflected by TB *iṣcäke* 'clay' also contains a reflex of this suffix.

Indeed, there is no trace of a suffix *-ka* for the 'brick' word in any Iranian language. The use of the *-ka-* suffix is extremely common in Iranian languages, which means that it could have been added to this word in an unattested language, although the fact it is lacking from Iranian languages in this word, coupled with the other arguments cited above, make the hypothesis that *iṣcäke* derives from an Iranian posited form **ištika-* rather unlikely.

One could imagine that the Tocharian *-äke* represents the *-ika-* part of **išti-ka-*, built on PIr. **išti-*, but, if my etymology of TB *epastye* 'skillfull' (< Pre-B **epəstiye*) is accepted, OSIr. **-ti-* did not become **-cə-*, although it can be argued that the preceding **-s-* blocked the palatalization. Another argument is that **i-* should have been rendered as **ye-*, were the word of Old Steppe Iranian origin (as in *yentuke* s.v., *yetse* s.v.). A counterargument to that is that TB *iścem*, which I argue below is of Old Steppe Iranian origin, does not start with *ye-*. One could suggest the *i-* in *iścem* was influenced by the *i-* in *iṣcäke*, of close or identical meaning, or that its initial *i-* remained unchanged because it was not accented. Why would Proto-Tocharian have borrowed two different words, with the same meaning, from the same language?

As to the initial *d-* and *g-* or *k-* of Burushaski in this word, I take it that they represent attempts by Burushaski speakers at reproducing an initial sound which was not readily available in their own phonological system. The presence of this initial sound can also be seen from the fact that the word was likely borrowed with an initial laryngeal in (Indo-)Iranian. Indeed: it was likely rendered as **ḥ* (Kümmel's notation), hence the *x-* in Middle and New Persian *xišt* (cf. Kümmel 2018: 166), but this initial sound was *a priori* not reflected in Tocharian.

3.2.1.3 TB *iścem* 'Clay'

The meaning of *iścem* can hardly be 'brick' as claimed by Adams (DTB: 72) as *iścemaṣṣe* means 'made of clay' (as Adams himself admits, cf. DTB: 72). If the meaning of *iścem* were 'brick', then *iścemaṣṣe* would naturally have meant 'made of brick'.

If Tocharian *iṣcäke* is indeed a direct borrowing from BMAC into Tocharian, one might be tempted to believe that this is also the case for Tocharian B *iścem* 'clay' and its derivative *iścemaṣṣe*. However, we do not know of a BMAC suffix or ending *-*am* (cf. Lubotsky 2001). A more straightforward etymology consists in taking it from OSIr. **išt̯ám* (neut.), as was done by Adams (DTB: 73). The root **išti-* 'brick' in Indo-Iranian is ultimately a BMAC borrowing (cf. Kümmel 2020: 257). The difference in meaning between 'clay' and 'brick' is problematic here and, although this only pushes the problem further, the shift can easily be attributed to Old Steppe Iranian. The word *iścem* should designate a different type of clay than *iṣcäke*.

An obvious counter-argument is that, in regard of TB *yetse*, TA *yats* 'skin' (s.v.) and TB *yentuke* 'Indian', both from Old Steppe Iranian, we would expect Tocharian B **yeścem*, as per the sound law PIr. **i-* > OSIr. **i̯a-* (or **i̯e-*), cf. p. 163 f. Three solutions can be proposed here: first, one can imagine that, if the word was accented on the last syllable (as a neuter noun), a different rule applied: unaccented **i-* would then remain **i-* in Old Steppe Iranian. Another solution consists in suggesting that, because the Iranian word started with a consonant (**h*$_2$-, that is **ḥ-* in Kümmel's notation), see above (also Kümmel 2018: 166), the **i* of the Old Steppe Iranian form remained as such, as it was not in absolute word-initial position. The third solution is that the initial *i-* of *iścem* would have been influenced by that of *iṣcäke*, of identical or very similar meaning.

3.2.1.4 Conclusion

In conclusion, I suggest that Tocharian B *iṣcäke* 'brick' was borrowed from BMAC directly, as proposed by Pinault (2002). TB *iścem* 'clay brick', however, was borrowed from Old Steppe Iranian **išt̯ám*, a neuter accented on the second syllable. This is similar to *weretem** 'debt, surety' (s.v.), also borrowed from an Old Steppe Iranian neuter **u̯rátam*.

3.2.1.5 Appendix: a Philological Discussion on Avestan *zəmōištuua*

An analysis of the Avestan form *zəmōištuua*, often cited in conjunction with the Tocharian forms cited above, seems in place, although not essential to the discussion. The word *zəmōištuua* appears in a passage of the Vendidad (VIII:10), namely:

> *duua dim nara isōiϑe vīzōištąm vīzuuārəntąm maγna anaiβi-vastra zəmōištuue vā zarštuue vā upaskanbəm*
> "deux hommes, vaillants et habiles, nus, sans vêtements, devront déposer le corps sur la brique d'argile ou de la pierre" (Lecoq 2016: 938).

Here *zəmōištuue upaskanbəm* is translated by 'brique d'argile', and in the Pahlavi *zəmōištuue* is translated as '*zamīgēn*', lit. 'earthen', thus 'clay brick' or 'earthen brick', although *upaskanbəm* does not designate a simple brick, a basic meaning similar, if not identical, to Tocharian *iścemaṣṣe*. Bartholomae (AiW: 1691) suggests the etymology to be **zama* + **ištiah*, but he notes "[m]an erwartete freilich bei dieser Et. die Schreibung *zəmō.iš*°", which is indeed true. The absence of such a spelling shows that the word *zəmōištuua* was not understood as a compound by the scribes. Duchesne-Guillemin writes: "[u]ne autre série a en 1er t[erme] un nom de matière. […]" and further quotes "*zəm.varəta-* « motte de terre » et l'obscur *zəmōištva* qui n'est même pas sûrement un c[om]p[osé]." (Duchesne-Guillemin 1936: 137).

The form *zəmōištuua* is not often found in manuscripts: Ave976 has *zəmōište*, (which stands for +*zəmōišti*), while all other Iranian manuscripts available in the Avestan Digital Archive[2] read *zəmōišti*. There is more variation in Indian manuscripts: some (such as ML630) read *zəmōišti* or *zimōisti* (M2), two manuscripts (G106 and Bh3) have *zimōištuuō* ("*vā zarštuue vā*"), B4 has *zimōistevā* ("*zarastavaevā*") G112 and Fires1 have *zəmōišta* ("*vā zruuašta*"), K10 has *zəmōištai*, corrected in *zəmōištuui*, ("*vā zarštuui vā*"). This kind of variation is common with hapaxes.

Avestan manuscripts with Pahlavi commentary agree more: MU1, E10, G34, T44, L4, F10 *zəmōištuuō*, K1, ML3, B1, *zəmōištuue*, M3 *zimōištuue*.[3]

Not a single manuscript has a separation point indicating a compound inside this word, and it is translated in Pahlavi as *zamīg gōn* 'earth coloured (thing), earthy (thing)' or *zamīgēn* 'earthen' and in Persian as "(on) the dry soil". Although much more common in Pahlavi manuscripts, the forms *zəmōištuuō*, etc. could be anticipated from the next word: *zarštuue*, and variants. Indeed: one notes that the ending often "rhymes" with that of the next word here, as can be seen from the list I made. For instance, *zimōistevā* precedes and rhymes with *zarastavaevā*, while the variant *zəmōišta* is followed by and rhymes with *zruuašta*.

If the word indeed designates dried bricks made of clay,[4] then it is unclear why it was not understood as a compound, since *zəmō-* is a normal first mem-

2 Ave977/978, Ave991, Ave1001, ML16226, RSPA230, HM2, YL1, ML15283, TU4, VJ.

3 The Pahlavi translation of F10 has ‹zmyg gw› (maybe originally +‹zmygyn›, since the points on the second ‹g› were added later) 'earthen, earth coloured', corrected by the Persian translator in ‹zmyg gw**n**› and translated in Persian as زمین گون را *zamīn gūn rā*, while the interlinear Persian translation under the Avestan text has زمین گون چیز را صاف کنید *zamīn gūn čīz rā ṣāf kunēd* that is "make you pure an earthen thing". The word was thus interpreted as a substantivized adjective.

4 It is unclear to me what other types of bricks could be opposed to these clay bricks: golden bricks?

ber of compounds. It is also unclear why the form was not **zəmō.ištiia-*. If the form *zəmōišti*, found in the Iranian manuscripts as well as in some Indian ones, is authentic, then the word is indeed a cognate of Old Persian *išti-* 'sun-dried brick'. It seems to me that comparative grammar can help us establish that *zəmōišti* is indeed the original form in the Avestan passage.

On the basis of comparative evidence, (cf. section 3.2.1.2 of the current chapter) what is reconstructible for Indo-Iranian is **ḥišt-*, often, but not always, suffixed with **-i-*. The form **ḥišt-* could either derive from **ḥišt-* or from **ḥist-* with RUKI, which also applied to BMAC loanwords in Indo-Iranian (Lubotsky 2001: 304).

In conclusion, if I am correct, the Avestan word *zəmōištuua* and variants should no longer be cited as a cognate of Old Persian *išti-* and Vedic *íṣṭakā-* f. both meaning 'brick', but rather its variant *zəmōišti*,[5] found in most Iranian manuscripts, should be cited. This variant seems to go back straightforwardly to **zəmō.išti* that is, literally, 'earth-brick', but it is unclear why it was no longer understood as a compound by Avesta copyists and Avestan grammarians.

3.2.2 *TB* eñcuwo, *TA* añcu* *'Iron'*

To the Tocharian B word *eñcuwo* 'iron' corresponds unattested Tocharian A *añcu** 'iron'. The latter can straightforwardly be extracted from the derived adjective *añcwāṣi* 'made of iron', cf. also TB *eñcuwañ(ñ)e* 'made of iron', with a rarer variant *iñcuwaññe*.[6] These words have remained difficult to etymologize. In the present discussion I will argue that this word is of Iranian origin. However, I chose to study it in this BMAC chapter because it has frequently been suggested that Tocharian B *eñcuwo* and its Tocharian A equivalent are of BMAC or substratal origin.

TB *eñcuwo* and TA *añcu** have no clear Indo-European origin. They have been connected to Iranian words of similar meaning, for example by Winter (1971: 222) who links them to Ossetic *ændon* 'steel'. Schwartz (1974: 409[33]) connects the Tocharian words to Khwarezmian *hnčw* 'spear-head', but suggests they are both of non-Indo-European origin. Isebaert (1980: 191–192) connects them to PIr. **spana-*/**safna-* 'iron', deriving both from the Proto-Indo-European root **ḱu̯e* (*sic*) "from **ḱeu-n-*/**ḱu̯-en-* 'to be bright, to shine' ['stralen, schitteren']". I could not find this root in the LIV[2], probably as it is mostly found in nouns.

Tremblay (2005: 424) assumes an Iranian origin, and further assumes that the Iranian word itself is inherited from Proto-Indo-European. He writes: PIE

5 See also Vendidad VIII:8 *ištiiehe vā zarštuuahe vā* where we find the element *ištiia-*.

6 On which see Peyrot (2008: 60). He convincingly arguments that *eñcuwo* is likely to be the oldest form, and *iñcuwaññe* a later variant.

*$h_2$ék-u̯on- 'cutting edge' > Old Sakan *anc̑u̯an- (a "nasalized variant" of Old Sakan *ac̑u̯an-), which was later borrowed into Tocharian.[7] This "nasalized variant" would have no reflex in Sakan languages, because Khotanese *hīśśana-* 'iron' shows no trace of the first nasal.

Later, Pinault (2006: 184–189) proposed to derive both the Tocharian words and Indo-Iranian *anću- (see Lubotsky, 2001: 304, 310), meaning 'Haoma, Soma plant' (the sacred plant and substance of both Zoroastrian and Vedic religions) from a substratal form *anću-. He argues that the substratum word designated a reddish, rusty colour, an attribute of both the twigs of the Haoma plant, and of iron.

Finally, Adams (DTB:84–85) proposes an Indo-European derivation of the Tocharian word (*h_1n̥-ǵ^{h}eu̯eh2(-n)- 'what is poured in'). He suggests that the word was borrowed from Proto-Tocharian into Proto-Iranian, but does not exclude a transfer in the other direction.

All of the etymologies stated above have their weaknesses: what Isebaert could not have known, since it was established much later, is that the Old Steppe Iranian reflex of Proto-Indo-European *-ḱu̯- is reflected in Proto-Tocharian as *tsw* rather than *ñcw* (see for instance *etswe* pp. 18-22). This also contradicts Tremblay's proposal. As to Adams' proposal, it is improbable that the Proto-Tocharian form *eñcuwo* spread to even one Iranian language (cf. Del Tomba 2023: 129), let alone so many, as he proposes. Indeed, this would be the only Proto-Tocharian word in Iranian. Furthermore, we see that the Iranians were in all aspects more technologically advanced than the Tocharians. That they would have borrowed a metal name from the latter is highly unexpected.

Together with Federico Dragoni and Michaël Peyrot, I have developed an alternative etymology in order to explain the Tocharian words for 'iron'. This etymology is presented in detail in Peyrot, Dragoni & Bernard (2022). In short, we propose that Proto-Tocharian *eñcuwo* derives from the Pre-Khotanese predecessor of Khotanese *hīśśāna-* 'iron'.

It was demonstrated that Khotanese *-śś-* may correspond, at an earlier stage, to Tocharian *-ñc-* (Chen & Bernard 2024). Indeed, the Tocharian word *śāñcapo*, which, as we show in Chen & Bernard (2024) means 'mustard', corresponds regularly to Pre-Khotanese *śanźapa- (or *śanźapāna*), which yielded Khotanese *śśaśvāna-* 'mustard'; compare the reconstruction of an Iranian preform *sinšapa- 'mustard' for the rest of Iranian, in Henning (1965: 44).

7 According to him *anc̑u̯an- (or rather a descendant thereof) was also borrowed by Khwar. *hnčw* 'spear-head', *hnǰw* 'iron-tip', because of the treatment of PIE *ḱu̯ in Sakan languages. "The initial *h* in *hnčw* is either a 'cockney aspiration', as in *hrs-* 'bear' < *ṛša- [...], or a closer assimilation to *handāna-* 'alloy'."

For the *-ī-* in the first syllable of the Khotanese word *hīśśana-*, we argue that it is due to an independently proven Khotanese sound change: when followed by a *yod* in the third syllable, accented *a* regularly became *-ī-*, as in Khotanese *ysīrra-* 'gold' < PIr. **ǰáraṇi̯a-* (Skjærvø 2004: II, 331). We thus suggest that Proto-Khotanese-Tumshuqese **hám-ćṷaṇi̯a-* became Pre-Khotanese **henśṷaṇi̯a-* first. This form **henśṷaṇi̯a-* was then borrowed as Proto-Tocharian **eñcuwañña* or **eñcuwañño*, with subsequent reinterpretation as an adjective: **eñcuwaññe* became an adjective 'made of iron', from which the form **eñcuwo* was then extracted by back-formation.[8]

We further suggest with caution that it was the speakers of early Khotanese who introduced iron into the Tarim Basin (cf. Peyrot, Dragoni & Bernard 2022: 419, with references). At the time, these speakers of early Khotanese were perhaps part of the Aqtala Culture, and their language might thus perhaps be conceived of as the language of the inhabitants of Jumbulaq Qum, one of the major sites of this culture, as presented by Debaine-Francfort and Idriss (2001). This is compatible with the current archaeological evidence.

According to us, the Khotanese word is ultimately a derivative from **ham-* 'together with' + **ćṷaH-*, which would in the context of iron have meant 'strike iron'. Iron would thus be the metal that is "struck together".

In conclusion, Tocharian B *eñcuwo* and Tocharian A *añcu** 'iron' derive from Pre-Khotanese **henśṷaṇi̯a-* 'iron'. This fits with the archaeological data, which indicate that iron was introduced from the west towards the east.

3.2.3 *TB* etre, *TA* atär *'Hero'*

A number of proposals have been put forward to explain this word as Iranian (most were summed up in Isebaert 1980: 191). I will discuss these to determine if they could possibly justify to consider TB *etre*, TA *atär* as loanwords from an Iranian language.

Pinault (2006: 171–175, reproduced in 2020a: 381–382) proposes to link these words to the Indic word *átharvan-*, Avestan *āϑrauuan-* 'priest' (on the Avestan form see de Vaan 2003: 65). The Indic and Avestan correspondence is irregular: Sanskrit *-ar-* corresponds here to Avestan *-ra-*. This irregularity is typical of BMAC loanwords borrowed into Indo-Iranian, cf. Lubotsky (2001: 303).

Although Pinault's connection is tempting, it is not assured, since a priest and a hero differ in many respects. Furthermore, a priest is not associated with the same type of strength or guile that is most often associated with heroes.

8 Another possibility that we suggest is that the Tocharian forms were borrowed from a pre-Khotanese form **henśṷana-* when the *yod* had already disappeared (cf. our paper for more detail).

Adams proposes that this word is from PIE *h_2*ot-ro-*, related to TA *ātäl* 'man', which he then derives from *h_2*et-lo*. The semantic shifts 'man' > 'hero' and conversely, although attested, are not evident and require a certain cultural context to work. More importantly, it is far from assured that this root described by Adams can be reconstructed for Indo-European at all.

A derivation from the otherwise unknown Iranian root *aϑ-* 'be violent' (Bailey 1975: 7[16]), or from Ir. **atara-* or **ātara-* (cf. Winter 1971: 218–219) have also been proposed. However, the Iranian root *aϑ-* 'to be violent' does not seem to exist; in any case I could not find any valid cognate, or any traces in the scientific literature. I do not see which words **atara-* or **ātara-* (**atara-* 'one of both'?) Winter refers to, but without a cognate, or an explanation, his theory remains weak.

Pinault's proposal that this word TB *etre*, TA *atär* 'hero' is of BMAC or Central Asian origin is the most convincing proposal made for these words. Whether one agrees with Pinault's connection of these words with Vedic *átharvan-* and Avestan *āϑrauuan-* / *aϑaurun* 'priest' or not depends on one's view of the underlying notions that make up the character of a priest and that of a hero. One could perhaps think of an original meaning *'leader': with the semantic shift *'leader of the sacrifice' on one hand, and the semantic shift *'leader' > 'hero' on the other hand. The latter shift has a parallel in Irish *nēath* 'hero' from the root **ni-* 'to lead' (cf. Buck 1949: 712). Even if the connection of PT **etre* to Vedic *átharvan-* and Avestan *āϑrauuan-* were to be rejected, it would not preclude a BMAC origin for this word. In lack of a convincing Indo-European or Iranian cognate, this remains the most plausible hypothesis to date.

In conclusion, Tocharian B *etre* and Tocharian A *atär*, which go back to Proto-Tocharian **etre*, are probably of BMAC origin, as proposed for the first time by Pinault (2006).

3.2.4 *TB* kenek, *TA* kanak *'Cotton Cloth, Cotton Fabric'*

Pinault (2001: 128–129)[9] has shown that the Tocharian A word *kanak*, and its Tocharian B cognate *kenek*, correspond to Old Uyghur *böz* 'Baumwollstoff'. He established its exact meaning as being 'étoffe de coton' (2001: 129); 'cotton cloth' (DToA: 92). It has variously been proposed that these Tocharian words derive from an Iranian form **kanaka-* (Pinault 2001: 129; Isebaert 2003; Tremblay 2005: 425; Peyrot 2018: 270–271).

9 Although Pinault (2001: 129), more specifically, connected it to Khwarezmian *kcynyk* 'Seidenstoff', deriving the Tocharian word from a form **kcenek* < **käcenek*. For a criticism of this etymology, see Lubotsky & Starostin (2003: 260).

However, there is—to my knowledge—not a single Iranian form that goes back to **kana-ka-*. Almost all attested Iranian forms either go back to **kanā-*, for instance Ossetic Digor *gænæ*, or to **kan(a)ba- ~ *kanafa-*, as do Khotanese *kaṃha-* 'hemp', New Persian *kanaf*, Khwarezmian *knbynk* 'linen', Sogdian *kynp'* 'hemp, flax'.[10] On the basis of the attested forms, **kanaka-* cannot be reconstructed for Proto-Iranian or Old Iranian.

Peyrot (2018: 270) noted this difficulty. He further wrote (2018: 271) that "in view of the TB vowels *e_e* for Iranian *a_a*, the borrowing must be relatively old; a Proto-Tocharian reconstruction **kenek* would theoretically be possible. The word is clearly a Wanderwort originally from the Middle East so that a unified reconstruction for Iranian cannot be given […]."

It is important to examine this hypothesis, and to see whether the origin of Proto-Tocharian **kenek* is Iranian or not, and, if not, what else it could be. If it is Iranian, it should indeed come from Old Steppe Iranian, because of the vocalism *e* for Iranian **a*. It has been proposed that the Iranian proto-forms **kanā-*, **kanaba- ~ *kanaf-* were borrowed from Uralic **kana-* 'hemp (narcoleptic)' (Katz 2003: 143 f.), but this is doubted by Kümmel (2020: 255) who believes that both the Iranian and the Uralic forms were borrowed from the same source.[11] Katz (2003: 143) further suggests that "Alanic **kœnœ*" was borrowed from Mari **kəne*.

If Katz' hypothesis of a Uralic origin is nevertheless correct, an Old Steppe Iranian form **kanaka-*, which would be a simple *-ka-* enlargement on this **kana-*, could have existed. Nevertheless, an Old Steppe Iranian form **kanaka-* would have yielded Proto-Tocharian ***kenke* TB *†keṅke* TA *†kaṅk* as per the established syncope rules (see section 2.7.2.7). Thus, even if such a word ever existed in Iranian, it could not have yielded TB *kenek* and TA *kanak*.

Indeed, as noted above, this word is a typical Wanderwort, found in some Semitic languages, in Proto-Germanic **hanipa-* 'hemp' (cf. Šorgo 2020: 440), Proto-Slavic **konopl̓a*, Greek κάνναβις 'hemp', but also in Kartvelian (Svan *kan* 'hemp') and in Abkhaz *akonə* 'hemp', etc. Witzel (1999: 55) proposed that Greek κάνναβις, Old High German *hanaf*, Dutch *hennep* all were borrowed from Scythian. This is perhaps possible (although one would need to explain the gemination of the Greek), but in that case the Scythian word needs to have been different from its Ossetic cognate *gænæ*, which is rather uneconomical.

---

10 According to Gharib (SD: 203), the Sogdian word was borrowed from Syriac *gnb'*.

11 Although the forms starting with *g-*, such as Ossetic Digor *gænæ* Iron *gæn* 'hemp', Kabard *ǵanä* 'shirt', would fit well with a language that does not have a voiced/unvoiced phonological opposition.

As discussed below (see section 3.2.1.2), the *-*ka* suffix that was observed by Lubotsky (2001: 304) in Indo-Iranian words of BMAC origin, is also found in a number of BMAC Tocharian words. A number of words presented in this chapter share the structure *CaCáCa*, unlike most Old Steppe Iranian words seen in the previous chapter, which rather appear to have a *CáCaCa*(*Ca*) structure.

As mentioned section 2.7.2.7, Proto-Tocharian got rid of trisyllabic loanwords with identical vowels through the apocope of the middle syllable (type OIr. **rataka-* → PT **retke* 'army'). I have also tried to show that Old Iranian loanwords into Tocharian, with a few explicable exceptions, had fixed first-syllable stress. In my view, it is possible that fixed middle syllable stress prevented such a reduction in this word: while **rátaka-* could easily become **retke*, it would have been more difficult for **kanáka-* to become †*keṅke*. Inherited examples of a similar syncope rule are found in, for instance, TB *sonop-* < **sonopo-* and TB *tsetserñu* < **tsetsereññu*. This is why I think that the apparently unsuitable structure with three syllables with identical vowels was in this case resolved differently, namely by dropping the final vowel, yielding PT **kenek*, which does not otherwise receive an easy explanation. Thus, on the basis of the apparent accentuation of the word and of its formation (with the BMAC nominal suffix *-ka-*), I believe that this word was borrowed, not from an Old Iranian language, but from the source language I call BMAC for convenience.[12]

For the semantics, I suggest that the word originally meant 'hemp cloth', and, possibly because cotton was more common in the Tarim basin or in Tocharian material culture, the meaning of these words shifted to 'cotton cloth' in Proto-Tocharian, or otherwise independently in Tocharian A and B.

In conclusion, Tocharian B *kenek* and Tocharian A *kanak*, both going back to Proto-Tocharian **kenek*, cannot be of Iranian origin, simply because there is no plausible Iranian form from which they could derive. Even an Old Steppe Iranian **kanaka-* should have yielded a different result, namely, Proto-Tocharian **kenke*. We could think of a consonant stem **kenek-* or **kanak-*, but there is no support for it among Iranian languages. Instead, the word is more likely to come from the Central Asian BMAC language, the vowels of which were rendered in Indo-Iranian as **a* and in Tocharian as **e* and **ə*; which also possessed a suffix **-ka-*, and which seems to have had second syllable stress. These features

12 Carling (2005: 55) writes that this word is "obviously a Turkish borrowing". I cannot see how this is obvious, especially since TB *kenek* and TA *kanak* do not have a Turkish structure or appearance, and, in my view, no certain Turkish words in Proto-Tocharian have been recorded to this day (the best candidate would be PT **kaun* 'sun, day', TB *kauṃ*, TA *koṃ* 'id.', but even this case is controversial, cf. Lubotsky & Starostin 2003: 257 f.).

together converge towards a possible borrowing from BMAC **kanáka-* 'hemp' into Proto-Tocharian at first as **keneke*, which became **kenek* through apocope.

3.2.5 *TB* kercapo *'Donkey'*

The etymology of Tocharian B *kercapo* 'ass, donkey' (Archaic TB *kercc̈apo*, in THT 343 b7), also *kercapañ̈e* 'pertaining to a donkey', *kercapiśke* 'PN' (with a name-forming affectionate diminutive) is not evident. This word is mentioned in multiple works, but I will only cite two extensively, Winter (1971) and Adams (DTB), because they represent two important and common views concerning this Tocharian B word.

Winter (1971: 222) writes:

> Urverwandtschaft mit aind. *gardabha-* ist ausgeschlossen; das *-c-* ebenso wie das *-a-* der tocharischen Form würden unerklärt bleiben. Gleichermaßen unmöglich ist die Annahme einer direkten Entlehnung aus dem Indischen: weder *-e-* noch *-c-* noch *-o-* wären begründbar. Anderseits kann B *kercapo* kaum auf ein echtiranisches Wort zurückgehen: aind. *gardabha-* scheint keine iranischen Entsprechungen zu haben. [...]

He further proposes that Sanskrit *gardabha-* was borrowed into an Iranian language, which should also be the source language of Tocharian B *eñcuwo* 'iron', and possibly also Tocharian B *witsako* 'root'. He suggests that this Iranian language is close to Ossetic ("scheint dem heutigen Ossetisch sehr nahezustehen").

Adams (DTB: 210) adds:

> [r]eflecting a PTch **kercäpā-*, which, except for the stem class, is the exact equivalent of Sanskrit *gardabhá-* (m.) 'donkey, ass' (< **gordebho-*) with the same **-bʰo-* which appears in other Indo-European designations of animals [...]. If, as has so often been suggested [...], *kercapo* is a borrowing from Indic *gardabhá-*, the borrowing must be very early, before the merger of the non-high vowels in Indo-Iranian (otherwise we would expect **kertepo* or the like with the first and second vowels alike and no palatalization). [...] Anreiter (apud Thomas, 1985b: 134) suggests that the Indic and Tocharian words are both borrowed from some third (and unknown) language.

Winter's argumentation is weakened by the fact that no cognate of Vedic *gardabhá-* has been found in Iranian, as he himself admits, and the idea of this borrowing is completely *ad hoc*. Besides, it is very unlikely that such a language

would have had *-č-* where Indic has *-d-* (cf. Pinault 2008: 394). Further, I believe that it has been demonstrated (s.v. *eñcuwo*; cf. Peyrot, Dragoni & Bernard 2022) that *eñcuwo* is a pre-Khotanese loanword into Tocharian. Therefore, Winter's hypothesis is no longer acceptable. As to Tocharian B *witsako*, s.v. Adams's interpretation is flawed as well, because of the simple fact that no other borrowing from Pre-Proto-Indo-Iranian, before the vocalic merger of **e* and **o* into **a*, into Proto-Tocharian is known so far. In Adams (2017: 1368) we also find the supplementary proposition that Tocharian B *kercapo* and Vedic *gardabhá-* are both inherited from Indo-European. No other cognate can be added to this comparison, and it is very difficult to see how it would have functioned formally (cf. Pinault 2008: 394).

Anreiter's proposal is more enticing. Tocharian B *kercapo* cannot derive from Sanskrit or any Indic language, for the reasons evoked by Winter, namely: the vocalism and consonantism do not match. They can hardly both be inherited, as no other cognates are found, and, more importantly, as the phonetics do not match either. Furthermore, *gardabhá-* is not necessarily analyzed as **garda-* + *-bha-* (from PIE **-bʰo-*), although one can suppose that the element **garda-* is due to a secondary recharacterization (see below). Below, I will investigate Anreiter's proposal, and try to propose a scenario for the borrowing.

An obvious candidate for Anreiter's third language is the BMAC language, which counts a number of words "borrowed independently by Common Tocharian and Indo-Iranian in the late Bronze Age", to cite Pinault (2006: 170).

However, precisely on this point, Pinault (2008: 392 f.) has a different opinion. He believes TB *kercapo* to be related to Tocharian B *koro**, which he translates as 'mulet'. Pinault takes *koro** as deriving from Old Steppe Iranian **xara-*. According to Pinault, the expected pre-form **kere* 'mulet' would have taken the *-o* ending, in analogy with words such as *okso* 'oxen', and, as Pinault suggests, would have undergone umlaut **kero* > *koro*. I do not know of any parallel to such an analogy. To note, only Old Steppe Iranian *ā*-stems, not *a*-stems, were made into *o*-stems in Tocharian (see previous chapter) and no Old Steppe Iranian word shows umlaut of **e* to **o* in Tocharian B.[13] According to Peyrot (2016: 328), there is no *o*-umlaut of Tocharian B *e* at all. Pinault's etymology of *koro** can thus not be accepted.

Pinault goes on to propose that the element *ker-* in *kercapo* also derives from **xara-*. Since donkeys are "shouting" animals, he suggests that *gardabhá-* was

13 Witzel (1999: 54) has also tried to connect *kercapo* "*ker-ca-po*" to **khara- ~ xara-* 'donkey', and mentions the Proto-Dravidian form **garda-* 'donkey'.

reanalyzed within Sanskrit as being related to the root *gard-* 'to shout'. The Tocharians, according to Pinault, would have taken this as an example and built the word *kercapo* on the basis of the Sanskrit form: he thus sees in *capo* a reflex of the root *tap-* 'proclamer, annoncer à haute voix', which has palatalized allomorphs.

This etymology leaves us with too many difficulties: if Tocharian B had a form **kere*, borrowed from Old Steppe Iranian **xara-*, why would it not have made a compound word ***kerecäpo* instead? The form *koro**, can certainly not be Old Steppe Iranian in origin, for the reasons evoked above. If *koro** was used, we would expect ***kor*(*o*)*cäpo* instead. An Iranian etymology for *koro** can perhaps be suggested: either a word related to Sogdian *γwr* 'wild ass', or to PIr. **xara-* 'donkey', through Khotanese *khara-* 'donkey': acc.sg. *kharu* → TB *koro**. If the meaning of *koro** was 'camel' as tentatively suggested by Adams (DTB: 218), then these Iranian etymologies are impossible.

A more important argument against Pinault's proposal, perhaps, is the fact that the very existence of the root **gard-* 'to shout' is doubtful (cf. EWAia[1]: 493). As professor Lubotsky informs me, the traditional etymology of Vedic *rā́sabha-* 'ass; donkey' from *ras-* 'to shout' is also very doubtful: the root *ras-* occurs very rarely, and the form *rā́sabha-* implies a form **rā́sa-* 'screamer', with wrong accentuation (instead of expected **rāsá-*).

I propose to see in Tocharian B *kercapo* 'donkey, ass' and Vedic *gardabhá-* 'idem' two different reflexes of a BMAC word: possibly the *-d-* in Sanskrit and the *-c-* in Tocharian go back to a single phoneme, perhaps a palatalized *d*, or perhaps they both go back to BMAC **gardeba-* or **gardepa-*, with inner-Tocharian palatalization.

The ending *-bhá* of Sanskrit can be analogical after multiple other animal names. It is also possible that it reflects the pronunciation of a BMAC phoneme, rendered as *p* in Tocharian. A more likely solution is that the original BMAC word was something like **gardyapa-* or rather **gardepa-*, with the BMAC *-pa* suffix, cf. Lubotsky (2001: 305). The suffix *-pa* was then replaced in Indic by the common animal names suffix *-bhá* found in, for example, Vedic *vṛ̥ṣabhá-* 'bull'.

As to the final *-o* of the Tocharian form, it is quite commonly found in substratum words: *mewiyo* 'tiger', *peñiyo* 'splendor', *witsako* 'root', and is most probably due to them being remade according to the morphological classes of Tocharian nouns.

It should be noted that Carling (2005: 54) also proposed a BMAC origin for both *gardabhá-* and *kercapo*, although this proposal of hers has been virtually never cited by scholarship. Her arguments rest on the difficult correspondence between the Tocharian and Sanskrit vowels, and on the fact that the Sanskrit

word has no Indo-European cognate. At the time she published her article, it had not yet been completely made clear that there exists no likely Indo-Iranian borrowing in Proto-Tocharian.

Professor Lubotsky has also suggested to me a possible connection of the BMAC etymon **gardepa-* 'donkey, ass' to BMAC **gṛda-* 'penis' (Lubotsky 2001: 307), certainly due to the fantastic size of donkey male instruments.

In conclusion, Tocharian B *kercapo* 'donkey, ass' seems to be related to Vedic *gardabhá-*, of identical meaning. There is no consensus as to the etymology of the Tocharian B form. Following Anreiter, I propose that both the Sanskrit and the Tocharian words derive, with different phonetic adaptations, from a common substratal (BMAC) source.

3.2.6 *TB* kroṅkśe *TA* kroṅśe *'Bee'*

For the Tocharian B word *kroṅkśe* 'bee' (with variants such as *krokśe* and *kroṅśe*), borrowed into Tocharian A as *kroṅśe*, several different etymologies have been proposed. We will examine them below.

First, one can cite Isebaert (1980: 148) who proposed to explain the Tocharian word as related to Sanskrit *krauñca-* 'crane'. He saw it as a contamination of **kroṅke* and **kroñśe*, from Prakrit **krŏṅ* and **krŏñca-*, hypothetical forms which would be, according to Isebaert, related to "Sanskrit *kruṅ*" and *kruñca-* (respectively). As to the semantics, it is according to him related to an onomatopoeic root "**kruñc-*, *kūj-*" 'agreeable to listen to' (which Isebaert cites from Thieme 1974: 295). This is ultimately the etymology I will follow, although differing in the detail.

Later, Hilmarsson proposed (1986: 34 f.) to connect the Tocharian word *kroṅkśe* to the Germanic word for 'honey', Old Icelandic *hunang*, OHG *honang* < **hunanga-* (cf. Kroonen 2013: 255–256, who connects Greek κνῆκος f. 'safflower' to the Germanic root, cf. Beekes 2010: 722–723). In Germanic we find an alternation between **hunanga-* (OHG *honang*, Dutch *honing*, etc.) and **hunaga-* (e.g. English *honey*, the Finnish loanword from Germanic *hunaja*).

Hilmarsson (1986: 35–37) further proposes to connect the Germanic and Tocharian words to Sanskrit *kāñcaná-* 'gold, golden' and Lat. *canicae* 'bran', among others. As Hilmarsson reminds us (1986: 35), the second nasal in the Germanic words has often been viewed as secondary by previous scholarship. The variant of the type **hunaga-* was often considered primary. The reason evoked by scholarship was generally that the common Germanic suffixes *-ing-* and *-ung-* influenced **hunaga-* and that it thus produced **hunanga-*. Hilmarsson objects to this view that the suffix *-aga-* was productive in Proto-Germanic, so that the forms of the type *hunanga-* need to be original. Hilmarsson thus convincingly argues that no influence from the suffixes *-ing-* and *-ung-* has to be assumed for the Proto-Germanic word for 'honey'.

Hilmarsson further postulates a Proto-Indo-European form **kn̥Honko/e-* (1986: 36) which would be the source of the Germanic root **hunanga-*, of Sanskrit *kāñcaná-* and of Pre-Proto-Tocharian **knonko-* > Proto-Tocharian **krœnko* > **kronko*. Finally, "the *ōn*-suffix was apophonic, its *e*-grade causing palatalization of preceding susceptible consonants. Through generalization of the *o*-vocalism and the palatal *-ś-* the attested Tocharian forms were reached." Hilmarsson also explains the retention of *-k-* as ("perhaps") a reflex of the velar pronunciation of the *-ṅ-*.

Hilmarsson's arguments concerning the etymology of *kroṅkśe* do not seem very compelling to me. First, it is difficult to believe that both generalization of the palatalization due to the *e*-grade of the *-ōn* suffix and generalization of the *o*-vocalism took place. If this word were really suffixed with *-ōn* we would not expect a final *-e* but a final *-o* (as demonstrated notably by Hilmarsson 1987). The retention of *-k-* as a reflex of the velar pronunciation of the *-ṅ-* seems completely *ad hoc* and without parallel. How could the **n* have remained velar after the palatalization of the *k*?

I therefore have to agree with Adams (DTB: 235) that Hilmarsson's demonstration requires "some very complicated phonological changes in Tocharian". Besides assuming an *ad hoc* dissimilation of *n* to *r* in **knonko-* > **krœnko*, **krœnko* (**krenko* in our notation) should have become †*kreṅke* in TB. The retention of *-k-* as a reflex of the velar pronunciation of the *-ṅ-* before a *ś* is completely unprecedented in Tocharian.

I also disagree about the semantic connections made here: firstly, the connection with Sanskrit *kāñcaná-* 'gold, golden' is doubtful, as honey is not always yellow. To support it there needs to exist a certain Indic phraseology connecting honey with gold, or with a golden colour. No such parallel is adduced by Hilmarsson.

Now, bees too can be yellow, at least partially, but they would not be called 'the yellow ones' (perhaps, at the most 'the yellow insects'). There does not seem to be another parallel, at least not among Indo-European languages (cf. Buck 1949: 192).

Hilmarsson is right in that bees are often lexically associated with honey. He cites English *honey-bee* and Icelandic *hunangsfluga* as examples. One can also adduce the Chinese example *mìfēng* 蜜蜂 [honey-bee] 'bee', and Greek μέλισσα 'bee' < *μέλιτ-i̯α 'the one of the honey'. It should nevertheless be noted that compounds like English *honey-bee* are only needed if in that language the word for bee is also used to designate related insects that do not produce honey. English, for instance, calls a number of insects *bee*, such as the bumblebee (French *bourdon*). In French, for example, *abeille* only designates honey producing insects, e.g. the definition from the Nouveau Petit Robert (2007: 4)

"abeille, n.f. [...] Insecte social hyménoptère (*apidés*) dit *mouche à miel* vivant en colonie et produisant la cire et le miel [...]."

As a French speaker, I would thus find it absurd to add the word *miel* 'honey' to *abeille* 'bee'. Some dialects that do not have the word *abeille* call them *mouche à miel* 'honey fly' (e.g. Bourbonnais [muʃ a mjɛl]), identical to the Icelandic compound evoked by Hilmarsson. As Peyrot and Meier (2017: 11) write: "although 'bee' and 'honey' are obviously contingent semantically, a direct change in meaning from the one to the other is not at all trivial." According to all these observations Hilmarsson's etymology of the Tocharian word for 'bee' needs to be abandoned, both on formal and semantic grounds.

In my opinion, Isebaert was partially correct in seeking a connection to Sanskrit *krauñca-* 'crane'. In Bernard (2020: 33 f.) I proposed, on the basis of a wide range of Indic and Iranian comparanda, to reconstruct a substratum root **kraṷ-* alternating with **kru-*, variously suffixed in Indo-Iranian to designate loud animals. The original sense of all suffixed forms seems to be 'loud animal', since that is the smallest common denominator of all the species of animals derivatives of this root designate. One of these suffixed forms contains the suffix *-ñc-* added both to the base **kraṷ-* and to the base **kru-*. I also suggested that this suffix is of substratal origin (2020: 34). This suffix can be seen in Skt. *krauñca-* and *kruñc-* 'crane'.

It is known that Tocharians were very familiar with honey and bees, since Chinese borrowed its word for 'honey' from Tocharian (see for instance Behr 2001: 359). I can see two reasons for calling bees the 'loud animal' *par excellence*. The first one is evident: groups of bees make a very loud continuous noise. This noise can be deafening, especially if one gets too close to a beehive. The Hebrew word דבורה (*dḇōrā*) 'bee' is probably related to the verb 'to speak' (root *d-b-r*).[14]

The second possibility is that the word initially designated the fly, the sound of which is not even necessary to describe. The semantic change 'fly' > 'bee' is attested. For example in most Normand dialects *mūk* (< Lat. *musca* 'fly') is used to designate bees. It is nevertheless unclear to me whether this meaning of *mūk* is likely to be back-formed on the compound *mūk a myèl* 'honey fly', also attested in Normandy. This weakens very much the latter hypothesis. Another possible example is Proto-Iranian **makš-* 'fly', borrowed into Proto-Finno-Ugric as **mekše* 'bee' (cf. Holopainen 2019: 139 f.; van Sluis 2022: 5).

14 I thank Dr. Benjamin Suchard for informing me of this parallel. If this word is a cognate of Aramaic *zibbūr* 'hornet', Classical Syriac *debbōrā* 'wasp, hornet' and Arabic *dabbūr* ~ *zanbūr* 'id.' (all from Proto-Semitic **ᵈzanbūr-*), then it was perhaps remotivated on the basis of the root *d-b-r* 'to speak'.

The former hypothesis somewhat finds an echo (although chronologically much later) in Buddhist phraseology, even in Tocharian: one can think of THT 571b4 *krokśäṃts weśeññа māka* "the voice of many bees". In my opinion PT **kroṅkśe* designated the swarm of bees initially, and later came to designate the bee in general, and the individual bee in particular. Such a semantic development is attested, for example in Romanian *albină* 'bee' < late Lat. *aluīna* 'beehive' (cf. Buck 1949: 192).

The Tocharian combination of *ṅk* and *ś* occurs in this word only. Furthermore, only one other Tocharian word contains the combination *k* + *ś*: *lyekśiye* 'millet', and it has no known etymology, cf. Peyrot (2018: 245). This naturally suggests a borrowing, rather than inheritance. I believe that in TAB *kroṅ*(*k*)*śe* the cluster *-ṅ*(*k*)*ś-* (which is variously written *ṅkś* ~ *ṅś* ~ *kś*) is a (Proto-)Tocharian rendering of the "substratal" sound which is spelled ‹ñc› in Sanskrit.

If one accepts this etymology which, as with all these non-Indo-European matters, cannot reach a high level of certainty, then Tocharian B *kroṅkśe* must go back to a proto-form **krVÑCa-* 'loud animal'. I write *ÑC* in capital letters to indicate that sound which is rendered as *-ñc-* in Sanskrit and ⟨*ṅ*(*k*)*ś*⟩ in Tocharian. One can evoke Armenian *kṙunk* 'crane', also of difficult etymological derivation (Martirosyan 2010: 377), however Armenian *k-* points to **g-* and this word must thus be left out of the discussion.

As to the *-o-* corresponding to *-u-* in the Sanskrit form, it is not an expected correspondence. Nevertheless, since we do not know the original value of the "substratal" vowel, it is possible that it was neither **o* nor **u* but a sound in between. Furthermore, perhaps the alternation *krauñca-* ~ *kruñc-* also found as **kraṷ-* ~ **kru-* in Iranian […] could reveal different strategies to adapt that original substratal phoneme.

Since the etymology of *krauñca-* and of *kroṅkśe* relies, I believe, on the notion that these were noisy animals, one could think that they were independently built on an onomatopoeia "*kronk*" or the like. This onomatopoeia, however, is very far away from the type of noise made by cranes or bees, and rather reminds us of a falling or breaking object.

As a conclusion, I wish to underline that, if the Tocharian and Indic forms are borrowed from a non-Indo-European language, it is possibly a different language than the one which yielded most other words of this list. The structure is disyllabic or perhaps monosyllabic with a thematic vowel, cf. Sanskrit *krauñca-* ~ *kruñc-* and it is quite different from that of other words presented and discussed here, which are mostly trisyllabic with second (middle) syllable stress, and usually contain a schwa. It is also unusual to see this word reflected in the Indic branch and apparently in Tocharian, without any reflex in Iranian. As far as I know, there is indeed no Iranian correspondent to the Sanskrit substratal

-ñc- suffix. However, other reflexes of the root **kraṷ-* are found in Iranian languages (cf. Bernard 2020: 31f.) and a BMAC origin cannot be excluded for all these words.

3.2.7 *TA* lepäś* *'Jackal'*

This word is only attested in the gen.pl. *lepśäśśi*. Malzahn (2014: 92–93) writes that "[…] *lepś* does not have the ring of an inherited word, and the jackal is *not* native to Central Asia". Although it is true that this word does not have the ring of an Indo-European inherited word, it is not scientifically correct that the jackal is not native to Central Asia (the golden jackal for instance is present in Central Asia). She further suggests that TA *lepäś** 'jackal' was a loan from an unattested Tocharian B word (that, for the sake of the argument, would have to be set up as **lepśe vel sim.*), itself from Sanskrit *lopaśa-* 'a kind of fox or jackal' *vel sim.* but that it was borrowed into (pre-)PT early enough to get its vowel *-o-* treated in the same way as PIE **-o-*. This is hardly possible, as for instance no Sanskrit vowel **o* can be shown to have been treated in Tocharian as PIE **-o-*, no matter how early the word was borrowed.

Palmér, Jakob, Thorsø & al. (2021) suggest that the Indo-Iranian word for 'fox' is a cognate of, for instance, Greek ἀλώπηξ 'fox', and make it go back to **h₂lop-eḱ-*. Nonetheless, there are many issues concerning the vocalism of this etymology, notably in the root and in the suffix. The fact that the word for 'fox' shows the same vocalic shortenings in Iranian languages as do a number of other borrowed animal names (and no inherited name) rather suggests that this is a borrowed word.[15]

It would be more plausible to consider TA *lepäś** as a direct BMAC loanword. If *lepäś** goes back to Proto-Tocharian, it points to a pre-form **laipəśe*, which would correspond to the general structure of the "substratum" words as described above. One can further speculate that **laipəśe* was a distant cognate of Sanskrit *lopaśa-* 'fox, jackal' < PIIr. **r(/l)aṷpaća-*, itself a BMAC loanword, with a different initial diphthong. If this is correct, the element **äśe* could be

15 Some of the issues found in their paper include the fact that if, as they claim, the suffix **-āća-* spread from the word for 'fox' to the other animal words, it needs to have spread after the thematicization to other words, when every individual language had generalized a short or long variant of the suffix: this poses a problem, given that not every language has a short or long variant for every *āća*-word (cf. Bernard 2020: 37f.). Palmér, Jakob, Thorsø & al. (2021: 241) write that "[a]s the suffix **-ăća-* was unique, **(H)rāpăća-* may have been analysed as containing the suffix **-ća-* '-like' (cf. Skt. *yuvaśa-* 'youthful') by the speakers of Proto-Indo-Iranian." The suffix *-aśa-* 'like' is absent from Iranian, and seems to be an Indian, if not Sanskrit, innovation. In any case, they do not bring forward any argument in favor of its antiquity, which should be necessary for such a claim.

interpreted as a reflex of the BMAC suffix *-*aća*-, which is often found with words denoting "dangerous animals" (on which see Bernard 2020: 38 f.).[16]

3.2.8 *TB* witsako *'Root'*

The Tocharian B feminine noun *witsako* 'root' is of great importance. For a long time it has been claimed to be of Iranian, and generally more precisely of Old Iranian origin. The first scholars to suggest an Iranian origin for *witsako* were Karl Bouda (apud Krause 1971: 37) and Winter (1971: 222), and they were followed by many others.[17] However, with the establishment of the features of Old Iranian loans in the second chapter of the present book, such claims should be thoroughly reviewed and reconsidered. This is what I intend to do below.

Indeed, the *communis opinio* about the Tocharian B word *witsako* 'root' is that it is a borrowing from an Iranian form, either from a (pre-)Proto-Ossetic preform of Ossetic Iron *widag* 'id.', or from a form closely related to it. The first to propose this was Winter (1971: 222); it was then accepted by most if not all scholars (see for example Hilmarsson 1986: 227; Abaev 1989: 106; Kim 1999: 124; Adams DTB: 658).

The Ossetic word *widag*, and its Scythian cognate Βιδάχης, have been connected before that to Av. *vaēⁱti-* 'willow, willow-twig' (cf. NP *bēd* 'willow') by Abaev (1949: 186), and every other scholar has since repeated this. It seems that the Avestan word for 'willow' should rather be seen as a cognate of Ossetic Iron *xæris*, Digoron *xærwes* 'willow', in my view from **xara-* 'donkey' and Iron *wis*, Digoron *wes* 'rod; copse, brushwood', which Morgenstierne (1938: 264; 1942: 269) derives from **u̯aitsa-*. Donkeys enjoy stripping the bark from willows, and they also enjoy eating willow in general.[18] Cheung (2002: 51), however, connects the Ossetic words for 'willow' to *xæræ* 'grey', which is possible, as some willows are indeed grey, although my own etymology is also possible.

To explain the Tocharian B sequence *-ts-* in this word, it is mostly assumed that the Tocharian form derives in some way from an Old Iranian **u̯aitikā-* or **u̯aitiakā* or the like, although it is also usually noted that these forms would not straightforwardly yield *witsako* (e.g. Isebaert 1980: 97, who suggests that TB *witsako* is a borrowing from its unattested TA cognate).[19] Some scholars, such

16 In Bernard (2020: 38) I proposed that **raṷpaća-* (as opposed to the variant **raṷpāća-*) derived from the substratal word **raṷpi-* suffixed with *-*aća*-. Perhaps **laiṗi-ća-* or the like would have been borrowed as PT **laiṗaśe* (see also § 3.4.3 in the present book).

17 Although this idea is not mentioned as originating from Bouda in the scholarly literature.

18 See for example https://www.thedonkeysanctuary.org.uk/for-owners/owners-resources/safe-trees-and-shrubs-for-donkeys (consulted on the 8th of March 2025).

19 So far, there are no commonly accepted Tocharian A loanwords into Tocharian B, which makes this hypothesis very unlikely.

as Winter (1971: 222), Tremblay (2005: 426) and Adams (DTB: 657–658), propose that the Tocharian B form derives directly from a Middle Iranian **wīδikā-*, itself derived from **u̯aitikā-*.

Kim (1999: 124–126) proposes that the Proto-Ossetic predecessor of Iron *widag*, Digoron *wedagæ*, which he reconstructs as **wędəga*, was borrowed into Tocharian before the assibilation of *d* to Proto-Tocharian **ts*. It would subsequently have undergone devoicing from **dz* to **ts*. At the same time, he derives this Proto-Ossetic **d* from Proto-Iranian **t*. While somewhat ingenious, his reasoning does not work. Proto-Ossetic **d* would not be reflected by a **ts* in Proto-Tocharian. For one thing, Old Steppe Iranian **d* corresponds to Proto-Tocharian **t*, for example in Old Steppe Iranian **pari-banda-* → Proto-Tocharian **perpente* 'burden' (s.v. TB *perpente*) or in Old Steppe Iranian **spaldaka-* → Proto-Tocharian **speltke* > TB *spel*(*t*)*ke*, TA *spaltäk* 'zeal' (s.v.). It never corresponds to Proto-Tocharian **ts*. It would thus be very difficult to imagine how Proto-Ossetic **d*, which would probably be later than Old Steppe Iranian, could yield Proto-Tocharian **ts*.

I believe it is now necessary to briefly discuss the etymology of the Ossetic words and related Iranian words. Iron *widag*, Digor *wedagæ* point to Proto-Ossetic **wedagæ-*. These Ossetic words are cognate with Sogdian *wyt'k* 'string' and Yaghnobi *wita* 'cord' (Cheung 2002: 242), but also with Pashto *wuláy* f. 'root, root-fibre', albeit Cheung (2017: 42) cites it as a possible Ossetic loanword into prehistoric Pashto. In any case, the forms cited above point to a reconstructed form **u̯aitāka-*, which is incompatible with Tocharian B *witsako* (/witsə́ko/, see below) because this rather points to a short middle vowel.[20]

These phonetic details concerning the vocalism and the consonantism of *witsako* disagree with the theory that TB *witsako* 'root' is borrowed from Iranian. Besides, I believe that the Iranian forms cited above (Iron *widag*, Digoron *wedagæ* and their Sogdian, Yaghnobi and Pashto cognates) need to be separated from Avestan *vaēʿti-* 'willow, willow-twig' and New Persian *bēd* 'willow', due to their semantic distance: a willow has roots, and roots can be willow-roots, but the semantic proximity does not get any closer.[21]

---

20 An interesting form is provided by Wanjī *wisk-* < **u̯aitika-* in the toponym *Wiskroγ* 'grape vine(yard?)' (cf. Lashkarbekov 2008: 76), with *-roγ* being a borrowing from Tajik *roγ* 'field'. Naturally, this cannot be a descendant of the donor form of Tocharian B *witsako*, for formal reasons (the intermediary form between **u̯aitika-* and *wisk-* is reconstructed as **viϑ-k* by Lashkarbekov) as well as semantic ones. On the other hand, grapevine is quite different from willows, and perhaps resembles roots, in the way that it develops and grows. This topic is quite intricate.

21 Although one could argue that the original meaning is 'string, cord' (as in Sogdian) which is somewhat closer.

Furthermore, going back to phonetics, I have established that Old Steppe Iranian *-*aḭ*- is reflected in Proto-Tocharian as *-*ey*- (s.v. *waipecce*), preserved in Archaic Tocharian B as *ey* (TB *eynāke*, *meyyā*), and reflected as Tocharian A *e* and classical Tocharian B /ay/. Old Steppe Iranian **-*u̯aitāka*- would thus become Tocharian B **-*waitke*, Tocharian A **-*wetäk*, or perhaps Tocharian B **-*waitake* /wáytake/, Tocharian A **-*wetak*. In any case, it would not have become Tocharian B *witsako*.

Moreover, as is well known, the Tocharian word is actually /witsóko/, as can be seen, for example, in PK AS 9Ba6 *witsäko*. As I hope to have shown, Old Steppe Iranian had fixed initial accent, with a few exceptions (see section 2.7.2.7). This accent was reflected in Proto-Tocharian. If one wishes to make the word "Proto-Ossetic" or the like, this does not solve the problem, as Ossetic also had fixed initial accent (Cheung 2002: 123). The Tocharian B word *witsako*, with its middle syllable stress, therefore does not seem to be of Old Steppe Iranian or Ossetic origin.

A Middle Iranian language could not serve as a better source, since it would imply both the preservation of initial *-*u̯*-, which excludes Khotanese, and a sound change *-*ē* > *-*ī*, which would exclude all known Middle Iranian languages but Khotanese and Tumshuqese. However, most importantly, it would not be of any use to explain the aberrant Tocharian *-ts*- corresponding to Old Iranian *-*t*- and Middle Iranian *-*d*-. This is because we know that Proto-Tocharian *-*ts* had already arisen when Tocharian came in contact with Old and Middle Iranian languages, and Tocharian simply devoiced Iranian *-*d* to *-*t* in all cases, while it reflects Old Steppe Iranian *-*tˢ* and *-*dᶻ* (< PIr. *-*ć* and *-*ʃ*) as *-*ts*. All the Iranian words that are close in meaning and form to our words for 'root', for instance Pashto *wuláy* 'root, root-fibre', go back to a *-*d* or *-*t*.

One could think, however, that the etymon *-*u̯aitsa*-, suffixed with *-ka*-, could have yielded the relevant Tocharian form. It has yielded, as mentioned above, Ossetic Iron *wis*, Digor *wes*, *yes* 'rod; copse, brushwood' (cf. Morgenstierne 1942: 269). These Ossetic words are, however, far removed semantically, and this would also imply that the *-*sa*- in *-*saka*- would first become a schwa in that given language. However, the stress would then remain on the first syllable since it is difficult for an *-*a* vowel in Iranian to both be weakened to schwa *and* take the accent. Both obstacles, coupled with the fact that the Ossetic words are quite isolated within Iranian, render this etymology for Tocharian B *witsako* quite unlikely.

The Iranian words cited above with the meanings 'root' or 'string, cord' have not yet found a satisfying Indo-European etymology, and lack identifiable Indo-European cognates. It thus seems that they are potentially borrowings

themselves. One could think of a Pre-Khotanese borrowing, but the expected cognate is not attested in Khotanese or in Tumshuqese, and furthermore, the accent would also be a problem as Tocharian words of Khotanese origin usually show initial accent (Dragoni 2023).

No known language can be mentioned as a potential donor, but in this particular case, the BMAC language could be evoked. Some Indo-Iranian substratum words designate flora, for example **u̯ŕćša-* 'tree' (Lubotsky 2001: 313). An interesting point here is that the Iranian forms mentioned above all point to a shared proto-form **u̯aitāka-*, as if the *-ka-* was part of the root itself. This suffix *-ka-* cannot be understood synchronically at any stage (as far as our knowledge goes) as an addition to an otherwise known word **u̯aitā-*, mostly because such a noun does not otherwise exist in Iranian. This makes it plausible that it is the BMAC suffix **-ka-* described by Lubotsky (2001: 304).

It is also possible that both the Iranian and Tocharian words derive from the BMAC language. Other such examples can be found, as Pinault (2003; 2006) proposed, corroborated by further examples (s.v. *kercapo*, *kroṅkśe*). In the case I am right and it is a BMAC loanword found both in Iranian and in Tocharian, one can think that, in the same way that BMAC **-dy-* became Proto-Tocharian **-c-* in TB *kercapo* 'donkey, ass' (s.v.), BMAC **-tV-* could have become Proto-Tocharian **-ts-*, for instance if the following vowel was **-ē-*. Alternatively, perhaps the form was originally **u̯aiṭiṭāka-* or the like, and **-tį̄-* became Proto-Tocharian **-ts-* while **-aiṭiṭā-* was dissimilated to **-aiṭā-* in Iranian. What is safer to assume is that the BMAC source word was adapted differently in Iranian and Tocharian.

In conclusion, after having shown the difficulties with the traditional etymologies of TB *witsako* 'root', which make it impossible to derive it from an Iranian language, I have presented a new possible etymology of this word. I suggest it is a BMAC word, itself connected with the Iranian forms Iron *widag*, Digor *wedagæ* 'root', Pashto *wuláy* 'root, root-fiber', Sogdian *wyt'k* 'string' and Yaghnobi *wita*, and thus, ultimately, a cognate of these Iranian forms that previous scholars wanted to derive it from.

I am aware of the fact that my BMAC derivation is tentative. Although it seems a good option to me, I should stress that if this derivation is not accepted, my rejection of the traditional derivation of the Tocharian word directly from Iranian still holds true. The supposed correspondences contradict the secure patterns of the adaptation of Old Steppe Iranian words in Tocharian, and no plausible Iranian candidate for the borrowing can be found.

3.2.9 *TB* śerwe, *TA* śaru *'Hunter'*

Tocharian B *śerwe* 'hunter' (adj. *śerwäññe*) and Tocharian A *śaru* 'idem' point to a Proto-Tocharian reconstructed form **śerwe* or **śerəwe*. From this noun, a

verb **śerw-* was built, with a subjunctive stem **śerwəy^{a}/$_{e}$* (cf. Peyrot 2013: 826), as found in the Tocharian B infinitive *śerītsi* 'to hunt' (cf. Pinault 2008: 588 f.).

Pinault (2006: 179–181) connects these words to a BMAC word reconstructed as **ćarṷa* by Lubotsky (2001: 310). Lubotsky does not give a particular meaning to this word, but assumes it is the name of the deity from which Vedic *Śarvá-* and Avestan *Sauruua-* 'name of a daēva' come. Pinault suggests an original meaning 'hunting, living in the forest', and notes that *Śarva-* is depicted as an archer in the Vedas.

Further cognates adduced by Pinault are Ossetic Digoron *sorun/surd*, Iron *suryn/syrd* 'to track, hunt' and Khotanese *hasura-* 'quarry, hunted beast'. This is accepted by Cheung (EDIV: 338) who posits a Proto-Iranian verb **sarṷ-* 'to hunt' and adds the Yazgholami verb *sard/sar-* 'to track, hunt' as an additional cognate.

Adams (DTB: 695) prefers to connect Tocharian B *śerwe*, Tocharian A *śaru* to Proto-Indo-European **ǵ^{h}ṷēr-* 'wild animal' with a secondary suffix **-ṷo-*. However, as Pinault (2006: 179) points out, the function of the suffix is problematic.

Another explanation, based on **ḱērṷ-o-* and a connection to the Proto-Indo-European word for 'horn', is mentioned (with relevant literature) by Pinault (2006: 179–180). This proposal is also found again in Jasanoff (2017: 79). The idea is that the word for horn served to make an animal name 'stag' and from this animal name the name of the 'hunter' was produced: or, as Jasanoff (op. cit.) writes "lit[erally] 'stag man'".

This idea is rightly rejected by Pinault (2006: 179–180) for obvious semantic reasons (Indo-Europeans did not hunt stags more than any other animals; as Pinault writes: "hunting was never limited to stags or horned animals, especially in Central Asia").

Pinault's explanation makes much more sense than any Proto-Indo-European connection made until now. He suggests a semantic path in Indo-Iranian from **'hunting, hunter' to **'living in the wild' > 'god of the wild, killer of living beings'. It is not necessary for this semantic development to have occurred within Iranian, as the source of Tocharian loanwords from BMAC is probably not the exact same as for Indo-Iranian, as I argue further below. The theonymic aspect of this word could have already arisen in the BMAC source language of Indo-Iranian, or, possibly, it both meant 'hunter' and designated the god of hunters, or a hunter god in the source language (both of Indo-Iranian and Tocharian, or only of Indo-Iranian).

In conclusion, I follow Pinault (2006: 179–181) in viewing the origin of Tocharian B *śerwe* 'hunter' and Tocharian A *śaru* 'id.' as BMAC, ultimately related to the theonym Vedic *Śarvá-* and the *Daimonym* Avestan *Sauruua-*. The Proto-Indo-European connections that were proposed for the Tocharian words are

rather weak, and the connection to the Indo-Iranian names are justified (at least for the Indic side) by the identification of *Śarvá-* to a hunting character.

3.2.10 *TB* ṣecake*, TA* śiśäk *'Lion'*

Tocharian B *ṣecake* and its Tocharian A equivalent *śiśäk*, both meaning 'lion', do not have a clear etymology. It is for instance impossible to reconstruct a single prototype for both forms, as Proto-Tocharian **e* does not yield Tocharian A *i*, but *a*, and PT **c* does not yield Tocharian A *ś* but *c* (see for example TB *epiyac*, TA *opyāc* 'memory'). On the basis of the Tocharian B form, one needs to reconstruct Proto-Tocharian **ṣecəke*, but **ṣecəke* would not yield Tocharian A *śiśäk*.

To explain the *ś-* in the Tocharian A form, the influence of TA *śiśri*, for which Sieg (1944: 16) proposed the meaning 'mane' has been evoked (e.g. DTB: 723). This is of course speculative, but not unlikely, if the meaning is correct, as few other animals have manes, and manes are thus often seen as stereotypical of lions. Note, however, that *śiśri* is a hapax and that its meaning could actually be different. Indeed, in the context where it occurs, the lion is lying down and his *śiśri* is "up" (*koc*), which is unexpected. Furthermore, as Pinault notes (2022: 525) the word for the lion's 'mane' is attested in Tocharian A, it is *kesār-śiśäk* (*kesār* is a borrowing from Sanskrit *kesara-* 'mane' and the compound could be a calque of Sanskrit *kesari-siṃha*, cf. Pinault, op. cit. or a compound where the second element simply specifies the meaning of the first one). Pinault (2022: 525) thus suggests a different meaning for *śiśri*, namely, 'whiskers' (also DToA: 471). He also analyzes (op. cit.) the form *śiśri* as a dual and proposes to derive this word to the Proto-Indo-European form **deiḱ-ero-* 'indicateur' < **deiḱ-* 'to show, to indicate'. This new meaning does not preclude the solution suggested above: if cats and other feline animals were not so common among Tocharians, whiskers could have been viewed as typical of lions, and thus **ṣecəke* 'lion' could have become TA *śiśäk* by contamination of TA *śiśri* 'whiskers'. To note, Carling and Pinault rather view *śiśäk* as a noun of Central Asian origin, "probably influenced at some stage by Chinese" (DToA: 471).

Indeed, in lack of an evident Indo-European etymology, TB *ṣecake* and TA *śiśäk* have often been compared to Middle Chinese *srij-tsiX* 獅子, modern Chinese *shīzi* (cf. DTB: 722). They also have been compared to Vedic *siṃhá-* 'lion' and Classical Armenian *inj*, Middle Armenian *unj* 'panther, leopard', (cf. e.g. Witzel 1999: 56; see the literature in DTB: 722 and Blažek & Schwarz 2017: 69). If these forms are related, it must be in a very complex and intricate way.

It has been variously argued that the Tocharian words were borrowed from Chinese, or the other way around. Since lions were originally present in the

West and were only later introduced to the Chinese (cf. Blažek & Schwarz 2017: 69), and there were no lions in China proper until the first centuries of our era at least (cf. Behr 2004: 6). I follow Pulleyblank (1962) in observing that the borrowing direction from Chinese to Tocharian for this word does not make much sense and that the other direction is preferable. One problem for the comparison is the fact that Tocharian *-ke* seems, superficially, to correspond to Chinese *-X*, which is not a consonant, but designates a tone.

As other examples of interesting similarities between Chinese words and Tocharian ones, one can cite TB *ṣitsok* 'millet alcohol', derived from Chinese *shǔ jiǔ* 黍酒 < MC *syoX tsjuwX*, and TB *śakuse* 'brandy', derived from *sù jiǔ* 粟酒 < MC *sjowk tsjuwX* (cf. for instance Peyrot 2019: 144, pace Ching 2010: 387[46]). There are thus other examples of Chinese *-X* corresponding to Tocharian *-k*, although this is obviously a problematic correspondence. Since, for geographical reasons, a borrowing from Chinese into Tocharian is unlikely for this word, it is possible that the Chinese word for 'lion' was borrowed from Tocharian, as suggested earlier in the literature.

It has also been brought forth that TB *ṣecake*, TA *śiśäk* were borrowed from an Old Iranian **šargu-* ~ **šargawa-* 'lion' (e.g. Kümmel 2020: 259). This is impossible, as Old Steppe Iranian **šargu-* would yield Proto-Tocharian **ṣerk*, and **šargawa-* would yield Proto-Tocharian **ṣerkwe*. Furthermore, the reconstruction of this root for Proto-Iranian seems more than doubtful to me (regarding its internal derivation, formal aspect and geographical distribution).

The only segment for which a Proto-Tocharian reconstruction is fitting for this word is the ending in **-əke* (Behr 2005: 10; Pinault 2015: 188). This ending was already observed for *iṣcäke* (s.v.) and I proposed that it was of BMAC origin in *iṣcäke*, in accordance with Lubotsky's discoveries (2001: 304). This might suggest that these words were borrowed from a BMAC language. One can refer to Behr (2005: 12), who saw the Tocharian words as borrowed from another language, and also wrote that Chinese borrowed its word for 'lion' from that same third language.

Given the possibility that Tocharian A *śiśäk* was influenced by *śiśri* 'mane' or 'whiskers', we can theoretically reconstruct a Proto-Tocharian form **ṣecəke* 'lion', a trisyllabic word with middle syllable stress, like most other words in this chapter.

In conclusion, TB *ṣecake* and TA *śiśäk* 'lion' are problematic words, with both internal and external issues. Unless new facts enlighten the situation and help us solve this equation, I do not think a clear etymology can be proposed. However, they are clear borrowings from another language, and their **-əke* suffix seems to suggest the source word was of a BMAC origin.

3.3 Other Possible BMAC Loanwords in Tocharian

In this section I will discuss three more words of possible BMAC origin found in Tocharian, in less detail than the previous word studies. Two of these words have already been suggested as BMAC loanwords in the literature and one, *kanti*, has not.

3.3.1 *TB* kanti *'Naan' or 'Groats'*

The meaning of TB *kanti* is not entirely certain. Thomas (1954: 749; 1957: 136) translates it as 'pastry' (in German: *das Gebäck*). Ching (2010: 263) writes, concerning this word:

> *kanti* is a kind of foodstuff or meal, as revealed in a weaver's story in Buddhist literature (THT 375 b5, cf. TEB (II), Text XV). However, its identification as a kind of 'baken' stuff as suggested by Sieg (1950: 213, cf. TEB (II): 178) hitherto lacks proof.[22]

Adams (DTB: 146) translates it as '± bread'. Indeed, in a great number of occurrences[23] it is used together with *yikṣye* 'flour'.[24] THT 433 a2 has, for instance, "for *kanti*, flour was spent: eight pecks" (adapted from Ching 2010: 276). For the discussion of the meaning of *kanti*, I will mention other passages as well. For clarity, THT 433 is an account document relative to a monastery.

THT 433 a4

ikäṃ-ñune śle trukāle kantiś yikṣye ok tom piś ṣaṅkäṃ

"On the 29th day, for the *kanti* together with allocation, flour: eight pecks, five pints." (adapted from Ching 2010: 276).

---

22 Ogihara (2009: 455) suggests that *kant*[*i*] in PK NS 114 a1 designates a sort of place. That could perhaps fit in that specific context, but certainly not in the other occurrences. To note, TB *kanti* is completely unrelated to TA *kanti* 'gong, cymbal' (cf. Carling 2005: 64; DToA: 93).

23 Kizil WD-II-4 a7ii; THT 433 a4, a6, a11, a16, a24, a26, a28, a31; THT 458 a2; THT 2698.a a3; THT 2802 and 2905 a3; THT 2879 a2; THT 2885 a2.

24 The meaning 'flour' of *yikṣye*, *yäkṣiye* is not entirely assured, but there are good arguments in its favor (see e.g. Peyrot 2018: 256 f.). I also want to add an etymological argument to the lot: Adams (DTB: 535) suggests it could be derived from *yäks-* 'to embrace, to entangle, squeeze (?)', but sees this meaning as "semantically distant from 'grind' or 'press down, crush' we might expect to lie behind 'flour'". This can easily be solved if one thinks of flour paste, of flour being used to prepare virtually any meal: it is entangled together, made together into one single unit of paste, through the addition of ingredients such as eggs, water, etc. The word *yikṣye* would thus originally have designated flour paste.

THT 433 a6

ñweṃne trukāle Sumaiśe wasa kantine yikṣye ok tom piś [*ṣ*]*aṅkäṃ*
"On the last (= 30th) day, Sumaiśe gave (his offerings as) allocation. The flour in *kanti*: eight pecks, five [pints]." (Ching 2010: 276)

THT 433 a14

traiy meñantse ne trukāle Tuṣi wasa kantine Ynaikentāñe Taiṣṣi śwāle wasa
"On the 3rd day of the month, Tuṣi gave (his offerings as) allocation, in (the form of) *kanti*. The Great Commissioner (?) of Ynaikentā gave (his offerings as) food." (Ching 2010: 276).

THT 433 a21

ṣkas meñantse ne sāṅkiś trukālene yikṣye tarya tom kantine
"On the 6th day of the month, in the allocation for the *saṃgha*, flour: three pecks, in *kanti*" (adapted from Ching 2010: 277).

THT 375 b5

[...] *eṣe śanom*(*pa*) *kr*(*e*)*nta śwatsanma kanti täṅktsi ārwer yāmormeṃ ceyna* (*aṣani*)*k*(*a*)*ṃne kaly*(*m*)*isa s*(*p*)*ā* – –
"having prepared good food and even *kanti* with his wife, (he) behaved properly towards those worthy ones." (adapted from CEToM).

From these and further examples, one can deduce that *kanti* was not considered as food (*śwātsi*, plural *śwatsanma*) proper. This does not preclude the identification of *kanti* with bread, as bread is often conceptualized as different from food, mostly because bread can hardly constitute a whole meal, but rather should come together with it. From THT 375 b5 one can also deduce that it is not something that was usually prepared by people in their kitchen: having prepared good food and **even** *kanti*. In our own context, we would readily think of bread, which is usually bought from a bakery, while the rest of food can be prepared for the guests. However, another, probably better way of interpreting it implies that *kanti* is a more precious meal, "having prepared good food and even having prepared *kanti*!" Another text might also suggest that *kanti* is specially important food.

Kizil WD-II-4 a7

pleṃtne kapyāres kantiṣeś yikṣiye la(*ś*) *towa 4*
"On the occasion of the full moon, four pecks of flour for the *kanti* of the workers"

Flour (*yikṣye*) is almost always provided for *kanti*, and as far as I could find, no other ingredient is mentioned for it. This also suggests it could be bread or its Indian equivalent, naan, which are both prepared mostly with flour, but it should be a special, precious type of naan.

Yeast is also a possible and common ingredient used for bread as well (there is no TB word with a secured meaning 'yeast', see DTB: 576 for a discussion of *räṅkāñi*). The fact that is not mentioned is not crucial: it does not have to be present in great quantities, and it can be reused from the same paste of flour if more flour is added to it day after day. It also appears that *kanti* was prepared with oil or ghee (cf. Ching 2010: 385). This would correspond to a more eastern type of bread, such as Indian naan, which is often prepared with ghee.

Flour being allocated "in *kanti*" (*kantine*) is more problematic. One can interpret it as the flour being used for the preparation of bread (as in THT 433 a4). Another way of interpreting the sentences is by glossing it as "flour [equivalent to X quantity of] naan". The word *kanti* could also technically refer to pastry, however one might wonder why so much pastry was produced for monasteries, and, more importantly, why sweet ingredients such as honey are not mentioned in regard to it. Naan *vel sim.* seems a more fitting translation for it.

Another possibility is that *kanti* designates groats, that is, the hulled kernel of cereal grains. In that case one is to understand the occurrences of *kantine yikṣye* as 'flour in (the form of) groat', which is called groat flour in English (Fr. *farine de gruau*—Modern French *gruau* corresponds to both gruel and groat). Groats need to be soaked in a liquid stuff in order for them to be easily chewed: it could also correspond to the ghee/oil used in order to prepare *kanti* (cf. Ching 2010: 385). One can also imagine that the ghee or oil was used as a topping for the *kanti*. In that case, TB *oṅkarño* would designate rice porridge, and *kanti* groats or groat flour. One wonders, nonetheless, if it could still fit well a context such as that of THT 375 b5. Would "even" groats be prepared for such prestigious guests? Or is the meaning of the sentence "they prepared good food, and even (such food as) groats"? An element possibly in favor of the meaning 'groats' for *kanti* is the absence of other ingredients than *yikṣye* provided for it.

The etymology of *kanti* is not evident. Adams (DTB: 146) connected it to Av. *gantuma-* 'wheat', New Persian *gandum* 'id.', Vedic *godhū́ma-* 'id.', etc. He suggests Indo-European cognates for the Iranian and Indian forms, such as Greek χόνδρος 'grain, grain of salt, seed, barley-grain, cartilage' (Beekes 2010: 1643), and tries to reconstruct an Indo-European word, a root-noun **gʰondʰ-*. The details are very unsatisfactory, the Greek word has no clear Indo-European etymology (cf. Beekes op. cit.). As Peyrot writes (2018: 256–257[13], with references), a con-

nection of *kanti* to *gantuma-* etc. is thinkable if the word means 'bread', but the connected word has so many variants that it is rather a Wanderwort (see Kümmel 2020: 259 with a similar view).[25]

One can conceive that the Indian form *godhūma-* 'wheat' contains the reflex of a specific vowel, which was reflected as **an* in Proto-Iranian and **aṷ* in Indian. The alternation between **t* in Proto-Iranian (cf. Morgenstierne 1973: 157 for the reconstruction with **-nt-* in this word) and **d^h* in Indo-Aryan seems also suggestive of two adaptation strategies for a single source phoneme, not dissimilar to what I suggested for TB *kroṅkśe* (s.v.). Some scholars have suggested a deformation due to folk etymology, *go-dhū́ma-* 'cow-smoke' (e.g. Witzel 1999: 29). I do not understand the motivation behind such a folk etymology, and I see an issue in that the word can certainly be reconstructed for Proto-Indo-Aryan (CDIAL: 230), so the folk etymology would have dated back to the Proto-Indo-Aryan period as well, which seems rather eerie to me.

One finds Sanskrit *-ndh-*: Iranian *-nt-* also in Ved. *gandhá-* 'smell': Av. *gaⁱnti-* 'bad smell' (NP *gand* 'stink'). Lubotsky (2001: 303) considers this as an example of an irregular correspondence and he suggests a BMAC origin for this word. I also want to suggest Proto-Iranian **gantúma-* (most probably *gantū́ma-*, in any case the stress is warranted by the Pashto, see p. 167): Proto-Indo-Aryan **godhū́ma-* 'wheat' represent BMAC loanwords. This would fit the archaeological reality of the BMAC homeland, and the known fact that the BMAC people practiced agriculture. From a formal point of view, the Indo-Aryan word fits perfectly a very common BMAC pattern as defined by Lubotsky (2001: 303) namely, trisyllabic nouns with a long middle syllable. For the Iranian part, a shortening due to the frequency of words ending in **-(t)uma* is thinkable. In any case, as usual, the middle syllable is the accented one, both for the Indo-Aryan and Iranian words.

I wish to suggest that TB *kanti* is indeed a cognate of PIr. **gantúma-* etc. as has been done before, but with the specific suggestion that it derives from another BMAC variant of the word: perhaps **gãDəy-* or the like. If the meaning of *kanti* was 'groats' then a word originally meaning 'cereal' is not out of place, think of English *cereal* designating breakfast made out of grain (id. with Fr. *céréales*) so basically the same process 'cereal (food)' > 'one specific cereal-based food'. For the variation between the Indo-Iranian forms on one side and the Tocharian on the other, it should be kept in mind that the Tocharian meaning is not identical to the Indo-Iranian one, and that it could thus reflect a different Wortbildung.

25 Kümmel maintains the more ancient comparison of these words with Proto-Semitic **ḥinta-* 'wheat' (+ Hittitte *kant-* 'a kind of grain, emmer', cf. Witzel 1999: 29). However I do not see how this comparison would work formally.

While situated very far from where Tocharians possibly had met BMAC peoples (see § 4.2), and much closer to where Indo-Iranians met them, in Togolok 1 (South Turkmenistan), a BMAC site, a quantity of cereal grains was found "as is typical for prehistoric and historic bread-baking cultures across West Asia" (Cerasetti, Arciero, Billings, Cattani & al. 2022: 104). The authors further mention that "the grinding and baking of wheat into bread was an important daily part of the dietary economy" (Cerasetti, Arciero, Billings, Cattani & al. 2022: 104). They add (op. cit.) that naked and hulled barley may have perhaps been ground into flour or consumed fermented or as porridge. The BMAC predecessor of TB *kanti* could thus have been made out of barley.

It is true that *kanti* looks similar in structure to the Ved. *gandhá-* 'smell': Av. *gaⁱnti-* 'bad smell' pair mentioned above (so one could take it as an argument that they derive from a common source language), however on the one hand it is not entirely sure what they go back to as the Vedic words ends with *-á* and the Avestan one with *-i*; on the other hand, homophones are possible within any language. One can also add that, for *gantuma-*: Vedic *godhúma-* the first-syllable vowel was probably different as well. This would explain why we have TB *-a-* corresponding to Ir. *-a-*, whereas in other alledged BMAC loanwords, IIr. **a* corresponds to PT (and TB) **e*. Provisionally I decided to spell the vowel as **ã*.

In conclusion, I suggest that TB *kanti* which, I believe, designated either groats or naan, derives from a BMAC variant of the same root as Indic **godhúma-* and PIr. **gantúma-*, both of which mean 'wheat'.

3.3.2 *TB* peñiyo, *TA* pañi *'Splendor'*

Pinault (2006: 181–183) reconstructs a BMAC word **paṇi* 'wealth', of which PT *peñiyo* would be a derivative. One major argument used by Pinault is the existence of the tribe name Ved. *Paṇi-* designating a rich but unholy race of men. Pinault's etymology works formally and semantically. It should nonetheless be noted that another etymology has been proposed for these words: Beekes (2010: 1546) has connected them to PIE **bʰeh₂-* 'to shine, appear'. Beekes's etymology is formally difficult (cf. Del Tomba 2023: 179) so that Pinault's proposal seems more likely.

3.3.3 *TB* mewiyo *'Tiger'*

Dragoni (2023: 156 f.) has suggested that Tocharian B *mewiyo* 'tiger' was borrowed from BMAC **maṷiia-* which, according to him, was also borrowed in Iranian: Sogdian *myw* and Khotanese *mūya-**. The final *-o* of the Tocharian B form could fit this hypothesis. No better Indo-European or Iranian origin has been put forward (for a critical review of the literature, see Dragoni 2023: 156 f.),

and this hypothesis would strengthen my suggestion that TA *lepäś** 'jackal' is a BMAC loanword (see above) as they both belong to the same semantic field.

3.4 Discussion

3.4.1 *Presentation of the Loanwords and Their Phonetics*

In this chapter I tried to analyze and discuss thirteen Tocharian words that have no clear Indo-European etymology.

These words are: TB *iṣcäke* 'brick'; TB *eñcuwo*, TA *añcu** 'iron'; TB *etre*, TA *ätär* 'hero'; TB *kercapo* 'donkey'; TB *kenek*, TA *kanak* 'cotton cloth, cotton fabric'; TB *kanti* 'groats or naan'; TB *kroṅkśe*, TA *kroṅśe* 'bee'; TB *peñiyo*, TA *pañi* 'splendor'; TB *mewiyo* 'tiger'; TA *lepäś** 'jackal'; TB *witsako* 'root'; TB *śerwe*, TA *śaru* 'hunter'; TB *ṣecake*, TA *śiśäk* 'lion'.

Out of these words, it was determined that TB *eñcuwo*, TA *añcu** 'iron' is of Iranian origin (cf. also Peyrot, Dragoni & Bernard 2022). I considered that all the others were borrowed, or possibly borrowed, at the Proto-Tocharian stage. Indeed, when cognates are found in both languages, a Proto-Tocharian prototype can be reconstructed for all these words, except for TB *ṣecake*, TA *śiśäk* 'lion', where the discrepancy could be explained by an influence of TA *śiśri* 'whiskers' on the Tocharian A form.

On the basis of the Old Steppe Iranian–Proto-Tocharian correspondences established in the previous chapter, I was able to reject an Iranian etymology for all these words when an Iranian etymology had been proposed. Besides, no convincing Iranian etymon has yet been found to explain TB *witsako* 'root' and TB *kenek*, TA *kanak* 'cotton cloth' as Iranian loanwords, nor does an Iranian etymon for TB *etre*, TA *ätär* 'hero' exist.

For a number of those words it was proposed that they were from the same language that is otherwise known as the BMAC language/substratum or Indo-Iranian substratum (Lubotsky 2001). For each and one of those words for which it was proposed, I have examined these proposals, and found them, if not convincing, entirely plausible (with the exception of B *eñcuwo*, TA *añcu** for which an Iranian etymology was ultimately found).

It is useful to compare, like Lubotsky (2001) did, the main features of these words in order to establish more general observations about BMAC loanwords in Tocharian. Out of the twelve words I assume to be of BMAC origin, nine follow the same syllabic pattern, namely, they are trisyllabic words with fixed second-syllable stress. These words are: TB *iṣcäke* 'brick'; TB *kercapo* 'donkey'; TB *kenek*, TA *kanak* 'cotton cloth, cotton fabric' (← **kanáka* 'hemp cloth'); TB *peñiyo*, TA

pañi 'splendor'; TB *mewiyo* 'tiger'; TB *witsako* 'root'; TB *śerwe* < **śeruwe*, TA *śaru* 'hunter'; TB *ṣecake*, TA *śiśäk* 'lion'. Only four words are disyllabic: TB *etre*, TA *atär* 'hero'; TB *kanti* 'naan' or 'groats'; TB *kroṅkśe*, TA *kroṅśe* 'bee'; TA *lepäś** 'jackal' (which could be from **laipəśe*).

This observation is to be compared to the fact that Old Steppe Iranian loanwords in Tocharian had fixed first-syllable stress (§ 2.7.2.7). The stress in these BMAC words needs to be old, or original, because, according to my findings, Proto-Tocharian preserved the original stress when borrowing words. One can also deduce from TB *kenek*, TA *kanak* 'cotton cloth, cotton fabric' which go back to BMAC **kanáka*, that the borrowings precede the syncope law of Proto-Tocharian (see § 2.7.2.7), and they have thus in any case not been borrowed much later than the Old Steppe Iranian loanwords.

The tendency of the Indo-Iranian substratum language to have words with three syllables was observed by Lubotsky (2001: 303). He noted that many of the substratal words in his list were "trisyllabic nouns with long middle syllable". This differs from our list, where the middle syllable is—in a small majority of cases—an accented schwa. If we look at the Vedic accent in those trisyllabic words listed by Lubotsky (2001: 305), we see that the accent is sometimes on the first, middle or last syllable, with no easily recognizable pattern. One can imagine that an original middle stress accent "lengthened" the vowel, and that the stress would later be placed on another syllable. Perhaps Tocharian speakers confused length with stress, or perhaps Proto-Indo-Iranians confused stress with length, leading to the different adaptation of the middle syllable in Proto-Indo-Iranian and Proto-Tocharian.

However, it is also possible that the source languages of Indo-Iranian and Proto-Tocharian were different, albeit related, with different stress patterns. In words that are, in my opinion, clear cognates, such as PT **kercápo*: Ved. *gardabhá-* (not listed in Lubotsky 2001, because there are no Iranian cognates), the stress is clearly different in Vedic and Tocharian, and there is no lengthening of the middle syllable in the Indic word, although a likely explanation in this case can be that Vedic speakers interpreted the ending as related to the animal name suffix *-bhá-*.

Another striking phenomenon is the fact that the BMAC vowel borrowed as **a* in Indo-Iranian was borrowed as Proto-Tocharian **e* (TB *e*, TA *a*), typically, Tocharian B *śerwe*, Tocharian A *śaru* 'hunter': Ved. *Śarvá-* 'name of a god', and PT **kercəpo*, TB *kercapo*: Ved. *gardabhá-*. This coincides with the adaptation of Old Steppe Iranian **a* as Proto-Tocharian **e* (see the previous chapter). In PT **kercəpo*, TB *kercapo*, compared to Ved. *gardabhá-*, it also appears that another vowel borrowed as **a* in Indic was borrowed as schwa in Tocharian.

I have argued that there was an unclear phoneme which was rendered by PT **ṅkś* and Indic *-ñc-*, in the word TB *kroṅkśe* 'bee': Sanskrit *krauñca-* 'crane' (going by, I argue, the notion of 'loud animal'). Further, the initial phoneme (perhaps a pharyngeal?) seen in PIIr. **ḥišt-* 'brick' was not rendered by Proto-Tocharian. Finally, the palatal phoneme provisionally set up as **ć* was rendered by Proto-Tocharian as **ś*, as in Proto-Tocharian **śerəwe* 'hunter' ← **ćarwa*, and by (Proto-)Indo-Iranian as **ć*, as in Proto-Indo-Iranian (or common Indo-Iranian) **ćarṷa-* 'hunting god'.

Out of all the substratal suffixes noted by Lubotsky (2001: 304), I can notably recognize the suffix *-(*ə*)*ka*, found in TB *iṣcäke* 'brick', *ṣecake* 'lion' and in PT **kenek* 'hemp cloth' (or 'cotton cloth') if from an earlier form **keneke* as I suggested. The substratal suffix **-pa-* (also discovered by Lubotsky), is found as *-po* in TB *kercapo* 'donkey'. As I wrote above, we can suppose that this BMAC word was remotivated in Vedic, with the replacement of the BMAC suffix **-pa-* with the animal name suffix **-bhá-* which is found for instance in *vṛṣabhá-* 'bull' (this could also explain the difference in stress between the Vedic and the Tocharian words).

3.4.2 *The Semantics of the BMAC Loanwords in Tocharian*

In order to get a clearer picture of the type of contact that took place between the BMAC language and Proto-Tocharian, it is necessary to analyze the semantic fields of the loanwords. We easily see that they mostly concern realia:

1. Animal names: TB *kercapo* 'donkey'; TB *kroṅkśe*, TA *kroṅśe* 'bee'; TB *mewiyo* 'tiger'; TA *lepäś** 'jackal'; TB *ṣecake*, TA *śiśäk* 'lion'.
2. Botanics: TB *witsako* 'root'; TB *kenek*, TA *kanak* 'cotton cloth' (possibly originally a plant name).
3. Legends, myths: TB *peñiyo*, TA *pañi* 'splendor'; TB *etre*, TA *ätär* 'hero'.
4. Hunting vocabulary: TB *śerwe*, TA *śaru* 'hunter'.
5. Construction material: TB *iṣcäke* 'clay or brick'.
6. Food: TB *kanti* 'naan' or 'groats'.

By far the largest category of BMAC loanwords studied in this chapter is constituted by animal names. This differs somewhat from the main semantic category found in the Indo-Iranian substratum loanwords (Lubotsky 2001: 307): names of body parts. Nevertheless, the second largest category, as per Lubotsky (2001), seems to be "wild animals", or perhaps, rather, animals found in the wild (since camels, donkeys and tortoises, and even pigeons do not need to be wild animals). Equally important for Indo-Iranian is the category "religion, cult", of which we find only a small trace in Tocharian, namely what I named "legends, myths", the notion of glory and the notion of hero. Then comes "clothing" to which corresponds TB *kenek*, TA *kanak* 'cotton cloth' (if it was not borrowed as a

plant name, differently from what I hesitantly suggested above). "Building technology" has the Tocharian cognate TB *iṣcäke* 'brick'. Finally, for the categories "artifacts", "water economy and irrigation", "cattle breeding" and "agriculture" I could find no corresponding Tocharian "substratal word".

There are of course much fewer Tocharian words in my list than Indo-Iranian words in Lubotsky's list. As this is a selective study, I did not include every possible example, although I did not find any example of the latter categories, even in other sources.[26] The evident explanation is that Tocharian A and B have a much smaller quantity of attested words than the very well attested and prolific Indo-Aryan and Iranian branches of Indo-European. This discrepancy in data will naturally lead to a bias in the comparison of the Tocharian and the Indo-Iranian "substratum" words. Nonetheless, the type of BMAC vocabulary in Tocharian in my view seems to indicate another type of contact, namely, contact between a more "primitive" BMAC culture than the one in contact with Indo-Iranians. It is also clear from the animal terms that these BMAC people were more familiar with the fauna of the region than the Tocharians, which might imply that the latter were only arriving in the region when they borrowed the words from the BMAC language.

An interesting point is that, according to my findings, TB *iṣcäke* 'clay' was borrowed from the BMAC language, while TB *iścem*, of a similar or identical meaning, was borrowed from Old Steppe Iranian. Iranians, in their turn, had borrowed this word from a BMAC language (cf. Kümmel 2020: 257). This might indicate that Tocharians had contact with a different BMAC people from the one Indo-Iranians were in contact with, and also, perhaps, that the type of clay designated by both words was different, so that Tocharians could have used both of them in a complementary manner, although there are not enough attestations of TB *iṣcäke* to be sure of this.

An intriguing fact, if my reasoning is correct, is that these BMAC speakers called bees 'the noisy ones'. This could suggest that they did not collect honey, as they did not name them after honey, which is their major attribute to honey-collecting peoples (although they perhaps only called swarms of bees so, and had a different word for individual bees). Indo-Europeans did

26 One could, however, suggest TB *ñemek* 'harvest' which has no clear etymology and resembles TB *kenek* 'cotton cloth, cotton fabric'. However, not only it has no Indo-Iranian equivalent, but it has received a very convincing etymology by Pinault (2020b: 214–215) who derives it from PIE **nem-* 'to take', with the **-ek* suffix of Tocharian (for which see Pinault 2020b). Pinault (op. cit.) further connects it to TA *ñomes* 'martingale' (the harnessing part of the plow) which for him originally meant 'pertaining to taking (the animal under control)'. I think one can also propose that it originally meant '(harness) pertaining to harvest'.

have a word for honey, **melit-*, so they were familiar with honey, but they also did not conceive of bees as 'honey-makers', which implies they were possibly not familiar with the process of honey-collecting (cf. van Sluis 2022). Last, any horse-related words are lacking. Absence of evidence is not evidence of absence, naturally, however we are bound to think that horses at such early times were the preserve of Indo-European speaking populations (cf. also § 2.2.3).

These BMAC speakers were familiar with hunting, which could imply that they were a hunter gatherer society, but could also simply imply that they hunted besides their agricultural activities. In Indo-Iranian, a cognate of TB *śerwe* is also found, yet we can be fairly certain that the BMAC speakers in contact with Indo-Iranians had an agricultural society. As far as I know, no agricultural society has completely abandoned hunting activities, even though they are sometimes reduced to a hobby, as they are currently in Western Europe. As far as I can see, the Tocharians did not borrow any names of sophisticated tools or weapons from them, as they did from Old Steppe Iranian. This is remarkable, as the BMAC civilization in the West of Central Asia exported weapons, notably decorated weapons (cf. Parpola 2022: 26). Two other interesting words, represented by TB *peñiyo*, TA *pañi* 'splendor'; TB *etre*, TA *atär* 'hero', seem to indicate that Tocharians could also have heard stories from that BMAC people, that this people also influenced their world view, to a more limited extent, of course, than Iranians did. This could perhaps be compared with "mythical" BMAC loanwords in Indo-Iranian, such as Vedic *gandharvá-*: Avestan *gaṇdərəβa-* 'a mythical being' (cf. Lubotsky 2001: 303).

Thus, although BMAC loanwords are also found in Indo-Iranian, the picture of the culture of the source language given by the borrowings in Tocharian is very different from what we can deduce from the ones found in Indo-Iranian. Most importantly, Tocharian cannot be shown to have borrowed words relative to "water economy and irrigation", "cattle breeding", or "agriculture", categories which we find listed in Lubotsky (2001). The same BMAC word which for Iranians meant 'clay brick' probably designated 'clay' in Tocharian. This perhaps shows that the BMAC speakers in contact with Indo-Iranians had a more evolved culture, with bricks made of clay, whereas BMAC speakers in contact with Tocharians had a technologically less advanced culture.

Interestingly, since there are no indications of words relative to weaponry or violence and war (unlike from Old Iranian), it is difficult to imagine that Tocharian speakers were in any type of military conflict with speakers of the BMAC language. If Tocharian borrowed its word for 'root' from them, this may show that they had a certain knowledge of botany, and perhaps thus of medicine, as roots often have medical uses.

The BMAC loanwords in Tocharian seem to be less sophisticated in nature than the Old Steppe Iranian ones, indicating possibly a less developed culture. The BMAC civilization in the Oxus had a rather complex political and administrative structure, with sophisticated seals (Francfort 2018) and a complex, rich mythology, filled with heroes and an organized divine hierarchy (on which see Francfort 2018a). This is partly reflected in the religious vocabulary borrowed by Indo-Iranian from the BMAC language (Lubotsky 2001: 304). This is not the image one naturally has by observing the BMAC loanwords in Tocharian, although this is naturally speculative, and it requires archaeological support. One could deduce that these borrowings from BMAC in the Pre-Proto-Tocharian period preceded those from the more militarized and more economically developed Old Iranian civilization. Indeed, it is difficult to accept the scenario that Indo-Iranian should have borrowed their words from this civilization, usually words for simple notions, for realia, then, hundreds of years later, an Iranian tribe came to the Tocharians, probably with military strength, and only then came that unknown people, or was that unknown people encountered, which had kept the same archaic phonology as during the Indo-Iranian borrowings, and that Tocharians borrowed some elements of their lexicon from them.

Furthermore, I do not see why they would not have borrowed the same words from the Iranians, who also possessed a similar if not more developed craftsmanship, and mastering of metals and construction materials, etc. In particular, it is difficult to understand why they would not have borrowed the animal names that they borrowed from the BMAC language directly from Old Steppe Iranian instead. I nevertheless do not think that the borrowings of BMAC words into Tocharian occurred much earlier than those into Indo-Iranian: they could have occurred more or less at the same time, but from a sister language.

It seems unlikely that all three, the Indo-Iranians, the Tocharians, and the BMAC people were at the same place at the same time. If Tocharians and Indo-Iranians were at the same place at the same time when the latter borrowed their BMAC loanwords, why did Tocharians not borrow any Indo-Iranian words at the same time? Of such words there is no trace. Many peoples borrowed from Proto-Indo-Iranian during the Proto-Indo-Iranian migration (for a survey, see Kümmel 2020). It seems more likely, in my opinion, that the Indo-Iranians were in contact with a specific group of BMAC speakers, much more advanced than the one Proto-Tocharian speakers were in contact with. BMAC contact should thus be parallel, but not necessarily simultaneous, for Indo-Iranians and Tocharians. This would perhaps explain why a word like 'hunter' was deified in a more complex civilization, while it kept its base meaning 'hunter' in the variant that was borrowed into Tocharian.

3.4.3 *Western and Eastern BMAC Varieties*

One can push the hypothesis discussed above further, and propose that there existed an Eastern and a Western BMAC language. Tocharians would thus have been in contact with the Eastern BMAC language and Indo-Iranians with the Western one. This needs of course much more research, but one can already mention a few of the differences that appear between both varieties. This distinction could be supported by archaeology (Zhang & al. 2021), which recently found traces of BMAC people in the region where Tocharian speakers where also found, although the place and time of borrowings can be different from this specific presence. I should specify that it is not so much that traces of "BMAC people" were found, since the BMAC, genetically, were a mixture of different haplogroups, but rather that a possible connection between the ancient, autochthonous Tarim Basin people and the BMAC people could be made. Francfort (2023: 59) makes a different conclusion based on Zhang & al.'s paper and considers that both Tocharians and BMAC people (people from the Oxus civilization) are excluded as ancient, autochthonous inhabitants of the Tarim Basin. In any case, these BMAC people need not have been in the Tarim Basin itself, as contact between them and the Tocharians could have occurred at any geographical point between the Steppes and Xinjiang. It is thus entirely possible to imagine a greater spread of BMAC people in the region of South Siberia and North Kazakhstan, for example.

It is possible, if not probable, that a BMAC language spoken so far away from the West was different from the language spoken in Bactriana-Margiana properly. Below, I will try to systematically present the variation we observe between Indo-Iranian BMAC loanwords and Tocharian BMAC loanwords, in order to present what type of variation one can find.

Western BMAC **d* or **t* vs. Eastern BMAC **ǰ* or **č*
- Vedic *gardabhá-* 'donkey' vs. Tocharian B *kercapo* 'donkey'
- PIr. **ištika-* 'brick' vs. Tocharian B *iṣcäke* 'clay'.

Western BMAC **d* vs. Eastern BMAC **dz* (?)
- PIr. **u̯ītāka-* ~ **u̯aiṭāka-* 'root': Iron *widag*, Digor *wedagæ* 'root', Pashto *wuláy* 'root, root-fibre', Sogdian *wyt'k* 'string' and Yaghnobi *wita* 'cord' vs. TB *witsako* 'root'.

Western BMAC **-ñc-* vs. Eastern BMAC **-nkś-*
- Sanskrit *krauñca-* 'crane' vs. TB *kroṅkśe* 'bee' (for the semantics, s.v. TB *kroṅkśe*).

Western BMAC *-*au̯*- vs. Eastern BMAC *-*aį̯*-
– PIIr. **r*(/*l*)*au̯pi*- 'marten', **r*(/*l*)*au̯paća*- 'fox, jackal' (< *'dangerous marten') vs. TA *lepäś** 'jackal' < PT **laipəśe*.[27]

Western BMAC *-*ba*- ~ *-*fa*- suffix (< *-*pa*-?) vs. Eastern BMAC *-*ka*- suffix
– Although the suffix *-*ka*- is known in Western BMAC, it is not found in Common Iranian **kanaba*- ~ **kanafa*- 'hemp' vs. PT **kenek* 'cotton cloth'.

There were also differences in meaning:
– Indo-Iranian **aϑrau̯a*- ~ **aϑaru̯a*- 'priest' vs. PT **etre* 'hero'.
– PIIr. **išti*- 'brick' vs. Tocharian B *iṣcäke* 'clay'.
– PIIr. **ćaru̯a*- 'hunter god' vs. PT **śerəwe* 'hunter'.
– Sanskrit *krauñca*- 'crane' vs. TB *kroṅkśe* 'bee' (for the semantics, s.v. TB *kroṅkśe*).

One can add a further example which does not directly concern Tocharian:

Western BMAC **paraću*- 'axe' vs. Eastern BMAC **paratu*- 'id.'
– Indo-Aryan **paraću*- 'hatchet' (Ved. *paraśú*-, etc.) vs. Eastern Iranian **paratu*- 'axe' (borrowed in PT **peretu*).

The alternation between **gãD-úma*- 'wheat' and **gãDəy*- 'naan or groats' (s.v. *kanti*) could be dialectal or dialect-internal: as long as a cognate of one or the other form is not found in one or the other language, it is impossible to determine. In general, the examples provided are too few to make developed conclusions concerning these differences, but they seem to indicate that, although there probably was a dialect continuum between BMAC varieties, they were sensibly different on a number of points.

Usually, Proto-Indo-Iranian borrowed from one single variety and Proto-Tocharian from another. However, the word for 'axe' was seemingly borrowed later, and from two different varieties: **paraću*- in Indo-Aryan (which, however, might be of a high antiquity) and **paratu*- in Eastern Iranian (which needs to be a later borrowing—cf. in general section 2.7.3.7). This should represent a later variety of Eastern BMAC that remained for a longer time in the

27 Very boldly, and strictly for the sake of speculation, one could imagine that there was a similar correspondence between **gãDau̯-ma*- > *gantúma*, *godhūma*- (s.v. *kanti*) and **ganDəy* > PT **kanti* (*au̯* : *aį̯*, *vel sim.*).

region. While one cannot be sure that BMAC is the donor language of this word, the BMAC civilization was known for its axes (cf. Parpola 2022: 26), including ceremonial axes that carried a very important symbolic power (cf. Francfort 2018a: 129 f.), so it definitely is a good candidate. Further, the Indian variant in itself could be used as an argument, although not decisive and mostly corroborated by the previous points, for a BMAC origin of that word (on the basis of the variation between Indian and Iranian adaptation of BMAC sounds observed by Lubotsky 2001). Another possible late, Eastern BMAC loanword in Iranian is represented by Avestan *xrafstra-* 'obnoxious animal', which might have been borrowed into Avestan at a later stage (cf. Bernard 2020: 41).

These examples are few, and in general the idea that both Proto-Indic and Proto-Iranian borrowed a series of words from the same variety of BMAC language can be upheld, while I hold that, on the basis of the phonetic and semantic differences mentioned above, Proto-Tocharian borrowed its "BMAC" vocabulary from a different variety thereof.

I am aware of the fact that this is an entirely new discussion topic, which is subject to caution, and needs much more research. More research would also include archaeological data in order to support or disprove, for example, the idea that Eastern BMAC, from which Proto-Tocharian borrowed its words, was less advanced technologically than Western BMAC, from which Indo-Iranian borrowed its words. While it is clear, from archaeological data, that "Western BMAC" and related peoples owned domestic animals—such as camels, sheep, goats, cattle, zebus, water buffaloes—and ate their meat rather than wild animals' (e.g. Cerasetti, Arciero, Billings, Cattani & al. 2022: 106 f.; see also Francfort 2009: 98), I am not aware what research can tell about Eastern BMAC food sources, especially at the time when contact with Tocharian should have occurred.

All in all, this new hypothesis has the advantage to explain and justify the fact that both Indo-Iranians and Proto-Tocharian speakers borrowed words from what seems to be the same language, but not at the same level of technological advancement, and certainly not at the same place.

Animal names were adapted to the local fauna: while 'dangerous (or obnoxious) martens' became foxes in the Iranian world (PIr. **raṷpaća-*) they designated jackals in the Indian and Tocharian worlds (Indo-Aryan **laṷpaća-*, PT **laipəśe*). Perhaps an interesting way to study these dialect differences would consist in integrating zoological studies to the lexical and etymological studies that could be done.

I have added this speculative perspective in order to try to make sense of some of the frequent differences in BMAC borrowings in Tocharian and Sanskrit

or Iranian. Naturally, as I explained throughout the chapter, they could also be due to different adaptations of the same phonemes. The semantic differences, however, could be indeed due to dialectal or linguistic difference, which could in their turn, suggest the existence of an Eastern vs. a Western BMAC variety, dialect or language.

CHAPTER 4

Conclusion

In the present book, I have discussed the oldest Iranian loanwords in Tocharian as well as possible loanwords from the so-called BMAC language.

4.1 Results

My starting point has been that there is no systematic study of the oldest layers of borrowing in Tocharian, despite the obvious relevance these layers have for our understanding of Tocharian as well as Old Iranian linguistic prehistory. Although a number of studies on Iranian and Tocharian contact exist, none systematically presents and analyzes the relevant material in full. It thus appeared necessary to approach the oldest layers of borrowing in Tocharian systematically, discussing all of the Old Iranian and possible BMAC loanwords in Tocharian, including those which were, accordingly to my analysis, wrongly attributed to Old Iranian, and some which are doubtful.

My investigation was structured along three research questions, which I repeat here:

1. Are the words belonging to the Old Iranian stage from one single language, as Schmidt thought, or multiple ones, as found elsewhere in the literature? Is there any reason to answer this question clearly?
2. If the Old Iranian layer is one single language, can its features be established? Is it homogeneous? How can we identify the words belonging to this stage and layer?
3. Are there any BMAC loanwords in Tocharian? Did they come directly from a BMAC language, or through, for instance, Iranian?

In the following, I will examine whether these questions have been answered, and how.

1. The first and second questions, concerning the Old Iranian stage of loanwords, which was first described by Schmidt (1985) in a seminal paper, I tried to answer in chapter 2. Schmidt grouped together a number of words which he considered to belong together, namely, words deriving from one stage of one dialect, and most of which I used as a departure point in this research. He proposed a number of sound correspondences such as Old Iranian **a* corresponding to Proto-Tocharian **e*. He also demonstrated that the reflexes of Proto-Indo-European **ḱ* and **ǵ* were **t^s* and **d^z* in that Iranian language. These

© KONINKLIJKE BRILL BV, LEIDEN, 2025 | DOI:10.1163/9789004732537_005
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

sounds go back to Proto-Iranian *ć and *ǰ, respectively, two sounds which had thus become affricates in Old Steppe Iranian. I re-examined these and other correspondences and tried to establish whether they were regular.

It appeared that the correspondence between Old Iranian *ā and Proto-Tocharian *a and Old Iranian *a and Proto-Tocharian *e, as in TB *kertte* 'sword' corresponding to Proto-Iranian **karta-* 'knife' or *waipecce* 'property, possession' corresponding to Proto-Iranian **hu̯aǰ-paϑi̯a-*, was essential for the selection of these words. This correspondence is not found in other Iranian loanwords in Tocharian, not even in Proto-Khotanese-Tumshuqese and Pre-Khotanese loanwords, which are very old and likewise belong to the Old Iranian stage (cf. Dragoni 2023). This correspondence, in combination with others, led to the rejection of a number of proposed loanwords from Old Iranian, such as TB *witsako* 'root', which does not fit the phonetic patterns of the Old Iranian layer.

On the way, I crossed three different words in Tocharian, which I assumed to have been calqued, or partially calqued, on Old Steppe Iranian words or phrases. These are: PT **kaun-ñəkte* 'sun-god' or 'sun-lord' for 'sun', PT **yok* 'hair, body hair, color' and PT **śaẃəṣye* 'world'. These examples, in my view, further demonstrate the deep influence of Old Steppe Iranian culture and language on the Proto-Tocharian society.

I also attempted to determine the chronological stage and the dialect affiliation of the Iranian source dialect more exactly. To this end I examined all the phonetic correspondences at hand, but also the morphology of the loanwords from an Iranian perspective, and tried to establish whether the portrait of one single language could be drawn. The answer was positive, as all phonetic and morphological features appear to form a coherent group, with no need to assume dialect differences or chronological developments within the Old Iranian source. At the same time, it is clear that this language was very archaic in a number of its traits: for instance, it had in all positions preserved *t^s and *d^z as reflexes of Proto-Indo-Iranian *ć and *ǰ.

2. The second question was whether the features of the Old Iranian source dialect can be established, if it was one homogeneous variety. As explained above, the correspondences between the Proto-Iranian reconstructed forms and the shape of the loanwords in Tocharian had to be established first. Once this was done, it was easier to establish the features of the Old Iranian source dialect itself.

It is useful to discuss the methodology I used in greater detail. Naturally, language contact is a very wide concept, and it is quite usual that language contact, especially when occurring due to population contact (versus, for example, elite domination), involves more than two varieties. If one looks at the French influence over English, it is quite obvious that multiple dialects and varieties

of French influenced English, both diachronically and synchronically. This is also the case for Persian and Arabic, and it is even true with dead languages (for example, Latin words were borrowed into French from Classical Latin, as well as vulgar Latin). It is thus entirely conceivable that a group of Iranian languages, dialects or varieties (with the vagueness these notions carry) went into contact with Proto-Tocharian, and that these yielded various Tocharian words, with different phonetic and semantic features. Not including this possibility would certainly cause a grievous bias to the present work. Nonetheless, starting from this assumption would also be damageable, as, if I started to consider each word for its own features as borrowed "on its own", it could soon be established that Tocharian borrowed from as many Old Iranian dialects as words. If one starts with the assumption that there were multiple sources, while in fact there only was one, it would become difficult to reach that conclusion, while, if one starts with the assumption that there was one source, but in fact there were multiple ones, it is easier to reach that conclusion. Epistemologically, it is necessary to start with the easiest solution: the existence of one single source variety.

I thus departed from the more economical assumption that all the Old Iranian words studied here were borrowed from one single source, while keeping in mind the possibility that it was not so. It was necessary to look at every problem, every unexpected outcome, compare it with the available data we have on Iranian languages, in order to see if it did not warrant a different source. In fact, many times I thought this was the case, and throughout my preliminary research, I often posited different Old Iranian stages or dialects in order to explain this or that word. Often, finding a different example (for instance the initial *ye-* in *yentuke* and *yetse*) solved the issue, this was also the case for TB *ekṣinek** 'dove', although the "other example" was found in Ossetic (see p. 171f. for more detail). Some examples, like PT **epiyac* 'memory', were problematic, while being of clear Iranian origin (in this specific case, because the origin of the *-c* is not obvious). In that case, the crucial point was to determine that the issue is not related to an internal Iranian problem: the Proto-Iranian form of this word is **abi-i̯āta-*, and there is no reason to interpret this "palatalization" as an Iranian phenomenon. With this methodology, I was able to ascertain the identity of the source of all these words, and thus to attribute their origin to a prehistorical contact situation relatively limited in place and time (that is, one single language → one single language contact). I also would like to add that the existence of unsolvable irregularities in a set of borrowed words does not even imply that there were multiple source languages (although that would be a logical deduction): it is also possible that the source language borrowed itself some forms, or that it analogically remodelled some words and, given that we work

in the fog of unattestedness, it is difficult to precisely know such things. Fortunately, it is not necessary to deal with this issue, as the features of Old Steppe Iranian words can be explained in a straightforward fashion—apart from PT **epiyac* 'memory' which I mentioned above.

To establish the features of the Old Iranian source dialect I have naturally focused on innovations rather than on archaisms. I have so identified a number of sound changes, such as the shortening of **ā* in front of *yod* and *waw*; the loss of **h* in intervocalic and word-initial (and possibly word-final) position; the simplification of the cluster **dm-* to **m-*; the change of **rd* to **ld*, etc. Morphological features were also proposed, as it became clear that Old Steppe Iranian preserved *u*-stems, and that the nominative singular of neuter stems was likely to be OSIr. **-ám*. I also observed that there were some important semantic changes.

A very important feature concerned the stress system of the Old Iranian source language. Indeed, apart from some suffixes and prefixes bearing fixed stress (identical to the stress seen in the Vedic cognates of these suffixes and prefixes for instance), Old Steppe Iranian seems to have had fixed initial stress, as far as can be judged from the evidence I found. This is not a unique feature, as it is noticeably shared by Ossetic, which is again a remarkable fact.

The discussion on the origin of PT **ekṣineke* 'dove', section 2.7.2.8 has led me to propose a very ancient shared sound change between Old Steppe Iranian and Ossetic, namely **anḭV* > **iḭnV*. If I am correct, this would demonstrate the existence of an Old Steppe Iranian–Ossetic node. Furthermore, Old Steppe Iranian and Ossetic share nine to eleven relevant phonetic isoglosses (see p. 182). But Old Steppe Iranian is not the ancestor of Ossetic, as the Old Steppe Iranian sound law **rd* > **ld* shows, but also more importantly Old Steppe Iranian **mānúḭa-* 'servant' vs. Proto-Ossetic **māniḭa-* 'husband'. Indeed, neither 'servant' nor 'husband' can reasonably derive from each other, but here it is clear that they were specializations of the Old Iranian word **mānúḭa-* 'the one of the house'. If this conclusion is correct, it could imply that no known Iranian language directly descends from Old Steppe Iranian, but that Old Steppe Iranian shares genetic kinship with Ossetic. The systematic comparison of both languages could possibly lead to the reconstruction of the Old Steppe Iranian–Ossetic branch, which could be tentatively be named "Scytho-Steppic".

Old Steppe Iranians were not an isolated people, they were probably not the only Iranian people in the region. They were part of a greater continuum of Iranian-speaking nomads, who are conveniently called "Scythians". Some of the Scythian peoples were already known in Antiquity, and they were mentioned for instance by Greek historians. The languages of Scythian peoples are not documented through texts, but are known to us through undirect sources (cf. Pin-

ault 2008a: 106). The Old Steppe Iranian words discussed in this book thus provide further undirect evidence of a Scythian language and its lexicon, and possibly of a greater antiquity than the remainder of Scythian languages. Scythian names recorded by Herodotus (died ca. 425 B.C.E.) possess the form **aspa-* for 'horse' (cf. Pinault 2008a: 108), as opposed to Old Steppe Iranian **at^{s}u̯a-*, which is clearly more archaic, as it preserves the Old Iranian sequence **-t^{s}u̯-*, which otherwise became **-sp-* in the Scythian languages Herodotus recorded words from.

The establishment of the features of Old Steppe Iranian led to the exclusion of some words, like Tocharian B *witsako* 'root', mentioned above, but also to a number of possible new loanwords, such as Tocharian B *ākteke* 'wonderful', *epastye* 'skilfull', *epe* 'or', *eprete* 'resolute, steadfast', Tocharian A *nātäk* 'lord' and others.

3. I have tried to answer my third research question in chapter 3: "Are there any BMAC loanwords in Tocharian? Did they come directly from a BMAC language, or through, for instance, Iranian?" It is thanks to the study of the Old Steppe Iranian vocabulary in Tocharian that this question could be answered. As explained above, the criteria to determine if a word was of Old Steppe Iranian origin or not led to the rejection of some words from that group. Some of these rejected words belong, in my view, to another language: the BMAC language (also called Oxus Civilization language) described by Lubotsky (2001), or a variety closely related to that BMAC language. Pinault (2002; 2003; 2006) also published specifically on BMAC loanwords in Tocharian.

These BMAC loanwords also shared features, but, unlike for Old Steppe Iranian, we have no documented BMAC-related languages to which we can confront our loanwords in order to reconstruct proto-forms. That is, the only support we may possibly have to show that a Tocharian word can be of BMAC origin is the existence of parallel borrowings into Indo-Iranian, "borrowing cognates". There is thus no way to verify independently how BMAC phonemes are represented in Tocharian, since we know too little about the phonological system of the BMAC language. For instance, in Vedic *gardabhá-* m. 'donkey, ass' corresponding to Tocharian B *kercapo* /kercə́po/ 'donkey',[1] we can see that Indo-Iranian **a* corresponds to **e* in the first syllable of the Tocharian word, but to **ə* in the second. It is attractive to posit different phonemes for the BMAC source on the basis of Tocharian, but such a contrast cannot be verified independently.

1 The TB word *kercapo* 'donkey' was previously believed to be related to Vedic *gardabhá-* 'donkey, ass' through way of inheritance. Carling (2005: 54) accepted this connection, but she explained the relation between these two words with the assumption of a common BMAC source.

Together with Michaël Peyrot and Federico Dragoni (Peyrot & Dragoni & Bernard, 2022), I have suggested that a word previously proposed as being of BMAC origin by Pinault (2006: 184–189) was instead a Pre-Khotanese loanword: TB *eñcuwo*, TA *añcu** 'iron'. More precisely, it would have been borrowed from the Pre-Khotanese ancestor of Khotanese *hīśśāna-* 'iron', which we reconstructed as **henśu̯aṇ̇a-*, subsequently analyzed as a *-ñña* adjective by Tocharian speakers, leading to the back-formation of the forms TB *eñcuwo*, TA *añcu**.

Because of significant and coherent phonetic differences, I have suggested that the BMAC loanwords in Tocharian do not come from exactly the same BMAC variety as those of Indo-Iranian. In view of the presumably large geographic distance, I provisionally termed these two varieties Western BMAC, the source of borrowings into Indo-Iranian, and Eastern BMAC, the source of borrowings into Tocharian. An example is Western BMAC **aw* vs. Eastern BMAC **ay*, as can be seen in Proto-Tocharian **laipəśe* 'jackal' vs. Proto-Indo-Iranian **laṷpaća-* 'fox, jackal' (Proto-Iranian **raṷpaća-*). Another example is Western BMAC **d* or **t* vs. Eastern BMAC **ǰ* or **č* in the examples Vedic *gardabhá-* 'donkey, ass' vs. Tocharian B *kercapo* 'donkey' and PIIr. **ištika-* 'brick', Burushaski *diṣṭík* 'brick' vs. Tocharian B *iṣcäke* 'clay'. There are also significant semantic differences, such as Western BMAC 'brick' for the latter etymon, versus Eastern BMAC 'clay'. These semantic differences coherently point towards a technologically less advanced society for the Eastern BMAC variety than for the Western BMAC variety. The fact that the source varieties appear to be different further suggests that BMAC loanwords in Tocharian were directly borrowed from a BMAC source, not by mediation of Old Steppe Iranian. From an archaeological point of view, it is difficult to imagine that the people of the Bactriana Margiana Archaeological Complex were identical to a people that was much further to the north, or to the east, speaking the exact same language, and this seems to be confirmed by the phonetic correspondences mentioned above. This naturally should be counterbalanced by the fact that we do not know where this people was and where specifically they entered in contact with Tocharians. All I can provide on this topic are speculations (see § 4.2).

As a conclusion, it appeared that the BMAC loanwords borrowed into Tocharian are sometimes similar to those borrowed into Iranian from the BMAC language, although less numerous, which is expected given the difference in the size of corpora. I also suggested that BMAC–Tocharian contacts were very early, as the words are less technical than what can be deduced from contact with Iranian and Indic languages. Thus, chronologically, Tocharian speakers would first have come into contact with speakers of the so-called BMAC language, and only later with speakers of Old Steppe Iranian. Furthermore, they show differences both in terms of semantics and phonetics with the BMAC loanwords found in Indo-Iranian languages.

In my investigation of Old Steppe Iranian and BMAC loanwords, I sometimes needed to discuss words that eventually turned out to be of different origin. This is notably the case with borrowings from Khotanese. Apart from TB *eñcuwo*, TA *añcu**, mentioned above, I have discussed TB *kamartīke* 'ruler' and TA *kākmärt* 'sovereignty'; TA *kāre* 'sword'; and TB *kāswo* 'skin disease' as well as TB *tāno* 'seed, grain'. All five must derive from Khotanese and are also included in Dragoni (2023). In the appendices 1 and 2 I also discussed two further Tocharian words: TB *yāmuttse*, TA *yāmutsi* 'parrot' which I suggest is of Chinese origin, and TB *patstsāṅk*, TA *pātsaṅk* 'window' which, I propose, is of Bactrian origin.

4.2 The Tocharian Way

My inquiry into the oldest layers of loanwords in the Tocharian vocabulary naturally leads to the question of when and where this contact between speakers of Tocharian and both speakers of Old Steppe Iranian and speakers of the BMAC language occurred. Even though the aims of this study were primarily of a linguistic nature, I venture to make a few notes on this issue.

As is well known, the Tocharians were an Indo-European people, and thus did not originate from Western China, or from Central Asia, but rather from the Ukrainian Steppes, where the Yamnaya Culture is strongly associated with speakers of Proto-Indo-European. Thus, we know where the Tocharians started, and that they finally arrived in the Tarim Basin, but many details of their trajectory remain to be clarified. In my view, it is possible and even likely that Tocharians followed an almost straight line, from the Yamnaya homeland almost directly to Afanasievo.

It is indeed important to note that Tocharians were, as far as we can tell, a group of people that came from the Yamnaya culture, and who migrated around 3400 BCE (cf. Solans 2024: 268 f.), as a group, carrying their language with them. All genetic studies point towards the presence of a Yamnaya-connected group in the Tarim Basin, who left major traces in the genetic make up of modern Xinjiang (e.g. Ning, Wang & al. 2019). It is thus not a Hungarian-like scenario (or a Roman Gaul scenario, in a way), where a dominating minority would have changed the language of a wider people, but a real group migration.

Yamnaya steppe people belong to a specific clan whose patrilinear lineage is R1b Z2013: Iranian populations have a R1a Z94 lineage, they are connected of course, but separate and more closely related to the Russian corded ware culture (Solans 2024). Ancient Steppe populations in the Xinjiang were R1b, so Yamnaya descendents, but nowadays the Xinjiang population has both R1a and R1b patrilinear markers (cf. Ning, Wang & al. 2019) so one can deduce that this

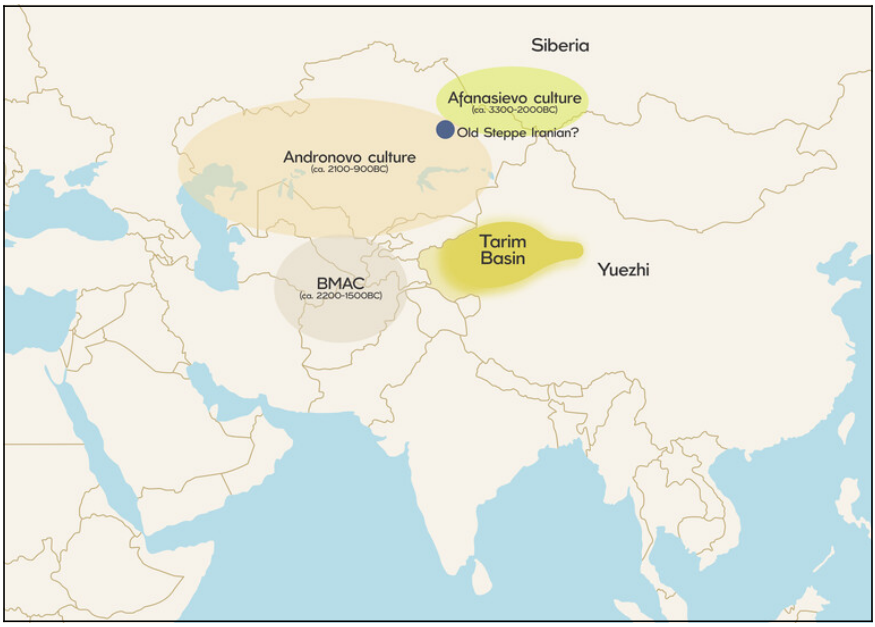

FIGURE 2 The Andronovo culture is commonly associated with Proto-Indo-Iranian, Afanasievo with Pre-Proto-Tocharian speakers, the Tarim Basin is where Tocharian manuscripts were found, and where Tocharians and Khotanese lived. The Old Steppe Iranian people were probably living in the east of the Andronovo culture. The Yuezhi are perhaps related to the Old Steppe Iranians. What I named "Eastern BMAC" might perhaps have been found alongside the eastern borders of the Andronovo culture.
© CHAMS BENOÎT BERNARD

population has Tocharian and Iranian admixture, which probably happened over time with the progressive settlement of Iranian peoples in the region. In my view, this should not be confused with the earliest Tocharian–Old Steppe Iranian contacts, which must have left little genetic trace in the Tocharian population when it arrived in the Tarim Basin.

In line with the focus of my linguistic investigation, I will here concentrate on the question whether the contacts with Old Steppe Iranian and with the so-called BMAC language took place in the Tarim Basin or elsewhere. For the contacts with Old Steppe Iranian ("the archaic stratum"), Peyrot (2018: 272, 280) assumes that the ancestors of the Tocharians had already arrived in the Tarim Basin when these took place. He tentatively locates Old Steppe Iranian north or east of the Tocharian area.

An important genetic study (Zhang & al. 2021) demonstrated that the famous Bronze Age Tarim Basin mummies, dating back to 2100–1700 BCE, were

not Indo-European and therefore cannot be identified as Tocharians, as had previously been assumed by many scholars. Indeed, these mummies rather belong to a genetically very different autochthonous population (Zhang & al. 2021: 260). Tocharian presence in the Tarim Basin thus does not need to be as old as these mummies, and there is no longer any need to assume that the Tocharians had already arrived in the Tarim Basin when the contacts with Old Steppe Iranian took place.

In my view, an important argument for dating and locating the contacts is the order in which the contacts with the BMAC language, Old Steppe Iranian and Proto-Khotanese-Tumshuqese, the earliest layer of contacts with Khotanese, took place. Indeed, I believe that a relative order of these contacts can be set up on linguistic grounds.

As I argue, the contacts with Old Steppe Iranian must be dated before the split of Proto-Tocharian in Tocharian A and Tocharian B. This is strongly suggested by the many examples of perfectly regular sound correspondences between Tocharian A and B loanwords from Old Steppe Iranian; by the law of syncope; and in particular, by the connection of this law to the stress system of Old Steppe Iranian; as well as by the reconstruction of Proto-Tocharian verbs based on Old Steppe Iranian nouns. Contacts with Proto-Khotanese-Tumshuqese occurred either slightly before the split of Proto-Tocharian into the daughter languages or were simultaneous with it (Dragoni 2023: 209 f.). The later Pre-Khotanese loanwords are to be dated after the split (Dragoni 2023: 209).

On the basis of the BMAC vocabulary, and on the basis of the fact that some BMAC items in Tocharian are reconstructible for Proto-Tocharian, I suggested that BMAC–Tocharian contact preceded the split of Proto-Tocharian in Tocharian A and Tocharian B too. Furthermore, the vocabulary of BMAC borrowings in Tocharian indicates a culture that was not extremely developed, possibly a hunter-gatherer society. We have, notably, a relatively large number of animal names, the word for 'hunter' and the word for 'root'. The Delta of the Oxus is suitable for agriculture (Francfort 2009: 98), but perhaps the situation was so different in Dzhungaria or in the region that the "BMAC people" there never managed to develop agricultural activities as their main alimentary resource.

The survival of populations speaking a BMAC related language into the first millennium BCE (perhaps even later) is likely on multiple grounds: the existence of elements of vocabulary (s.v. TB *peret*) borrowed from a language related to the BMAC language described by Lubotsky (2001); a possible grammatical substratum in Iranian languages of a specific zone (Bonmann, ongoing research); genetic studies finding BMAC components in various Iranian populations (cf. Solans 2024: 271). One of the two Tarim Basin languages discovered

by Huard (2024a) might be a descendant of the ancient BMAC language. The same arguments can also be used, together with the ones I have adduced (§ 3.4) to argue for the existence of another branch of the BMAC language and culture, albeit different, more to the east than what has been suggested until now.

It seems the most plausible that Tocharian first came into contact with the so-called BMAC language and borrowed some words for relatively primitive concepts, but also for local realities, such as animal names. Only after did they come into contact with Old Steppe Iranian and finally with Proto-Khotanese-Tumshuqese, both of which were apparently technologically more advanced, and were the source of vocabulary referring to more developed concepts.

If we assumed, alternatively, that Tocharian came in contact with Old Steppe Iranian first, and borrowed the words for 'army', 'castle', 'sword', 'lord', 'slave' and even basic words like 'group', why would they later need to borrow basic animal names, and words such as 'root' and 'hunter' from the language of another, much less technologically advanced group? In other words, the speakers of Old Steppe Iranian certainly had words for 'donkey', 'lion', 'root', so why would Tocharians have borrowed these words later from a seemingly less prestigious source language?

Thus, I assume that Tocharian came into contact with the BMAC language first, then with Old Steppe Iranian, and then with Proto-Khotanese-Tumshuqese peoples. It is attractive to identify early speakers of Khotanese with the Aqtala culture (cf. Peyrot 2018: 275 f.). The speakers of Old Steppe Iranian are tentatively located to the north or east of the Tocharian area by Peyrot (2018: 280). In any case, it is in my view very unlikely that Old Steppe Iranians were in the Tarim Basin at the time the Tocharians were there or before. On the basis of the semantics of the borrowed vocabulary, I hypothesize that the Old Steppe Iranians probably conquered or subdued the Tocharians in some way. If the Old Steppe Iranians were such an important culture, how should they have disappeared from the Tarim basin without leaving any trace at all? A possible counterargument to this is the newly deciphered Kushan script, called "eteo-Tocharian" by its decipherers (Bonmann, Halfmann, Korobzow & Bobomulloev 2023). This language was most likely that of the Kushan, and thus of their Yuezhi ancestors. There is a real possibility that it represents a descendant of Old Steppe Iranian, or of a closely related language.

With the above considerations in mind, one can suggest the following scenario for the arrival of Tocharians in the Tarim Basin: Tocharians, arriving to eastern Central Asia through South Siberia, first encountered the so-called BMAC speakers possibly in Dzhungaria. Tocharians learned from these BMAC speakers about a number of animals in the region, about plants, about food-gathering techniques, and perhaps about some food items. Nonetheless, there

is a problem: what if the BMAC people that Tocharians were in contact with are actually the BMAC related people found by Zhang & al. (2021)? Then, it would follow that the scenarii described above need to be rethought. For speculation's sake, let us imagine that Tocharian–BMAC contact occurred in the Tarim Basin. Then either contact with Old Steppe Iranian occurred after, and necessarily in the Tarim Basin (because there is no reasonable scenario where the Tocharians would have left and then come back again),—in which case we would really want traces of Old Steppe Iranians in the Tarim Basin—or it occurred before, whether in the Tarim Basin or elsewhere.

One of the conclusions of this book (section 3.4.3) was that the BMAC language in contact with Tocharian was not identical to the one Indo-Iranians were in contact with. This can be shown by constant phonetic differences in the adaptation of words, but also by the differences in meanings found in BMAC words in Indo-Iranian and in Tocharian. As a consequence, Tocharians did not meet the BMAC people in the same place as the Indo-Iranians did, but more to the east, probably when they migrated south from Siberia into Dzhungaria, before they crossed the Tian Shan mountains (on which migration, see e.g. Ning, Wang & al. 2019: 2530).

There is yet another possibility: that the BMAC speakers lived among Old Steppe Iranians, and that Tocharians, dominated by the Old Steppe Iranians, also had frequent contact with BMAC speakers. In which case, it is possible to understand the borrowing of more simple and more concrete day-to-day elements, animal names, etc. simultaneous to the other type of borrowings which occurred with the Old Steppe Iranians. If this is correct, however, one must renounce to **peretu* as a more certain Old Steppe Iranian loanword. As of now, this hypothesis, which needs to be worked out together with archaeological and archaeogenetical data, seems weaker than the hypothesis I have discussed above. I present arguments against it in § 3.4.2.

In the late Bronze Age, that is, 1500 to 800 BCE, we find archaeological cultures distinguished by roller pottery in the Steppes. These cultures had acquired horse-riding techniques around the 16th century BCE (cf. Parpola 2022: 48–49). Parpola (2022: 49) further adds that these horse riding peoples were the "immediate ancestors of the Iranian-speaking mounted horsemen known as Scythians, Sarmatians and Sakas […]". In all likelihood, the Old Steppe Iranians were part of this roller pottery culture, as were the ancestors of the Ossetes, who are most closely related to the Old Steppe Iranians from a linguistic point of view (see for example p. 182). We also know that a drought occurred between 1200 and 800 BCE on the Eurasian Steppes where the Old Steppe Iranians likely were (cf. van Geel et al. 2004). This drought perhaps pushed the Old Steppe Iranians (and possibly other Iranian peoples) into the periphery of the steppe, so

that they reached Dzhungaria, where they encountered the Tocharians (cf. Peyrot 2022a). This would be in agreement with the archaeological data presented by Parpola (2022).[2] These contacts would all have occurred before Tocharians entered the Tarim. It should be kept in mind that horse-riding did not arrive in the east of the Tian Shan and in the Tarim Basin before the late 12th century BCE (Wertmann, Yibulayinmu, Wagner & al. 2023: 2). It is possible that horse-riding was brought to the Tarim Basin, either by the Old Steppe Iranians (before or after they entered in contact with the Tocharians) or by the Tocharians themselves.

The Old Steppe Iranians and the Scythians have split from their common ancestor, which I suggest to name "Scytho-Steppic" (cf. fn. 158 p. 188). Speculatively, one may place this split in the last centuries of the second millennium BCE. Whether the drought I mentioned is the cause of this split or not is not an object of knowledge at this stage.

Proto-Tocharian proper, that is, the last stage of the language as descended from Proto-Indo-European, right before its separation into two branches (Pre-TB and Pre-TA), is estimated by most scholars at around 500 BCE, some estimate it is earlier, between 1000 BCE and 500 BCE, and some a little later, between 500 BCE and the 1st century CE (cf. Peyrot 2022: 87). This is more of an internal Tocharian issue, and language contact cannot provide a definitive solution to this dating question. This is because it is not obvious from the data whether the Old Steppe Iranian layer of vocabulary was borrowed in Proto-Tocharian or in Pre-Proto-Tocharian, namely, right before language division or much earlier.

What I have been consistently calling "Proto-Tocharian" is a practically useful way of designating various phases of the language, from the moment the characterist Tocharian sound changes occurred from Proto-Indo-European up to the separation into Pre-TB and Pre-TA. Certainly, it predates by much Buddhism which was introduced after the separation of both languages (cf. Peyrot 2022: 87 with references). Peyrot suggested that the Old Steppe Iranian loanwords could be associated with the presence of Andronovo related groups in Northern Xinjiang in the 13th to the 9th century BCE (Peyrot 2018: 280; 2022: 87). This corresponds to the drought described above, but also to the roller pottery horse riders mentioned by Parpola. To sum up, we do not have hard evidence for the dating of these contacts, however a series of arguments point towards the second half of the 2nd millennium and the first quarter of the 1st millennium BCE.

2 It is possible that it is during this migration that the Old Steppe Iranians, the ancestors of the Sogdians, of the Ossetes, and of various other Iranian peoples, borrowed the word **paratu-* 'axe', possibly from Eastern BMAC people.

Schoubben (2024: 269–270[626]) convincingly ascribes the Pre-Bactrian loanwords (only found in Tocharian B, not in Tocharian A, cf. Bernard 2024: 289–290) to Hellenistic times (323 BCE–31 BCE). They are congruent with the introduction of vine culture in the Xinjiang in the 3rd century BCE (cf. Schoubben, loc. cit.). This is a useful *terminus a quo*. The consequences are many, and a separate study is needed to fully consider the archaeological and historical information concerning Tocharian—Bactrian contact. For one, it confirms that the Tocharian split took place far back in the 1st millennium BCE: Tocharians had already met with Proto-Tumshuqese-Khotanese speakers, then split, then Tocharian B speakers were in contact with Pre-Bactrian speakers. This process took at least a few centuries, and if we want to account for Proto-Tocharian–Old Steppe Iranian contact, we have to project Proto-Tocharian back to at least about 800 BCE, if not 1000 BCE. In any case, this means that placing Proto-Tocharian at a later date, between 500 BCE and the 1st century CE is unlikely.

The Old Steppe Iranians very likely conquered or subdued the Tocharians in some way, as the type of vocabulary (military, social, and even the grammatical loanword *epe*) seems to suggest. Although there probably was at least a part of the population which was bilingual Tocharian–Old Steppe Iranian, as is suggested by the ease with which they integrated the loanwords, the influence of the Old Steppe Iranian language seems to have been rather limited in time. Notably, I have observed no heterogeneity in this layer of vocabulary, which I take as an indication that the borrowings occurred in a rather short period. Furthermore, although a word such as TB *epe* 'or' shows the strong cultural domination of Old Steppe Iranians on Tocharian (as grammatical borrowings usually do), the semantic fields of Old Steppe Iranian loanwords are relatively limited. For instance, we have no evidence for the borrowing of words for food, only two examples of borrowed animal names, namely *etswe* 'mule' and *ekṣinek** 'dove'. More food and animal names might well be expected if the contacts extended over a longer period.

While Old Steppe Iranians seemed to have been a horse-riding military society, the image we can draw of them based on the type of vocabulary borrowed by the Tocharians is that of a rather complex society, with courtyards, doors, armies, servants and lords. A society where people owned goods, prepared alcoholic beverages, perhaps even burrowed canals. They weighed on the Tocharians to the point they influenced the way the latter named the world.

It is crucial to examine the relationship of both Old Steppe Iranians and Proto-Tocharian speakers to horses. As mentioned above, Old Steppe Iranians might be part of a greater Scythian-like group of horse-riding populations. Wertmann, Yibulayinmu, Wagner & al.'s (2023) discovery of the earliest saddle found to date, a leather saddle, in the Yanghai cemetery in Turfan is telling

(see also § 2.2.3): they themselves attribute the saddles of that cemetery to "a community of mounted pastoralists living in the steppes or mountains in the near or distant western or northern neighborhood" (more specifically: either the woman in the tomb is from that community, or she acquired the saddle from them). This is a recent discovery, and time will tell if following such a track is fruitful or not, nonetheless, I have positive doubts that this community represents either Old Steppe Iranians or Tocharians, with a preference for Old Steppe Iranians, as they would have had a wider geographical extension. It would be crucial, in my view, to do a DNA analysis of the woman in the tomb that contained that saddle.

A population to the south of the Tarim Basin is unlikely to have had any influence on the horse elements of this cemetery: no horse-riding populations are found at such an early date in this zone. This certainly confirms the intuition that Tocharians did not enter the Tarim Basin through the south—this never was a strong hypothesis, but it is still useful to note. Besides, Wertmann, Yibulayinmu, Wagner & al. (2023: 14) have noted the presence of horse-related material (whips, cheek-pieces, bridles, and a bit) in graves dating back to the earliest period of the cemetery: ca. 1300–1000 BCE. While for the abovementioned saddle, a Tocharian origin is possible, in my view such an early date rather corresponds to Old Steppe Iranians.

Strikingly, the Yanghai saddle shows no trace of iron, while later models of saddles did (such as the Subeixi saddle, a later model, also discussed by Wertmann, Yibulayinmu, Wagner & al. 2023: 9 f., 13). This is coherent with the suggestion that iron was introduced to Tocharians, and to the rest of the Tarim Basin by the ancestors of the Khotanese (cf. Peyrot, Dragoni & Bernard 2022; see also p. 224 f.), and that Old Steppe Iranians either did not know of it, or did not use it prominently.

After the contacts with the Old Steppe Iranians, the Tocharians entered the Tarim Basin and encountered early speakers of Khotanese. It seems that the split of Proto-Tocharian into the two daughter languages Tocharian A and Tocharian B took place only then, after their arrival in the Tarim Basin. The contacts with Khotanese and its prestage lasted for centuries, way into the historical period, since we find in Tocharian traces of multiple stages of these languages: Proto-Khotanese-Tumshuqese, Pre-Khotanese, Old Khotanese, Late Khotanese and Tumshuqese (Dragoni 2023). As a pure matter of speculation, one can imagine that the Tocharians penetrated the Tarim Basin through the East, besides the Tian Shan mountain range, and did not directly cross the mountains. They would then have remained in this northern part of the Tarim Basin, where they still lived in the historical period. Old Steppe Iranians could have also gone to the south, or perhaps to the west, and might have mixed them-

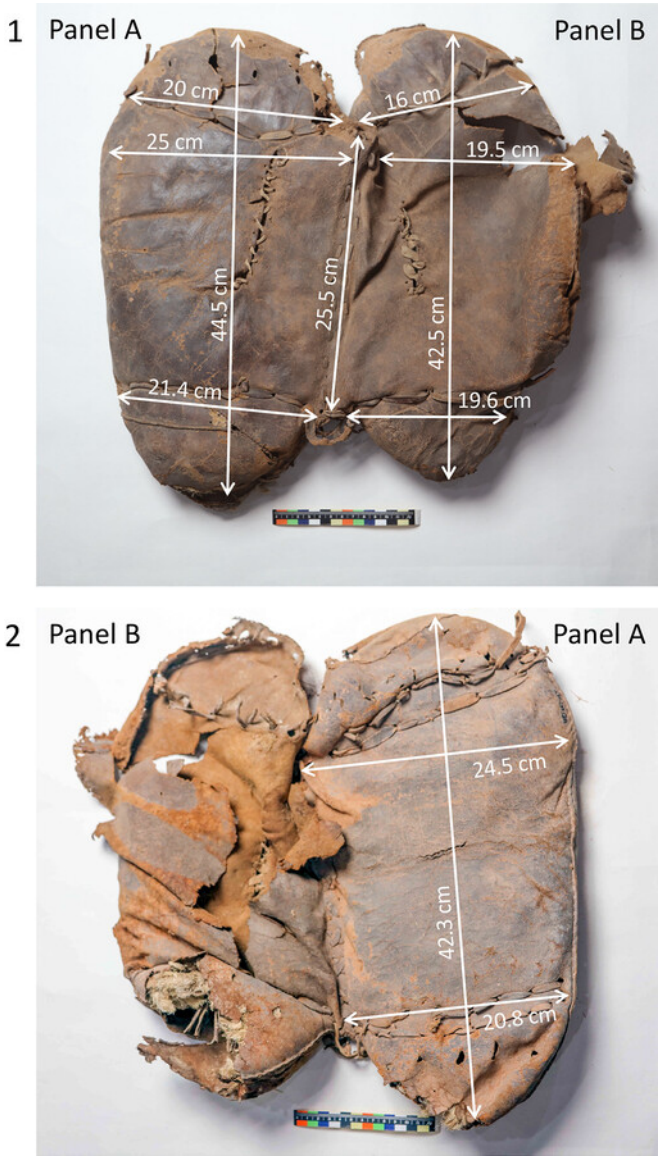

FIGURE 3 The saddle found in the tomb at the Yanghai cemetery
© WERTMANN, YIBULAYINMU, WAGNER & AL. 2023,
PHOTO BY PATRICK WERTMANN

selves and blent into the Scythian peoples there, who were, after all, their not so-distant cousins.

In my view, it is possible to partially identify the descendants of Old Steppe Iranian speakers to the Yuezhi (月氏), a nomadic pastoral people described by the Chinese, living in China during the 1st millennium BCE. The Old Steppe Iranians, or perhaps a greater federation of Steppe Iranians, woul have continued their eastwards march, and adopted an (even?) more nomadic lifestyle due to difficulties encountered in their homeland. A number of Yuezhi migrated to the west afterwards, founding the Kushan empire, and the decipherment of the Kushan script done by Bonmann, Halfmann, Korobzow & Bobomulloev (2023) and the subsequent study of this language are thus crucial in order to shed a stronger light on this question. In my view, the identification of the Tocharians to the Yuezhi, as has been done in the past, is unwarranted, however, it is also unwarranted to identify all the Yuezhi, who more probably were a federation, or perhaps even simply sparse groups with little in common, to one specific group. One is reminded of how the Chinese nowadays call every Muslim Iranian ethnic group living in or near China "Tajik", even if they are not technically Tajiks (e.g. the Sariqoli).

An interesting problem to solve, in the future, that would bear consequence on the scenarii enunciated above, is how the Tocharians entered the Tarim Basin. Was it to the north-west of the Tian Shan mountains, or through the west, reaching immediately Turfan and Ürümqi? As noted above, one can exclude that they have crossed the Kunlun Shan mountains to the south, neither their settlement pattern nor the archaeological data would support this. Even if it is established that Tocharians came through Dzhungaria, the steps they took afterwards, and how they settled in the Tarim Basin itself are important information. Whatever solution is brought forward, it needs to account for all the linguistic facts of which we are certain. It also needs to reckon with the fact that the pre-Bactrian–Tocharian B (pre-Tocharian B, in fact) contacts occurred outside of the reach of Tocharian A, in the last centuries of the 1st millennium BCE (cf. Schoubben 2024: 269–270[626]; Bernard 2024: 289–290).

I should stress again that the whole scenario is principally informed by my linguistic findings, combined with those of Dragoni (2023) and Schoubben (2024). Obviously, additional archaeological and genetic evidence is needed to see if it can be further confirmed or rather needs to be revised. However, at this point it is as coherent from the linguistic point of view as I could make it. In my view, a task for future research is to search for possible descendants of the Old Steppe Iranians, which will probably yield more insights on the prehistory of this hitherto unknown ancient Iranian population. It would also be profitable to examine in detail the possible relationship between an archaeological cul-

ture attributed to the Old Steppe Iranians—if it is possible to find one—and its close ancient Scythian relatives. In particular archaeological data connected to horse-riding and its spread in the entire region needs to be connected systematically to the genetic and linguistic information we have on these ancient Iranian peoples and the Tocharians. The present book is but a first step towards a better understanding of these ancient peoples of the Steppe which much later ended up, I believe, scattered on the silk road.

APPENDIX 1

The Tocharian Word for 'Parrot' and Its Origin

> […]
> Therefore it
> Sports in high places,
> Perches in remote regions,
> In flight, reluctant to flock at random,
> On the wing, careful in choosing its grove.
> Purple feet, cinnabar beak,
> Green jacket, kingfisher collar.
> Elegant-looking, many-colored;
> Fine-sounding: "Caw! Caw!"
>
> MI HENG (禰衡)'s Rhapsody on a parrot 鸚鵡賦 (198 A.D.) (transl. Graham 1979: 45).

In the present study, which is not entirely dependent on the topic of the rest of this book, I will investigate the etymology of a Tocharian A and B word, namely TB *yāmuttse* and TA *yāmutsi* 'parrot'.[1] This appendix might be seen as relevant to the book for three reasons: first, it concerns a word that has been claimed to be of Sogdian origin, while I claimed (fn. 29 p. 50), that there was only one known Tocharian B word of probable Sogdian etymology (namely, TAB *menāk* 'comparison, example' from Sogdian *myn'k*). As my colleague Dr. Niels Schoubben (p.c.) writes: "*menāk* can also be borrowed from Bactrian, as the relevant word is attested as Manichean Bactrian *myn'g* 'resembling, like'." Thus, once a Sogdian origin for TB *yāmuttse* is completely eliminated, there remains no Tocharian B word of possible Sogdian origin, which fact is worthy of consideration.

Second, as a word starting with *y-* and ending in TB *-e*, its etymology needs to be investigated in order to eliminate it (or not) from the list of possible Old

1 The Dictionary of Tocharian A (DToA), which was published after the redaction of this study, proposes an identical etymology to the one suggested here. However, as no extensive study on this Tocharian word was published yet, I decided to include it in the book. On this word, cf. also Chen 2019: 209.

© KONINKLIJKE BRILL BV, LEIDEN, 2025 | DOI:10.1163/9789004732537_006
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

Steppe Iranian loanwords in Tocharian. The reason I chose to discuss this word in an appendix is that neither the scholarship nor myself have considered an Old Iranian origin for this word to be a departure point for understanding its etymology, so that it appeared more meaningful to consider this word on its own.

1 The Meaning of TB *yāmuttse* and TA *yāmutsi*

TB *yāmuttse* and TA *yāmutsi* have been translated variously in the past. One common translation is '(a kind of) waterfowl' (Lubotsky & Starostin 2003: 262; DTB: 532). TA *yāmutsi* was translated by Bailey (1937: 915) as 'parrot'. Sieg and Siegling (1949: 48–49) have claimed that it corresponds to Sanskrit *haṃsa-*, which claim has been followed by Schwentner (1958). Sieg and Siegling as well as Schwentner translate the Sanskrit as 'flamingo', but *haṃsa-* means 'goose' or 'swan' (SED: 1286; cf. Huard 2020: 216). The Sanskrit word for 'goose' (*haṃsa-*) is rendered by TB *kents* 'goose'.

As proposed by Chen (2019: 209) and Huard (2020: 216), to the restoration (*yām*)*utts*(*i*)*nts*(*o*) in THT 29 a3, a reading [k] · [n]ts· and a restoration *kentsanto* should be preferred. I thus believe this occurrence of *yāmuttse* should be deleted.

In at least three Tocharian A texts (A 70 b6, A 92 b2 and A 301 a4) *yāmutsi* is closely preceded by *kokāñ*, 'ruddy shelducks'. In two of those texts (A 92 b2 and A 301 a4) *yāmutsi* is immediately preced by *ṣpārāñ*, which is evidently the plural of *ṣpār**, the name of an animal which has not yet been identified (see Poucha 1939: 89; Poucha 1955: 355 suggests it is a sparrow, Lat. *passer*).[2]

The fact that these animals are mentioned together in lists seems to indicate they shared some degree of similarity. It could thus be possible that, since *kokāñ* is a waterfowl, both *ṣpārāñ* and *yāmutsi* are also waterfowls. Nevertheless, it has been remarked upon in the literature that a Tocharian B–Old Uyghur bilingual text has shown that the Old Uyghur equivalent of the TB word *yāmuttse* is Old Uyghur *toti* 'parrot', with cognates such as Khwarezmian *twtyk*, Sogdian *twty*, New Persian *tūtī* ‹ṭwṭy› (Dari *tūtā* ‹ṭwṭ'›)[3] all 'parrot' (cf. Zieme 2005: 290;

2 This had already been suggested before—in any case, it had been suggested that they are related—cf. Bailey 1937: 915[1] who writes: "[e]tymologically Engl. *sparrow* has been compared with Agnean *ṣpār-* ['sparrow' is Skt. *caṭaka*]. Would such a meaning suit here?". By "here" Bailey refers to these two texts.

3 Note one of the earliest recorded New Persian forms, dating back to the late 10th century, ‹ṭwṭk› *ṭūṭak* or *ṭōṭak* (cf. Minorsky 1955: 257). Whatever its etymology, this word and its cog-

Peyrot, Pinault and Wilkens 2019: 78). It thus clearly appears that Tocharian B *yāmuttse* should mean 'parrot' and it is extremely likely that it is also the meaning of Tocharian A *yāmutsi*, as already proposed by Bailey (1937: 915).

The Tocharian B form, previously restored as *yāmuttsi* on the basis of the Tocharian A form should be read correctly as *yāmuttse* (cf. Peyrot, Pinault and Wilkens 2019: 78).

2 The Etymology of TB *yāmuttse* and TA *yāmutsi*

Neither the immediate nor the ultimate etymology of TB *yāmuttse* and TA *yāmutsi* 'parrot' are obvious. Parrots were not, as far as I could find, known to Proto-Indo-European speakers, and there is no inherited Indo-European name for the parrot. The forms *yāmuttse* and *yāmutsi* do not display a regular correspondence, which might suggest they have been borrowed independently in each language.

Many scholars, among which Poucha (1939: 89) and Lubotsky and Starostin (2003: 262 f.) propose to see the Tocharian words as borrowings from Chinese. This would not be surprising from a cultural perspective: although coming from the "Far West", parrots took a very important place in Chinese culture and poetry, already early on (Schafer 1959).[4]

The Chinese word from which the Tocharian words would be borrowed is *yīngwǔ*, often suffixed with the Chinese suffix *-zi* (on which see below). Poucha (1939: 89 ff.) proposes that this word was either borrowed directly from Chinese into Tocharian, or through an Iranian intermediary (this last solution is also suggested by Lubotsky and Starostin 2003: 263).

Poucha (1939: 88) directs to two passages of the famous Sogdian *Sūtra of Causes and Effects* which contain the word *ʾym'wtsy* (MacKenzie 1970: 9–11)

nates do not go back to Greek ψίτταχος 'parrot', as even Hasandoust admits (cf. Hasandoust 2014: 1957). That ψίτταχος comes from Sanskrit *śúka-* 'parrot', as claimed by some (cf. Hasandoust op. cit. with references to the literature) seems as unlikely. For the sake of completeness, I should also like to say that *tūtī*, etc. is unlikely to represent the sound of the parrot. The parrot's song, in the wild, is more akin to a whistle. The ‹ṭ› of the Persian word, coupled with the alternance between *-ī* in Iran and *-ā* seem, in the absolute, to underlie an Arabic */ṭūṭē/ < **ṭūṭaya* (with *-ē* from *alīf maqṣūrah* being rendered as *-i* < *-ī* in Iran and *-ā* in Afghanistan). The greatest obstacle to this etymology is naturally that no such Arabic word was recorded, and the Sogdian cognate makes it unlikely as well.

4 Parrots were already known in China in the second and third century BC. Their provenance was Lóngshān, on the Shaanxi-Gansu border (today in a rather eastern part of modern China). Precisely "[t]hey made their nests high in the Lung Mountains on what is now the Shensi-Kansu border, near the caravan road which leads into Central Asia" (Schafer 1959: 272).

which corresponds to Mandarin Chinese *yīngwǔzi* 鸚鵡子 'parrot' and which clearly designates a parrot in at least one of the two contexts.

The Chinese word *yīngwǔzi* was found as early as the 3rd century CE (in the translation of the *Avadānaśataka* by Zhi Qian, an Indo-Scythian immigrant, cf. *Taishō shinshū Daizōkyō* 1988 no. 200, 4.231a16).[5] The form *yīngwǔ* (without a suffix), which is also the form commonly used in modern Mandarin, is found even earlier, already at least in the 2nd century C.E. (cf. Graham 1979).

This word has no etymology in Chinese, and is likely to be a borrowing from another language, although it is unclear which. Philippe Hattat, with whom I reflected upon this word for a rather long time, suggests, convincingly in my view, to see the first element as descending from Proto-Sino-Tibetan **qreŋ* 'parrot'. A Proto-Sino-Tibetan **q-* would regularly correspond to Chinese *ʔ-* and, according to Pulleyblank (1991: 374) the first syllable of *yīngwǔ* 'parrot' in Early Middle Chinese was *ʔəɨjŋ/ʔɛ:jŋ* (Late MC *ʔja:jŋ*, Early Mandarin *jiŋ*). This does not mean that the Chinese word is inherited, as in all likelihood it is not, but it might be borrowed from a Sino-Tibetan language, more to the South and to the West, that is, closer to the Asian homeland of parrots.[6]

The Early Middle Chinese was *ʔəɨjŋ/ʔɛ:jŋ* corresponds to both Korean *aengmu-* in *aengmusae* 앵무새 'parrot'[7] and Sogdian *ʾym'wtsy*, both from Chinese. The Sogdian word has been read as *ēmūtsi* ~ *ēmutsi* (e.g. Poucha 1939: 88; Lubotsky and Starostin 2003: 262) or *imutsi* (SD: 86), but it would perhaps be more coherent to read it as **aymutsi*, if the Chinese word was not borrowed into Sogdian before the sound law PIr. **aɨ̯* > *ē*. Since this sound change occurred in all varieties of Sogdian, it must be Proto-Sogdian.

Pulleyblank (1991: 326) writes that the second syllable was pronounced as *muə̆ʿ* in Early Middle Chinese. The first two syllables of the word were thus **ʔəɨjŋmuə̆ʿ* in Early Middle Chinese. The cluster **ŋm* must have been simplified to *m* in Sogdian, thus yielding **aymutsi*. In Sogdian, a sequence *-ŋgm-* might have existed (cf. ‹ʾnkmʾn›, ‹ʾngmʾn› 'in public', ‹ʾngmʾny› 'publicly', read as *angmān* by Schwartz 1968: 144, but as *angamān* by Gharib SD: 41), but, even if Schwartz's reading is correct, the simplification might still have taken place

5 Checked from https://cbetaonline.dila.edu.tw/zh/T04n0200_p0231a16
I thank Dr. Ruixuan Chen for giving me this reference.

6 Is there a connection of any sort between **qreŋ* and Proto-West-Bahnaric **kla:ŋ* 'bird of prey'? I also entertained the thought, for some time, that the second element in the Chinese word went back to something somehow related to Proto-Austronesian **manuk* 'bird' but it is not possible as it would not have yielded the Chinese form we have at hand.

7 The syllable *sae* does not derive from Chinese *-zi*. The Korean word is composed of *aengmu* 앵무 'parrot' and *sae* 새 'bird', which is very commonly used to make bird names, such as 사다새 *sadasae* 'pelican', *dugyeonsae* 두견새 'cuckoo' from *dugyeon* 'cuckoo' + *sae* 'bird'.

as /ŋgm/ is simply different from /ŋm/, which would be much harder to realize to people who are not used to this combination. Furthermore, in Sogdian, there was no phoneme /ŋ/: [ŋ] was simply the allophonic pronunciation of /n/ in front of /g/. It is thus theoretically possible that the *-m-* found in the Tocharian A and B forms is taken from the Sogdian form.

The only two problems that remain with this word are the discrepancy between Early Middle Chinese *ʔəɨjŋmuə̆'tsɨ* (鸚鵡子) and the word-initial *yā-* element of Tocharian A and B; and the discrepancy seen in the difference of final vowel between Tocharian B *yāmuttse* and Tocharian A *yāmutsi*. The latter problem only has two solutions that I can think of: either Tocharian B *-ttse* and Tocharian A *-tsi* in this word represent different adaptations of the Early Middle Chinese *-tsɨ/-tsi* (Pulleyblank) or *-tsiX* (Baxter & Sagart) (子) diminutive suffix (Pulleyblank 1991: 420; Baxter & Sagart 2014: 159, Old Chinese **tsəʔ* according to Baxter & Sagart); or the Tocharian A form was reinterpreted as a *-tsi* feminine (of an unattested **yāmuts*). I do not see any motivation for the latter hypothesis: the first one is more likely. An easy solution to explain the Tocharian B ending is that it is secondary and analogical of other words. Indeed: many Tocharian B words end in *-tse*. Such a refection, if it happened (of *-*tsi* → *-tse*) could have happened at any stage.

The question concerning the first part of the word is difficult to solve. Poucha (1939: 89–90) has tried to solve it, but the state of the knowledge of historical phonology of both Tocharian and Chinese—in his days—kept him from making convincing proposals.[8] In what follows, I will suggest a few scenarios myself, but, if I should say the conclusion in advance, it is that I am unsure which one fits better, and that the precise detail concerning these two Tocharian words and their exact provenance remains doubtful.

The initial *y-* of TB *yāmuttse*, TA *yāmutsi* is not an inner-Tocharian problem: it is either a Chinese one, or an Iranian one. On the Iranian side, both Sogdian and Bactrian show traces of an ancient change *$a\underset{\cdot}{i}$- > *$\underset{\cdot}{i}a$-. For Sogdian, one can think of *ytkw* 'bridge' < OIr. **haiṭuka-*, for Bactrian ιωγο 'one' < OIr. **aiu̯aka-* (cf. Bernard 2019: 42). However, these changes (still unexplained in the general frame, since they seem to be "sporadic", which simply means irregular) are too old to be applied to the borrowed vocabulary. Sogdian itself does not attest a form with initial *y-* in the word *ʾym'wtsy*, and this word is not attested in Bactrian.

8 Poucha (op. cit.) suggests that there was an initial epenthetic *y-* as in TB *yakwe*, TA *yuk* 'horse', hence **ēmūtsi* > *yāmutsi*. Naturally, for *yakwe* and *yuk*, it is the sound law *$(h_1)e$- > **yä-*, a sound law which had ceased to operate since a long time at the time when TB *yāmuttse*, TA *yāmutsi* were borrowed.

First, only Sogdian, of all Iranian languages, attests this word (*ʾym'wtsy*). Second, it is only found in one Buddhist text, the *Sutra of Causes and Effects*. The Sogdian word for 'parrot', in other texts, is *twty* (SD: 393). Pelliot has thus suggested—convincingly in my view—that this is a proof that the text was translated from Chinese (Gauthiot and Pelliot 1926: 56). In fact Gauthiot and Pelliot (1926: VIII) propose to see in *ʾym'wtsy* and in four other words, not borrowings from Chinese, but Chinese transcriptions by the Sogdian translator. All in all, it is unlikely that Bactrian, unless in a specific and similar Buddhist context, would have borrowed such a word from Chinese, even more when parrots were also known to Bactrians from elsewhere than China. We do not have texts of this type in Bactrian, that is, translations of Chinese texts with a great number of borrowings (borrowings from Chinese in Bactrian are rare and restricted semantically to the political sphere). The Khotanese word for 'parrot' is *īmūka-* (DKS: 33). Khot. *īmūka-* is hardly a convincing etymon for Tocharian B *yāmuttse*, Tocharian A *yāmutsi*, although it might be a cognate thereof.

No other Iranian language that I know of seems to present a form similar to the one mentioned above for the word 'parrot'. All in all and as far as we know, it does not seem there is a plausible phonetic scenario to derive the Tocharian forms from Iranian forms. Probably, the solution must come from Chinese.

The Late Middle Chinese pronunciation of this word must have been *ʔja:jŋvuă'-* (cf. Pulleyblank 1991: 374, 326). Perhaps a specific dialect of Chinese, in contact with Tocharian, would have had already an initial *ʔja:* at the time of contact with Tocharian. One can thus hypothesize that *ʔja:jŋvuă'tsɨ* was, at some period of time and in some specific dialect, **(ʔ)jajŋvuătsɨ*, simplified to **yajŋvuătsɨ*.

As to the second part, that is, Tocharian *-m-* corresponding to Chinese *-ŋvu-* or *-ŋm-*, it is either a simplification due to the fact that Tocharian did not possess a phoneme /ŋ/ (compare Middle Chinese 青 *tsʰɛjŋ*, modern Chinese *qīng* 'blue' → TAB *tseṃ* 'blue', cf. Lubotsky & Starostin 2003: 264) or a different phenomenon: the assimilation of the labial feature of the glide to the nasal **ŋ*.[9] A somewhat comparable phenomenon is observed in TB *mätstsa-* 'to die of hunger, to starve', cognate of TA *nätswā-* 'id.', where it has been assumed that the *m-* of the TB form derives from **n-* assimilated to the *-w-* (cf. DTB: 493).

The **-ɨ* vowel of the *-zi* suffix (or **-iX* according to the Baxter/Sagart reconstruction) would have been adapted as *-e* in Tocharian B and *-i* in Tocharian A.

9 An identical example of this hypothetical sound change is found in Batley English (Yorkshire) [lʌmwɛd͡ʒ] 'language', a form I was able to record myself in a middle-aged Batley English speaker.

The internal Chinese etymology of *yīngwǔzi* 鸚鵡子 'parrot' is unknown. It is likely to be a loanword, since its form cannot be explained from a Chinese point of view, and parrots are not initially indigenous to the territories from where Chinese was initially spoken (cf. Schafer 1959). Besides, there are other problems with this word: its reconstruction is disputed, and its characters (鸚 and 鵡) are only found in this particular word. There are problems found in other apparent cognates. In particular, I would like to point out that besides Japanese *ōmu* 'parrot' (← **ʔaŋmu-*) and *inko* (a later borrowing which reflects the Modern Mandarin pronunciation *yīng*), one finds the reading *yō* for the character 鸚, which indicates an onset with a glide. This could indicate various readings of this word in Chinese dialects.

In conclusion, although the details of the derivation of TB *yāmuttse* TA *yāmutsi* 'parrot' remain mysterious, I believe one should rather look at a direct borrowing from some variety of Chinese, without an Iranian intermediary. The only clear cognate of the Chinese word *yīngwǔ-zi* 鸚鵡子 'parrot' in an Iranian language appears in a Chinese translation into Sogdian. In Sogdian, it is a marginal word, not likely to have been used outside of literary contexts, and unlikely to have been borrowed by Tocharian, not only because of this. Its form ‹*ʾymʾwtsy*› could reveal an initial *ay-* or *ē-*, but not *ya-*. Another problem is caused by the different final syllable *-ttse* in TB and *-tsi* in A, which could perhaps indicate two independent borrowings from Chinese.

APPENDIX 2

On the Etymology of TB *patstsāṅk* and TA *pātsaṅk* 'Window'

> Lo! in yon brilliant window-niche
> How statue-like I see thee stand,
> The agate lamp within thy hand!
> Ah, Psyche, from the regions which
> Are Holy-Land!
>
> EDGAR ALLAN POE, *To Helen*

The first etymology suggested for Tocharian B *patstsāṅk* and Tocharian A *pātsaṅk* 'window' was made by Poucha (1939: 216). He proposed that these two words derive from an earlier Chinese form related to "Northern Chinese" *bōlí* (玻璃) *chūang* (窗) (transcription adapted) 'window' (with the element *bōlí* 'glass'). More precisely he suggested they go back to Old Chinese *"p'uâ-ṭṣ'âng"* 'Glas-Fenster'. Poucha admits that this collocation so far is not found in Old Chinese literature, and ends by saying that this is an important etymology, as it would indicate knowledge of glass windows among Tocharians.

Poucha's etymology unfortunately does not work on the phonetic level, and is very unlikely from a historical point of view. Glass-windows appeared rather late in Asia, and they came from the West. In Eastern Asia, windows were in general filled with paper (cf. *Encyclopedia Britannica online*, s.v. *window*).[1] Phonetically there is no way to obtain *-ts-* from **-č-* in Chinese. If from Chinese, one would expect *-ts-* to reflect Chinese *-ts-*, as in TB *ṣitsok* 'millet alcohol' (see p. 220); TB *yāmuttse* 'parrot' and TA *yāmutsi*, see Appendix 1. Because glass-windows were imported from the West, the Tocharians, had they acquired them, should have borrowed their word for it from a Western language rather than from Chinese.

1 See https://www.britannica.com/technology/window (consulted on the 4th of February 2022).

© KONINKLIJKE BRILL BV, LEIDEN, 2025 | DOI:10.1163/9789004732537_007
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

Another proposal which has been made concerning the words TB *patstsāṅk*, TA *pātsaṅk* is found in Wilkens's *Handwörterbuch des Altuigurischen* (Wilkens 2021). In it, Wilkens suggests (2021: 548) that it is a cognate of Old Uyghur *pačiŋ* 'fence', from Chinese 爸城 *ba cheng*, Late Middle Chinese *pa: ʂʱiajŋ*. The Chinese word (*bāchéng*) means 'stockade'. This etymology is very difficult to accept, notably because of the semantic difference between a stockade and a window. Besides, I do not understand how it would work phonetically.

Isebaert (1980: 50–51) proposed that Tocharian B *patstsāṅk* and Tocharian A *pātsaṅk* go back to a single proto-form, itself going back to Middle Iranian **patihāng* < **pati-frānaka-*. This derivation is commonly accepted (cf. DTB: 388; DToA: 274), but it is phonetically impossible, as only Proto-Indo-European **ti* yields Tocharian *ts*, and even Old Steppe Iranian **-ti-* only yields Proto-Tocharian **tə* (cf. **abí-stH-(i̯)i̯a-* > **abí-sti-i̯a-* > PT **epə́stəye*). This proto-form of Isebaert's would not explain the gemination in Tocharian B either, which gemination might be old.

This proposal of Isebaert's is motivated by the Armenian word *patowhan* 'window', deriving from **pati-frāna-* 'ouverture d'aération' (cf. Benveniste 1957–1958: 63–64). The Armenian word obviously derives from a Middle Iranian source which underwent the **-fr-* > **-hr-* sound change (preserved in dialectal forms: Agulis *ptə́rhan*, Łarabał *ptṙhan*, cf. Martirosyan, summer school class, Leiden 2019). Martirosyan (idem) proposes thus a proto-form **paturhan*, simplified within Armenian. Possibly the Classical Armenian form is a direct loan from a later stage of the Iranian language form (**patəhān*?). Tremblay (2005: 435) goes even further than Isebaert, and proposes that the Tocharian words directly derive from specifically Parthian **patihān*.

A further problem with this Iranian derivation, apart from the impossiblity to explain the *-ts*(*ts*)- found in Tocharian, resides in the length of the vowels. Tocharian B does not reveal anything concerning the length of the **a* in this word, other than that the second syllable was accented, but Tocharian A could show that the second **a* is short (cf. Schwartz 1974: 406). Possibly, this means that the Iranian form these words go to—if there is any—should be **pă̆tsang* or **pă̆dzang*. Of course, one could easily object that an original TA **pātsāṅk* or **patsāṅk*, if borrowed early enough, could have been shortened into *pātsaṅk* (Krause & Thomas 1960: 45 f.). An internal Tocharian argument, however, cannot be invoked to explain away the *-ts*(*ts*)-.

I propose that this word is actually of Bactrian origin. In favor of this hypothesis: 1. the sequence *-ts-* is very common in Bactrian (it derives from an earlier **č*, **tVs*, **tVč*, and maybe **ǰ*). 2. The *-aṅk* ending of Tocharian usually corresponds to the Bactrian ending -αγγο, as demonstrated by Schwartz (1974: 399–411).

A possible source of this word should be an etymon **pāta-* 'protected by' or **pāti-* 'against' + **čanaka-*. Semantically, it seems very difficult since **čanah-* means 'pleasure'. One could perhaps also think of **pāti-āsānaka-* 'against (= towards/opposite to) the (sun)rise', implying that windows are placed in the direction of the sun. The phonetics would thus be **pāti-āsānaka-* > Bactrian *παδιασαγγ → TAB **patyasaṅk* with syncope (*a_a_a > a_ɔ_a) > **patstsaṅk*. Perhaps even *παδιασαγγο > *παδισαγγο > *patstsāṅk*. A possible parallel for the latter change is found in Bactrian μιροσανο 'sunrise', which derives from μιυροασανο. Unfortunately, this change is quite late, too late for a Tocharian loanword: it is attested in the documents V and W, which date back to the eighth century of our era. The semantics are not very strong either.

As the above suggestions are unconvincing, I would like to suggest another option, which might sound surprising at first. In my view, the Tocharian words discussed above derive from a Middle Iranian reflex of PIr. **pāti-ǰanaka-* 'onto (= for) murder'. There exists at least two parallels for such an etymology.

First, in French *une meurtrière* 'murderer (fem.)' is a type of window found in strongholds, castles, defensive walls and structures, precisely called as such because one can use it to kill without being killed. In English, this type of window is called *arrowslit* but also *loophole*: it is a narrow but long window, large enough for arrows to be shot out of.

Second, in Latin, *fenestra* which later became French *fenêtre*, German *Fenster* 'window' originally meant 'trou pratiqué dans une paroi, meurtrière' in Latin (Ernout & Meillet 1967: 225) then took on the meaning 'window' not only in Romance languages but also in Latin itself.

Of course, this etymology is speculative to an extent, but one might wonder if such a word did not exist in Bactrian as well. Indeed: arrowslits were found in citadels in Bactria (e.g. Zadiyan Qala, cf. de la Vaissière, Marquis & al. 2015: 246, 250; Shaydullaev 2011: 95). Bactrian loopholes varied in size, in Bactria they were 20 cm wide, equally wide throughout the wall; in Kampyrtepa their size varies from 15 to 30 cm (Shaydullaev 2011: 95[4]). From Shaydullaev's drawing of the last remaining loophole of Jandavlattepa (Shaydullaev 2011: 95 fig. 2.2., 4) one can see that it was constituted of two upper bricks put against together in a triangular shape, with 22 cm width at the basis, and below, bricks facing each other with again some width in between ("a loophole in the shape of an arrow" as per Shaydullaev's own words). This shape reminds one of early European church and castle windows, with the only difference that these were wider.

The Bactrian reflex of **pāti-ǰanaka-*,*παδοζιγγο, would yield the Tocharian forms quite straightforwardly. It is similar to the borrowing of Bactrian φρομιγγο as TAB *pärmaṅk* 'hope'. Then, one can imagine that the meaning would expand to the more general meaning of 'window' in Tocharian. Note that a change from

'arrowslit, loophole' to 'window' is much more likely than a shift in the opposite direction. Dr. Romain Garnier suggests to me that Lat. *fenestra* itself would reflect Proto-Italic *χ^{w}*en-estro-* 'striking place' ultimately going back to the root *$g^{u}{}^{h}$*en-* 'to strike, to kill', the same root from which Iranian **ǰan-* derives. That the Tocharian word designated at least other types of windows than arrowslits is guaranteed by, for instance, A 253 b3, where it is mentioned that adorned women jump from the high windows of a wondrous palace, only to bounce back up again. This clearly designates wider types of windows, as human adults, no matter how thin, could not jump through most arrowslits.

In conclusion, there is no clear, evident etymology for Tocharian B *patstsāṅk*, Tocharian A *pātsaṅk* 'window'. In any case, Isebaert's etymology from Middle Iranian **patihāng*, also found in Tremblay (2005: 435) cannot be upheld. The word has a seemingly Bactrian appearance, containing probably the Bactrian suffix -αγγο, reflected in Tocharian as *-aṅk*. I suggest that this word ultimately derives from the Bactrian descendant of an Old Iranian **pati-ǰanaka-* or **pāti-ǰanaka-*, parallel to French *meurtrière* (and possibly Latin *fenestra*), which would have been generalized to other types of windows in Tocharian.

Bibliography

Book Abbreviations

AiW Christian Bartholomae (1904). *Altiranisches Wörterbuch*. Strassburg: Verlag von Karl J. Trübner.

BHSD Franklin Edgerton (1953). *Buddhist Hybrid Sanskrit Grammar and Dictionary, Vol. II. Dictionary*. New Haven.

CDIAL Sir Ralph L. Turner (1962–1966). *A Comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.

CPD David N. MacKenzie (1971). *A Concise Pahlavi Dictionary*. London: Oxford University Press.

DKS Harold W. Bailey (1979). *Dictionary of Khotan Saka*. Cambridge–London–New York–Melbourne: Cambridge University Press.

DMMPP Desmond Durkin-Meisterernst (2004). *Dictionary of Manichaean Middle Persian and Parthian*. Turnhout: Brepols.

DTB Douglas Q. Adams (2013). *A Dictionary of Tocharian B. Revised and Greatly Enlarged*. Amsterdam–New-York: Rodopi.

DToA Gerd Carling & Georges-Jean Pinault (2023). *Dictionary and Thesaurus of Tocharian A*. Wiesbaden: Harrassowitz Verlag.

EDIV Johnny Cheung (2007). *Etymological Dictionary of the Iranian Verb*. Leiden–Boston: Brill.

EWAia Manfred Mayrhofer (1992–2001). *Etymologisches Wörterbuch des Altindoarischen*. 3 vols. Heidelberg: Carl Winter.

LIV^2 Helmut Rix & al. (2001). *Lexikon der Indogermanischen Verben*. Wiesbaden: Dr. Ludwig Reichert.

M-W Sir Monier Monier-Williams (1899). *A Sanskrit-English Dictionary. Etymologically and philologically arranged with special reference to cognate Indo-European languages*. Oxford: Oxford University Press.

SD Badr-uz-Zamān Gharib (1995). *Sogdian Dictionary: Sogdian–Persian–English*. Tehran: Farhangan.

TEB II Werner Thomas, with the help of Wolfgang Krause (1964). *Tocharisches Elementarbuch. Band II. Texte und Glossar*. Heidelberg: Carl Winter.

References

Abaev, Vasilij Ivanovič (1949). *Osetinskij jazyk i fol'klor*. (Ossetic language and folklore). Moscow–Leningrad: Izdatelstvo nauk USSR.

Abaev, Vasilij Ivanovič (1965). *Skifo-evropejskie izoglossy na styke vostoka i zapada* (Scythian-European isologlosses at the junction of the East and the West). Moscow.

Abaev, Vasilij Ivanovič (1973). *Istoriko-ètimologičeskij slovar' osetinskogo jazyka II*. (Historical and etymological dictionary of the Ossetic language. Volume II). Moscow–Leningrad: Izdatelstvo nauk USSR.

Abaev, Vasilij Ivanovič (1989). *Istoriko-ètimologičeskij slovar' osetinskogo jazyka IV*. (Historical and etymological dictionary of the Ossetic language. Volume IV). Moscow–Leningrad: Izdatelstvo nauk USSR.

Adams, Douglas Q. (1999). *A dictionary of Tocharian B*. Amsterdam–Atlanta: Rodopi.

Adams, Douglas Q. (2017). "Tocharian" In: Mate Kapović (ed.), *The Indo-European Languages, Second Edition*. London–New York: Routledge, 452–475.

Adams, Douglas Q. (2017a). "The lexicon of Tocharian" In: Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*. Berlin–Boston: de Gruyter, 1365–1388.

Asmussen, Jes P. (1965). *Xuāstvānīft. Studies in Manichaeism*. Translated by Niels Haislund. Copenhagen: Prostant Apud Munksgaard.

Back, Michael (1978). *Die sassanidischen Staatinschriften*. Leiden–Tehran–Liège: Bibliothèque Pahlavi.

Bailey, Harold W. (1937). "Ttaugara" *Bulletin of the School of Oriental Studies* 8, 883–922.

Bailey, Harold W. (1958). "Languages of the Saka" In: *Iranistik, IV:1 Linguistik*. Leiden–Köln: Brill, 131–154.

Bailey, Harold W. (1961). Review of: E. Benveniste (1959), *Études sur la langue ossète*. *Journal of the Royal Asiatic Society of Great Britain and Ireland* 1/2, 53–56.

Bailey, Harold W. (1967). *Indo-Scythian Studies. Khotanese Texts volume VI. Prolexis to the Book of Zambasta*. Cambridge: Cambridge University Press.

Bailey, Harold W. (1975). "The second stratum of the Indo-Iranian gods" In: John R. Hinnels (ed.), *Mithraic Studies. Proceedings of the first International congress of Mithraic Studies. Volume II*. Manchester: Manchester University Press, 1–20.

Bailey, Harold W. (1982). *The Culture of the Sakas in Ancient Khotan*. (Colombia Lectures on Indian studies edited by Ehsan Yarshater). New-York: Delmar.

Bailly, Anatole (2000). *Dictionnaire grec français*. (New edition, original edition 1894). Paris: Hachette.

Bartholomae, Christian (1895–1904). "Vorgeschichte der Iranischen Sprachen" In: Wilhelm Geiger & Ernst Kunst (eds.), *Grundriss der iranischen Philologie. Erster Band. I. Abteilung*. Strassburg: Verlag von Karl J. Trübner.

Bartholomae, Christian (1906). *Zum Altiranischen Wörterbuch*. Strassburg: Verlag von Karl J. Trübner.

Bartholomae, Christian (1916). *Zur Kenntnis der mitteliranischen Mundarten. I*. Heidelberg: Carl Winter.

Baxter, William H. & Laurent Sagart (2014). "Old Chinese reconstruction, version of

20 September 2014". Online text, retrievable on https://sites.lsa.umich.edu/ocbaxtersagart/wp-content/uploads/sites/1415/2025/03/BaxterSagartOCbyMandarinMC2014-09-20.pdf

Beekes, Robert S.P. (1988). *A Grammar of Gatha-Avestan*. Leiden: Brill.

Beekes, Robert S.P. (2010). *Etymological Dictionary of Greek*. Leiden–Boston: Brill.

Beekes, Robert S.P. (2011). *Comparative Indo-European Linguistics. An Introduction. Second Edition. Revised and Corrected by Michiel de Vaan*. Amsterdam–Philadelphia: John Benjamins.

Behr, Wolfgang (2001). Review of: Jan Ulenbrook (1998), *Zum Alteurasischen. Eine Sprachvergleichung*. *Oriens* 36:1, 356–361.

Behr, Wolfgang (2004). "*Hic sunt leones*—two ancient Eurasian migratory terms in Chinese revisited (1)" *International Journal of Central Asian Studies* 9, 1–24.

Behr, Wolfgang (2005). "*Hic sunt leones*—two ancient Eurasian migratory terms in Chinese revisited (2)" *International Journal of Central Asian Studies* 10, 1–28.

Benveniste, Émile (1946). *Vessantara Jātaka. Texte sogdien édité, traduit et commenté*. Paris: Librairie Orientaliste Paul Geuthner.

Benveniste, Émile (1957–1958). "Mots d'emprunt iraniens en arménien" *Bulletin de la Société de Linguistique de Paris* 53/1, 55–71.

Benveniste, Émile (1966). *Titres et Noms propres en iranien ancien*. Paris: Klincksieck.

Benzing, Johannes (1983). *Chwaresmischer Wortindex*. Wiesbaden: Otto Harrassowitz.

Berger, Hermann (1998). *Die Burushaski-Sprache von Hunza und Nager*. Wiesbaden: Otto Harrassowitz.

Bernard, Chams B. (2019). "On the etymology of Persian *yak* 'one'" *Wékʷos* 4, 41–55.

Bernard, Chams B. (2020). "Some plant and animal names in Gavruni" In: Romain Garnier (ed.), *Loanwords and Substrata. Proceedings of the Colloquium held in Limoges* (5th–7th June, 2018). Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck, 27–61.

Bernard, Chams B. (2022). "Une étymologie brillante: sur l'origine de PIr. **xšaiṭa-* "brillant, chef"" *Wékʷos* 6, 25–29.

Bernard, Chams B. (2022a). "Once more on 'one': elements in favor of MP *yak* and Kim's law of Ossetic" *Wékʷos* 6, 31–37.

Bernard, Chams B. (2024). Review of: Carling, Gerd & Georges-Jean Pinault, *Dictionary and Thesaurus of Tocharian A*. *Wékʷos* 7, 285–304.

Bernard, Chams B. & Chen Ruixuan (2022). "A fall in the pit. On Tocharian B *koṣko*, *koṣkīye*" *Indo-Iranian Journal* 65, 1–31.

Bielmeier, Roland (1987). "Zum Namen der kaukasischen Iberer" In: Piotr O. Scholz & Reinhard Stempel (eds.), *Nubia et Oriens Christianus: Festschrift für C. Detlef G. Müller zum 60. Geburtstag*. Köln: J. Dinter, 99–105.

Blažek, Václav & Michal Schwarz (2017). *Early Indo-Europeans in Central Asia and China. Cultural relations as reflected in language*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.

Bonmann, Svenja, Natalie, Korobzow, Bobomullo Bobomulloev (2023). "A Partial Decipherment of the Unknown Kushan Script" *Transactions of the Philological Society* 121, 293–329.

Boyce, Mary (1954). *The Manichaean hymn cycles in Parthian*. London: Oxford University Press.

Buck, Carl D. (1949). *A Dictionary of Selected Synonyms in the Principal Indo-European Languages. A contribution to the history of ideas*. Chicago–London: University of Chicago Press.

Buyaner, David B. (2021). "K voprosu o predystorii iranskix sibiljantov (na materiale predpolagaemyx iranskix zaimstvovanij v armjanskom i toxarskom) [On the prehistory of Iranian sibilants (based on putative Iranian loanwords in Armenian and Tocharian)]" *Rodnoy Yazyk, Festschrift for Joy Edelman*. Moscow: Institute of Linguistics of the Russian Academy of Sciences, 64–86.

Canevascini, Giotto (1993). *The Khotanese Saṅghātasūtra. A critical edition*. Wiesbaden: Dr. Ludwig Reichert.

Cantera, Alberto (2017). "The phonology of Iranian" In: Jared Klein, Joseph Brian & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*. Berlin–Boston: de Gruyter, 481–503.

Carling, Gerd (2000). "New look at the Tocharian B medical manuscript IOL Toch 306 (Stein Ch.00316.a2) of the British Library—Oriental and India Office Collections" *Historische Sprachforschung* 116, 75–95.

Carling, Gerd (2005). "Proto-Tocharian, Common Tocharian, and Tocharian—on the value of linguistic connections in a reconstructed language" In: K. Jones-Bley, M.E. Huld & A. Demma Volpe (eds.), *Proceedings of the Sixteenth Annual UCLA Indo-European Conference, Los Angeles, November 5th–6th, 2005*. Washington, D.C.: Institute for the Study of Man, 47–70.

Carling, Gerd (2009). *Dictionary and Thesaurus of Tocharian A*. Part 1: A–J. Compiled by Gerd Carling, in collaboration with Georges-Jean Pinault and Werner Winter. Wiesbaden: Otto Harrassowitz.

Carling, Gerd (2020). "A dangerous story: the linguistic behavior of the category sharp cutting implements" In: Romain Garnier (ed.), *Loanwords and Substrata. Proceedings of the Colloquium held in Limoges* (5th–7th June, 2018). Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck, 121–149.

Cerasetti, Barbara, Roberto Arciero, Traci N. Billings, Maurizio Cattani & al. (2022). "The Rise and Decline of the Desert Cities. The Last Stages of the BMAC at Togolok 1 (Southern Turkmenistan)". In: Christoph Baumer, Mirko Novák and Susanne Rutishauser (eds.), *Cultures in Contact. Central Asia as Focus of Trade, Cultural Exchange and Knowledge Transmission*. Wiesbaden: Harrassowitz Verlag, 89–115.

Chavannes, Édouard (1911). *Cinq cents contes et apologues extraits du Tripiṭaka chinois. Tome II*. Paris: Ernest Leroux.

Chen, Ruixuan (2019). Review of: Olav Hackstein, Hiromi Habata & Christoph Bross (2019), *Tocharische Texte zur Buddhalegende. Kratylos* 65, 203–211.

Chen, Ruixuan & Chams Bernard (2024). "A Spicy etymology. Remarks on Tocharian AB *śāñcapo*" *Studia Indica* 1, 423–444.

Cheung, Johnny (2002). *Studies in the Historical Development of the Ossetic Vocalism*. Wiesbaden: Dr. Ludwig Reichert.

Cheung, Johnny (2005). Review of: Georg Morgenstierne (2003), *A New Etymological Vocabulary of Pashto. Bulletin of the School of Oriental and African Studies* 68: 1, 127–129.

Cheung, Johnny (2010). "Selected Pashto Problems I: The Accent in Pashto" *Persica* 23, 109–121.

Cheung, Johnny (2011). "Selected Pashto Problems II. Historical Phonology 1: On Vocalism and Etyma" *Iran and the Caucasus* 15, 169–205.

Cheung, Johnny (2013). "Pashto Problems III. Ancient Loanwords from Early New Persian and Indo-Aryan and the Historical Contacts of Pashto & its Speakers" In: Sergei Tokhtasev & Pavel Lurje (eds.), *Commentationes Iranicae. Vladimiro f. Aaron Livschits nonagenario donum natalicium*. Saint-Petersburg: Nestor-Historia, 614–630.

Cheung, Johnny (2015). "The Persian Verbal Suffixes *-ān* and *-andeh* (*-andag*)". In: Uwe Bläsing, Victoria Arakelova & Matthias Weinreich (eds.), *Studies on Iran and the Caucasus: Presented to Prof. Garnik S. Asatrian on the Occasion of his 60th Birthday*. Leiden–Boston: Brill, 271–290.

Cheung, Johnny (2017). "On the Origin of the Terms "Afghan" and "Pashtun" (Again)" In: Enrico Morano, Elio Provasi & Adriano V. Rossi (eds.), *Studia Philologica Iranica. Gherardo Gnoli Memorial Volume*. Roma: Scienze e Lettere, 31–49.

Ching, Chao-jung (2010). *Secular documents in Tocharian: Buddhist economy and society in the Kucha region*. Paris: PhD dissertation.

Ching, Chao-jung (2011). "Silk in ancient Kucha: on the Toch. B word *kaum** found in the documents of the Tang period" *Tocharian and Indo-European Studies* 12, 63–82.

Ching, Chao-jung & Ogihara Hirotoshi (2020). *Qiūcí shíkū tíjì* [Cave inscriptions in ancient Kucha]. Shànghǎi: Zhōngxī Shūjú.

Chung, Jin-il (2004). *Das Upasaṃpadā-vastu: Vorschriften für die bud-dhistische Mönchsordination im Vinaya der Sarvāstivāda-Tradition: Sanskrit-Version und chinesische Version*. Göttingen: Vandenhoeck & Ruprecht.

Ciancaglini, Claudia A. (2008). *Iranian Loanwords in Syriac*. Wiesbaden: Dr. Ludwig Reichert.

Ciancaglini, Claudia A. (2015). "Allomorphic Variability in the Middle Persian continuants of the Old Iranian suffix **-ka-*" In: Uwe Bläsing, Victoria Arakelova & Matthias Weinreich (eds.), *Studies on Iran and the Caucasus: Presented to Prof. Garnik S. Asatrian on the Occasion of his 60th Birthday*. Leiden–Boston: Brill, 291–308.

Clauson, Sir Gerard (1972). *An Etymological Dictionary of Pre-Thirteenth-Century Turkish*. Oxford: Clarendon Press.

Couvreur, Walter (1964). "Nieuwe Koetsjische fragmenten van het Araṇemi-jātaka" *Orientalia Gandensia* 1, 237–249.

Debaine-Francfort, Corinne & Abduressul Idriss (2001). "Djoumboulak Koum, une cité fortifiée." In: Corinne Debaine-Francfort & Abduressul Idriss (eds.), *Keriya, mémoires d'un fleuve: Archéologie et civilisation des oasis du Taklamakan*. Suilly-la-Tour: Findakly, 120–136.

Degener, Almuth (1989). *Khotanische Suffixe*. Stuttgart: Franz Steiner Verlag Wiesbaden GmbH.

Del Olmo Lete, Gregorio & Joaquín Sanmartín (2003). *A Dictionary of the Ugaritic Language in the Alphabetic Tradition*. Translated by Wilfred G.E. Watson. Leiden–Boston: Brill.

Del Tomba, Alessandro (2020). "The development of PIE *-*oi̯* in Tocharian" *Rivista degli studi orientali*, nuova serie 93(1/2), 21–34.

Del Tomba, Alessandro (2023). *The Tocharian Gender System. A Diachronic Study in Nominal Morphology*. Leiden–Boston: Brill.

Derksen, Rick (2008). *Etymological Dictionary of the Slavic Inherited Lexicon*. Leiden–Boston: Brill.

Dragoni, Federico (2021). "Materia Medica Tocharo-Hvatanica" *Bulletin of the School of Oriental and African Studies* 84, 295–319.

Dragoni, Federico (2023). *Watañi lāntaṃ. Khotanese and Tumshuqese Loanwords in Tocharian*. Wiesbaden: Dr. Ludwig Reichert.

Duchesne-Guillemin, Jacques (1936). *Les composés de l'Avesta*. Liège: Faculté de Philosophie et Lettres.

Duchesne-Guillemin, Jacques 1941. "Tocharica" *Bulletin de la Société de Linguistique* 41, 140–182.

Dunkel, George E. (2014). *Lexikon der indogermanischen Partikeln und Pronominalstämme. Band 2. Lexikon*. Heidelberg: Universitätsverlag Winter.

Eilers, Wilhelm (1962). "Iranisches Lehngut im Arabischen Lexikon: über einige Berufsnamen und Titel" *Indo-Iranian Journal* 5, 203–232.

Emmerick, Ronald E. (1968). *The Book of Zambasta. A Khotanese poem on Buddhism*. London: Oxford University Press.

Emmerick, Ronald E. (1970). "Nanda the Merchant" *Bulletin of the School of Oriental and African Studies* 33, 72–81.

Emmerick, Ronald E. (1989). "Khotanese and Tumshuqese" In: Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*. Wiesbaden: Dr. Ludwig Reichert, 204–229.

Ernout, Alfred & Antoine Meillet (1967). *Dictionnaire étymologique de la langue latine. Histoire des mots. Quatrième édition, deuxième tirage augmenté de corrections nouvelles*. Paris: Klincksieck.

Fellner, Hannes (2006). "On the Development of Labiovelars in Tocharian" In: Karlene Jones-Bley, Martin E. Huld, Angela Della Volpe & Miriam Robbins Dexter (eds.), *Proceedings of the Seventeenth Annual UCLA Indo-European Conference, Los Angeles 2005*. Washington: Institute for the Study of Man, 51–65.

Fellner, Hannes (2007). "The Expeditions to Tocharistan" In: Melanie Malzahn (ed.), *Instrumenta Tocharica*. Heidelberg: Carl Winter, 13–36.

Filippone, Ella (2017). "On the Meaning of Avestan *nāuuiia-* and Pahlavi **nāydāg** " In: Enrico Morano, Elio Provasi & Adriano V. Rossi (eds.) *Studia Philologica Iranica. Gherardo Gnoli Memorial Volume*. Rome: Ismeo, 99–138.

Filliozat, Jean (1948). *Fragments de textes koutchéens de médecine et de magie. Texte, parallèles sanskrits et tibétains, traduction et glossaire*. Paris: Librairie d'Amérique et d'Orient Adrien-Maisonneuve.

Francfort, Henri-Paul (2009). "L'âge du bronze en Asie centrale. La civilization de l'Oxus" *Anthropology of the Middle East* 4, 91–111.

Francfort, Henri-Paul (2018). "Die Siegel der Oxus-Zivilisation" In: Matthias Wemhoff, Manfred Nawroth, Rainer-Maria Weiss & Alfried Wieczorek (eds.) *Margiana. Ein Königreich der Bronzezeit in Turkmenistan*. Petersberg: Michael Imhof Verlag, 133–139.

Francfort, Henri-Paul (2018a). "Symbole der Macht" In: Matthias Wemhoff, Manfred Nawroth, Rainer-Maria Weiss & Alfried Wieczorek (eds.) *Margiana. Ein Königreich der Bronzezeit in Turkmenistan*. Petersberg: Michael Imhof Verlag, 133–139.

Francfort, Henri-Paul (2023). "Behind and Beyond the Archaeology of the Silk Road: Laboratory Analyses in Eurasia, Some Results, Discussions, and Interpretations for Protohistory and Antiquity" *Acta Via Serica* 8, 53–78.

Frisk, Hjalmar (1960). *Griechisches Etymologisches Wörterbuch*. Heidelberg: Carl Winter.

Garnier, Romain (2019). Review of: Manfred Brust (2017), *Historische Laut- und Formenlehre des Altpersischen. Mit einem etymologischen Glossar*. *Wékʷos* 4, 366–373

Garnier, Romain & Jacques François (2020). Review of: Guus Kroonen, James P. Mallory & Bernard Comrie (2018, eds.), *Talking Neolithic. Proceedings of the workshop on Indo-European origins held at the Max Planck Institute for Evolutionary Anthropology*. *Bulletin de la Société de Linguistique* 115:2, 48–63.

Gauthiot, Robert E. (1912). "Une version sogdienne du *Vessantara Jātaka*; publiée en transcription et avec traduction" *Journal Asiatique* 19, 429–510.

Gauthiot, Robert E. (1913). *Essai sur le vocalisme du sogdien*. Paris: Librairie Paul Geuthner.

Gauthiot, Robert, and Paul Pelliot. 1926. *Le Sûtra Des Causes et Des Effets Du Bien et Du Mal*. 2,1: *Transcription, Traduction, Commentaire et Index*. Paris: Geuthner.

van Geel, Bas, N.A. Bokovenko, N.D. Burova, K.V. Chugunov, V.A. Dergachev, V.G. Dirksen, & al. (2004). "Climate change and the expansion of the Scythian culture after 850 BC, a hypothesis" *Journal of Archaeological Science* 31, 1735–1742.

Geiger, Wilhelm (1901). *Die Sprache der Afghānen, das Paṣtō*. Wilhelm Geiger & Ernst Kunst (eds.), *Grundriss der iranischen Philologie. Erster Band. 2. Abteilung*. Strassburg: Verlag von Karl J. Trübner, 201–230.

George, Ken (2020). *An Gerlyver Meur. Cornish-English English-Cornish Dictionary*, 3rd edition. Llandysul: the Cornish Language Board.

Gershevitch, Ilya (1961). *A Grammar of Manichean Sogdian*. Oxford: Basil Blackwell.

Gershevitch, Ilya (1971). "Iranian words containing *-ăn*" In: Clifford E. Bosworth (ed.), *Iran and Islam, in memory of the late Vladimir Minorsky*. Edinburgh: Edinburgh University Press, 267–291.

Gholami, Saloumeh (2014). *Selected Features of Bactrian Grammar*. Wiesbaden: Harrassowitz.

Graham, William T. Jr. (1979) "Mi Heng's *Rhapsody on a Parrot*" *Harvard Journal of Asiatic Studies* 39, 39–54.

Hackstein, Olav, Hiromi Habata & Christoph Bross (2019). *Tocharische Texte zur Buddhalegende*. Dettelbach: Münchener Studien zur Sprachwissenschaft.

Hackstein, Olav, Hiromi Habata & Christoph Bross (2021). "Diverse Wiedergeburten des Bodhisattva: die Bahubuddha-Geschichte in tocharischer Überlieferung (B400–401)" *Münchener Studien zur Sprachwissenschaft* 73, 51–95.

Hansen, Olaf (1940). "Tocharisch-iranische Beziehungen" *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 94, 139–164.

Hartmann, Markus (2013). *Das Genussystem des Tocharischen*. Hamburg: Baar-Verlag.

Hasandoust, Mohammad (2014). *Farhang-e Riše-šenāxti-ye Zabān-e Fārsi* (Etymological Dictionary of the Persian Language). Tehran: Āṯār.

Hellquist, Elof (1922). *Svensk etymologisk ordbok*. Lund: Gleerups Förlag.

Henkelman, Wouter F.M. (2018). "Bactrians in Persepolis—Persians in Bactria" In: Johanna Lhuillier & Nikolaus Boroffka (eds.), *A Millennium of History. The Iron Age in southern Central Asia (2nd and 1st Millennia BC). Proceedings of the conference held in Berlin (June 23–25, 2014) Dedicated to the memory of Viktor Ivanovich Sarianidi*. Berlin: Dietrich Reimer.

Henning, Walter B. (1940). *Sogdica*. London: the Royal Asiatic Society.

Henning, Walter B. (1946). "The Sogdian texts of Paris" *Bulletin of the School of Oriental and African Studies* 11, 713–740.

Henning, Walter B. (1965). "A grain of mustard" *Annali dell'Istituto Universitario Orientale di Napoli, Sezione linguistica* 6, 29–47.

Hilmarsson, Jörundur G. (1986). *Studies in Tocharian Phonology, Morphology and Etymology with Special Emphasis on the* o*-Vocalism*. Leiden: PhD dissertation.

Hilmarsson, Jörundur G. (1987). "The element *-ai*(-) in the Tocharian nominal flexion" *Die Sprache* 33, 34–55.

Hilmarsson, Jörundur G. (1991). "Tocharian etymological notes 1–13" *Tocharian and Indo-European Studies* 5, 137–183.

Hilmarsson, Jörundur G. (1996). *Materials for a Tocharian Historical and Etymological Dictionary. Edited by Alexander Lubotsky and Guðrún Þórhallsdóttir*. Reykjavík: Málvísindastofnun Háskóla Íslands.

Hoernle, Rudolf A.F. (1916). *Manuscript remains of Buddhist Literature found in Estern Turkestan. Volume I. Parts I and II Manuscripts in Sanskrit, Khotanese, Kuchean, Tibetan and Chinese*. Oxford: Clarendon Press.

Hoffmann, Karl (1992). *Aufsätze zur Indoiranistik. Band III*. Wiesbaden: Dr. Ludwig Reichert.

Hoffmann, Karl and Bernhard Forssman (2004). *Avestische Laut- und Flexionslehre. Durchgesehene und erweiterte Auflage*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.

Holopainen, Sampsa (2019). *Indo-Iranian Borrowings in Uralic. Critical Overview of the Sound Substitutions and Distribution Criterion*. Helsinki: PhD dissertation.

Holthausen, Ferdinand (1963). *Altenglisches etymologisches Wörterbuch*. Heidelberg: Carl Winter.

Horn, Paul (1893). *Grundriss der Neupersischen Etymologie*. Strassburg: Verlag von Karl J. Trübner.

Horn, Paul (1898–1901). "Neupersische Schriftsprache" In: Wilhelm Geiger & Ernst Kunst (eds.), *Grundriss der iranischen Philologie. Erster Band. 2. Abteilung*. Strassburg: Verlag von Karl J. Trübner, 1–200.

Hosek, Lauren, Robin J. James & William T.T. Taylor (2024). "Tracing horseback riding and transport in the human skeleton". *Science Advances* 10, 1–8.

Huard, Athanaric (2020). "On Tocharian B *kents** and the origin of PIE **ǵʰans-** *Wékʷos* 5, 215–262.

Huard, Athanaric (2022). *Recherches sur les textes de méditation en tokharien*. Paris: EPHE, unpublished PhD thesis.

Huard, Athanaric (2024). "Qu'est-ce qu'un Bouddha? La notion de *farnah* dans les langues bouddhiques d'Asie centrale" handout of a presentation for the *Journée du monde indien* (13th of June 2024).

Huard, Athanaric (2024a). "Two unidentified languages from ancient Kucha" *Wékʷos* 7, 145–198.

Hübschmann, Heinrich (1895). *Persische Studien*. Strassburg: Verlag von Karl J. Trübner.

Hübschmann, Heinrich (1897). *Armenische Grammatik*. Leipzig: Verlag von Breitkopf & Härtel.

Hurford, James R. (1974). "Exclusive or Inclusive Disjunction" *Foundations of Language* 11:3, 409–411.

Isebaert, Lambert (1980). *De Indo-Iraanse Bestanddelen in de Tocharische Woordenschat. Vraagstukken van fonische productinterferentie, met bijzondere aandacht voor de Indo-Iraanse diafonen* a, ā. Leuven: PhD dissertation.

Isebaert, Lambert (2003). "Trois mots iraniens en tokharien" In Alois van Tongerloo

(ed.), *Iranica Selecta, Studies in Honour of Professor Wojciech Skalmowski*. Turnhout: Brepols, 115–120.

Isebaert, Lambert (2019). "Emprunt et loi phonétique. Le cas de tokharien B *ñ(y)ās* / A *ñās* « désir » et B *ñ(y)ātse* / A *ñātse* « danger, détresse »" In: Natalia Bolatti Guzzo & Piotr Taracha (eds.), *"And I knew Twelve Languages" A Tribute to Massimo Poetto on the Occasion of His 70th Birthday*. Agade Bis: University of Warsaw, 266–273.

Itkin, Ilya (2002). "The Linguistic Features of Tocharian A Manuscript *Maitreyāvadāna-vyākaraṇa*" *Manuscripta Orientalia* 8:3, 11–16.

Jasanoff, Jay (2017). *The Prehistory of the Balto-Slavic Accent*. Leiden–Boston: Brill.

Ji, Xianlin, in collaboration with Werner Winter & Georges-Jean Pinault (1998). *Fragments of the Tocharian A Maitreyasamiti-Nātaka of the Xinjiang Museum, China*. Berlin–New-York: Mouton de Gruyter.

Joki, Aulis J. (1973). *Uralier und Indogermanen. Die älteren Berührungen zwischen den uralischen und indogermanischen Sprachen*. Helsinki: Suomalais-Ugrilainen Seura.

Josephson, Judith (2011). "Definiteness and deixis in Middle Persian" In: Mauro Maggi & Paola Orsatti (eds.), *The Persian language in history*. Wiesbaden: Dr. Ludwig Reichert, 23–39.

Justi, Ferdinand (1905). Review of: Christian Bartholomae (1904). *Altiranisches Wörterbuch*. *Anzeiger für Indogermanische Sprach- und Altertumskunde* 17/1-2-3, 84–131.

Karulis, Konstantins (1992). *Latviešu etimoloģijas vārdnīca*. Riga: Avots.

Katz, Hartmut (2003). *Studien zu den älteren indoiranischen Lehnwörter in den uralischen Sprachen. Aus dem Nachlaß herausgegeben von Paul Widmer, Anna Widmer und Gerson Klumpp*. Heidelberg: Carl Winter.

Kellens, Jean (1974). *Les Noms-racines de l'Avesta*. Wiesbaden: Dr. Ludwig Reichert

Kellens, Jean (1984). *Le Verbe avestique*. Wiesbaden: Dr. Ludwig Reichert.

Kent, Roland G. (1953). *Old Persian: Grammar. Texts. Lexicon*. New Haven: American Oriental Society.

Kim, Ronald (1999). "Observations on the Absolute and Relative Chronology of Tocharian Loanwords and Sound Changes" *Tocharian and Indo-European Studies* 8, 111–138.

Kim, Ronald (2003). "On the Historical Phonology of Ossetic: The Origin of the Oblique Case Suffix" *Journal of the American Oriental Society* 123, 43–72.

Kim, Ronald (2007). "Vowel Weakening in Tocharian A Preterite Participles and Abstract Nouns" *Departemental Bulletin Paper*, *Kyoto University*, 1–30.

Kim, Ronald (2019). "Adjectival suppletion in Tocharian" In: Harald Bichlmeier and Velizar Sadovski (eds.), *Diachronic Perspectives on Suppletion*. Hamburg: Baar-Verlag, 173–192.

Klingenschmitt, Gert (1975). "Altindisch *śaśvat-*" *Münchener Studien zur Sprachwissenschaft* 33, 67–78.

Klingenschmitt, Gert (2000). "Mittelpersisch" In: Bernhard Forssman & Robert Plath (eds.), *Indoarisch, Iranisch und die Indogermanistik, Arbeitstagung der Indogerman-*

ischen Gesellschaft vom 2. bis 5. Oktober 1997 in Erlangen. Wiesbaden: Dr. Ludwig Reichert, 191–229.

Kloekhorst, Alwin (2008). *Etymological Dictionary of the Hittite Inherited Lexicon*. Leiden–Boston: Brill.

Koivulehto, Jorma (2001). "The earliest contacts between Indo-European and Uralic speakers in the light of lexical loans". In: Chr. Carpelan, A. Parpola & P. Koskikallio (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations. Papers presented at an international symposium held at the Tvärminne Research Station of the University of Helsinki 8–10 January 1999*, Mémoires de la Société Finno-ougrienne, Vol. 242. Helsinki: Société Finno-ougrienne, 235–264.

Kölligan, Daniel (2020). "Deaffrication in Armenian?" In: Romain Garnier (ed.), *Loanwords and Substrata. Proceedings of the Colloquium held in Limoges* (5th–7th June, 2018). Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck, 215–235.

Korn, Agnes (2005). *Towards a Historical Grammar of Balochi*. Wiesbaden: Dr. Ludwig Reichert.

Korn, Agnes (2009). "Lengthening of *i* and *u* in Persian" In: Werner Sundermann, Almut Hintze & François de Blois (eds.), *Exegisti monumenta Festschrift in Honour of Nicholas Sims-Williams*. Wiesbaden: Harrassowitz Verlag, 197–213.

Korn, Agnes (2016). "A partial tree of Central Iranian. A new look at Iranian subphyla." *Indogermanische Forschungen* 121, 401–434.

Korn, Agnes (2019). "Isoglosses and subdivisions of Iranian" *Journal of Historical Linguistics* 9:2, 239–281.

Korn, Agnes (2021). "Contributions to a relative chronology of Persian: The non-change of postconsontal *y* and *w* in Middle Persian in context" *Indo-European Linguistics* 9, 85–127.

Krause, Wolfgang (1954). "Das Numerussystem des Tocharischen" *Nachrichten der Akademie der Wissenschaften in Göttingen* 1, 1–12.

Krause, Wolfgang (1971). *Handbuch der Orientalistik. Tocharisch*. Leiden–Köln: Brill.

Krause, Wolfgang & Werner Thomas (1960). *Tocharisches Elementarbuch. Band I*. Heidelberg: Carl Winter.

Kroonen, Guus (2013). *Etymological Dictionary of Proto-Germanic*. Leiden–Boston: Brill.

Kümmel, Martin (2007). *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Dr. Ludwig Reichert.

Kümmel, Martin (2018). "The Survival of Laryngeals in Iranian" In: Lucien van Beek, Alwin Kloekhorst, Guus Kroonen, Michaël Peyrot, Tijmen Pronk (eds.), *Farnah: Indo-Iranian and Indo-European Studies in Honor of Sasha Lubotsky*. Ann Arbor–New York: Beech Stave, 162–172.

Kümmel, Martin (2020). "Substrata of Indo-Iranic and related questions" In: Romain Garnier (ed.), *Loanwords and Substrata. Proceedings of the Colloquium held in Limoges* (5th–7th June, 2018). Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck, 237–277.

Kuryłowicz, Jerzy (1964). "L'accentuation en vieil-iranien" In: Georges Redard (ed.), *Indo-Iranica. Mélanges présentés à Georg Morgenstierne à l'occasion de son soixante-dixième anniversaire*. Wiesbaden: Otto Harrassowitz, 103–107.

Kuryłowicz, Jerzy (1975). "L'accent du mot en v.-iranien" In: *Monumentum H.S. Nyberg. Volume 1*, 499–507.

Ladefoged, Peter and Ian Maddieson (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Lashkarbekov, Bogšo B. (2008). "Starovandžskij jazyk (vanǰivor)" In: *Osnovy Iranskogo jazykoznanija* (Foundations of Iranian linguistics). Moscow: Vostočnaja Literatura, 61–109.

Lazard, Gilbert (1963). *La langue des plus anciens monuments de la prose persane*. Paris: C. Klincksieck.

Lecoq, Pierre (2016). *Les Livres de l'Avesta*. Paris: Les Éditions du Cerf.

Librado, Pablo, Naveed Khan, Antoines Fages, Mariya A. Kusliy, Tomasz Suchan & al. (2021). "The origins and spread of domestic horses from the Western Eurasian steppes" *Nature* 598, 634–652.

Lidén, Evar (1916). *Studien zur tocharischen Sprachgeschichte*. Göteborg: Wetergren & Kerber.

Lubotsky, Alexander (1981). "Gr. *pḗgnumi*: Skt. *pajrá-* and loss of laryngeals before mediae in Indo-Iranian" *Münchener Studien zur Sprachwissenschaft* 40, 133–138.

Lubotsky, Alexander (1989). "Against a Proto-Indo-European phoneme **a*" In: Theo Venneman (ed.), *The New Sound of Indo-European*. Berlin–New-York: de Gruyter, 53–66.

Lubotsky, Alexander (1998). "Avestan *xᵛarənah-*: the etymology and concept" In: Wolfgang Meid (ed.), *Sprache und Kultur. Akten der x. Fachtagung der Indogermanischen Gesellschaft Innsbruck, 22–28. September 1996*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.

Lubotsky, Alexander (2001). "The Indo-Iranian substratum" In: Chr. Carpelan, A. Parpola & P. Koskikallio (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations. Papers presented at an international symposium held at the Tvärminne Research Station of the University of Helsinki 8–10 January 1999*, Mémoires de la Société Finno-ougrienne, Vol. 242. Helsinki: Société Finno-ougrienne.

Lubotsky, Alexander (2001a). "Reflexes of Proto-Indo-European **sk* in Indo-Iranian" *Incontri linguistici* 24, 25–57.

Lubotsky, Alexander (2021). "Sanskrit roots *hi-* 'to impel, hurl', *hiṣ-* 'to injure, harm', *hīḍ-*

'to make/be angry' and the Indo-European root enlargements *-s-* and *-d-*" *Historische Sprachforschung* 131, 227–235.

Lubotsky, Alexander and Sergei Starostin (2003). "Turkish and Chinese loan words in Tocharian" In: Brigitte L.M. Bauer & Georges-Jean Pinault (eds.), *Language in Time and Space. A Festschrift for Werner Winter on the Occasion of his 80th birthday*. Berlin–New-York: de Gruyter, 257–270.

Lurje, Pavel (2010). *Personal names in Sogdian texts*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

MacKenzie, David N. (1970). *The 'Sūtra of Causes and Effects of Actions' in Sogdian*. London: Oxford University Press.

MacKenzie, David N. (1990). *The Khwarezmian Element in the Qunyat Al-Munya*. London: School of Oriental and African Studies.

Maggi, Mauro (2003). "New Persian glosses in East Syriac texts of the eighth to tenth centuries" In Paul Ludwig (ed.): *Persian origins: Early Judaeo-Persian and the emergence of New Persian: Collected papers of the symposium, Göttingen 1999*. Wiesbaden: Harrassowitz, 111–145.

Mahamedi, Hamid (1994). "Davān" In: *Encyclopædia Iranica*, Vol. 7, Fasc. 2, pp. 129–132.

Malzahn, Melanie (2007). "Tocharian Desire" In: Alan J. Nassbaum (ed.), *Verba Docenti: Studies in historical and Indo-European linguistics presented to Jay H. Jasanoff by students, colleagues, and friends*. Ann Arbor: Beech Stave Press, 237–249.

Malzahn, Melanie (2014). "Tocharian A *śorki* 'fear' and two other TA scary words" *Tocharian and Indo-European Studies* 15, 87–94.

Martirosyan, Hrach (2010). *Etymological Dictionary of the Armenian Inherited Lexicon*. Leiden–Boston: Brill.

Mayrhofer, Manfred (1983). "Lassen sich Vorstufen des Uriranischen nachweisen?" *Anzeiger phil.-hist. Klasse der Österreichischen Akademie der Wissenschaften* 120, 249–255.

Mayrhofer, Manfred (1989). "Vorgeschichte der iranischen Sprachen; Uriranisch" In: Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*. Wiesbaden: Dr. Ludwig Reichert, 4–24.

Meillet, Antoine (1913). "Le tokharien" *Indogermanisches Jahrbuch* 1, 1–19.

Meillet, Antoine (1921). *Linguistique historique et linguistique générale*. Paris: Champion.

Ménage, Gilles (1694). *Dictionnaire étymologique, ou Origines de la langue françoise. Nouvelle Edition revue & augmentée par l'Auteur*. Paris: Jean Anisson.

Meunier, Fanny (2013). "Typologie des locutions en *yām-* du tokharien" *Tocharian and Indo-European Studies* 14, 123–185.

Meunier, Fanny (2015). *Recherches sur le génitif en tokharien*. Paris: PhD dissertation.

Milizia, Paolo (2015). "Greek λοιμός, Middle Persian *rēm*, and the Avestan root **raṷ-*" *Indogermanische Forschungen* 80, 87–106.

Minorsky, Vladimir (1955). "Addenda to the *Ḥudūd al-ʿālam*" *Bulletin of the School of Oriental and African Studies* 17, 250–270.

Monchi-Zadeh, Davoud (1990). *Wörter aus Xurāsān und ihre Herkunft*. Leiden: Brill.

Morano, Enrico (1982). "The Sogdian Hymns of Stellung Jesu". *East and* West 32, 9–43

Morgenstierne, Georg (1927). *An Etymological Vocabulary of Pashto*. Oslo: Kommisjon Hos Jacob Dybwad.

Morgenstierne, Georg (1930). "The Name *Munjān* and Some Other Names of Places and Peoples in the Hindu Kush" *Bulletin of the School of Oriental Studies* 6, 439–444.

Morgenstierne, Georg (1932). "Supplementary notes on Ormuri" *Norsk Tidskrift for Sprogvidenskap* 5, 5–36.

Morgenstierne, Georg (1938). *Indo-Iranian Frontier Languages. Vol. II. Iranian Pamir Languages* (*Yidgha-Munji, Sanglechi-Ishkashmi and Wakhi*). Oslo: Aschehoug & Co. (Instituttet for Sammenlignende Kulturforskning).

Morgenstierne, Georg (1942). "Iranica" *Norsk Tidsskrift for Sprogvidenskap* 12, 258–271.

Morgenstierne, Georg (1973). *Irano-Dardica*. Wiesbaden: Dr. Ludwig Reichert.

Morgenstierne, Georg (2003). *A New Etymological Vocabulary of Pashto*. Compiled and Edited by J. Elfenbein, D.N. MacKenzie and Nicholas Sims-Williams. Wiesbaden: Dr. Ludwig Reichert.

Narten, Johanna (1968). "Ved. *iláyati* und seine Sippe" *Indo-Iranian Journal* 10–14, 239–250.

Ning, Chao, Chuan-Chao Wang, Shizhu Gao & al. (2019). "Ancient Genomes Reveal Yamnaya-Related Ancestry and a Potential Source of Indo-European Speakers in Iron Age Tianshan" *Current Biology Report* 29, 2526–2532.

Nouveau Petit Robert (2007). Alain Rey, Josette Rey-Debove & le collectif Le Robert (eds.). Paris: Le Robert.

Novák, Ľubomír (2013). *Problem of Archaism and Innovation in the Eastern Iranian Languages*. Prague: PhD dissertation.

Nyberg, Henrik S. (1931). *Hilfsbuch des Pehlevi, II: Glossar*. Uppsala: Otto Harrassowitz.

Nyberg, Henrik S. (1974). *A Manual of Pahlavi. II Ideograms, Glossary, Abbreviations, Index, Grammatical Survey, Corrigenda to Part I*. Wiesbaden: Otto Harrassowitz.

Oettinger, Norbert (1983). *Untersuchungen zur avestischen Sprache am Beispiel des Ardvīsūr-Yašt*. Munich: Habilitation thesis.

Ogihara, Hirotoshi (2009). *Researches about Vinaya-texts in tocharian A and B*. Paris: PhD dissertation.

Ogihara, Hirotoshi (2016). "Konponsetsuissaiuburitsu yakuji" ni kanrensuru niten no tokarago B danpen ni tsuite.『根本説一切有部律薬事』に関連する二点のトカラ語B断片について". In: Masaharu Arakawa & Mikio Shibata (eds.), *Shirukurōdo to kindai nihon no kaikō. Saiiki kodai shiryō to nihon kindai bukkyō*. Tōkyō: Bensei Shuppan, 258–276.

Ogihara, Hirotoshi (2019). "Tokarago B Buddhastotra san Futsu shi shahon saiken no

kokoromi 『トカラ語 B «Buddhastotra»（讃仏詩）写本再建の試み" *Tokyo University Linguistic Papers* 41, 1–46.

Olsen, Birgit A. (1999). *The Noun in Biblical Armenian. Origin and Word-Formation—with special emphasis on the Indo-European heritage*. Berlin–New-York: de Gruyter.

Palmér, Axel, Anthony Jakob, Rasmus Thorsø, Paulus van Sluis, Cid Swanenvleugel & Guus Kroonen (2021). "Proto-Indo-European 'fox' and the reconstruction of an athematic *ḱ*-stem" *Indo-European Linguistics* 9, 234–263.

Pan, Tao (2021). "Tocharian A *ārkiśoṣi* 'world with radiance' and Chinese *suo po shi jie* 'world of *sabhā*'" *Acta Asiatica Varsoviensia* 34, 263–294.

Parpola, Asko (2022). "Formation of the Indo-Iranian languages. Locations and dates according to archaeological evidence" In: Peter M. Scharf (ed.) *Indian linguistic studies in honor of George Cardona. Volume 11. Historical linguistics, Vedic, etc*. Providence: The Sanskrit Library, 1–83.

Peyrot, Michaël (2008). *Variation and change in Tocharian B*. Amsterdam–New-York: Rodopi.

Peyrot, Michaël (2010). "Proto-Tocharian Syntax and the Status of Tocharian A" *Journal of Indo-European Studies* 38, 132–146.

Peyrot, Michaël (2012). "The Tocharian A match of the Tocharian B obl.sg. *-ai*" *Tocharian and Indo-European Studies* 13, 181–220.

Peyrot, Michaël (2013). *The Tocharian Subjunctive. A Study in Syntax*. Leiden–Boston: Brill.

Peyrot, Michaël (2015). Review of: Dieter Maue (2015), *Alttürkische Handschriften Teil 19: Dokumente in Brāhmī und tibetischer Schrift Teil 2*. *Tocharian and Indo-European Studies* 16, 215–226.

Peyrot, Michaël (2015a). "Tocharian Language" In: *Encyclopædia Iranica*, online edition, available at http://www.iranicaonline.org/articles/tocharian-language.

Peyrot, Michaël (2016). "Language contact in Central Asia: On the etymology of Tocharian B *yolo* 'bad'" In: Bjarne S.S. Hansen, Benedicte Nielsen Whitehead, Thomas O. Olander & Birgit A. Olsen (eds.), *Etymology and the European Lexicon. Proceedings of the 14th Fachtagung der Indogermanischen Gesellschaft, 17–22 September 2012, Copenhagen*. Wiesbaden: Dr. Ludwig Reichert.

Peyrot, Michaël (2018). "Tocharian Agricultural Terminology: Between Inheritance and Language Contact" In: Guus Kroonen, James P. Mallory & Bernard Comrie (eds.), *Talking Neolithic: Proceedings of the workshop on Indo-European origins held at the Max Planck Institute for Evolutionary Anthropology, Leipzig, December 2–3, 2013*, Institute for the Study of Man, 247–277.

Peyrot, Michaël (2018a). "Tocharian B *etswe* 'mule' and Eastern East Iranian" In: Lucien van Beek, Alwin Kloekhorst, Guus Kroonen, Michaël Peyrot, Tijmen Pronk (eds.), *Farnah: Indo-Iranian and Indo-European Studies in Honor of Sasha Lubotsky*. Ann Arbor–New York: Beech Stave, 162–172.

Peyrot, Michaël (2019). Review of: Václav Blažek & Michal Schwarz (2017), *Early Indo-Europeans in Central Asia and China. Cultural relations as reflected in language*. *Tocharian and Indo-European Studies* 19, 143–146.

Peyrot, Michaël (2020). "The simplification of the Archaic Tocharian B clusters *ltk* and *rtk*—with a note on the vowel assimilation in Late Tocharian B *oṅkorño*" *Tocharian and Indo-European Studies* 20, 191–200.

Peyrot, Michaël (2021). "On a festive occasion. The inflection of Tocharian B *ekṣalye* and morphologically related words." In: Hannes A. Fellner, Melanie Malzahn & Michaël Peyrot (eds.), *lyuke wmer ra. Indo-European Studies in Honor of Georges-Jean Pinault*. Ann Arbor: Beech Stave, 450–461.

Peyrot, Michaël (2022). "Tocharian" In: Thomas O. Olander (ed.), *The Indo-European language family. A phylogenetic perspective*. Cambridge: Cambridge University, 83–101.

Peyrot, Michaël (2022a). "The tertiary homeland of Tocharian: On the drivers and the chronology of the trajectory to the Tarim Basin", lecture at "The Secondary Homelands of the Indo-European Languages (IG-AT2022)", 6 September 2022.

Peyrot, Michaël, Federico Dragoni & Chams Bernard (2022). "The spread of iron in Central Asia: On the etymology of the word for 'iron' in Iranian and Tocharian" *Bulletin of the School of Oriental and African Studies* 85, 403–422.

Peyrot, Michaël & Kristin Meier (2017). "The word for 'Honey' in Chinese, Tocharian and Sino-Vietnamese" *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 167, 7–22.

Peyrot, Michaël, Georges-Jean Pinault & Jens Wilkens (2019). "Vernaculars of the Silk Road—A Tocharian B–Old Uyghur Bilingual" *Journal Asiatique 307*, 65–90.

Pfeifer & al. (online). *Etymologisches Wörterbuch des Deutschen (1993), digitalisierte und von Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache*, https://www.dwds.de/wb/etymwb/Hof.

Pinault, Georges-Jean (1987). "Sites divers de la région de Koutcha. Épigraphie koutchéenne. I. Laissez-passer de caravanes.—II. Graffites et inscriptions." In: *Mission Paul Pelliot, VIII. Sites divers la région de Koutcha*. Paris: Collège de France, 59–169.

Pinault, Georges-Jean (1988). "Révision des fragments en tokharien B de la légende de Mahāprabhāsa" In: Peter Kosta with the help of Gabriele Lerch & Olivier Peter (eds.), *Studia Indogermanica et Slavica. Festgabe für Werner Thomas zum 65. Geburtstag*. München: Verlag Otto Sagner, 175–210.

Pinault, Georges-Jean (1989). "Introduction au tokharien" *Lalies* n°7. Paris: Presses de l'École Normale Supérieure, 3–224.

Pinault, Georges-Jean (1994). "Lumières tokhariennes sur l'indo-européen" In: Jens Elgård Rasmussen (ed.), *In honorem Holger Pedersen. Kolloquium der Indogermanischen Gesellschaft vom 26. bis 28. März 1993 in Kopenhagen*. Wiesbaden: Reichert Verlag, 365–396.

Pinault, Georges-Jean (1995). "The rendering of Buddhist terminology in Tocharian" *Journal of Dunhuang and Turfan Studies* 1, 9–35.

Pinault, Georges-Jean (1998). "Tocharian languages and Pre-Buddhist Culture" In: Victor A. Mair (ed.) *The Bronze Age and Early Iron Age Peoples of Central Asia*. Washington: Institute for the Study of Man, 358–371.

Pinault, Georges-Jean (2001). "Nouveautés lexicales et morphologiques dans le manuscript de Yanqi du Maitreyasamiti-Nāṭaka en tokharien A" In: Stefan Wild & Hartmut Schild (eds.) *Akten des 27. Deutschen Orientalistentages* (*Bonn—28. September bis 2. Oktober 1998*). *Norm und Abweichung*. Würzburg: Ergon Verlag, 121–136.

Pinault, Georges-Jean (2002). "Tokh. B k_u*cañ̃ñe*, A k_u*ciṃ* et skr. *tokharika*" *Indo-Iranian Journal* 45, 311–345.

Pinault, Georges-Jean (2002a). "Tocharian and Indo-Iranian: relations between two linguistic areas" In: Nicholas Sims-Williams (ed.), *Indo-Iranian Languages and Peoples*. Oxford: Oxford University Press, 243–284.

Pinault, Georges-Jean (2003). "Sanskrit *kalyāṇa-* interprété à la lumière des contacts en Asie Centrale" *Bulletin de la Société de Linguistique de Paris* 98:1, 123–161.

Pinault, Georges-Jean (2003a). "Sur les thèmes indo-européens en **-u-*: dérivation et étymologie" In: Eva Tichy, Dagmar S. Wodtko & Britta Irslinger (eds.), *Indogermanisches Nomen. Derivation, Flexion und Ablaut. Akten der Arbeitstagung der Indogermanischen Gesellschaft Freiburg, 19. bis 22. September 2001*. Bremen: Hempen Verlag, 153–188.

Pinault, Georges-Jean (2006). "Further links between the Indo-Iranian substratum and the BMAC language". In: Bertil Tikkanen and Heinrich Hettrich (eds.), *Themes and Tasks in Old and Middle Indo-Aryan Linguistics*. Delhi: Motilal Banarsidass, 167–196.

Pinault, Georges-Jean (2006a). "Retour sur le numéral "un" en tokharien" *Indogermanische Forschungen* 111, 71–97.

Pinault, Georges-Jean (2008). *Chrestomathie tokharienne. Textes et grammaire*. Leuven–Paris: Peeters.

Pinault, Georges-Jean (2008a). "La langue des Scythes et le nom des Arimaspes" *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres*, 152, 105–138.

Pinault, Georges-Jean (2008b). "Bilingual hymn to Mani. Analysis of the Tocharian B parts" *Studies on the Inner Asian Languages* 23, 93–120.

Pinault, Georges-Jean (2011). "Let us now praise famous gems" *Tocharian and Indo-European Studies* 12, 155–220.

Pinault, Georges-Jean (2015). "The formation of Buddhist languages, as exemplified by the Tocharian evidence" In: Melanie Malzahn, Michaël Peyrot, Hannes Fellner & Theresa-Susanna Illés (eds.), *Tocharian Texts in Context. International Conference on Tocharian Manuscripts and Silk Road Culture*. Bremen: Hempen Verlag, 159–185.

Pinault, Georges-Jean (2016). "Ariyāramna, the Pious Lord" In: Céline Redard (ed.), *Des Contrées avestiques à Mahabad, via Bisotun. Études offertes en hommage à Pierre Lecoq*. Neuchâtel: Recherches et Publications, 113–130.

Pinault, Georges-Jean (2016a). "Reassessing Varuṇa's name" In: Hans H. Hock (ed.), *Veda and Vedic Literature. Select Papers from the Panel on "Veda and Vedic Literature" at the 16th World Sanskrit Conference (28 June–2 July 2015) Bangkok, Thailand*. Bangkok: Sanskrit Studies Centers (DK Publishers Distributors), 53–71.

Pinault, Georges-Jean (2016b). "Les Tokhariens, passeurs et interprètes du bouddhisme" In: Michel Espagne, Svetlana Gorshenina, Frantz Grenet, Shahin Mustafayev et Claude Rapin (eds.) *Asie Centrale. Transferts culturels le long de la Route de la soie*. Paris: Vendémiaire, 167–200.

Pinault, Georges-Jean (2017). "Genesis of the PIE gerundival suffix **-etó-*" In: Claire Le Feuvre, Daniel Petit & Georges-Jean Pinault (eds.), *Verbal Adjectives and Participles in Indo-European Languages. Proceedings of the conference of the Society for Indo-European Studies (Indogermanische Gesellschaft), Paris, 24th to 26th September 2014*. Bremen: Hempen Verlag, 343–375.

Pinault, Georges-Jean (2019). *Vedica. 1. Etymologica*. Paris: Les Cent Chemins.

Pinault, Georges-Jean (2020). "Tocharian Taxonomy of Wealth" In: Harald Bichlmeier, Ondřej Šefčík & Roman Sukač (eds.), *Etymologus. Festschrift for Václav Blažek*. Hamburg: Baar-Verlag, 323–337.

Pinault, Georges-Jean (2020a). "Tocharian lexicon in the light of contact phenomena" In: Romain Garnier (ed.), *Loanwords and Substrata. Proceedings of the Colloquium held in Limoges* (5th–7th June, 2018). Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck, 367–401.

Pinault, Georges-Jean (2020b). "Tocharian nouns of the Latin *ferōx*-type" *Tocharian and Indo-European Studies* 20, 201–227.

Pinault, Georges-Jean (2020c). "The Dharma of the Tocharians" In: Vincent Tournier, Vincent Eltschinger & Marta Sernesi (eds.) *Archaeologies of the Written: Indian, Tibetan and Buddhist Studies in Honour of Cristina Scherrer-Schaub*. Napoli: Unior Press, 461–492.

Pinault, Georges-Jean (2022). "Philologie des textes bouddhiques d'Asie Centrale" *Annuaire de l'École Pratique des Hautes Études. Chroniques de la section des Sciences historiques et philologiques—année 2020–2021*, 520–528.

Pisowicz, Andrzej (1985). *Origins of the New and Middle Persian Phonological Systems*. Kraków: Uniwersytet Jagielloński.

Poucha, Pavel (1939). "Tocharische Etymologien" *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 93, 204–216.

Poucha, Pavel (1955). *Thesaurus Linguae Tocharicæ Dialecti A. Institutiones Linguae Tocharicæ. Pars 1*. Praha: Státní Pedagogické Nakladatelství.

Provasi, Elio (2003). "Sogdian *farn*" In: Carlo G. Cereti, Mauro Maggi & Elio Provasi

(eds.), *Religious themes and texts of pre-Islamic Iran and Central Asia. Studies in honour of Professor Gherardo Gnoli on the occasion of his 65th birthday on 6th December 2002*. Wiesbaden: Dr. Ludwig Reichert, 305–322.

Provasi, Elio (2013). "The Diamond Earth. A Manichaean Sogdian Eulogy" In: Sergei Tokhtasev & Pavel Lurje (eds.), *Commentationes Iranicae. Vladimiro f. Aaron Livschits nonagenario donum natalicium*. Saint-Petersburg: Nestor-Historia, 378–393.

Pulleyblank, Edwin G. (1962). "The Consonantal System of Old Chinese" *Asia Major* 9, 58–144.

Pulleyblank, Edwin G. (1991). *A Lexicon of Reconstructed Pronunciation in Early Middle Chinese, Late Middle Chinese and Early Mandarin*. Vancouver: UBC Press.

Rastogi, Nalin & R.C. Rastogi (1984). "Leprosy in ancient India" *International Journal of Leprosy and Other Mycobacterial Diseases* 52/4, 541–543.

Remmer, Ulla (2006). *Frauennamen im Rigveda und im Avesta. Studien zur Onomastik des ältesten Indischen und Iranischen*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Ringe, Donald A. (1996). *On the Chronology of Sound Changes in Tocharian. Volume 1: From Proto-Indo-European to Proto-Tocharian*. New Haven: American Oriental Society.

Ronsard, Pierre de (1553). *Livret de Folastries, A Janot Parisien. Plus, quelques Epigrames grecs & des Dithyrambes chantes au Bouc de E. Jodëlle Poëte Tragiq*. Paris: veuve Maurice de La Porte.

Rossi, Adriano V. (2010). "Possiamo continuare a connettere ir. **ādu-* a ie. 'mangiare'?" In: Antonio Mendicino & Giovanni Belluscio (eds.), *Scritti in onore di Eric Pratt Hamp per il suo 90. compleanno*. Cosenza: Università della Calabria, 331–344.

Rotman, Andy (2008). *Divine Stories: Divyāvadāna*. Part 1. Boston: Wisdom Publications.

Schafer, Edward H. (1959). "Parrots in Medieval China" In: *Studia Serica Bernhard Karlgren Dedicata*. Copenhagen: Ejnar Munskgaard, 271–282.

Schmidt, Klaus T. (1974). *Die Gebrauchsweisen des Mediums im Tocharischen*. Göttingen: PhD dissertation.

Schmidt, Klaus T. (1985). "Zu einigen der ältesten iranischen Lehnwörter im Tocharischen" In: Ursula Pieper & Gerhard Stickel (eds.), *Studia Linguistica Diachronica et Synchronica, Werner Winter Sexagenario Anno MCMLXXXIII gratis animis ab eius collegis, amicis discipulisque oblata*. Berlin–New York–Amsterdam: de Gruyter, 757–767.

Schmidt, Klaus T. (1986). *Fragmente eines buddhistischen Ordinationsrituals in westtocharischen Sprache: aus der Schule der Sarvāstivādins*. Habilitation thesis.

Schmidt, Klaus T. (2018). *K.T. Schmidt: Nachgelassene Schriften. Bearbeitet und herausgegeben von Stefan Zimmer*. Bremen: Hempen Verlag.

Schmitt, Rüdiger (1989, ed.), *Compendium Linguarum Iranicarum*. Wiesbaden: Dr. Ludwig Reichert.

Schoubben, Niels (2024). *Traces of language contact in Niya Prakrit. Bactrian and other foreign elements*. Leiden: PhD dissertation.

Schulze, Wilhelm (1932). "Ratäk" *Zeitschrift für vergleichende Sprachforschung* 59: 212.

Schwartz, Martin (1969). Review of: Harold W. Bailey (1967), *Indo-Scythian Studies, being Khotanese Texts, Vol. VI: Prolexis to the Book of Zambasta*. *Journal of the American Oriental Society* 89, 444–447.

Schwartz, Martin (1974). "Irano-Tocharica". In: Philippe Gignoux & Ahmad Tafazzoli (eds.), *Mémorial Jean de Menasce*. Louvain: Imprimerie Orientaliste, 399–411.

Schwartz, Martin (2002). "How Zarathushtra Generated the Gathic Corpus: Intertextual and Intertextual Composition" *Bulletin of the Asia Institute*, New Series, 16, 53–64.

Schwentner, Ernst (1956). "Toch. B *ekṣinek* 'Taube'" *Zeitschrift für vergleichende Sprachforschung* 72, 238.

Schwentner, Ernst (1958). "Toch. A *yāmutsi*, B *yāmuttsi* 'Flamingo'" *Zeitschrift für vergleichende Sprachforschung* 75, 206.

Seržant, Ilja A. (2008). "Die idg. Wurzel **kelh*$_{1}$- "etw. bewegen" und **kelh*$_{3}$- "sich erheben"." *Indogermanische Forschungen* 113:59–75.

Shapira, Dan (1998). *Studies in Zoroastrian Exegesis: Zand*. Hebrew University of Jerusalem: PhD dissertation.

Shaydullaev, Shapulat (2011). "Excavations in Sector 04—preliminary report" In: Kazim Abdullaev & Ladislav Stančo (eds.), *Jandavlattepa. The Excavation Report for Seasons 2002–2006*. Vol. 1. Prague: Univerzita Karlova v Praze, Karolinum Press, 94–98.

Sieg, Emil (1944). *Übersetzungen aus dem Tocharischen. Teil I*. Munich: Bayerische Akademie der Wissenschaften. Philosophisch-Historische Klasse.

Sieg, Emil (1950). "Geschäfliche Aufzeichnungen in Tocharisch B aus der Berliner Sammlung" *Miscellanea Academica Berolinensia* 2, 208–223.

Sieg, Emil (1952). *Übersetzungen aus dem Tocharischen II, Aus dem Nachlass herausgegeben von Werner Thomas*. Berlin: Abhandlungen der Deutschen Akademie der Wissenschaften zu Berlin, Klasse für Sprachen, Literatur und Kunst, Jahrgang 1951, Nr. 1.

Sieg, Emil & Wilhelm Siegling (1921). *Tocharische Sprachreste, I. Band. Die Texte. A. Transcription*, Berlin–Leipzig: de Gruyter.

Sieg, Emil & Wilhelm Siegling (1949). *Tocharische Sprachreste Sprache B. Die Udānālaṅkāra-Fragmente. Übersetzung und Glossar*. Göttingen: Vandenhoeck & Ruprecht.

Sieg, Emil & Wilhelm Siegling (1953). *Tocharische Sprachreste. Sprache B. Aus dem Nachlaß herausgegeben von Werner Thomas*. Göttingen: Vandenhoeck & Ruprecht.

Sieg, Emil, Wilhelm Siegling & Wilhelm Schulze (1931). *Tocharische Grammatik*. Göttingen: Vandenhoeck & Ruprecht.

Sims-Williams, Nicholas (1985). *The Christian Sogdian Manuscript C2*. Berlin: Akademie-Verlag.

Sims-Williams, Nicholas (1989). "Sogdian" In: Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*. Wiesbaden: Dr. Ludwig Reichert, 173–192.

Sims-Williams, Nicholas (1994). "Bactrian ownership inscriptions" *Bulletin of the Asia Institute* 7, 173–179.

Sims-Williams, Nicholas (1997). *New light on ancient Afghanistan: the decipherment of Bactrian*. London: School of Oriental and African Studies.

Sims-Williams, Nicholas (1998). "The Iranian languages" In: A. Giacalone Ramat & P. Ramat (eds.), *The Indo-European languages*. London: Routledge, 125–153.

Sims-Williams, Nicholas (2002). "Ancient Afghanistan and its invaders: Linguistic evidence from the Bactrian documents and inscriptions" In: Nicholas Sims-Williams (ed.), *Indo-Iranian Languages and Peoples* (*Proceedings of the British Academy 116*). Oxford: Oxford University Press, 225–242.

Sims-Williams, Nicholas (2007). *Bactrian Documents from Northern Afghanistan. II: Letters and Buddhist Texts*. London: The Nour Foundation.

Sims-Williams, Nicholas (2009). "Avestan *fradaϑāi*, Bactrian φαρο, and their cognates" In: Éric Pirart & Xavier Tremblay (eds.), *Zarathushtra entre l'Inde et l'Iran. Études indo-iraniennes et indo-européennes offertes à Jean Kellens à l'occasion de son 65e anniversaire*. Wiesbaden: Dr. Ludwig Reichert, 279–287.

Sims-Williams, Nicholas (2014). "The story of Daniel and other texts from the manuscript E29" In: Nicholas Sims-Williams (ed.), *Biblical and other Christian Sogdian texts from the Turfan Collection*. Turnhout: Brepols, 75–100.

Sims-Williams, Nicholas (2017). "Iranian" In: Mate Kapović (ed.), *The Indo-European Languages*. Second Edition. London–New-York: Routledge, 263–286.

Sims-Williams, Nicholas (2020). *A Dictionary: Christian Sogdian, Syriac and English*. Wiesbaden: Dr. Ludwig Reichert.

Skjærvø, Prods O. (1994). "Of Lice and Men and the Manichean Anthropology" In: R. Söhnen-Thieme and O. v. Hinüber (eds.), *Festschrift George Buddruss*. Reinbek: Wezler, 269–286.

Skjærvø, Prods O. (2004). *This Most Excellent Shine of Gold, King of Kings of Sutras, The Khotanese Suvarṇabhāsottamasūtra*. Volume I and volume II. Harvard: Harvard University.

Sluis, Paulus van (2022). "Beekeeping in Celtic and Indo-European" *Studia Celtica* 61, 1–28.

Solans, Gabriel (2024). "Paléogénétique: que peut-elle dire aux linguistes sur les langues indo-européennes?" *Wékʷos* 7, 247–281.

Steblin-Kamenskij, Ivan M. (1999). *Ètimologičeskij slovar' vaxanskogo jazyka* [Etymological dictionary of the Wakhi language]. Saint-Petersburg: Peterburgskoe Vostokovedenie.

Steingass, Francis J. (1892). *A Comprehensive Persian-English Dictionary, Including the Arabic words and phrases to be met with in Persian literature, being Johnson and*

Richardson's Persian, Arabic and English Dictionary, revised, enlarged, and entirely reconstructed. London: Routledge and Kegan Paul.

Szemerényi, Oswald (1951). "Sogdicisms in the Avesta" In: F. Altheim (ed.), *Aus Spätantike und Christentum*. Tübingen: Max Niemeyer, 153–166.

Šorgo, Aljoša (2020). "Characteristics of lexemes of a substratum origin in Proto-Germanic" In: Romain Garnier (ed.), *Loanwords and Substrata. Proceedings of the Colloquium held in Limoges* (5th–7th June, 2018). Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck, 427–472.

Tarbé, Prosper (1851). *Recherches sur l'histoire du langage et des patois de Champagne. Tome second*. Reims: Regnier.

Tavernier, Jan (2007). *Iranica in the Achaemenid Period* (*ca. 550–330 B.C.*). *Lexicon of Old Iranian Proper Names and Loanwords, Attested in Non-Iranian Texts*. Leuven–Paris–Dudley: Peeters & Departement Oosterse Studies.

Thieme, Paul (1974). "Atharva-Veda 5.23.4" In: Manfred Mayrhofer, Wolfgang Meid, Bernfried Schlerath & Rüdiger Schmitt (eds.), *Antiquitates Indogermanicae. Gedenkschrift für Hermann Güntert*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck, 295–300.

Thomas, Werner (1954). "Die Infinitive im Tocharischen" In: Johannes Schubert & Ulrich Schneider (eds.), *Asiatica. Festschrift Friedrich Weller, Zum 65. Geburtstag gewidmet von seinen Freunden, Kollegen und Schülern*. Leipzig: Harrassowitz, 701–764.

Thomas, Werner (1957). *Das Gebrauch der Vergangenheitstempora im Tocharischen*. Wiesbaden: Otto Harrassowitz.

Thomas, Werner (1966). "Bemerkungen zum Gebrauch von toch. A *yo* und B *wai*" *Central Asiatic Journal* 11, 264–274.

Trautmann, Martin, Alin Frînculeasa, Bianca Preda-Bălănică, Marta Petruneac, Marin Focşăneanu & al. (2023). "First bioanthropological evidence for Yamnaya horsemanship" *Science Advances* 9, 1–13.

Tremblay, Xavier (2001). *Pour une histoire de la Sérinde. Le manichéisme parmi les peuples et religions d'Asie Centrale*. Wien: Österreichische Akademie der Wissenschaften.

Tremblay, Xavier (2005). "Irano-Tocharica et Tocharo-Iranica" *Bulletin of the School of Oriental and African Studies* 68, 421–449.

Utas, Bo (1965–1966). "Old Persian Miscellanea" *Orientalia Suecana* 14–15, 118–140.

de Vaan, Michiel (2003). *The Avestan Vowels*. Amsterdam–New-York: Rodopi.

de Vaan, Michiel (2008). *Etymological Dictionary of Latin and the other Italic Languages*. Leiden–Boston: Brill.

Vahman, Fereydun & Garnik S. Asatrian (1987). *West Iranian Dialect Materials from the Collection of D.L. Lorimer. Vol. 1. Materials on the Ethnography of the Baxtiaris*. Copenhagen: University Publishers.

de la Vaissière, Étienne, Philippe Marquis & Julio Bendezu Sarmiento (2015). "A Kushan military camp near Bactra" In: in Harry Falk (ed.), *Kushan Histories. Literary Sources and Selected Papers from a Symposium at Berlin, December 5th to 7th, 2013*. Bremen: Hempen Verlag, 241–254.

Van Windekens, Albert J. (1963). "Études de phonétique tokharienne IV" *Orbis* 12, 463–468.

Van Windekens, Albert J. (1976). *Le tokharien confronté avec les autres langues indo-européennes. Volume I. La phonétique et le vocabulaire*. Louvain: Centre International de Dialectologie Générale.

Van Windekens, Albert J. (1977). "Recherches complémentaires sur le vocabulaire tokharien I" *Orbis* 26, 141–148.

Viredaz, Rémy (2017). "Trois étymologies ossètes: *rasyg*, *ævzist*, *ærxᵒy*" *Studia Iranica* 46, 31–48.

Watson, Wilfred G.E. (2013). "Indo-European and Semitic: Two-way traffic" In: Juan P. Monferrer-Sala & Wilfred G.E. Watson (eds.), *Archaism and Innovation in the Semitic Languages*. Córdoba: Oriens Academic, 163–193.

Weiss, Michael (2015). "The rite stuff: Lat. *rīte*, *rītus*, TB *rittetär*, TA *ritwatär*, and Av. *raēϑβa-*" *Tocharian and Indo-European Studies* 16, 181–198.

Wertmann, Patrick, Maria Yibulayinmu, Mayke Wagner, Chris Taylor, Samira Müller & al. (2023). "The earliest directly dated saddle for horse-riding from a mid-1st millennium BCE female burial in Northwest China" *Archaeological Research in Asia* 35, 1–16.

Widmer, Paul (2007). "Zur Bedeutung von indoiranisch **nāu̯i̯a-*" *Indo-Iranian Journal* 50, 215–228.

Wikander, Ola (2022). "The Horse that Said *W*: Northwest Semitic *ss*(*w*), an Indo-European (Luwian) Plural, ⟨s̀⟩, and Poetic Motifs in Ugarit and the Hebrew Bible" *Ugarit-Forschungen* 53, 423–459.

Wilkens, Jens (2021). *Handwörterbuch des Altuigurischen. Altuigurisch–Deutsch–Türkisch*. Göttingen: Universitätsverlag.

Williams, Nicholas J.A. (2011). "Middle and Late Cornish" In: Elmar Ternes (ed.), *Brythonic Celtic—Britannisches Keltisch: from medieval British to Modern Breton*, Münchner Forschungen zur historischen Sprachwissenschaft 11, Bremen: Hempen, 237–357.

Winter, Werner (1962). "Nominal and Pronominal Dual in Tocharian" *Language* 38:2, 111–134.

Winter, Werner (1965). "Tocharian evidence" In: Werner Winter (ed.), *Evidence for laryngeals*. London–The Hague–Paris: Mouton & Co., 190–211.

Winter, Werner (1970). "Some Widespread Indo-European Titles" In: George Cardona, Henri M. Hoenigswald & Alfred Senn (eds.), *Indo-European and Indo-Europeans: Papers Presented at the Third Indo-European Conference at the University of Pennsylvania*. Philadelphia: University of Pennsylvania Press, 49–54.

Winter, Werner (1971). "Baktrische Lehnwörter im Tocharischen" *Donum Indogermanicum. Festgabe für Anton Scherer zum 70. Geburtstag*. Heidelberg: Carl Winter, 217–223.

Winter, Werner (1979). "On the importance of links not missing" In: Bela Brogyanyi (ed.), Studies in *Diachronic, Synchronic and Typological Linguistics. Festschrift for Oswald Szemerényi on the Occasion of his 65th Birthday. Volume II*. Amsterdam: John Benjamins, 989–994.

Winter, Werner (1984). "B *śka*, A *śka*; B *ecce*, A *aci*" In: Andrzej Pietrzak (ed.), *Studia Tocharica*. Poznań: Wydawnictwo Naukowe Uniwersytetu Im. Adam Mickiewicza, 117–123.

Winter, Werner (1994). "Syncope in Tocharian A" In: Jens E. Rasmussen (ed.), *In honorem Holger Pedersen. Kolloquium der indogermanischen Gesellschaft vom 26. bis 28. März 1993 in Kopenhagen*. Wiesbaden: Dr. Ludwig Reichert, 401–415.

Witzel, Michael (1999). "Substrate Languages in Old Indo-Aryan (Ṛgvedic, Middle and Late Vedic)" *Electronic Journal of Vedic Studies* 5–1, 1–67.

Yoshida, Yutaka (2009). "Sogdian" In: Gernot Windfuhr (ed.), *The Iranian languages*. London–New York: Routledge, 279–335.

Zhang, Fan, Chao Ning, Ashley Scott, Qiaomei Fu, Rasmus Bjørn, Wenying Li & al. (2021). "The genomic origins of the Bronze Age Tarim Basin mummies" *Nature* 599, 256–261.

Zin, Monika (2014). "Imagery of Hell in South, South East and Central Asia" *Rocznik Orientalistyczny*, 67, 269–296.

Index

A dagger (†) signifies that the form mentioned in the index is a ghost word or reconstruction.

For all languages transcribed in the Latin script, I follow the Latin alphabet. Characters with macron, haček or acute are cited after those without, schwa (ə) is considered as if the last letter of the alphabet.

The order of languages is:

– Proto-Indo-European
– Indo-Iranian languages: Avestan, Ossetic, Scythian, Khwarezmian, Bactrian, Parthian, Wanjī, Middle Persian, New Persian, Balochi, Vedic Sanskrit, Sanskrit
– Other Indo-European languages: Greek, Latin, French, Dutch
– Non-Indo-European languages: Chinese, Old Uyghur, Elamite, Aramaic/Syriac, Arabic

Proto-Indo-European

†*bʰag-* 37
**bʰeg-* 37–38
**bʰeh₂-* 225
†*bʰeh₂g-* 37–38
**-bʰo-* 207
**deh₁-* ~ **dh₁u-* 129
**deh₁u̯-* 128
**dei̯k-ero-* 219
**dem-* 175
**(d)ḱmto-* 103
**du̯oi-mo-* 146
**dʰu̯or-* 96
**gel-* 32–33*n*
**gʰerdʰ-* 112*n*81
**gʰordʰiu̯o-* 112
**gʰortiu̯o-* 112
**gʰortos* 114
**gʰr̥-dʰh₁-ó* 112*n*82
**ǵʰu̯ēr-* 218
**(s)gʷes-* 135
**Hu̯erdʰ-* 99
**h₁ed-net-* 126*n*107
**h₁ei̯-* 110
†*h₁ēd-so-* 61
**h₁éḱu̯o-* 19, 106n, 163
**h₁éti-* 106*n*
†*h₁n̥-ǵʰeu̯eh2(-n)-* 201
**h₁opi-* 80
**h₁oti-* 105
**h₁ou̯ē-* 81
**h₂ed-* 131
**h₂eu̯-* 86
**h₂eǵ-* 94
**h₂ék-u̯on-* 201
**h₂emǵʰ-* 15, 94–95
**h₂ot-ro-* 203
**h₃neh₂-* 124
**h₃nogʰʷo-* 97
**keh₂u̯-* 142
**ḱes-/*ḱseu̯-* 136
**ḱeu-n-/*ḱu̯-en-* 200
**ḱēru̯-o-* 218
**ḱh₂s-u̯o-* 136
**kʷieh₁-* 158
**lei̯H-* 65
**loi̯-* 66, 68
**melit-* 230
**moh₁iu̯o-* 57
†*moh₁iu̯i̯eh₂-* 57
**neḱu-* 97
**nem-* 229*n*
**nēteh₂-* 126*n*107
**nēt-h₂-o-* 124
**néi̯H-to-* 92
†*n-h₁dˢ-tu̯-i̯e/o-* 97
**niH-ti̯o-* 92
**noi̯u̯o-* 35*n*17
**n̥-d(h₃)-o-* 16
**oi̯u̯o-* 25–26
†*peleḱu-* 39–43
**(H)rei̯th₂-* 63–65
**sep-* 74
**spel-d-ske/o-* 73–74
**spol-* 74

Proto-Indo-European (*cont.*)
**-sth$_2$-* 79
**-u̯e* 90*n*
**u̯erH-* 70*n*
**u̯oi* 146

Indo-Iranian Languages

Avestan
aēnah- 23
aiiah- 195
airime 134
armaē šad- 134
axti- 22
ązah- 15, 188
āfš 132
ārmaiti- 134
āϑrauuan- 203
dānō.karš 143
dəmāna- 52, 175
fradaϑa- 48
gaēϑā- 150
gaⁱnti- 224–225
gandərəβa- 230
gantuma- 223
gaona- 149
gərəδa- 115
hacā 106
hvarə.xšaēta- 148
izaēna- 61
ištiia- 196
kamərəδa- 53
kapasti- 138
karəta- 27
kasuuiš 138
kaša- 31
kušaiti 142
maiiā- 57, 174, 186
nas 125
nasu- 97
nauuāza- 174
nāuuiia- 34
niiāzata- 94
nmāna- 175
raēϑβa- 64–65, 186
ratu 108
Sauruua- 218
spərəδa- 74
šiiāta- 72
taša- 40
uruuata- 70
vaβžaka *23*
vaēⁱti- 214
varka- 100
varəδa- 180
varəϑa- 40*n*20
vīspō.xᵛāϑrəm 94*n*71
vərəzāna- 99
vərəzāniia- 99
xrafstra- 234
xšaēta- 148
xumba 110
xᵛanaŋ- 122*n*100
xᵛaēpaiϑiia- 72
zaēna- 75, 186, 190
zaiia- 75
zaoϑra- 128
zəmōištuua- 196, 198–200

Ossetic
D = Digoron, I = Iron
D *æfsijnæ* I *æfsin* 171
ændon 200
ævi 81
æxsin 171
D *æxsijnæg* I *æxsinæg* 79, 171–173
cadæg 158
D *fælmæcun* I *fælmæcyn* 157
færæt 39, 186
D *finkæ* I *fynk* 119
D *gænæ* I *gæn* 204
D *ir* I *iræ* 14
kard 28–29, 43, 187–188
kært 112*n*83
D *læncæ* I *lænk* 118–120, 169, 184, 187–188
D *mud* I *myd* 98
D *mojnæ* I *moj* 172, 189–190
D *radæ* I *rad* 185
D *sujnæ* I *sin* 173*n*
D *sorun/surd* I *suryn/syrd* 218
D *sufcæ* I *syfc* 157
wat 160

Ossetic (*cont.*)
D *wedagæ* I *widag* 214–215, 217
D *wes* I *wis* 214, 216
D *xærwes* I *xæris* 214
xon- 122*n*100
D *xu* I $x_o y$ 161
zældæ 129–130, 165

Scythian
αδιγόρ 131
Βιδάχης 214
Φαραδου 41*n*23
**aspa-* 240

Khwarezmian
āγadik 78
ʾyδʾβrc 164
ʾγd 78
δm 176
δmʾs- 176*n*
γrδmʾn 176*n*
hncy 158
hnčw 200, 201*n*
kcynyk 203*n*
knbynk 204
mǀʾfxwʾny- 122*n*100
mǀkš- 142
pdyk 39–40, 186
rāk 160
sʾk 160
štyc 196
twtyk 254
wʾnk 120, 122*n*100
wrδc 178
xubisk 161
yδʾβrc 164

Bactrian
αβδδινο 170
αβιβταδο 170
αρμαυ- 134
βαγο 38
ιωγο 257
καμιρδο 54–55, 180–181
νιιαζο 95
νιβαλμο 167, 176
*παδοζιγγο 262
πιδοοαυανο 160
σαζαμανο 170
σπαχνιιο 56
φαρο 47–48
φρομιγγο 262
χαγγαρο 28
χοαν- 122*n*100
χοβιγο 72
**kōžgā* 139
**madu-* 145
mynʾg 253
**sabōli* 56

Parthian
āγad 78
†*ʾgtygysʾ* 77
bar 96
frhʾh 47–48
gyānβar 147
kwš- 142
niyāz 93, 95
**patihān* 261
prdk 41*n*23
xwn- 122*n*100
žafrān 126

Wanjī
amγal 179
devur 96
γain 181
kain 181
kin- 181
kirmγaz 181
main 181
pal 179
weǰ 38
wisk- 215*n*20
zil 179

Middle Persian
afsānag 160*n*
awistād 80
ayāb 86, 86*n*
ayyād 17
‹ʾynykyh› 22–23
āyandagīh 22
āz 93
bānūg 175*n*
-ēw 25
dānag 143, 144*n*
garōdmān 176*n*

Middle Persian (*cont.*)
gāl 115–116
grōh 32
**grōhag* 32
gul 180
hazārag 103
kamāl 180
**kanag* 138
**kanagīg* 138
‹kynyk› 138
kōf 32
kwhš- 142
nihān 125–126, 126*n*106, 185
nihuft- 126
niyān 125
niyāz 93, 95
niyāzōmandīh 94*n*72
prdk-y 41*n*23
radag 62, 185
rēm 68
sadag 103
sāl 177
šād 73
šh 126
twhš- 142
wāng 120, 122*n*100
wiyābān 126
xwad 162
xwān- 122*n*100
yak 172*n*
zamīgēn 199
zēn 186, 190

New Persian
afsāna 160*n*
astar 18*n*4
az 106
āb 132
ārām 134
āz 93
āzār 22
ʿālam 150
bāng 120
bēd 215
dasta 104
dāna 143, 144*n*
-ē 25
-ihå 27
Inǰigān 164
gand 224
gandum 223
gul 180
gurōh(*a*) 32
hōšāz 15*n*2
kabōtar 79
kamāl 180
kana 138
kanaf 204
kand- 138*n*122
kārd 187
kīna-kaš 138
kōh 33
mān 52
mul 53
nahān 125, 185
niyāz 93, 95
o yå 85
panāh 125, 185
panhān 125
parwanda 51, 187
sad 104
sada 103
sar 104
sāl 177
šād 73
tabar 40
tūtī 254, 254*n*3
ustād 80
wan-dāna 143*n*126
xanǰar 28*n*10
xišt 197
xumb 110
xuršēd 148
xwad 162
yak 172*n*147
yā 86
yād 17
zi 106

Balochi
ēyōk 172*n*
grōp 32, 185
gwanǰ- 120
gwānk 120, 122*n*100
hīz 61
īt 195
kumb 110

Vedic Sanskrit

abhí 169
áṃhas- 15
arámati- 134, 134*n*
átharvan- 203
áyas- 195
énas- 24
evá 26
dhānā́- 143
dvā́ra- 96
gandhá- 224
gandharvá- 230
gardabhá- 206–209, 240
godhū́ma- 223–225
gr̥há- 112
hásta- 31
hetí- 190
íṣṭakā- 195, 200
jásu- 135
kāñcaná- 209–210
māyā́- 57, 173–174
mūrdhán- 53
paraśú- 39–41
nāthá- 123–125
rā́sabha- 208
saparyáti 74
sin̐há- 219
Śarvá- 218
śúka- 255*n*3
valká- 100
vratám 70–71
vr̥ṣabhá- 228
vāṇá- 120–121

Sanskrit

B = Buddhist Sanskrit

abhijñā 46–47
anīti 92*n*68
astra- 76
āgata- 78
bhoga- 189
bodhi 44
gaṇa- 32
gardabha- 206
godhūma- 224
gr̥hya- 112
haṃsa- 254
B *iṣṭā-* 195
iṣu- 75, 76
iṣukāra 75*n*
jagat- 150, 150*n*
kacchapa- 123
kacchū- 138
kesara- 219
kilāsa- 136, 140
koṭi 43
krauñca- 209–212, 228
kumbha- 109–111
B *kuṣṭha-* 135–136, 140
Lankā-pura 118
lepa- 66
loka- 150
lopaśa- 213
B *makuṭa-* 29
mukuṭa- 29
nāthaka- 123
nidhāna- 125
ogha- 34
B *phala* 44
pratijña- 107
pratimā- 107
sūcīmukha- 108*n*
surā- 98
tr̥ṇa- 131
tūbarika 194
upādāna 16
varga- 32
vimanas- 146
virodha- 32*n*
vyāma- 31
yojayati 128
yuj- 62, 67

Other Indo-European Languages

Greek

ἀλώπηξ 213
ἄναξ 123
ἀριθμός 63
αὐλή 114*n*85
δόμος 175
δύναμαι 128
ζῦθος 128*n*111

Greek (*cont.*)
κάνναβις 204
κνῆκος 209
κρύπτω 31
λοιμός 68
μέλισσα 210
ὀνίνημι 124
πέλεκυς 39–41
φαγεῖν 37–38
χόνδρος 223
χόρτος 112
ψίττακος 254*n*3

Latin
aluīna 212
angor 15
angō 94*n*71
angustia 94
aut 84
cohors 114
contaminare 66
fenestra 262–263
hortus 112
pellō 73
quietus 158
rīte 63–64
sedeō 133
sepeliō 74
sēdāre 133
super(*i*) 106
vel 84

French
abeille 210
apparition 78
bruit 121
centaine 103
céréales 224
cour 113–115
domestique 115*n*89
fenêtre 262
gruau 223
hâve 137
meurtrière 262–263
mouche à miel 211
ou bien 85–86
rivière 133
vert 130

Dutch
eng 94
groentje 130
hennep 204
hof 114
honing 209
kop 104
melaats 135
ofwel 85
vadem 31

Non-Indo-European Languages

Chinese
MC = Middle Chinese
bāchéng 爸城 261
bōlí chūang 玻璃窗 260
fú 福 44
jiǔ 酒 98
mìfēng 蜜蜂 210
shǔ jiǔ 黍酒 220
MC *srij-tsiX* 獅子 219
sù jiǔ 粟酒 220
tōng 通 47
MC *ts*^{h}*ɛjŋ* 青 258
MC *tsian*^{h} 箭 76
xū 鬚 102
yīngwǔzi 鸚鵡子 256–259
Yuezhi 月氏 251

Old Uyghur
ap 90–92, 191
arhant kutı 45
ay-tängri 148
baltō 43, 186
böz 203
burhan kutı 45
išič 194
karšı 111
katır 18–21, 18*n*4
koltı 43

Old Uyghur (*cont.*)
kün-tängri 148
kut 44–45
pačiŋ 261
toti 254
yir-tängri 148

Elamite
dabara 40
kur-da-bat-ti-iš 115
kurtaš 36n, 116
kur-ti-e-bar-šá 115
kur-ti-ia-ma 115
Muštimarda 53*n*
Ú-mi-ia-áš-šá 57

Aramaic/Syriac
debbōrā 211*n*14
gnb' 204*n*10
grwhq' 32–33
plq 41
zibbūr 211*n*14

Arabic
T = Tunisian
T *dawla* 91*n*62
kaʔs 104*n*75
T *ou bien* 85
T *wala* 85–86, 91, 191
ward 180
zanbūr 211*n*14

"How did the Tocharians reach China?", "Who did they meet on the way?" are some of the most intriguing questions in Indo-European studies. This book is zooming in on a specific part of the question: on their way to China, Tocharians were in contact with an Iranian people living in the south Siberian Steppes, and with a people related to the Oxus Civilization (BMAC). This Iranian people spoke a specific language, called here "Old Steppe Iranian". They gave Tocharians many words, such as *mañiye* 'servant', *etswe* 'burden-carrying horse' or 'mule', *pāke* 'portion, share'. The BMAC-related people gave the Tocharians other words such as *etre* 'hero' and *kercapo* 'donkey'. This book reconstructs features of the language of both these peoples, and examines how they influenced the Tocharians. Based on the latest archaeological findings, it also suggests a reconstruction of the chronology and the way the Tocharians followed before entering the Tarim Basin.

Winner of the 2nd prize for the best dissertation of the Indogermanische Gesellschaft prize for the best Indo-European studies dissertation.

Chams Benoît Bernard, Ph.D. (2023), is a postdoctoral researcher at Cologne University and a linguist specialized in Indo-Iranian and Tocharian languages. He works on literature, poetry, and has a particular interest in the history of languages, and etymology. He has published articles and works on various topics, such as Middle and New Persian, Tocharian etymology, Judeo-Persian, Indo-European and Semitic languages as well as religious studies.

ISBN 978-90-04-73252-0
ISSN 0926-5856
brill.com/lsie