Universiteit
Leiden
The Netherlands

# Tocharian and Samoyed: on the question of Uralic substrate influence in Tocharian

Warries, A.R.

**Citation**
Warries, A. R. (2025, June 18). *Tocharian and Samoyed: on the question of Uralic substrate influence in Tocharian*. Retrieved from https://hdl.handle.net/1887/4250485

| | |
|---|---|
| Version: | Publisher's Version |
| License: | Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden |
| Downloaded from: | https://hdl.handle.net/1887/4250485 |

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

# Tocharian and Samoyed

## On the question of Uralic substrate influence in Tocharian

Abel Warries

# Tocharian and Samoyed

## On the question of Uralic substrate influence in Tocharian

Tocharian and Samoyed
On the question of Uralic substrate influence in Tocharian

Abel Warries
PhD dissertation
Leiden University, The Netherlands

Copyright © 2025 Abel Warries. All rights reserved.

Cover image: pastel painting by Helen van Hoorn.

This research was carried out as part of the research project “The Tocharian Trek: A linguistic reconstruction of the migration of the Tocharians from Europe to China”, funded by the Europen Research Council under the European Union’s Horizon 2020 Excellent Science programme (Grant agreement ID: 758855).

# Tocharian and Samoyed

## On the question of Uralic substrate influence in Tocharian

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 18 juni 2025
klokke 13:00 uur

door
Abel Radu Warries
geboren te Amsterdam
in 1997

| | |
|---|---|
| Promotores: | Prof. dr. A.M. Lubotsky |
| | Prof. dr. M. Peyrot |
| Promotiecommissie: | Dr. O. Khanina (University of Helsinki) |
| | Prof. dr. A. Kloekhorst |
| | Prof. dr. G.J. Kroonen |
| | Dr. M. Zhivlov (National Research University HSE, Moscow) |

# Acknowledgements

This dissertation is the product of ten years at Leiden University, in the company of many inspiring friends and colleagues for whom I have the utmost appreciation. First and foremost, I would like to thank my supervisors, Michaël Peyrot and Sasha Lubotsky, for their encouragement and for sharing their insights throughout these years. They not only guided me throughout these arduous years of PhD research, but also helped form me as a linguist and a researcher. I am also grateful to the rest of the teaching staff at the Leiden Linguistics programme, in particular the CIEL section: Tijmen Pronk, who guided me when I wrote my BA thesis, Guus Kroonen, Alwin Kloekhorst, and Lucien van Beek.

I would not have stayed in Leiden for this long if I had not met so many wonderful people here. With Anouk Mudde and Lotte Meester I shared a love for Celtic, with Virginia Oogjen a budding interest in Uralic, and with Femke Montagne the world of birds (may we see many more together). When I embarked on my PhD, I also came to the world of natural history at Naturalis, which ended up being a continuous part of my life over the last five (coming on six) years. I would like to thank everyone at the museum for their kindness, and especially my PB colleagues for lifting my spirits in the difficult final stages of working on this dissertation. I sincerely thank Martijn Knapen and Thom van Hugte for their continued friendship over the years and lament the many kilometers of treacherous train tracks that separate us these days. Many thanks also to my longest Leiden friend and now paranymph Cid Swanenvleugel, without whom this day would have been unimaginable.

I have been fortunate to share the Tocharian part of this journey with Alessandro del Tomba, Chams Bernard, Federico Dragoni, Olga Lundysheva, Xiaoqiang Meng, and Yezi Mu, and I especially thank my fellow project members and friends Louise Friis, Niels Schoubben and Stefan Norbruis, and my office mates Ahmed Sosal and Saskia Dunn. Over the years, it has furthermore been a great pleasure and a privilege to work (and have lunch and borrels) together with such excellent people as Andrew Wigman, Anthony Jakob, Axel Palmér, Bente de Graeve, Carmen Spiers, Hester Groot, Jesse Wichers Schreur, Jiang Wu, Lis Kerr, Natasja Delbar, Nora Julmi, Olga Nozdracheva, Paulus van Sluis, Rasmus Torsø, Simon Poulsen, Sophia Nauta, Tobias Søborg, and Xander Vertegaal. As the last of the pandemic-time Indo-Europeanist PhD group to now leave the nest, I am eternally grateful for everyone's continued encouragement and support.

The Uralic world and the Samoyed part of this journey opened up to me more properly at a later stage while working on this PhD project. I am grateful for the teachings of Gerson Klumpp, Josefina Budzisch, Mikhail Zhivlov, Tapani Salminen, Valentin Gusev, and I especially thank Olesya Khanina for welcoming me at Helsinki University during the spring of 2023 and for the many thought-provoking discussions. My sincere thanks also to Juho Pystynen, Kaisla Kaheinen, Petri Kallio, and Sampsa Holopainen for

discussing with me all aspects of historical Uralic and Samoyed linguistics, and for their helpful feedback. My stay in Helsinki was further brightened by the friendship of Simone and Tuomo.

On this long and winding road I would further like to give mòran taing to the team of the Sabhal Mòr Ostaig summer courses, especially to my teachers Alec Bhaltois, Catrìona Nic an t-Saoir, and Muriel Fisher, and also to the many other kind people I met during my visits to Eilean a' Cheò. I have the fondest memories of days and late nights spent in merry Gàidhlig conversation with Eilís, Emma, Jennie, Josie, Lewis, Liam, Marcel, Melvin, Natalie, Nathaniel, and many more.

In the one free week between the end of the MA and the start of the PhD, I took a short vacation to Vienna and met Eugenio Loi. I could never have imagined then that we would still be such close friends all these years later, and I am overjoyed that he agreed to be here by my side as one of my paranymphs. I am also grateful to Simon Rey for his continued friendship, and I look forward to continuing our walks, talks and museum visits for many years to come.

Last, but certainly not least, I come to the people who have been in my life the longest. Joosje, Koen, Marion, and René have been like a second family for many years, and I have the greatest appreciation for their continued friendship and hospitality. To my parents, Helen and Michael, I owe the biggest debt of gratitude. Without their help and advice and their belief in me I would be standing here. Thank you both.

# Table of Contents


**1 Introduction ..... 1**

**1.1 A Uralic piece to the Tocharian puzzle ..... 1**

**1.2 Samoyed and the other Uralic languages ..... 3**

**1.3 Archaeological background ..... 7**

**1.4 Methodological considerations ..... 10**

**1.5 Research questions ..... 13**

**2 Relative chronology of Tocharian sound changes ..... 17**

**2.1 Introduction ..... 17**

**2.2 Tocharian and Proto-Indo-European phonology ..... 18**

**2.3 Early consonant developments ..... 21**

2.3.1 Laryngeals ..... 22
2.3.2 Syllabic resonants ..... 25
2.3.3 Palatovelars and labiovelars ..... 27
2.3.4 The stop system ..... 36
2.3.5 Summary ..... 49

**2.4 Palatalization ..... 51**

2.4.1 Palatalization as it affected the consonant system ..... 51
2.4.2 Developments of PIE **i* and **e* ..... 54
2.4.3 Developments of PIE **u* ..... 55
2.4.4 Palatalization as it affected the vowel system ..... 57
2.4.5 Developments of PIE **ō* and **eh₂/ā* ..... 59
2.4.6 Summary ..... 63

**2.5 Late vowel developments ..... 65**

2.5.1 *u*-umlaut of pre-PT **e* ..... 65
2.5.2 *o*-umlaut of pre-PT **u* ..... 66
2.5.3 *o*-umlaut of pre-PT **e* ..... 67
2.5.4 *o*-umlaut of pre-PT **a* ..... 69
2.5.5 *a*-umlaut of pre-PT **e* ..... 70
2.5.6 Effects of labial consonants on vowels ..... 72
2.5.7 Summary ..... 83

**2.6 Information from language contact ..... 83**

2.6.1 Early Iranian languages ..... 83
2.6.2 "BMAC language" ..... 86
2.6.3 Summary ..... 86

**2.7 Overview of Tocharian relative chronology ..... 87**

**3 Relative chronology of Samoyed sound changes ..... 95**

**3.1 Introduction ..... 95**

**3.2 Samoyed and Proto-Uralic phonology ..... 95**

**3.3 Loss of final PU *-*i* ..... 98**

**3.4 Changes affecting single consonants ..... 99**
3.4.1 PU **δ*, **s* and **š* ..... 99
3.4.2 PU **l*, **δ′* and **j* ..... 101
3.4.3 PU **w* and **γ* ..... 104
3.4.4 Summary ..... 106

**3.5 Changes affecting consonant clusters ..... 107**
3.5.1 PU **rk* and **lk* ..... 107
3.5.2 PU **kC*, **Ck* and geminates ..... 108
3.5.3 PU pre-consonantal glides ..... 110
3.5.4 Summary ..... 110

**3.6 Changes affecting non-low vowels ..... 111**
3.6.1 PU **u*, **i* and **ü* ..... 111
3.6.2 PU **ę* and PS **ę* and **į* ..... 116
3.6.3 PU **o* ..... 118
3.6.4 Summary ..... 121

**3.7 Changes affecting PU **a* ..... 121**
3.7.1 PU second-syllable **-a* to **-å* and *ə̂* ..... 122
3.7.2 PU first-syllable **a* to **å* and **a* ..... 124
3.7.3 PU second-syllable **-a* to **-ä* ..... 126

**3.8 Information from language contact: Yukaghir ..... 126**

**3.9 Overview of Samoyed relative chronology ..... 128**

**4 Phonological comparison ..... 131**

**4.1 Introduction ..... 131**

**4.2 Comparison of the consonant systems ..... 132**
4.2.1 The pre-Proto-Tocharian and pre-Proto-Samoyed stop systems ..... 132

4.2.2 Stop systems of other languages in the region ..... 134

**4.3 Comparison of the vowel systems ..... 136**
4.3.1 The pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems ..... 136
4.3.2 Vowel systems of other languages in the region ..... 139

**4.4 Overview and conclusions ..... 141**

**5 The Tocharian accent ..... 143**

**5.1 Accentuation in Tocharian ..... 143**
5.1.1 Basics of the Tocharian B accent ..... 144
5.1.2 Problems with the Tocharian B accent ..... 145

**5.2 Epenthesis and syncope ..... 147**
5.2.1 General remarks on epenthesis ..... 147
5.2.2 Epenthesis in *na*-presents ..... 149
5.2.3 Epenthesis in *s*-preterites ..... 162
5.2.4 Epenthesis in *Cst* and *Csk* ..... 163
5.2.5 Syncope ..... 164
5.2.6 The accent through the lens of syncope and epenthesis ..... 165
5.2.7 Intermediary summary ..... 167

**5.3 The accent in finite verbs ..... 168**
5.3.1 Mobile and fixed accent in the Tocharian B verb ..... 168
5.3.2 The accent in the thematic present [2] and *s*-present [8] ..... 172
5.3.3 The accent in the *s*-preterite [3] ..... 174
5.3.4 The accent in the subjunctives [1] and [5] ..... 179
5.3.5 The accent in the causative present-subjunctive [9] and preterite [2] .. 196
5.3.6 The phonologization and morphologization of the accent ..... 199

**5.4 The accent in other categories ..... 200**
5.4.1 Vowel contractions and the accent in the preterite participle ..... 200
5.4.2 Nasal prefixes and the accent ..... 208
5.4.3 The accent in kinship terms ..... 214
5.4.4 The accent in the secondary cases ..... 215
5.4.5 The accentuation of compounds ..... 217
5.4.6 The accentuation of loanwords ..... 218

**5.5 Tocharian A accentuation ..... 219**
5.5.1 Apocope, vowel weakening and syncope in Tocharian A ..... 219
5.5.2 Vowel contractions in Tocharian A ..... 221
5.5.3 Indirect indications of pre-Tocharian A accent ..... 222

**5.6 Conclusions on the accent ..... 223**

**6 Agglutinative case systems ..... 225**

**6.1 Introduction ..... 225**

**6.2 The Tocharian case system ..... 228**
6.2.1 The functions of the Tocharian secondary cases ..... 229
6.2.2 The forms of the Tocharian cases ..... 234

**6.3 The Samoyed case system ..... 241**
6.3.1 The Samoyed cases and their functions ..... 241
6.3.2 The Proto-Samoyed and Proto-Uralic case systems ..... 246

**6.4 Comparison of the Tocharian and Samoyed case systems ..... 250**
6.4.1 Case inventory ..... 250
6.4.2 Functional comparison ..... 251
6.4.3 Comparison of development ..... 252

**6.5 Conclusions on the agglutinative case systems ..... 253**

**7 Participial systems ..... 255**

**7.1 Introduction ..... 255**
7.1.1 Participles and their orientation ..... 256

**7.2 Tocharian ..... 260**
7.2.1 The *nt*-participle ..... 260
7.2.2 The *m*-participle ..... 262
7.2.3 The preterite participle ..... 268
7.2.4 The gerundives: necessitative and potential participles ..... 281
7.2.5 The privative: a negative participle ..... 288
7.2.6 Summary ..... 290

**7.3 Samoyed ..... 291**
7.3.1 Uralic and areal context ..... 297

**7.4 Conclusions on the participial systems ..... 298**

**8 Object marking on the verb ..... 301**

**8.1 Introduction ..... 301**

**8.2 The Tocharian pronoun suffixes ..... 303**
8.2.1 Functions of the pronoun suffixes ..... 304
8.2.2 Origins and status of the pronoun suffixes ..... 307
8.2.3 Similar pronominal elements in other Indo-European languages ..... 309

**8.3 The Uralic objective conjugation ..... 312**
8.3.1 General overview of the objective conjugation ..... 312
8.3.2 Formal reconstruction of the objective conjugation ..... 314
8.3.3 Functional reconstruction of the objective conjugation ..... 350

**8.4 Comparison between Tocharian and Uralic ..... 362**

**9 Potential loanwords ..... 367**

**9.1 Introduction ..... 367**

**9.2 Earlier comparisons ..... 368**
9.2.1 pre-PT **septə(m)* 'seven' ⇒ pre-PS **säjt³wə* 'id.' ..... 368
9.2.2 pre-PT **wesa* 'gold' and PS **wäsa* 'metal, iron' ..... 369
9.2.3 pre-PT **mēnē-* 'moon' ⇒ PS ±**mänä/iwjə* 'full moon' ..... 370
9.2.4 pre-PT **monko-* 'lack, deficit' ⇒ PS **måŋkâ* 'poverty, need' ..... 370
9.2.5 Pre-PT **kwënə(m)* acc. of 'dog' ⇒ pre-PS **wẹn* 'dog' ..... 371

**9.3 New comparisons ..... 372**
9.3.1 pre-PT **lonko-* 'steep slope; cliff' ⇒ PS **lâŋkâ* 'steep bank; steep' ..... 372
9.3.2 pre-PT **lenk-* 'dangle, sway, hang' ⇒ pre-PS **laŋkå-* 'hang, dangle' ..... 372
9.3.3 pre-PT **der-* 'split, separate' ⇒ pre-PS **tar-* 'divide, separate' ..... 373
9.3.4 pre-PT **newo-* 'new'⇒ Pre-PS **ńuwV* 'child, young boy' ..... 374
9.3.5 pre-PT ?**mëywya* 'power, strength' ⇒ pre-PS **mẹjâ* 'sturdy, solid' ..... 374

**9.4 Discussion ..... 376**

**10 Conclusions ..... 379**

**10.1 Overview ..... 379**

**10.2 Phonology ..... 380**

**10.3 Accent ..... 382**

**10.4 Morphology ..... 383**

**10.5 Lexicon ..... 385**

**10.6 Evaluation ..... 386**

**Etymological reference list 1: Tocharian ..... 389**

**Etymological reference list 2: Samoyed ..... 413**

**Bibliography ..... 457**

**Abbreviations of languages**

| | | | |
|---|---|---|---|
| Av. | Avestan | Md. | Mordvin |
| E | Erzya (Mordvin) | Ms. | Mansi |
| EM | Erzya and Moksha | Mt. | Mator |
| En. | Enets | Ne. | Nenets |
| EnF | Forest Enets | NeF | Forest Nenets |
| EnT | Tundra Enets | NeT | Tundra Nenets |
| Goth. | Gothic | Ng. | Nganasan |
| Gr. | Greek | OSIr. | Old Steppe Iranian |
| Fi. | Finnish | PIE | Proto-Indo-European |
| FU | Finno-Ugric | PNe. | Proto-Nenets |
| Hitt. | Hittite | PS | Proto-Samoyed |
| Hu. | Hungarian | PSaa. | Proto-Saami |
| IIr. | Indo-Iranian | PT | Proto-Tocharian |
| Kh. | Khanty | PTrk. | Proto-Turkic |
| KhKaz. | Kazym Khanty | PU | Proto-Uralic |
| KhO | Obdorsk Khanty | PYuk. | Proto-Yukaghir |
| KhSur. | Surgut Khanty | PYen. | Proto-Yeniseian |
| KhTrj. | Tremjugan Khanty | SaaN | North Saami |
| KhVj. | Vasjugan Khanty | SaaS | South Saami |
| KhVVj. | Vakh-Vasjugan Khanty | Sk. | Selkup |
| Khot. | Khotanese | Skt. | Sanskrit |
| KoZ | Komi Zyrian | TA | Tocharian A |
| Lat. | Latin | TB | Tocharian B |
| M | Moksha (Mordvin) | Udm. | Udmurt |

**Abbreviations of grammatical terms and glosses**

| | | | |
|---|---|---|---|
| 1 | first person | agn. | agent noun |
| 2 | second person | all. | allative |
| 3 | third person | aor. | aorist |
| A | agent | caus. | causative |
| abl. | ablative | cneg. | connegative |
| abs. | absolutive | com. | comitative |
| acc. | accusative | conj. | conjunction |
| act. | active | conn. | sentence-connector |
| adj. | adjective | dat. | dative |
| adjz. | adjectivized | def. | definite |

| | | | |
|---|---|---|---|
| dem. | demonstrative | opt. | optative |
| dist. | distal (demonstrative) | ord. | ordinal |
| dst. | (pre)destinative | P | patient |
| ela. | elative | pass. | passive |
| emph. | emphatic | per. | perlative |
| erg. | ergative | pers. | personal (pronoun) |
| ext. | extension | pf. | perfect |
| f. | feminine | pfv. | perfective |
| freq. | frequentative | pl. | plural |
| fut. | future | PN | proper noun |
| gen. | genitive | poss. | possessive |
| ger. | gerundive | priv. | privative |
| ill. | illative | pro. | prolative |
| imp. | imperative | prog. | progressive |
| indef. | indefinite | prohib. | prohibitive |
| inf. | infinitive | pron. | pronoun |
| ins. | instrumental | prs. | present |
| ipf. | imperfect | prt. | preterite |
| ipfv. | imperfective | pst. | past |
| itr. | intransitive | ptc. | participle |
| iter. | iterative | ptcl. | particle |
| lat. | lative | refl. | reflexive |
| loc. | locative | rel. | relative (pronoun) |
| m. | masculine | rprt. | reportative |
| mid. | middle | S | subject |
| n. | neuter | sbc. | subjective conjugation |
| necess. | necessitative | sbj. | subjunctive |
| neg. | negative | sg. | singular |
| nom. | nominative | vn. | verbal noun |
| obc. | objective conjugation | voc. | vocative |
| obl. | oblique | | |

**Other abbreviations**

| | | | |
|---|---|---|---|
| arch. | archaic | id. | same meaning as previous |
| der. | derivation | lit. | literally |
| dial. | dialectal | p.c. | personal communication |
| fn. | footnote | trsl. | translation |
| ibid. | same reference as previous | | |

**General symbols**

| | |
|---|---|
| [_] | phonetic representation |
| /_/ | phonological representation |
| \|_\| | morphological representation |
| ‹_› | infix in morphological representation |
| ⟨_⟩ | orthographic representation |
| *_ | reconstructed form |
| **_ | non-existent or incorrect form |
| _* | expected form, e.g., in a partially attested paradigm |
| A > B | A developed into B by sound change |
| A >> B | A developed into B by analogy |
| A → B | B is derived from A by derivation or semantic shift |
| A ⇒ B | B is derived from A by borrowing |

**Special symbols**

| | |
|---|---|
| *ċ* | provisional notation for an intermediary stage between PIE *$t\dot{i}$ and PT *$t^s$, see 2.3.4.2 |
| *ḋ* | provisional notation for an unclear intermediary development of PIE **d* in early pre-Proto-Tocharian, see 2.3.4.4 |
| *$\dot{z}^h$* | provisional notation for an intermediary stage between PIE *$d^h\dot{i}$ and PT *$t^s$, see 2.3.4.2 |

**Cover symbols**

| | |
|---|---|
| *C* | any consonant |
| *N* | any nasal |
| *n*¹ | Proto-Samoyed **n* or **ń* |
| *P* | any labial or labialized consonant |
| *R* | any resonant (*l*, *m*, *n*, *r*) |
| *t*¹ | Proto-Samoyed **c* or **t* |
| *t*² | Proto-Samoyed **c* or **s* or **t* |
| *t*³ | Proto-Samoyed **c* or **k* or **s* or **t* |
| *V* | any vowel |

# 1 Introduction

This dissertation is about the hypothesis that early Tocharian was in contact with a form of Uralic. In this introduction I will first give an overview of the overarching context in which the research of this dissertation has been conducted in section 1.1., followed in section 1.2 by some further introductory information on the Samoyed languages and their place within Uralic. The archaeological background of the main research question will be provided in section 1.3, and the methodological background of my linguistic research in section 1.4. Finally, section 1.5 will give an overview of the structure of the dissertation and introduce the most important research questions that require an answer as intermediate steps toward the final, overarching conclusion.

## 1.1 A Uralic piece to the Tocharian puzzle

Tocharian is in many ways an outlier among the Indo-European languages. On the eastern edge of the Indo-European linguistic area, it does not have any apparent close connections with other branches and shows an array of both archaic and innovative features. The Tocharian language group contains two closely related languages: Tocharian A (or "Agnean", "East Tocharian") and Tocharian B (or "Kučean", "West Tocharian"). The speakers of Tocharian A and B lived along the northern Silk Road in the Tarim Basin until their language disappeared around 1000 CE. The oldest Tocharian B texts date to around 400 CE, so that we have only a narrow window of approximately 600 years on this language. Tocharian A attestations begin approximately 250 years later. Further limiting our knowledge of the language is the fact that the contents of the sources is overwhelmingly Buddhist literature, and that most manuscripts are damaged. The great majority of texts are more properly characterized truly as fragments. There are approximately 9000 fragments of Tocharian B texts and 2000 of Tocharian A texts (Peyrot 2013: 1-2).

Much of the prehistory of the Tocharians is still shrouded in mystery, especially as we cast our view farther back in time. When the earliest attested texts were written around 400 CE, both Tocharian languages were already in the Tarim Basin as separate entities, and they were in contact with neighbouring languages like Middle Chinese, early Turkic and Middle Iranian languages such as Khotanese (e.g., Lubotsky & Starostin 2003; Peyrot 2018b; Dragoni 2023). The contact with Iranian languages stretches even farther back in time to the (pre-)Proto-Tocharian period, as has been demonstrated on the basis of prehistoric borrowings from Old Iranian into Tocharian (Bernard 2023). Interactions with Iranian continued in the form of contacts with Khotanese in the Tarim Basin itself (Dragoni 2023). Loanwords from another, unknown language have been related to the

language of the BMAC, which is otherwise only known in the form of loanwords into Indo-Iranian (Lubotsky 2001; Bernard 2023).

Beyond that, not much more is known with certainty about the early prehistory of the ancestors of Tocharian. However, since they spoke an Indo-European language, they must derive from the Proto-Indo-European speech community, which originated in the Pontic-Caspian Steppe (e.g., Mallory 1989; Anthony 2013; Anthony & Ringe 2015; Allentoft 2015). The potential route of the Tocharians from the Proto-Indo-European homeland to the Tarim Basin has been partially charted on the basis of archaeological and genetic evidence (e.g., Mallory & Mair 2000; Kroonen, Barjamovic & Peyrot 2018: 7–9), but much uncertainty still remains (see especially Mallory 2015).

A pertinent part of this discussion is the linguistic relationship between Tocharian and the other Indo-European languages. Anatolian is considered by a growing consensus among scholars to have been the first branch of Indo-European to split off from the protolanguage (see, e.g., Sturtevant 1933: 30, Kloekhorst 2008: 7–11; Oettinger 2013/14), and Tocharian has received a similar treatment due a number of lexical and grammatical quirks. This has given rise to the "Tocharian second" hypothesis. According to the Tocharian second hypothesis, it is thought that Tocharian was the second branch of Indo-European to split after Anatolian, based on a variety of features that appear to set Tocharian apart from the other non-Anatolian Indo-European languages (e.g., Schmidt 1992, Winter 1997, Carling 2005).

There is no consensus about the validity and applicability of each individual feature (for discussions, see Malzahn 2016; Peyrot 2019b; Friis 2024: 20–36), but if Tocharian was indeed the second branch to split from Proto-Indo-European, any Tocharian feature that happens to not be found in Anatolian could in principle represent an archaism in Tocharian that was lost or altered in the other branches of Indo-European. The directionality of a change cannot always be determined with certainty, however, and in such cases, we must remain cautious of assuming either Tocharian archaism or Tocharian innovation until a scenario can be presented that accords better with either option.

In general, many aspects of Tocharian that make it stand out are considered to be particularly innovative rather than archaic. A number of such features that appear unusual from an Indo-European perspective have been associated with language contact, particularly often with Uralic substrate influence (e.g., Krause 1951; Kallio 2001; Bednarczuk 2015; Peyrot 2019a). Uralic languages were historically spoken in southern Siberia along part of the route the Tocharians may have taken to the Tarim Basin, so that a demonstration of Tocharian-Uralic contact can provide a further anchor point for our understanding of early Tocharian prehistory.

A particular relationship with the easternmost branch of Uralic, the Samoyed branch has been advanced by Peyrot (2019a). Seeing as early Samoyed was spoken in an area of

southwestern Siberia almost directly north of the Tarim Basin (e.g, Janhunen 1998: 457), this would fit well in geographical terms. Closer investigation of the contact hypothesis is the topic of this dissertation. I will scrutinize the various arguments and make explicit what we are comparing and what we can base ourselves on, in order to further our understanding of this hypothetical language contact and to see if part of the Tocharian puzzle can indeed be filled in with Uralic pieces.

## 1.2 Samoyed and the other Uralic languages

There are nine uncontroversial branches of Uralic: Saami, Finnic, Mordvin, Mari, Permic, Mansi, Khanty, Hungarian, and Samoyed. According to the traditional phylogeny, Uralic is as a generally consistent westward-branching, language family. In this model, Samoyed, the easternmost branch, is more distant to all the other branches than the other branches are to one other (e.g., Janhunen 1981; 2000: 59–60). However, according to recent scholarship, the old model of binary splits cannot be taken for granted (K. Häkkinen 1984; Salminen 2002), and no new model has been accepted as an adequate replacement yet (see Aikio 2022a: 3–4). The phylogenetic position of Samoyed thus remains debated, like that of Tocharian.

While the phylogenetic reality of a primary split between Samoyed and Finno-Ugric is uncertain, it remains useful to differentiate between these two groups, and it is possible to use the term Finno-Ugric to refer to all branches of Uralic except for Samoyed (cf. e.g., Grünthal et al. 2022: 492). Further grouped together, perhaps in a true genealogical sense, are the "Ugric" languages Mansi, Khanty and Hungarian, of which Mansi and Khanty are together referred to as "Ob-Ugric". It is debated to what extent the similarities between these languages are due to shared innovation or due to parallel innovation and later convergence.

The Samoyed branch itself is usually divided into eight languages: Nganasan, Tundra Enets, Forest Enets, Tundra Nenets, Forest Nenets, Selkup, Kamas, and Mator. Of these, Kamas and Mator are extinct. Especially Mator is attested very incompletely and we have onlly little information on morphology or syntax (Helimski 1997 is a comprehensive treatment of the Mator material). Kamas was documented more extensively, and its grammar is known much better (see, e.g., Klumpp 2022). Selkup is subdivided into dialects in up to five main groups: northern (Taz) Selkup, central (Tym) Selkup, southern Selkup with the Middle Ob and Upper Ob dialects, and Ket Selkup (see Klumpp & Budzisch 2023: 898 with references). Tundra and Forest Nenets are closely related, but distinct, and the situation with Tundra and Forest Enets is of a similar nature (see, e.g., Mus 2023 on Nenets and Khanina & Shluinsky 2023 on Enets for more details).

The Nenets and Enets languages together with Nganasan are conventionally known as the northern Samoyed languages, while Selkup, Kamas, and Mator make up the

southern Samoyed languages. Kamas and Mator are sometimes referred to together as Sayan Samoyed, so named for the Sayan Mountains where they were spoken (Janhunen 1998: 458–459). These subgroupings are perhaps areal rather than genealogical, and the true phylogenetic structure of the Samoyed branch is still uncertain. The divide between northern and southern Samoyed languages is sometimes taken as a true phylogenetic divide, but it has also been argued that Nganasan and Mator, the northernmost and the southernmost language, represent particularly early offshoots (Janhunen 1991; Janhunen 1998: 459). The issue requires further research.

The traditional position of Samoyed as the first branch to split off from Proto-Uralic has led to its image as the odd one out, or "the Anatolian of Uralic" from an Indo-European point of reference. However, several important aberrant features of Samoyed concern the lexicon, and there it at least partially concerns Samoyed-specific innovations. For example, the word for 'hand' can be reconstructed as PU **käti* (illustrated in 1. below), but Samoyed instead has PS **utå* (in 2. below). This cannot be regarded as a meaningful isogloss, however, since PS **utå* is indicated by its phonology to be innovative: the Samoyed vowel combination **u-å* does not occur in inherited vocabulary (see subsection 3.6.1 for more details). The aberrant word for 'hand' thus has no bearing on the phylogenetic position of Samoyed. Many other differences in the basic vocabulary of Samoyed and Finno-Ugric may well be of a similar nature. The reputation of Samoyed as an outlier is also partially the result of its later documentation and the research history (Aikio 2022a: 4; Saarikivi 2022: 48).

1. PU **käti* > (e.g.) SaaN *giehta*, Fi. *käsi*, MdE *ked'*, MariE *kid*, Hu. *kéz*
2. PS **utå* > (e.g.) Ng. *d'ütü*, EnT *uδa*, NeT *ŋuda*, NeF *ŋŭta*, SkTaz *utį*, Mt. *uda*

A further striking difference between the Samoyed and Finno-Ugric languages can be found in the numeral system. Only the words for 'two (2)' and 'seven (7)' seem to be shared between Samoyed and the other Uralic languages, with the Samoyed word for 'ten (10)' additionally corresponding to 'five (5)' elsewhere. This is illustrated in Table 1.1.

Table 1.1: The Samoyed and Finno-Ugric numeral systems. Especially among the Finno-Ugric languages, there are a number of unclear phonological correspondences (see Aikio 2022: 25 for the reconstructions). The formations of '8' and '9' vary even within Finno-Ugric.

| | Samoyed | Finno-Ugric |
|---|---|---|
| 2 | **kitä*[(a)] | ~ **kAktA* |
| 3 | **nakur* | ≠ **kolmi/kulmi/kurmi* |
| 4 | **tättə* | ≠ **neljä* |
| 5 | **sə̂mpə̂laŋkə̂* | **wij(i)t(t)i* |
| 6 | **mə̂ktut* | ≠ **kuw(V)t(t)i* |
| 7 | **säjt³wə* | ~ **ćäjćimä/ćäććimä* |
| 10 | **wüət* | ≠ **loka* |

(a) The correspondence between FU **-kt-* and PS **-t-* is regular, but the vowels are unclear, even within Finno-Ugric.

On the basis of this discrepancy, Janhunen (2000: 60–61) has argued that Proto-Uralic did not have these numerals at all, and that they were created in Samoyed and Finno-Ugric independently. Since the Finno-Ugric numerals are related, this would be a significant shared innovation. However, as pointed out by Aikio (2002: 31), the Samoyed numerals for 'three (3)', 'four (4)' and 'six (6)' can be identified as innovative within Samoyed on the basis of their phonology. The second-syllable **u* of **nakur* and **mə̂ktut* would be difficult to derive from Proto-Uralic, and the internal **-kt-* of **mə̂ktut* should have regularly developed to PS **-t-* if this word were inherited from Proto-Uralic (see also 3.5.2). Similarly, the geminate of **tättə* should have been simplified to a singleton PS **-t-*. Furthermore, 'five (5)' looks to be a compound of some sort and may be a relatively recent creation, displacing or replacing the original word for 'five (5)' as found in the other Uralic languages as that came to mean 'ten (10)'. The Samoyed numeral system is thus demonstrably innovative, while the Finno-Ugric languages may at least in part preserve a more archaic situation (Aikio 2002: 31). As in the case of the word for 'hand', this does not directly support an early split of Samoyed from Proto-Uralic, it only points to strong lexical influence from other languages on Samoyed in its position at the eastern periphery of the Uralic linguistic area (Aikio 2002: 31–32).

The case for an early split of Samoyed is better made on the basis of examples like the word for 'hare', which contains an additional suffix in the Finno-Ugric languages that is absent in Samoyed. The Proto-Samoyed form is *ńåmå* from PU **ńoma* (in 3.), while other Uralic languages reflect **ńoma-la* instead (in 4.) (Janhunen 2007: 221).

3. PU **ńoma* > PS **ńåmå* > e.g., Ng. *ńomu*, NeT *nyawa* 'hare'
4. PU **ńomala* > e.g., SaaN *njoammil*, MdE *ńumolo*, Hu. *nyúl* 'hare'

A potential phonological argument may be found in the reflexes of PS **o* opposed to FU **u* in roots of the shape **C{o/u}Ci*, depending on which is the archaic state (see the discussion in 3.6.3). A systematic investigation of all the arguments may result in the conclusion that was Samoyed indeed the first to split off, but this remains a task of future research.

The Uralic language family is spread out over a large area of Eurasia, from the Scandinavian peninsula in the west to the Yenisei River in the east. Along their entire distribution the Uralic languages border on Indo-European languages, which have influenced and continue to influence each branch of Uralic in various ways. Contact between Saami and Germanic or between Finnic and both Germanic and Baltic is well-established and can be extensively shown with both ancient and modern loanwords (e.g., Aikio 2006b; Jakob 2023). There has also been much contact between Uralic languages and Indo-Iranian languages, dating far back to some of the earliest stages of Uralic that can be reconstructed (Holopainen 2019). On the easternmost fringes of the Uralic language family, the Samoyed languages have not been impacted as heavily with Indo-Iranian influence, although some loanwords have been suggested (Janhunen 1983, see also Holopainen 2019). It is as yet unclear how the relatively low number of Iranian loanwords in Samoyed is to be understood in relation to its phylogenetic position. Loan contacts with early Tocharian have also been proposed, both into Samoyed specifically (Janhunen 1983, Napol'skikh 2001, Kallio 2004) or even into Proto-Uralic (Bjørn 2022). It is difficult to verify the Tocharian origins of many of these words, however, as will be discussed in chapter 9 of this dissertation.

The last word on the early prehistory of Uralic has not yet been written, and the search for the precise location of the Proto-Uralic homeland continues (recently Grünthal et al. 2022; Saarikivi 2022; J. Häkkinen 2023). However, the discussion has narrowed to the region ranging from the western slopes of the southern or central Ural Mountains to west or central south Siberia. Genetic evidence supports the Siberian origins of Uralic (Tambets et al. 2018, Saag et al. 2019, Childebayeva et al. 2024). Archaeological correlates for Uralic and the dispersal of the Uralic languages have recently been sought in the Seima-Turbino phenomenon of around 2200–2000 BCE (dates per Marchenko et al. 2017), which is spread out across much of the known southern range of the Uralic linguistic area (Kallio 2006: 16–17, Parpola 2012: 156–160; Grünthal et al. 2022, and Childebayeva et al. 2024 for a genetic perspective).

Where pre-Proto-Uralic should have been located before that is another open question. For the Samoyed branch especially, longstanding associations with parts of the area to the north of the Tarim Basin are assumed (Janhunen 2022). If the Proto-Uralic homeland should be located closer to the Ural Mountains, however, we must take into account the time needed for pre-Proto-Samoyed to move farther to the east. In this way,

the interpretation of the Uralic homeland interacts with the question of Tocharian-Uralic contacts and their location in time and space.

The split of Proto-Uralic used to be estimated around 4000 BCE, but recent scholarship has shifted the date closer to 2500 BCE (Kallio 2006; Grünthal et al. 2022; J. Häkkinen 2023). The shallower time depth is based in part on the general abandonment of the idea that Samoyed represents a significantly early offshoot to the family tree, and in part on a reappraisal of the loan contacts between early Uralic and early Indo-Iranian (Holopainen 2019).

The potential of loan relationships between Proto-Indo-European and Proto-Uralic has shaped many analyses of the positions of these linguistic entities in time and space (e.g., Mallory 1989: 147–149; Anthony 2007: 74–76; J. Häkkinen 2012; 2023), but a number of the arguments for such early contacts between the two protolanguages themselves should be treated with caution or might have to be abandoned altogether, in accordance with more recent insights (Simon 2020; Holopainen 2021; Aikio 2022a: 25–26). The dating and position of early Uralic is thus intimately connected with our ideas about early Indo-European, and in some ways dependent on it. As our understanding of the relationships between these two language families changes, so do our hypotheses regarding their ancient past. After all, the Uralic and Indo-European puzzles are part of the same, larger puzzle of Eurasian prehistory over the last few millennia.

## 1.3 Archaeological background

Our linguistic investigation into the past cannot be separated from the physical world. The speakers of the ancient languages reconstructed with the comparative method have left traces here and there in the archaeological record, and while an exact identification of a linguistic entity with an archaeological culture is far from straightforward, it is possible to find correlates that explain (aspects of) the relevant linguistic facts. Tocharian represents a branch of Indo-European on the eastern periphery of the language family. Since Tocharian is not known to form a sub-branch with any of the other Indo-European branches (Peyrot 2022: 83–84), its early prehistory needs to be considered separately starting immediately from its separation from Proto-Indo-European. In accordance with the Tocharian second hypothesis, an early spatial separation of Tocharian from the rest of the early Indo-European linguistic area might be expected.

According to current insights, the speakers of Proto-Indo-European were associated with the Yamnaya ("pit grave") Culture located in the Pontic-Caspian steppe. Tocharian presence in the Tarim Basin, far to the east, should thus be the result of a prehistoric migration from this area. The Yamnaya Culture dates to around 3300–2600 BCE (dates per Morgunova & Khokhlova 2013) and consists mainly of burial sites. With advances in

ancient DNA research, the associated genetic signatures have been uncovered as well (e.g., Allentoft et al. 2015, Damgaard et al. 2018). This provides additional insights into the connections between Yamnaya and related archaeological cultures. The southern Siberian Afanasievo Culture of around 3100–2500 BCE (dates per Poliakov et al. 2019) is closely related to the Yamnaya culture and shows a quite exact genetic match (Damgaard et al. 2018; Narasimhan et al. 2019). Dating to around 2800 BCE, there is evidence of an extension of Afanasievo southward into the Dzungar Basin north of the Tarim Basin (Zhang et al. 2021). The archaeological and genetic connections between Yamnaya and Afanasievo have led to an association between the Afanasievo Culture and the Tocharians (Mallory & Mair 2000; Anthony 2007; Kroonen, Barjamovic & Peyrot 2018, etc.), which we can refer to as the "Afanasievo-Tocharian" hypothesis.

If a direct descent of the Afanasievo Culture to people in the northern Tarim Basin could be shown, that would strengthen the Afanasievo-Tocharian hypothesis. Genetically, this connection is not evident. The characteristic Y-haplogroup represented by most individuals of the Afanasievo culture is R1b1a1a2. While the Shirenzigou site in the Tianshan mountains on the north-eastern side of the Tarim Basin (see Figure 1 below) also shows this Y-haplogroup, the relevant samples are dated to 200–100 BCE (Ning et al. 2019). This could indicate a continued southward movement from Afanasievo and thus represent a Tocharian migration to the northeast of the Tarim Basim, but with more than 2000 years, the difference in time is too great to be certain of a direct connection.

Furthermore, more ancient genetic samples from the Tarim Basin Xiaohe cemetery (see Figure 1 below) have not shown any particular association with Afanasievo. Analysis of some individuals from the oldest layers of Xiaohe dating to around 2000 BCE were more consistent with the ancient DNA of the Andronovo Culture, whose area extended from the west to the Altai area following both the Afanasievo Culture and its successor, the Okunevo Culture (Li et al. 2010; Mallory 2015: 44). These individuals had the Y-haplogroup R1a1a, rather than Afanasievo-associated R1b1a1a2, and also showed signs of extensive admixture with southern Siberian populations (Li et al. 2010). A more recent analysis of ancient Xiaohe genetic material instead characterized their overall genetic signature as autochthonous to the Tarim Basin, showing admixture between Ancient North Eurasians and Northeast Asians, rather than a connection with populations associated with Indo-European languages, so that a connection with early Tocharian becomes uncertain (Zhang et al. 2021).

Thus, apart from the rather late Shirenzigou evidence of 200-100 BCE (Ning et al. 2019), the trail from the Afanasievo culture in the direction of the Tarim Basin only extends to around 2800 BCE in the Dzungar Basin (Zhang et al. 2021). Linguistic evidence for early Tocharian in the Dzungar basin has been roughly estimated to around 1000 BCE, almost two millennia after Afanasievo genetics enter the area, in the form of contacts of Proto-Tocharian with Old Iranian coming from the steppe (Bernard 2023: 251–252).

Whether the Proto-Tocharian speakers in the Dzungar Basin at that time represent the initial Afanasievo migration of 2800 BCE is uncertain, however. They might also represent a later, otherwise undetected migration into the area from the west, or a migration from the northern Afanasievo area later than the one around 2800 BCE currently evidenced by the genetic record.

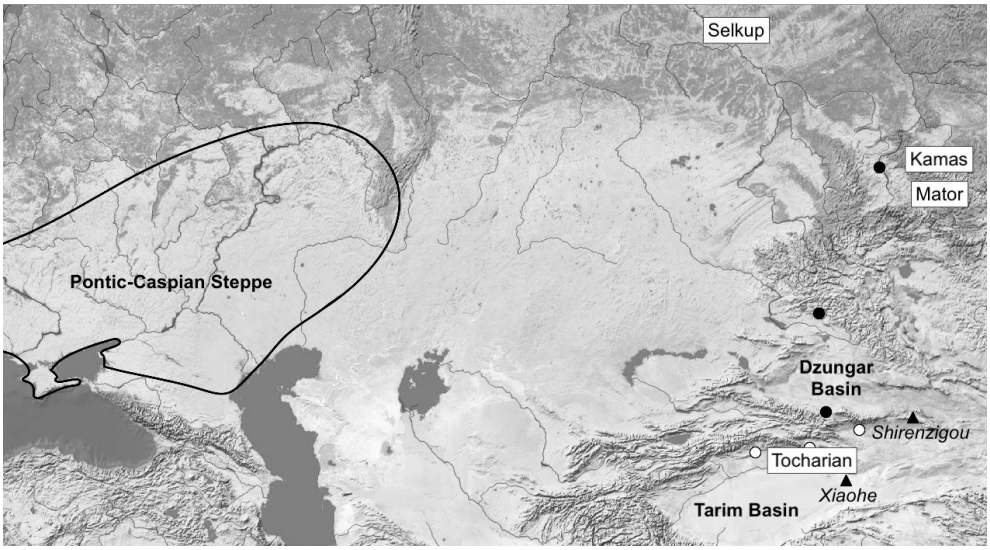


Figure 1: Geographical locations of Afanasievo sites sampled by Zhang et al (2021). marked with black dots; main find spots of Tocharian texts marked with white dots; Yamnaya Culture area approximated with a black line; location of other relevant elements marked with labels.

One great advantage that the Afanasievo-Tocharian hypothesis has is that it accounts for the phylogenetic distance between Tocharian and Indo-Iranian: with its early date starting around 3300 BCE, the establishment of the Afanasievo culture clearly precedes the Indo-Iranian migration to the steppe around 2000 BCE (Kroonen, Barjamovic & Peyrot 2018: 7–9). While this makes the Afanasievo culture the best fit for Tocharian on the basis of currently available evidence, the match is not in all aspects perfect one. Kroonen et al. (2022: 31) point out that Tocharian A *āre** 'plough (n.)' shows that Tocharian shares in the semantic innovation of the root PIE *$h_2erh_3$- from 'grind, crush, break up' to 'plough'. This development is found in all other branches where this root is attested except for Anatolian, which preserves the original, non-agricultural meaning (Kloekhorst 2008: 9–10). Interestingly, TA *āre** appears to be an independent formation within Tocharian (Peyrot 2018b: 262–263), unrelated to the instrument noun *$h_2erh_3$-*tro*- found elsewhere. Still, the specialized meaning of the root could point to the agricultural character of pre-Proto-Tocharian society. The etymology for TB *pīsäl*, TA *psäl* 'chaff', proposed by Kroonen et al. to derive from a PIE **peis*- 'grind, thresh' (2022: 16) also

suggests continued agriculture in pre-Proto-Tocharian society. Since there is no evidence for farming associated with the Afanasievo Culture, Kroonen et al. (2022: 33) argue that it may be not as good a match for early Tocharian as commonly believed. On the other hand, many other agricultural terms in Tocharian were adopted more recently from neighbouring languages (Peyrot 2018b), so that earlier pre-Proto-Tocharian agriculture was perhaps more limited or different from the farming practices of the eventual Tocharians themselves.

Since the palaeolinguistic evidence against an association with the Afanasievo Culture is so far limited to the semantic innovation of a single root, *$*h_2erh_3$-*, and potentially the etymological identification of TB *pīsäl*, TA *psäl* 'chaff', a complete dismissal of the Afanasievo-Tocharian hypothesis is probably premature. However, it is good to keep in mind that not all the evidence fits perfectly. It is theoretically possible that there were other impulses of migration from the western steppe to southern Siberia that are simply not (yet) evidenced at this moment in time, and that we therefore cannot weigh against the Afanasievo Culture. The Afanasievo-Tocharian hypothesis thus remains the most appropriate explanation for the presence of Tocharian in the Tarim Basin at the present moment, albeit in part only due to a lack of known alternatives.

This dissertation is concerned with the Afanasievo-Tocharian hypothesis in two ways. First, a Tocharian-Uralic contact hypothesis can only be understood if the ancestors of the Tocharians migrated via southern Siberia, like individuals coming from the (pre-)Yamnaya Culture to form the Afanasievo Culture. Evidence for Tocharian-Uralic linguistic contact would thus support this migration route. Second, the time depth of any linguistic contact in southern Siberia should match the time depth of the migration through southern Siberia. This latter point will remain difficult to specify in absolute terms, but nevertheless needs to be taken into account. If linguistic contacts between the ancestor of Tocharian and speakers of Uralic could be demonstrated, that would provide a further link on the migration of the Tocharians, locating them to the north of the Dzungar Basin at some point in the past, at least in partial accordance with the "Afanasievo-Tocharian" hypothesis.

## 1.4 Methodological considerations

Demonstrating a prehistoric language contact situation is a complicated matter. The traces that language contact can leave behind are varied, and may apply to any aspect of the linguistic system. Furthermore, since languages will change over time in any case, it is difficult to be certain whether particular changes were indeed due to influence from another language, or simply result from internal linguistic changes. Aside from these difficulties, in a prehistoric language contact situation, the arguments are necessarily based on a comparison between the reconstructed stages of two languages, which can

only be accessed indirectly. Due diligence must be paid to the chronologies of changes in the two languages separately to avoid anachronistic comparisons.

Broadly speaking, language contact may leave two types of traces: loanwords and structural influence. The establishment of language contact in the distant past is the clearest in the first case. Borrowed lexical material can constitute a relatively unambiguous trace of influence of a donor language, on the recipient. If the existence of loanwords can be proven to a satisfactory degree, a contact hypothesis in even the deep past may be widely accepted. For example, this is the case for the early contacts between Indo-Iranian and especially Finno-Ugric Uralic (Holopainen 2019), and for the early contacts between Tocharian and Iranian (Bernard 2023).

Without the concrete evidence from loanwords, prehistoric language contact is not easy to prove. The absence of loanwords does not, however, provide decisive proof against prehistoric contact (Thomason 2001: 91–92). Absence of evidence is not evidence of absence, and moreover, there are certain circumstances that can lead to a dearth of loanwords. For instance, it may be the case that the loanwords that were there have been replaced by newer (loan)words since they entered the language. This is especially likely when the contact would have taken place in the very distant past. In the case of an incompletely attested language like Tocharian, an additional problem could that the relevant words are simply not found in the available sources, either due to accident, or because they belonged to a register that was not written down. For example, words in the semantic domain of local plants and animals, where borrowings from a substrate language are likely (Thomason & Kaufman 1988: 39), may have fallen out of use due to changes in the environment when the Tocharians migrated into the Tarim Basin and settled there.

More significantly, not all types of linguistic contact are likely to leave traces in the form of loanwords. In the case of substrate interference as caused by language shift (Thomason & Kaufman 1988; Thomason 2001), the main effects of language shift may be found in the phonology and the grammar, rather than in the lexicon. A group of people that shift from one language to another, the target language, may take with them a foreign pronunciation and grammatical patterns in their version of the target language. A number of features native to the original language of a shifting population may end up being introduced into the speech community at large, thus altering the target language (Matras 2009: 57–58).

When loanwords are almost entirely absent, our investigation of language contact can only be carried out by looking at structural features. Thomason has put forward a set of requirements in the form of questions, according to which such a contact hypothesis can be tested (Thomason 2001: 93–95). Based on the result of this test, we may evaluate the prehistoric contact hypothesis at hand.

1) Is there a plausible historical scenario according to which the two speech communities were in contact, and was the type of contact conducive to contact induced change in one direction or the other? In this case, we do not have access to any information on the prehistorical contact in the case of the Tocharian-Uralic contact hypothesis, so that this particular question cannot be answered in detail. We can only argue that contact between the two groups was plausible (see section 1.3), but the exact nature of this plausible contact remains unknown.

2) Are there multiple features in the receiving language under investigation that can be reasonably connected to the suspected source language? If the two languages share only a single feature, the probability that this similarity came about as a result of chance is higher, and the argument that they resulted from a contact situation between precisely these languages is correspondingly not compelling. The more features can be shown to be shared, the more convincing the argument for contact-induced change will be.

3) Is the feature innovative in the target language, i.e., not inherited from its ancestor? If one or more of the features under investigation turn out to have been originally present in the target language before contact, it cannot have been the result of contact. Whether a feature is archaic or innovative in the target language can only be established by comparison with related language, preferably outside the same linguistic area.

4) Conversely, we must determine that the feature was present in the source language before contact, i.e., that it is inherited from the ancestor of the source language. If a feature is innovative in the source language, it cannot have been passed on to the target language in the distant past before it had come into existence.

In this dissertation, I will attempt to answer these questions for the hypothesis that pre-Proto-Tocharian has undergone contact induced change due to early Uralic, particularly pre-Proto-Samoyed. The choice to focus on the Samoyed branch of Uralic rests on previous research pointing to contact of early Samoyed with early Tocharian (Janhunen 1983, Kallio 2001, Kallio 2004, Peyrot 2019a). Comparison with pre-Proto-Samoyed also provides a bounded linguistic entity, defined at the most recent end by the reconstruction of Proto-Samoyed, and most distantly by Proto-Uralic. This is entirely parallel to the way that pre-Proto-Tocharian is bounded by Proto-Tocharian and Proto-Indo-European. Since the absolute time depths involved are unclear, it is possible that pre-Proto-Uralic could provide a potential alternative candidate for early contact with pre-Proto-Tocharian, but at present, a reconstruction of pre-Proto-Uralic in any detail is not yet possible. A contact hypothesis between pre-Proto-Tocharian and pre-Proto-Uralic is thus difficult to properly investigate in the same way that a contact hypothesis between pre-Proto-Tocharian and pre-Proto-Samoyed can be investigated. It is furthermore possible that early contacts between Tocharian and Uralic involved an extinct branch of Uralic, either a para-Samoyed or perhaps even a para-Uralic branch. We have no direct access to either, so that pre-Proto-Samoyed must serve as the only

available proxy at this moment. That being said, it is in some cases impossible to meaningfully distinguish between Proto-Uralic and early Samoyed and I will use the label Uralic/Samoyed to reflect this.

The question in (1) above has been satisfied as much as currently possible: contact between an early form of Tocharian and early Uralic/Samoyed can be understood in accordance with what we know about the respective language groups and associated archaeological and genetic information (see section 1.3). Since further elaboration on this point is not feasible due to the great time depth and the state of the non-linguistic evidence at this moment, the linguistic investigation may instead inform our understanding of the prehistoric contact scenario if it is judged strong enough.

The breadth of features investigated in this dissertation is in accordance with (2) and will be further elaborated in the next subsection. In order to establish whether the features are really shared between the receiving and source languages, reconstructions of the relevant pre-Proto-Samoyed and pre-Proto-Tocharian features must be undertaken independent from one another. This goes for every aspect of the comparison: phonology, morphology, and syntax. Only then will it be possible to say if the independent histories of the two language families suggest an intersection and an interaction at some point in time in accordance with points (3) and (4).

## 1.5 Research questions

The main aim of this dissertation is to refine our understanding of the Tocharian-Uralic/Samoyed contact hypothesis. I do this by investigating a number of interrelated research questions, all variations on a common theme.

1) Can the significant innovations in Tocharian consonant and/or vowel phonology be understood as the result of influence from Uralic/Samoyed?

Tocharian phonology is quite innovative from the Indo-European point of view, and substrate influence from Uralic/Samoyed has been considered as one possible driving factor in the more significant changes in both the consonant system (Kallio 2001; Bednarczuk 2015; Peyrot 2019a) and the vowel system (Peyrot 2019a; Warries 2022). In order to provide a well-informed view on the validity of these comparisons, it is necessary to understand the developments of the Tocharian and Samoyed phonological systems independently from one another. Only then can we try to establish whether the phonological systems showed significant overlap in their organization at some point in the past, which could indicate a convergence, and therefore, potential contact. To this end, chapters 2 and 3 treat the relative chronologies of sound changes of Tocharian and Samoyed independently. The final comparison is provided in chapter 4.

The historical grammar of Tocharian is full of controversial issues, and in order to arrive at the relative chronology necessary for a comparison with Samoyed, all the changes that affected the system need to receive an interpretation. Without a detailed understanding how both the vowel and consonant system changed through time, and how the two interacted, any external comparison would be problematic. Important secondary research questions here on the Tocharian side are: i) How should the development of the Tocharian vowel system be understood in relation to palatalization? ii) Which umlaut developments can be reconstructed for Proto-Tocharian, and how do they affect our understanding of the changes in the vowel system? iii) What are the regular developments of PIE *$\bar{o}$ and *$eh_2$ in Tocharian? iv) When did the reflexes of PIE *$o$ and *$\bar{e}$ merge? v) How can the development of PIE *$d$ in Tocharian be understood? These questions are addressed in chapter 2, embedded in an overview of the other relevant Tocharian sound changes.

On the Samoyed side, the same detailed view on the diachronic development of the phonology is necessary. Thus, various unsolved or controversial issues that affect our understanding of the development of the Samoyed phonological system through time need to be treated and evaluated. Particularly important questions are: i) How can the discrepancy between Samoyed *$o$ and Finno-Ugric *$u$ in stems of the type *$C\{o/u\}Ci$ be interpreted in a relative chronology of Samoyed sound changes? ii) How can the differentiation of PU *$a$-$a$ into PS *$å$-$å$ and *$a$-$ə̂$ be understood in pre-Proto-Samoyed? iii) Can the differentiation of PU *$ę$ into PS *$ę$ and *$į$ be dated relative to other changes? iv) What are the developments of roots with the shape PU *$CVwi$ and *$CVɣi$? v) Which sound changes are expected to have caused paradigmatic alternations, and to what extent do we see traces of this? Chapter 3 covers these questions along with the necessary context of the other Samoyed sound changes to provide a complete picture.

2) Does the potential for phonological influence from Uralic/Samoyed on Tocharian extend to the accent?

The place of the accent cannot be determined for Tocharian A, but Tocharian B shows a general second-syllable accent, unless the second syllable is also the final syllable. This differs from the first-syllable accent of Uralic/Samoyed. However, it is possible that the Tocharian B accentuation constitutes a later innovation that obscures an earlier system with first-syllable accent, which shifted to a second-syllable accent. Such an accent shift is assumed by Marggraf (1970: 21), Peyrot (2013: 507–515) and Jasanoff (2015), although the details of their analyses differ on very important points. No earlier first-syllable accent is assumed in other theories on the prehistory of the Tocharian accent (Ringe 1987; Winter 1993, 1994b; Malzahn 2010 *passim*), and the topic as a whole has received a number of different interpretations, which sometimes overlap only partially.

The place of the accent in prehistoric Tocharian must include an explanation for not only the eventual accentuation in Tocharian B, but also rise of accented epenthetic vowels and the loss of earlier second-syllable vowels inherited from Proto-Indo-European. In order to better understand the prehistory of the Tocharian accent, I will attempt to provide an answer to the following connected secondary research questions: i) In which phonological environments did epenthesis occur regularly? ii) How can the accented epenthetic vowels of Tocharian B be understood? iii) How did syncope interact with the accent? iv) How did the accent interact with vowel contractions? v) How can the irregularly accented verbal categories of Tocharian B be analysed in a framework with phonologically regular accentual developments? vi) How is the accentuation of the secondary cases in Tocharian B to be interpreted? In the background is the question of how the accent of Proto-Tocharian should be reconstructed, and which effects in Tocharian A can provide any indications. Can the usual assumption that the Proto-Tocharian accent was like the Tocharian B accent with innovations on the Tocharian A side only be substantiated? These questions are addressed in chapter 5.

3) Can the Tocharian case system be understood as the result of influence from Uralic/Samoyed?

In particular, the Tocharian local cases display an innovative agglutinative character, and include in their ranks new cases like the perlative. The connection of the agglutinative Tocharian case system with some type of substrate influence has a long history (Krause 1951; K.H. Schmidt 1990; Thomas 1994, Bednarczuk 2015, Peyrot 2019a), but there is disagreement as to which substrate language should be considered as the source (e.g., Akao 2020 argues for Turkic influence). The case systems of Tocharian and Uralic/Samoyed are discussed in chapter 6.

4) Can the Tocharian participial system be understood as the result of influence from Uralic/Samoyed?

This aspect of the Tocharian-Uralic/Samoyed contact hypothesis is new. Tocharian has participles for the present and the past, and also gerundives and a privative. A number of these participles are neither active nor passive, but rather display contextual orientation, which is atypical from an Indo-European perspective. Similar types of participles are found in Samoyed, and in Uralic languages more generally. In chapter 7, I compare the participial system of Tocharian to the participial systems that are found in Uralic/Samoyed.

5) Can the innovative Tocharian use of pronoun suffixes attached to the finite verb be understood in comparison with the Uralic/Samoyed objective conjugation?

Similarities between Tocharian and Uralic in this respect have been pointed out before (Peyrot 2019a; Georg 2023), but it is also clear that there are many differences between the two language groups. In chapter 8, I will investigate whether these can be reconciled. Importantly, the age of the objective conjugation in Uralic is a controversial question (see 8.3 for more details), leading to a natural secondary research question on this topic: can the objective conjugation be reconstructed for Proto-Uralic, and in what form? This question needs to be addressed before a comparison with Tocharian can be attempted.

6) Finally, in chapter 9, I treat potential loanwords from early Tocharian into early Samoyed and/or Uralic to see if there is any confirmation of the contact hypothesis in the form of lexical borrowings.

Loanwords can be used to support the information gleaned from our discussions of Tocharian grammatical innovations have a great potential to bolster the Tocharian-Uralic/Samoyed contact hypothesis. Unfortunately the evidence remains rather meagre (cf. Janhunen 1983; Napol'skikh 2001; Kallio 2004).

I evaluate the various features on both a case-by-case basis and with consideration for the aggregate. If substrate influence from early Uralic/Samoyed indeed applied to early Tocharian, that carries with it implications for our understanding of the prehistory of the languages on both sides of the comparison. The aim of this dissertation is to better inform our understanding of the past of the Tocharians and their placement in time and space. In order to carry out this study, a detailed consideration of both language families is necessary, spanning the different topics, from phonology to verbal morphology and participial syntax. The underlying details on both sides will be presented at each stage to make the comparison as transparent as possible. My hope is that this will not only contribute to our understanding of the Tocharian-Uralic/Samoyed contact hypothesis, but that the individual discussions that resulted from this investigation can contribute to our understanding of the historical grammar of both Tocharian and Samoyed, and highlight uncertain aspects where further investigation is needed.

# 2 Relative chronology of Tocharian sound changes

To be able to properly compare the phonological systems of pre-Proto-Tocharian and pre-Proto-Samoyed, we need to have an intricate understanding of their developments through time. Otherwise, the comparison remains impressionistic. In this chapter, I will present the most important sound changes that occurred from Proto-Indo-European to Proto-Tocharian. There are still many points of disagreement among specialists, so where space allows, I will treat the differing interpretations of the data and argue which I find to be best supported. In a few cases, I will offer an alternative solution.

In chapter 3, I will treat the developments of Samoyed in the same way. For both languages, the aim is to obtain a relative chronology of phonological developments that is as secure as possible with the data from presently available etymologies and to reconstruct the developments of the phonological system through time. Only then can a comparison be attempted between pre-Proto-Tocharian and pre-Proto-Samoyed, which will be the topic of chapter 4. In the present chapter on Tocharian phonological changes, the possible indications of contact between Tocharian and Samoyed will not be considered, as the internal relative chronologies should each stand on their own.

## 2.1 Introduction

A large number of the developments from Proto-Indo-European to Tocharian are well-established and have been studied in considerable detail. Ringe (1996) gives a useful overview of most sound changes and places them in a relative chronology. However, not all the information in Ringe's book is still representative of the more recent research in the field, and the relative chronology there cannot be used as a starting point for our purposes without re-evaluation. Pinault (2008: 413–460) also discusses various aspects of Tocharian historical phonology, and many other sources contain treatments of the general picture (e.g., Malzahn 2010: 1–22), provide a dedicated overview (Hackstein 2017), or treat the developments of specific sounds or sequences (e.g., Hilmarsson 1993; Kim 1999; Pronk 2009), so that there are various sources of information to draw on. The purpose of this chapter on Tocharian sound changes is to synthesise an improved relative chronology of the Tocharian phonological system. In order to do this within the confines of this dissertation, those changes that are thought to have impacted the overall phonological system will be given a preferential treatment, while a few minor developments that are less informative and would unnecessarily overcomplicate matters are left aside.

The general Tocharian phonological system and the ways in which it differed from that of Proto-Indo-European will be considered in subsection 2.2. Then, various

phonological changes will be covered in more detail by topic, starting with general changes to the Proto-Indo-European consonant system in subsection 2.3, followed by a section dedicated to palatalization and its relationship to developments in both the consonants and the vowels in subsection 2.4. Subsequently, some late changes that took place in the vowel system are treated in subsection 2.5. A number of sound laws will have to be mentioned in the discussion of others, and therefore may need to be brought up earlier than they themselves appear with a dedicated discussion; cross-references will be provided where appropriate. At the end of each subsection, there is a brief summary of the changes discussed. The numbers that accompany the cited etyma in this chapter correspond to etymological reference list 1, which contains more extensive etymological information and references. It can be found at the end of this dissertation.

## 2.2 Tocharian and Proto-Indo-European phonology

The phoneme inventories of Tocharian A and B are very similar, especially as regards the consonants. The differences between Tocharian and Proto-Indo-European, on the other hand, are significant. The reconstructed Proto-Indo-European consonant inventory is presented here in Table 2.1.

Table 2.1: The Proto-Indo-European consonant system.

| PIE | labial | dental | palatal | palatovelar | velar | labiovelar | "laryngeal" |
|---|---|---|---|---|---|---|---|
| stop | *p (b) b*$^{h}$ | *t d d*$^{h}$ | | *ḱ ǵ ǵ*$^{h}$ | *k g g*$^{h}$ | *k*$^{w}$ *g*$^{w}$ *g*$^{wh}$ | |
| | | | | | | | *h*$_1$ *h*$_2$ *h*$_3$ |
| sibilant | | *s* | | | | | |
| nasal | *m* | *n* | | | | | |
| liquid | | *l r* | | | | | |
| glide | | | *i̯* | | | *u̯* | |

Proto-Indo-European had a large number of obstruent phonemes, with a three-way opposition in the stops between, e.g., **t* vs. **d* vs. **d*$^{h}$. In the labial series, the evidence for PIE **b* is extremely limited, so that its proper existence as a member of the PIE consonant system is doubtful (see, e.g., Olander 2020, with literature).

A single sibilant **s* can be reconstructed, as well as four resonants **m*, **n*, **l*, **r*, and the glides **u̯* and **i̯*. There are also the three laryngeals **h*$_1$ *h*$_2$ *h*$_3$, whose original phonetic realization is a matter of debate (e.g., Kümmel 2012; Kloekhorst 2018). They often behave alike in their development in the historical phonology of the various Indo-European languages, including Tocharian. Taken as a class of consonants, the laryngeals can collectively be written with a capital **H* in those cases where the comparative material

does not allow us to determine which of the three was present, or where differentiation is not necessary for the issue at hand.

The traditional notation of stops as voiceless, voiced and voiced aspirate is adhered to throughout this dissertation in reconstructed forms, without taking a definite stance in the longstanding discussions about the phonetic interpretation of these series. A main point of contention is the interpretation of the traditionally simple voiced series with **d*, **ǵ*, **g*, **gʷ* as glottalized **ʔd*, **ʔǵ*, **ʔg*, **ʔgʷ* instead of plain voiced (Gamkrelidze & Ivanov 1972, Hopper 1973). This interpretation would make the almost complete absence of PIE **b* in the consonant system more understandable: the absence of a glottalized bilabial in a glottalized series is cross-linguistically more common than the absence of a voiced bilabial in a voiced series (ibid., and, e.g., Kortlandt 1985). The idea that the traditional series of voiced stops was originally glottalized is known as the Glottalic Theory.[1]

Additionally, combining the Glottalic Theory with evidence from Anatolian, Kloekhorst has proposed an interpretation of the original stop system as consisting only of voiceless stops, with geminates **pː*, **tː*, **ḱː*, **kː*, **kːʷ* instead of the traditional series **p*, **t*, **ḱ*, **k*, **kʷ* and glottalized voiceless stops **ʔt*, **ʔḱ*, **ʔk*, **ʔkʷ* instead of the traditional **d*, **ǵ*, **g*, **gʷ* (Kloekhorst 2016). Due to the uncertain position of Tocharian within Indo-European, the potential relevance of this alternative reconstruction for Tocharian is still unclear, as will be discussed in 2.3.4.4. The three competing interpretations are summarized in Table 2.2 for reference.

Table 2.2: Three alternative reconstructions of the Proto-Indo-European (or Proto-Indo-Anatolian) stop system.

| | labial | dental | palatovelar | velar | labiovelar |
|---|---|---|---|---|---|
| Traditional | *p (b) bʰ* | *t d dʰ* | *ḱ ǵ ǵʰ* | *k g gʰ* | *kʷ gʷ gʷʰ* |
| Glottalic | *p (ʔb) b* | *t ʔd d* | *ḱ ʔǵ ǵ* | *k ʔg g* | *kʷ ʔgʷ gʷ* |
| Kloekhorst | *pː (ʔp) p* | *tː ʔt t* | *ḱː ʔḱ ḱ* | *kː ʔk k* | *kːʷ ʔkʷ kʷ* |

The consonant system of Tocharian A and B is in any case rather different from that of Proto-Indo-European, as can been seen below in Table 2.3. Additional reconstructed Proto-Tocharian phonemes are marked with an asterisk. The velar nasal *ṅ* is dependent on a following velar stop.

[1] If the traditionally plain voiced series was characterized by a glottalic pronunciation, the aspiration (or breathy voice) traditionally assumed for the series **bʰ*, **dʰ*, **ǵʰ*, **gʰ*, **gʷʰ* would become redundant (see, e.g., Kümmel 2012 for an overview with further references).

Table 2.3: The Proto-Tocharian, Tocharian A and Tocharian B consonant inventories. Consonants exclusive to Proto-Tocharian marked with an asterisk.

| TAB~PT | labial | dental | retroflex | palatal | velar | labiovelar |
|---|---|---|---|---|---|---|
| stop | *p* **p*$^{y}$ | *t* | | | *k* | *k*$^{w}$ |
| affricate | | *ts* | | *c* | | |
| sibilant | | *s* | *ṣ* | *ś* | | |
| nasal | *m* **m*$^{y}$ | *n* | | *ñ* | (*ṅ* =[ŋ]) | |
| liquid | | *l r* | | *ly* | | |
| glide | *w* **w*$^{y}$ | | | *y* | | |

Tocharian phonology does not make distinctions based on such features as voicing (**t* vs. **d*) or aspiration (**d* vs. **d*$^{h}$ or **t* vs. **t*$^{h}$) at all, and no traces of potential glottalization (**d* vs. *ˀ*d* or **t* vs. *ˀ*t*) are found either.[2] Instead, Tocharian extensively employs palatalization in a great number of oppositions, so that almost every consonant has a palatal or palatalized counterpart. The only exceptions are *r*, *y* and the labials *p*, *m* and *w*; the palatalized counterpart to both *k* and *k*$^{w}$ is the *ś*. The labials do need to be reconstructed with palatalized counterparts **p*$^{y}$, **m*$^{y}$ and **w*$^{y}$ for Proto-Tocharian on account of some special developments. For example, the correspondence between TA *w* and TB *y* in certain words like TA *want*, TB *yente* 'wind' are understood as distinct developments of a palatalized labial glide **w*$^{y}$ to TA *w* and TB *y*; thus, PT **w*$^{y}$*ente* 'wind' (e.g., Hackstein 2017: 1326).

The vowel systems of Proto-Indo-European and Tocharian are also different, as shown in Table 2.4. The great changes that took place cannot be easily seen from looking at a general overview of the systems, and most vowels will need to be discussed individually.

Table 2.4: Proto-Indo-European and Proto-Tocharian vowel inventories.

| | front | back | | front | central | back |
|---|---|---|---|---|---|---|
| high | *i* | *u* | PIE > PT | *i* | *ə* | *u* |
| mid | *e ē* | *o ō* | | *e* | | *o* |
| low | | | | | *a* | *å* |

General innovative features of the Tocharian vowel system include the reduced central vowel /ə/ and a lack of length distinctions (i.e., no **ē* or **ō*). The length mark in Tocharian orthography marks either a qualitative opposition (Tocharian A), or the placement of

[2] Geminate stops in Tocharian do not correspond to PIE **t*, and are thus independent from Kloekhorst's reconstruction of this phoneme as an original geminate **t*:.

the accent (Tocharian B). At some point in Tocharian prehistory, front vowels palatalized preceding consonants, creating the pervasive palatalization opposition in the Tocharian consonant system. In the process, some front and back vowels, such as PIE *$\bar{e}$ and *$o$, merged phonologically (see 2.4).

In Proto-Indo-European, the high vowels *$i$ and *$u$ were the vocalic allophones of the consonants *$i̯$ and *$u̯$. The diacritic "$\underset{\wedge}{\ }$" indicates non-syllabicity of phonemes that may otherwise occur as syllabic nuclei. The opposite is marked with the diacritic "$\underset{\circ}{\ }$", as in the syllabic resonants *$m̥$, *$n̥$, *$l̥$ and *$r̥$, and the syllabic laryngeals *$h₁̥$, *$h₂̥$, *$h₃̥$, which could all function as syllabic nuclei as well as more canonical consonants. While syllabification was non-phonemic and automatic in Proto-Indo-European itself, making these diacritics strictly speaking anachronistic (e.g., Beekes 2011: 120), they are a useful tool for elucidating the developments in the later languages and will be extensively employed for this purpose throughout the present chapter.

This brief overview illustrates that Tocharian phonology underwent considerable modifications from its Proto-Indo-European ancestor. A number of developments are also unusual within Indo-European, such as the complete loss of distinctive manners of articulation in the stop system, which is not paralleled in any other branch of Indo-European. Some of the changes in the vowel system are also striking, like the shift from PIE *$o$ to PT *$e$ and the complete loss of distinctive vowel length. In the rest of this chapter, the general picture sketched here will become more detailed, and a number of important remaining problems will be discussed.

## 2.3 Early consonant developments

The developments of the Tocharian consonants are unique within Indo-European, and they provide an important point of comparison with Uralic/Samoyed. It is therefore important to gain a proper understanding of the various changes and the many difficulties involved. Overviews of the developments of laryngeals (subsection 2.3.1) and syllabic resonants (subsection 2.3.2) will be given first in this account of early pre-Proto-Tocharian phonology. Both led to the creation of new vowel phonemes and belong to some of the first changes in the pre-Proto-Tocharian phonological system. The developments of palatovelars and labiovelars constitute an interesting aspect of Tocharian in relation to its geographical position in the east, where satemization and its associated changes are more commonly observed in other Indo-European branches (see subsection 2.3.3). Possible interactions between labiovelars and various vowel phonemes could furthermore provide insight into the relative chronology of Tocharian sound changes. The discussion of the developments of the stop system as a whole (subsection 2.3.4) addresses some of the most characteristic and simultaneously most puzzling aspects of Tocharian historical phonology, chief among them the development of PIE *$d$.

### 2.3.1 Laryngeals

The three laryngeals *$h_1$, *$h_2$ and *$h_3$ constitute a famous part of the Proto-Indo-European consonant inventory. Some Anatolian languages, such as Hittite, preserve them in certain phonological positions, but in most other branches, the Proto-Indo-European laryngeals were eventually lost as distinct consonants. In Tocharian, the general tendency is for any laryngeal *$H$ that stood between two consonants to become PT *$a$. This can be schematically represented as a development from *$CH̥C$ to *$CaC$ (Ringe 1996: 22–25, 32–37, Hackstein 2017: 1316–1318); cf. the development of PIE *$ph_2tēr$ to PT *$pat^{y}er$ 'father' (TB *pācer* /pácer/, TA *pācar*). In a similar fashion, *$Cih_{2/3}C$ became *$CyaC$. With *$Cih_1C$ the result is generally assumed to have been *$CīC$ instead (Ringe 1996: 25; Hackstein 2017: 1314), but there are not many examples to confirm this development. In word-final position, at least *$Cih_2$ also became *$Cya$, but the developments of *$Cih_1$ and *$Cih_3$ in this position are uncertain.

When a laryngeal occurred next to a PIE vowel *$e$ or *$o$, it is normally assumed that a number of fairly standard changes occurred as also seen in other branches of Indo-European (cf., e.g., Ringe 1996: 7–9; Beekes 2011: 147), at least as an initial step. Further changes to these vowels are given in more detail below in other (sub)sections. PIE *$e$ could be coloured by *$h_2$ to become *$a$, either simple *$a$ from the sequence *$h_2e$, or long *$ā$ from the reverse sequence *$eh_2$. On the other hand, *$h_1$ only caused lengthening in *$eh_1$, but had no colouring effects, thus yielding *$ē$. The vowel PIE *$o$ was not coloured by any laryngeal.

The developments of *$h_3e$ and *$eh_3$ are uncertain. There are only very few examples with these sequences reflected in Tocharian, and they are contradictory and depend on different interpretations. Based on the effects of *$h_1$ and *$h_2$ as well as parallel developments in other Indo-European languages, it is normally assumed that *$h_3e$ became *$o$ and that *$eh_3$ became *$ō$ (e.g., Ringe 1996: 7–9, Hackstein 2017: 1313–1314); cf. subsections 2.4.5.1 and 2.4.5.2. Since instances of apparent *$h_3e$ to *$o$ could in principle reflect *$h_3o$ with an *o*-grade, the involvement of the laryngeal in colouring the vowel is uncertain, and alternative interpretations are possible. Thus, while the developments of these sequences as presented here are often assumed, some of them remain uncertain.[3]

- PIE *$Ch_1C$ > *$a$ (e.g., 78. PIE *$-mh_1no-$ > *$-mano-$, the mid.prs.ptc. suffix)

[3] Kim has proposed some different developments of sequences with a final laryngeal specifically at the end of a word: there, the laryngeal might be lost under certain conditions. The data is difficult to unambiguously interpret, and both word-final position and the number of syllables may play a role (Kim 2018: 98–112). Kim's conclusions, if accurate, would affect the precise development of particular words and suffixes, but they have no bearing on the further general vowel developments of Tocharian, and will be left out of considerations here.

- PIE *$Ch_2C$ > *$a$ (e.g., 91. PIE *$p\underset{\circ}{h_2}tēr$ > *$patēr$ 'father')
- PIE *$Ch_3C$ > *$a$ (e.g., 23. PIE *$ǵnh_3$- > *$ǵna$- 'know')

- ? PIE *$Cih_1$ > *$ī$ (e.g., 13. PIE *$d\underset{\circ}{u}idḱ\underset{\circ}{n}t$ → *$d\underset{\circ}{u}ih_1ḱ\underset{\circ}{n}t$ > *$wīkəmt$ 'twenty')[4]
- PIE *$Cih_2$ > *$ya$ (e.g., 18. PIE *$d^hi\underset{\circ}{h_2}g^w$- > *$d^hyag^w$- 'pierce, bite')
- PIE *$Cih_3$ > *$ya$ (e.g., 28. PIE *$g^wi\underset{\circ}{h_3}u$- > *$g^wyaw$- 'live')
- PIE *$Cuh_2$ > *$wa$ (e.g., 106. PIE *$s\underset{\circ}{u}\underset{\circ}{h_2}dro$- > *$swadro$- 'sweet')[5]

- PIE *#$h_1e$ > *$e$ (e.g., 35. PIE *$h_1e\underset{\circ}{ḱ}uo$- > *$ekwo$- 'horse')
- PIE *#$h_2e$ > *$a$ (e.g., 45. PIE *$h_2eǵ$- > *$ag$- 'drive')
- ? PIE *#$h_3e$ > *$a$ (e.g., 54. PIE ?*$h_3em(e)so$- > *$amso$- 'shoulder')
- ? PIE *#$h_3e$ > *$o$ (e.g., 56. PIE ?*$h_3eros$- > *$oros$- 'appearance')
- PIE *#$Ho$ > *$o$ (e.g., 58. PIE *$h_3ok^ws$- > *$ok^ws$- 'eye')
- PIE *#$h_1i$ > *$i$ (e.g., 37. PIE *$h_1itr$- → ?*$h_1itōr$ > *$itōr$ 'path')

- PIE *$eh_1$ > *$ē$ (e.g., 77. PIE *$meh_1n$-$s$- > *$mēn$-$ē(n)$- 'moon')
- PIE *$eh_2$ > *$ā$ (e.g., 2. PIE *$b^hreh_2tēr$ > *$b^hrātēr$ 'brother')
- ? PIE *$eh_3$ > ??? (no reliable examples, see subsection 2.4.5.1)
- PIE *$oh_1$ > *$ō$ (e.g., 15. PIE *$duoh_1$ > $dwō$ 'two (2)')
- ? PIE *$oh_2$ > ??? (no reliable examples, see subsection 2.4.5.1)
- ? PIE *$oh_3$ > ??? (no known examples)

The development and effects of especially *$h_3$ are difficult to establish due to a general lack of reliable examples. However, we can tentatively summarize the following trends in the developments of the three laryngeals:

- *$Ch_1C$ = *$Ch_2C$ = *$Ch_3C$, namely > *$CaC$
- *$Cih_1C$ ≠ *$Cih_2C$ = *$Cih_3C$, namely > $CīC$ and $CyaC$ respectively
- *$h_1e$ ≠ *$h_2e$, namely > *$e$- and *$a$- respectively; *$h_3e$ is uncertain
- *$eh_1$ ≠ *$eh_2$, namely > *$ē$ and *$ā$ respectively; *$eh_3$ is uncertain

[4] This change has very few examples, but this seems to be the best way to get the TAB *i* in this word (Ringe 1996: 25; Hackstein 2017: 1314). A possible parallel is the optative suffix PT *-*i*- from PIE *-$ih_1$- potentially via *$ī$ (Ringe 1996: 25; Hackstein 2017: 1314), and perhaps TA *wir* 'young, juvenile' if correctly etymologized and reconstructed as from PIE *$uih_1ro$- (Hackstein 2017: 1314). The problem with the word for 'twenty' itself is that it did not have a real *$h_1$ originally. On the other hand, there is no parallel for a development from PIE *$id$ to pre-PT *$ī$ either.

[5] I do not know any reliable examples with PIE *$uh_1$ or *$uh_3$. We might expect the former to become *$ū$ if the development from *$ih_1$ to *$ī$ is correct, and the latter *$wa$ on the basis of comparison with *$ih_{2/3}$ and *$uh_2$, but this is by no means assured.

We see one point of general agreement between all three: between two consonants, the laryngeals behave the same way and become **a*. In other environments there is mostly disagreement. After **i*, only **$h_2$* and **$h_3$* certainly became **a*. **$h_1$* rather seems to have caused lengthening, although this is based on few, partially uncertain examples. For the sequences of the type **HeH* and **eHH*, only clear information on **$h_1$* and **$h_2$* is reliable, and they yield different results: **$h_1$* gives an *e*-vowel, and **$h_2$* causes a change to a long *a*-vowel.

Laryngeals in other positions were often simply lost, for instance in most cases in word-initial position (in **$h_1$rud$^h$ro-*, **$h_2$ster-*, **$h_2$ueh$_1$nto-* below), in several consonant clusters with resonants (in **plh$_1$no-*, **prHwo-*, **solh$_2$uo-* below), and between vowels (in **tuHom* below). In intervocalic position the resulting hiatus could be filled up with a glide. Only PIE **$h_2$ntb$^h$oh$_1$* 'both' shows vocalization of a word-initial laryngeal as **a* (Ringe 1966: 15).

44. PIE **$h_1$rud$^h$ro-* > early pre-PT **rud$^h$ro-* 'red'
52. PIE **$h_2$ster-* > early pre-PT **ster-* 'star'
53. PIE **$h_2$ueh$_1$nto-* > early pre-PT **wēnto-* 'wind'
92. PIE **pl̥h$_1$no-* > early pre-PT **pəlno-* 'full'
93. PIE **pr̥Hu̯o-* > early pre-PT **pərwo-* 'first'
102. PIE **solh$_2$u̯o-* > early pre-PT **solwo-* 'whole'
113. PIE **tuHom* > early pre-PT **tuwom* 'you'

50. PIE **h$_2$ntb$^h$oh$_1$* > **antb$^h$ō* 'both' (or perhaps **h$_2$entb$^h$oh$_1$* > *antb$^h$ō*)

While the laryngeals were eventually lost in Tocharian, they had not yet disappeared in all positions when the syllabic resonants vocalized. This can be seen in the development of PIE **u̯l̥h$_2$ōnts*, the ancestor of PT **wəlo* 'king', which became post-PIE **wəlōnts* rather than ***wlōnts* (see Lubotsky 1994 on the reconstruction of this word). The latter pre-form would have resulted from ***u̯lōnts* with loss of the laryngeal before vocalization of **l̥* as **əl*, and this should have yielded PT ***lo* rather than **wəlo*; cf. the development of **u̯l̥h$_2$ntos* > **wlantos* > PT **lante*, the genitive of the same word. The syllabic resonants are the topic of the next subsection.

The development of PIE **eh$_1$* and **oh$_1$* seems to be identical to that of the corresponding long vowels **ē* and **ō*, so that an early merger of **eh$_1$* with **ē* and of **oh$_1$* with **ō* can be assumed, before any of the subsequent vowel changes. The elimination of the laryngeals as separate segments in other environments is difficult to date. The PIE sequence **eh$_2$* is the only source of pre-PT **ā*, so that no comparison with primary "**ā*" is possible. The situation with pre-PT **a* from PIE **h$_2$e* and from vocalized laryngeals is similar. However, after colouring (**eh$_2$* to **ah$_2$*, **h$_2$e* to **h$_2$a*, etc.), the three laryngeals at least cannot be phonologically differentiated anymore, so that an interpretation with a

single laryngeal phoneme **H* becomes possible. The laryngeal disappeared in cases where it lengthened the (coloured or uncoloured) preceding vowel before any other vowel changes had taken place. In principle, a phoneme **H* could have persisted in other environments for an indeterminate time.

### 2.3.2 Syllabic resonants

In Proto-Indo-European, the resonants **m*, **n*, **l* and **r* (collectively, **R*) could act both as consonants and as vowels. That is to say, they could function as the onset or coda of a syllable, but also as the nucleus. When they function as the nucleus of a syllable, they are known as syllabic resonants. Syllabic resonants, collectively written as **R̥*, stopped being syllabic at some point in pre-Proto-Tocharian. The general pattern is that a short vowel **ə* was inserted before the resonant (Ringe 1996: 67–69; Hackstein 2017: 1319–1320).[6] This means that syllabic resonants constitute one of the sources for the important innovative Proto-Tocharian vowel **ə*. The phonetic interpretation of this vowel as it developed initially will be further discussed in 2.4.4. The list below contains some representative examples of this development.

3. PIE **bʰr̥ǵʰros* > early pre-PT **bərgros* 'tall'
4. PIE **dekm̥* > early pre-PT **dekəm* 'ten (10)'
9. PIE **dn̥ǵʰueh₂* > early pre-PT **dəngʰwā*/*gʰəndwā* [metathesis] 'tongue'
13. PIE **duih₁km̥t* > early pre-PT **wīkəmt* 'twenty (20)'
60. PIE **ḱm̥tom* > early pre-PT **kəmtom* 'hundred (100)'
92. PIE **pl̥h₁nos* > early pre-PT **pəlnos* 'full'
93. PIE **pr̥Huos* > early pre-PT **pərwos* 'first'
98. PIE **septm̥* > early pre-PT **septəm* 'seven (7)'
118. PIE **u̯l̥h₂ōnts* > early pre-PT **wəlōnts* 'king'

In word-initial position, the development was different: instead of **ə*, the basic prop vowel was (pre-)PT **e* (Ringe 1996: 99–100; Hackstein 2017: 1319–1320). This (pre-)PT **e* was different from PIE **e*, which became PT **yə-*. Instead, it merged with PIE **o*, which also became PT **e*. It is difficult to determine whether this means that word-initial

[6] It has been suggested that the syllabic resonant **R̥* at first vocalized as **uR*, which later became PT **əR* with the general change from **u* to **ə* (e.g., Burlak 2000: 124). However, examples like PIE **ḱm̥tom* to PT **kənte* 'hundred (100)' show that this is impossible at least for **m̥*, since an intermediate ***kumtom* should have yielded PT ***kʷənte* with an initial labiovelar; this is not the case (cf. subsection 2.4.3). Since the difference between **u* and **ə* can only be seen in this type of context, next to a velar, the other examples do not provide any indication either way. See also Ringe (1991) for a discussion of this issue.

syllabic resonants were first vocalized as **o*, and alternative interpretations are possible. The examples below show these developments.

- PIE **n̥C* > PT **e(n)-* 'negative prefix' (also **o(n)-* and **a(n)-* by umlaut)
- PIE **h₁n̥C* > PT **e(n)-* 'in; adverbializing prefix' (also **o(n)-* and **a(n)-* due to umlaut); next to full grade **h₁enC* > PT **yə(n)-* 'id.'

41. PIE **h₁n̥ḱ-* > PT **enk-* > TB *eṅk-*, TA *eṃts-* 'seize, grip'
43. PIE **h₁r̥gʷ-* > PT der. **erkent* > TB *erkent*, TA *arkant* 'black, dark'
79. PIE **n̥ḱu̯o-* 'mortal' > PT **enkʷe* > TB *eṅkwe*, TA *oṅk* 'man'

The negative prefix **n̥-* and the local prefix **h₁n̥-* 'in' are the most common sources for word-initial syllabic resonants. The developments of PIE **n̥ḱu̯o-* and **h₁n̥ḱ-* suggest that the presence or absence of an initial **h₁* had no influence on the vocalization of the nasal. This contrasts with the development of **h̥₂ntbʰoh₁* 'both' with a vocalized **h₂* in the previous subsection, unless this reflects a less likely full-grade form **h₂entbʰoh₁* instead. The only instances of a non-nasal "initial" syllabic resonant are found in reflexes of the root **h₁r̥gʷ-* 'black, dark', which also vocalized with a vowel **e* as **er-*, similar to the **(h₁)n̥* in **n̥ḱu̯o-* 'mortal → man' and **h₁n̥ḱ-* 'seize, grip'.

Ringe points out that it is difficult to be sure when the prop vowel **ə* became a proper phonemic vowel in its own right, since a sequence *[əR] could originally simply have been the automatic phonetic representation of */R̥/. However, when some final nasals were lost, **ə* from word-final **n̥* and **m̥* would have found itself alone and fully phonemic (Ringe 1996: 67). It is impossible to tell whether initial syllabic resonants immediately yielded a separate vowel **ë* (non-front, non-rounded), but this may have remained non-phonemic until **o* became **ë* as well, as it would have been in allophonic distribution with **ə* until then.

However, the phonemic vocalization of the syllabic resonants, and thus the appearance of **ə* in the vowel system, can also be connected with another development. In PT **wəlo* 'king' from PIE **u̯l̥h₂ōnts*, we see that a syllabic resonant could occur in front of regular vowels where a laryngeal was lost; this is not the usual situation. Therefore, instead of positing a stage like hypothetical **wl̥ōnts*, with a vocalic resonant directly followed by a vowel, it is preferable to think of this form as **wəlōnts* with the new vowel **ə* in the first syllable as soon as the laryngeal disappeared in that position. It remains difficult to accurately determine when exactly consonantal laryngeals disappeared, however, even if all of their most salient interactions with the vowels took place at an early stage (see 2.3.1).

### 2.3.3 Palatovelars and labiovelars

There are three series of velars commonly reconstructed for Proto-Indo-European: palatovelars **ḱ*, **ǵ*, **ǵʰ*, plain velars **k*, **g*, **gʰ* and labiovelars **kʷ*, **gʷ*, **gʷʰ*. The palatovelars developed into palatal affricates in especially a number of eastern branches (e.g., in Indo-Iranian and Balto-Slavic), whereas they merged with the plain velars in other, mostly western branches (e.g., in Italic and Germanic). Languages belonging to the latter group are known conventionally as *centum*-languages, based on Latin *centum* /kentum/ from PIE **(d)ḱmtom* 'hundred'. The first group contains the so-called *satem*-languages, based on Avestan *satəm* from the same PIE **(d)ḱmtom* 'hundred'.

In Tocharian the development of the palatovelars is straightforward, but the labiovelars are disputed. They are preserved in some cases and lost in others, without an uncontroversial set of conditioning factors. Proposed interpretations involve the surrounding vowels, and thus potentially influence our understanding of the relative chronology in both the stop system and the vowel system. In this subsection, the early Tocharian developments of the palatovelars will be summarized and the complicated data of the labiovelars will receive a more detailed treatment.

#### 2.3.3.1 Developments of PIE palatovelars

The Tocharian words for 'hundred' are TA *känt* and TB *kante*, going back to PT **kənte* from PIE **(d)ḱmtom* (Krause & Thomas 1960: 64-65). This is just one of the examples that show that Tocharian depalatalized its palatovelars, and is to be counted among the *centum*-languages, unlike its geographically closest Indo-European neighbours of the Indo-Iranian branch. The status of Tocharian as a *centum*-language is further apparent from words like PIE **ḱu̯ō* > PT **ku* > TAB *ku* 'dog'; cf. Skt. *śvā́* 'id.', PIE **h₁eḱu̯os* 'horse' > PT **yəkʷe* > TB *yakwe*; TA *yuk* 'horse'; cf. Skt. *áśva-* 'id.', PIE **bʰr̥ǵʰro-* > PT **pərkre* 'tall, high'; cf. Skt. *bṛhánt-* (Ringe 1996: 39–41 cites many more examples).

The depalatalization of the palatovelars and their merger with plain velars in Tocharian may have occurred very early, since they do not show any palatal effect, nor any other distinctive behaviour (Ringe 1996: 39–42). Unless we can make a phonemic distinction between palatovelars as in PIE **ǵombʰo-* 'tooth' and secondarily palatalized velars, **ǵombʰo-* should probably have depalatalized to **gombʰo-* by the time that PIE **gʷih₃u-* 'live' was palatalized via **gyaw-* to PT **kʸaw-*.

#### 2.3.3.2 Developments of PIE labiovelars

While the development of palatovelars to plain velars is entirely straightforward, the Tocharian development of the Proto-Indo-European labiovelars is quite complicated, as is their representation in the Tocharian script (see, e.g., Hilmarsson 1993; Kim 1999; Pinault 2008: 445–447; Fellner 2005). The majority of original Proto-Indo-European

labiovelars lost their labialization in Tocharian, but a few did remain. The proposed conditions for the delabialization of labiovelars intersect with the developments of PIE **o*, **e*, **i* and **u* and must thus be discussed in order to gain a more complete understanding of the chronology of some developments in the vowel system as well. The followings list contains the words where a labiovelar was preserved and attested in either Tocharian language.[7]

29. PIE **gʷm̥-sk-* 'come' > PT **kʷəmnəsk-* > TB *kənməsk-*[8], TA *kumnäs-* 'come'
64. PIE **kʷekʷlo-* 'wheel; circle' > PT **kʷəkʷle* > TB *kokale*, TA *kukäl* 'wagon'
67. PIE **kʷi-so* 'who' > PT **kʷəse* > TB *kᵤse*, TA *kus* 'who'
68. PIE **kʷr̥i̯h₂-* 'buy' → **kʷr̥i̯h₂-wor* > **kʷəryawer* > PT **kʷəryor* > TB *karyor*, TA *kuryar* 'trade, commerce' (the labiovelar must have been delabialized analogically in TB)
88. PIE **pekʷ-* 'ripen' → *pekʷ-l̥* > PT **pʸəkʷəl*, pl. *pʸəkʷəla* > TB *pikul*, pl. *pikwala*, TA *pikᵤl*, pl. *puklā* 'year'
101. PIE **sokʷo-* 'sap'> **sokʷo-* > PT **sekʷe* > TB *sekwe*, TA *saku* 'pus'
125. PIE **u̯l̥kʷo-* 'wolf' > **wəlkʷo-* > PT **wəlkʷe* > TB *walkwe* 'wolf (?)'

These examples display a variety of phonological contexts in which a labiovelar **Kʷ* was apparently retained: **Kʷm̥*, **Kʷl*, **Kʷi*, **Kʷr̥*, **Kʷl̥*, and **Kʷo*. The initial labiovelar of PIE **kʷekʷlo-* 'wheel' must have been restored, however, since other evidence shows that the sequence **Kʷe* merged with **Ke* at some point: both ended up as palatalized TAB *śə*. The merger between these sequences is clear from the three examples below, but the exact mechanism and intermediary steps remains uncertain.

27. PIE **gʷenh₂* > PT **kʸəna* > TB *śana* /śə́na/, TA *śäṃ* 'wife'
45. PIE **h₂eǵ-e-* > **age-* > PT **akʸə-tər* > TAB *āśtär* 'lead' prs.mid.3sg.

---

[7] Ringe and Kim also list TA *kᵤñaś* 'quarrel, strife', connected with the root **gʷʰn-* 'hit' (Ringe 1996: 107; Kim 1999: 150). Dragoni has instead proposed this to be borrowing from Khotanese *gūrāś-* 'to quarrel' or the infinitive *gūrāśä*, although the shift from **r* to **ñ* is unclear (Dragoni 2023: 92–93). It is better to disregard this word in the current discussion.

[8] Delabialization in the sequence PIE **gʷm-* is posited by Pinault to account for the non-labialized reflex of 'come' in TB *kəmnəsk-*; this would have been regular in the forms such as the PIE **gʷmont*, which yielded prt.3pl. TB *kameṃ* /kə́men/ via **kmen* through a secondary restoration of the root syllable with an additional **ə*. The resulting root shape **kəm-* could have been generalized to the derivative *kəmnəsk-* in the prehistory of Tocharian B specifically; Tocharian A *kumnäs-* retains the regular labiovelar reflex of PIE **gʷm̥-sk-*, and no reflex corresponding to the Tocharian B preterite is known (Pinault 2008: 446). In any case, TA *kumnäs-* shows that the labiovelar was retained in this word.

65. PIE **kʷetuores* > PT **kʸətwerə* > TB *śtwer*, TA *śtwar* 'four (4)'

The first labiovelar of PIE **kʷekʷlo-* should thus regularly have been palatalized, yielding something like TB ***śk(w)ale*, TA ***śukäl*. Since this did not happen, the **e* must have been affected by the flanking labiovelars. An intermediate stage **kukʷlos* could be posited, similar to Greek κύκλος 'circle, ring, wheel' (Hilmarsson 1986: 61), or a secondary re-labialization due to assimilation of from expected **kekʷlos*.[9] Since the Proto-Tocharian reconstruction required for the vowel development in TB *kokale* includes two labiovelars, PT **kʷəkʷle*, it follows that a sequence PIE **kʷl* was preserved. Many other examples show a loss of labiovelars in additional contexts, without palatalization to **ś*, as listed below.

16. PIE **dʰegʷʰ-* > PT **tˢək-* > TB *tsək-*, TA *tsäk-* 'burn'
18. PIE **dʰi̯o̯h₂gʷ-* > PT **tˢak+a-* > TB *tsaka-*, TA *tsākā-* 'pierce'
26. PIE **gʷeh₃u-* > **kow* > PT **kew* > TB *keᵤ*, TA *ko* 'cow'
30. PIE **gʷreh₂-* → **gʷrḭ̯h₂-mr̥-* > PT **kramər* > TB *krāmär* 'weight, heaviness', der. *kramartstse* 'heavy', TA der. *krāmarts* 'heavy'
31. PIE **gʷr̥h₂u̯o-* > PT **kərwenʸe* > TB *kärweñe*, TA *kärwaṃs-* 'stone'
32. PIE **gʷʰen-* 'hit' → **gʷʰn̥-sk-* > PT **kəska-* > TB *kəska-* 'scatter, strike apart'
33. PIE **hₓeh₃gʷʰ-* > PT **yok-* > TB *yok-*, TA *yok-* 'drink'
38. PIE **hₓleng*ʷʰ- > PT **lənk-* > TB *lənk-*, TA *länk-* 'hang, dangle'
43. PIE **hₓregʷ-* → **hₓr̥gʷ-ont-* > **erkont-* > PT **erkent-* > TB *erkeṃt*, TA nom.pl. *arkaś* 'black'
58. PIE **h₃okʷ-s-* > **ok(s)* > PT **ek* > TB *ek*, TA *ak* 'eye'
65. PIE **kʷel-* → **kʷolo-* > **kolo* > PT **kele* > TB *kele* 'navel'
68. PIE **kʷo-* → **kʷo-tos* > **koto* > PT **kete* > TB *kete* 'whose'
70. PIE **kʷrinh₂-* > **krəna-* > PT **kərna-* > TB *kərna-* 'buy' (present stem)
82. PIE **nokʷtu-* > PT **noktə* > TA obl. *nokte* 'at night', adv. *noktiṃ* 'last night'
83. PIE **nokʷteu̯i̯o* > PT **nektʸəwʸe* > TB *nekciye*, TA *nakcu* 'at night, last night'
87. PIE **pekʷ-* > PT **pək-* > TB *pak-*, TA *päk-* 'cook, ripen'
90. PIE **penkʷto-* > PT **pʸənkte* > TB *piṅkte*, TA *pänt* 'fifth'
108. PIE **tekʷ-* 'run, flow' → **tekʷos-* > **tekos* > PT **tʸəke* > TB *cake* 'river'
127. PIE **u̯ōkʷs*, acc.sg. *u̯okʷṃ* > **wok* > PT **wek* > TB *wek*, TA *wak* 'voice'

[9] The labiovelars in this word were only finally lost in Tocharian B *kokale* in the process of a minor sound change converting PT **ə* to TB *o* between two labial consonants (see subsection 2.5.6). TA *kukäl* reflects the original PT **ə* affected differently by at least one preserved labiovelar in that language.

The precise conditioning of this loss of labialization has been interpreted in various ways, and clearly no single rule can cover all of the material. For a number of words, it has been assumed that a delabialization took place next to PIE **o*. This is interpreted as a change from **K$^{w}$o* to **ko* that took place before PIE **o* became unrounded to **ë* and eventual PT **e* in the relative chronology of sound changes (Ringe 1996: 42; Kim 1999: 149–150). This development would not only have taken place word-initially, but also word-internally (Ringe 1996: 42; Kim 1999: 159-164), and, according to Hackstein, between *o* and a following consonant (Hackstein 2017: 1325).[10] The relevant examples are as follows:

26. PIE **g$^{w}$eh$_{3}$u-* > **kow* > PT **kew* > TB *ke$_{u}$*, TA *ko* 'cow'
43. PIE **h$_{x}$rg$^{w}$-* → **h$_{x}$rg$^{w}$-ont-* > **erkont-* > PT **erkent-* > TB *erkeṃt*, TA nom.pl. *arkaś* 'black'
65. PIE **k$^{w}$el-* → **k$^{w}$olo-* > **kolo* > PT **kele* > TB *kele* 'navel'
68. PIE **k$^{w}$o-* → **k$^{w}$o-tos* > **koto* > PT **kete* > TB *kete* 'whose'
108. PIE **tek$^{w}$-* 'run, flow' → **tek$^{w}$os-* > **tekos* > PT **t$^{y}$əke* > TB *cake* 'river'

However, we have already seen above that there are two examples where the sequence **K$^{w}$o* did seem to preserve the labiovelar, namely in TB *walkwe* /wə́lk$^{w}$e/ 'wolf (?)' and TB *sekwe* 'pus'. Kim tentatively ascribes the retained labiovelar in *walkwe* to influence from the neighbouring **l* (Kim 1999: 166-167), but since the rationale would be that *l* is cross-linguistically a relatively more consonantal and that *k$^{w}$* in consonant clusters was preserved, which is not demonstrably the case (see below), this does not seem to be a fully cogent explanation. An assimilation to the initial **w-* might theoretically be possible, but cannot be demonstrated. TB *walkwe* thus remains puzzling, and provides an argument against a sound change from **K$^{w}$o* to **ko*. TB *sekwe*, TA *saku* 'pus', if properly derived from PIE **sok$^{w}$o-* 'juice, resin', would also show an unexpected retention of the labiovelar in the sequence **K$^{w}$o*.[11]

[10] Potentially, TB obl. *kolmai*, TA *koläm* 'basin' belongs here as well. Huard (2022: 411-416), with his new interpretation 'basin' rather than 'boat', derives this word from a PIE **g$^{w}$ol-u-* with an added suffix *-me*. Earlier etymologies depart from **kolmeh$_{2}$* or **kolmō-*. I have decided to exclude this word from consideration on account of the uncertainties involved in its etymological interpretation.

[11] Pinault assumes that the words where labiovelars were preserved in front of apparent **o* originally had a different nominative in **-ku* from earlier **-kwō(n)*, which was restructured, but (re)introduced **k$^{w}$* into the entire paradigm (Pinault 2008: 425). However, as there is no actual trace of an original nominative in **-ku*, this explanation is unconvincing.

Of the words that appear to show a development from **K$^{w}$o* to PT **ke*, the only example that is not a formation specific to Tocharian is TB *ke$_{u}$* 'cow' from PIE **g$^{w}$eh$_{3}$u-*. Here, it could also be assumed that delabialization occurred due to the following PIE **e* before laryngeal colouring, so that this example would align with the general merger of **K* and **K$^{w}$* before **e* instead (discussed above).

Four of the examples where **o* appears to have caused delabialization are formations specific to Tocharian, so that they could be considered less reliable and informative than TB *walkwe* 'wolf (?)' and *sekwe* 'pus'. The stem **k$^{w}$el-* from which TB *kele* 'navel' is derived could have lost the labiovelar in front of the **e*, before the formation of **k$^{(w)}$olo-* 'navel', and the expected labiovelar in **tek$^{w}$os-* 'river' may have similarly become delabialized in case forms such as gen.g. **tek$^{w}$es-os*, as suggested to me by Alexander Lubotsky (p.c., October 2024).

The missing labiovelar in TB *erkent* 'black' may not be directly understandable in the same way, but the related TB *orkamo*, TA *orkäm* 'dark; darkness' also lacks the labiovelar, so that this could be considered a feature of the root in Tocharian independent of the following **o* in the antecedent of the formation *erkent*. In root-final position, a number of other labiovelars were lost as well, without a clear single identifiable cause due to the inherent variation in following elements (see below). The lack of a labiovelar in TB *kete* 'whose' is unclear to me, but as a new formation, it may have been taken over from other forms of the paradigm as well. The original gen.sg. PIE **k$^{w}$eso* may have lost the labiovelar as well before it was replaced if **K$^{w}$e* was merged with **Ke*. Labiovelars also disappeared in a number of consonant clusters, as seen in the examples below.

58. PIE **h$_{3}$ok$^{w}$-s-* > **ok(s)* > PT **ek* > TB *ek*, TA *ak* 'eye'
82. PIE **nok$^{w}$tu-* > PT **noktə* > TA obl. *nokte* 'at night', adv. *noktiṃ* 'last night'
83. PIE **nok$^{w}$teu̯i̯o* > PT **nekt$^{y}$əw$^{y}$e* > TB *nekciye*, TA *nakcu* 'at night, last night'
90. PIE **penk$^{w}$to-* > PT **p$^{y}$ənkte* > TB *piṅkte*, TA *pänt* 'fifth'
127. PIE **u̯ōk$^{w}$s*, acc.sg. *u̯ok$^{w}$m̥* > **wok* > PT **wek* > TB *wek*, TA *wak* 'voice'

This environment is deemed preserverative by Kim instead, as he assumes that the trace of an earlier labiovelar in descendants of PIE **nok$^{w}$t-* 'night' is seen in the rounded vowel in TA *nokte* 'at night', and *noktiṃ* 'last night' (Kim 1999: 163–165). However, on account of the lack of a similar rounding in TA *nakcu* 'at night, last night' and the absence of a labiovelar in the TB cognate *nekciye* 'id.', the vocalism in Tocharian A should be explained differently (Pinault 2008: 424). Early umlaut due to a following **u* in an original nominative PIE **nok$^{w}$tu-* may be assumed instead of a labializing effect of an alleged preserved labiovelar (Pinault 1990: 181–190).

In the majority of examples, either the preceding **o* or the cluster could be regarded as the reason for delabialization (Hackstein 2017: 1325). However, it is still uncertain if **o*

caused delabialization at all (see above on **K$^{w}$o*). Furthermore there is at least one example where a labiovelar disappeared in a cluster without a neighbouring **o*, namely in the ordinal numeral PIE **penk$^{w}$to-* 'fifth'. This word does not retain any labialization in TB *piṅkte* (not ***piṅk$_{u}$te*), let alone in TA *pänt* where the velar stop is lost altogether due to a cluster simplification. However, an intermediate delabialization in the corresponding cardinal PIE **penk$^{w}$e* could also be assumed to have had an analogical effect.

At first sight, TB *laṅk$_{u}$tse* 'easy, light (not heavy)', derived from a PIE root **h$_{1}$leng$^{wh}$-* 'move lightly', would appear to indicate a retention before a consonant. However, TB *laṅk$_{u}$tse* should preferably be derived from a *u*-stem **h$_{1}$l̥ŋg$^{(w)h}$u-* also seen in Gr. ἐλαχύς 'small' and Skt. *laghú-* 'light', *raghú-* 'swift'. Consequently, the cluster TB /-k$^{w}$t$^{s}$-/ in this word probably arose through syncope from pre-PT **lənkut$^{s}$e*. The syncope could have taken place after PIE **penk$^{w}$to*, with a primary cluster, had already been simplified to pre-PT **penkto*.

We must furthermore assume that TB *akwatse* 'sharp' was secondarily derived from or affected by an unattested **ak$^{w}$ə* (PIE **h$_{2}$eḱu-*, see Pinault 2008: 445) after syncope had become inactive, since otherwise a quasi-PIE formation **h$_{2}$eḱut̥io* should have yielded TB ***āk$_{u}$tse* like TB *laṅk$_{u}$tse* from earlier **h$_{1}$lngʰu-tio-*. The labiovelar in the interrogative pronoun TB *mäkte* 'how', from **mo-k$^{w}$i-tod* 'what' (Pinault 2010; Peyrot 2018a: 80-81) may have been lost irregularly; TB *mäkte* 'how' rather than ***makte* /mə́kte/ is in any case not stressed in the same way as normal content words.

Delabialization may also be assumed in the cluster **K$^{w}$r*, *pace* Kim (1999: 151), who posits a continuation of **K$^{w}$r* as such until the Proto-Tocharian period on account of TA *kuryar* 'trade, commerce'. I think that this form is better interpreted as the result of **k$^{w}$* before a syllabic **r̥*, which was vocalized to **k$^{w}$ər* before the simplification of **k$^{w}$r* took place; cf. **k$^{w}$r̥i̯h$_{2}$wer* in the first list of this subsection. The lack of a labiovelar reflex in TB *karyor* 'trade, commerce' is likely due to analogy with the present stem PT **kərna-*, via **krəna-* from PIE **k$^{w}$rinh$_{2}$-* 'buy', which is one of the two examples showing a delabialization in a cluster **K$^{w}$r*. In that form, the initial labiovelar would necessarily have been in contact with the non-syllabic **r* until relatively late Tocharian liquid metathesis. The other example, PT **kramər* 'weight, heaviness' is dubious in its formation, as it is without exact parallels in the other Indo-European languages, but if it is correctly derived from the same PIE root **g$^{w}$reh$_{2}$-* 'heavy' as in Gr. βαρύς 'id.' and Lat. *gravis* 'id.', the labialization on the initial **g$^{w}$r-* was clearly lost, pointing to delabialization in this type of cluster.

70. PIE **k$^{w}$rinh$_{2}$-* > **krəna-* > PT **kərna-* > TB *kərna-* 'buy' (present stem)
30. PIE **g$^{w}$reh$_{2}$-* → **g$^{w}$r̥h$_{2}$-mr̥-* > PT **kramər* > TB *krāmär* 'weight, heaviness', der. *kramartstse* 'heavy', TA der. *krāmarts* 'heavy'

In one contrasting word, TB *kärweñe*; TA *kärwaṃs-* 'stone', a labiovelar before an original syllabic **r̥* was delabialized, but this was probably the result of dissimilation with the **w* in the following syllable instead (so Ringe *apud* Kim 1999: 151 fn. 26; Pinault 2008: 425).

31. PIE **gʷr̥h₂u̯o-* > PT **kərwenʸe* > TB *kärweñe*, TA *kärwaṃs-* 'stone'

The lack of a labiovelar in TB *kəska-* 'scatter, strike apart' is difficult to explain departing simply from **gʷʰn̥-sk-*. This would be expected to develop as **gʷm̥-sk-*, where Tocharian A shows a retention of the labiovelar in *kumnäs-* 'come'.

32. PIE **gʷʰen-* 'hit' → **gʷʰn̥-sk-* > PT **kəska-* > TB *kəska-* 'scatter, strike apart'

However, it may be imagined that the *sk*-present was influenced by the original root present, as reflected in Skt. 3sg. *hánti*, 3pl. *ghnánti* and Hitt. 3sg. *kuenzi*, 3pl. *kunanzi* from PIE **gʷʰenti*, **gʷʰnenti*.[12] Either **gʷʰe* or **gʷʰne* —or perhaps both— may have undergone regular delabialization and influenced the *sk*-present.

The remaining examples of delabialization occur in root-final position. There is no single rule that can capture all instances, since the following elements were variable, but understandable causes may be outlined.

16. PIE **dʰegʷʰ-* > PT **tˢək-* > TB *tsək-*, TA *tsäk-* 'burn'
18. PIE **dʰi̯h₂gʷ-* > PT **tˢak+a-* > TB *tsaka-*, TA *tsākā-* 'pierce'
33. PIE **h₁eh₁gʷʰ-* > PT **yok-* > TB *yok-*, TA *yok-* 'drink'
38. PIE **h₁lengʷʰ-* > PT **lənk-* > TB *lənk-*, TA *länk-* 'hang, dangle'
43. PIE **h₁regʷ-* → **h₂u̯gʷ-ont-* > **erkont-* > PT **erkent-* > TB *erkeṃt*, TA nom.pl. *arkaś* 'black'
87. PIE **pekʷ-* > PT **pək-* > TB *pək-*, TA *päk-* 'cook, ripen'
108. PIE **tekʷ-* 'run, flow' → **tekʷos-* > **tekos* > PT **tʸəke* > TB *cake* 'river'

Pinault proposes that some of the labiovelars in question were lost in consonant clusters, as in the cases discussed above, for example in the *s*-presents of PT **pək-* 'cook, let ripen (act.); boil, ripen (mid.)' from PIE **pekʷ-* and PT **tˢək-* 'burn' from PIE **dʰegʷʰ-* (Pinault 2008: 424). This would also work for the root PIE **dʰih₂gʷ-*, which must have been in

[12] The parallel *sk*-present found in Hittite *kuu̯aške/a-* ipfv. 'kill, slay' may reflect a secondary formation from the present on account of the vocalism (see Hackstein 1995: 200-201), although Kloekhorst (2008: 485-486) has argued that it can be regularly derived from PIE **gʷʰn̥-ske/o-*. If the Hittite and Tocharian formations are both old, analogical influence from the earlier root present in Tocharian is still a possibility.

direct contact with suffixal consonants both as an original root present (LIV: 142) and in its eventual nasal present inflection in TB prs. *tˢakna-* 'bite, pierce'. PT **lənk-* 'hang' from PIE **h₁leng$^{wh}$-* forms a root present as well in Tocharian, and can thus have lost the labiovelar in contact with a following consonant.

Kim rather ascribes the disappearance of the labiovelars to a delabialization in the sequence **-K$^{w}$o*, which occurred in part of the paradigm of the original thematic presents of some of these verbs. This **-K$^{w}$o-* alternated with **-K$^{w}$e-*, where the labialization was lost due to palatalization to **-k$^{y}$e-* and, eventually, TAB *-śä-*. Due to these two changes, the root-final labiovelars would have been completely eliminated from the paradigm (Kim 1999: 160). This may work for *d$^{h}$eg$^{wh}$-* and *pek$^{w}$-*, which can be reconstructed with a thematic present in Proto-Indo-European (LIV: 133, 468), but the explanations put forward by Pinault and Kim are not mutually exclusive. It is also possible that the labiovelar was delabialized only in **K$^{w}$e*, and that the delabialization was analogically spread from there (cf. above).

While attested without a labiovelar, the verb TAB *yok-* 'drink' seems to have initially escaped the effects of these processes. Though complicated in terms of its diachronic derivation (see Peyrot 2013: 448-449), the development of the verb PIE **h₁eh₁g$^{wh}$-* 'drink' to TAB *yok-* is most easily explained if there was still a labiovelar present at a relatively advanced stage to cause a rounding of the vowel. Otherwise, the expected outcome is ***yek-*.[13] The root-final labiovelar in **h₁eh₁g$^{wh}$-* should thus probably have avoided delabialization in at least part of the paradigm.

It is clear that the developments of original labiovelars from Proto-Indo-European in Tocharian is a complicated matter. At a shallow Tocharian level, it is also necessary to account of the verb TB *k$^{w}$əĺpa-*, TA *k$^{w}$älĺpa-* 'desire'. On account of the palatalized lateral, this must derive via metathesis from earlier **k$^{w}$l$^{y}$əpa-*. The viability of the sequence **k$^{w}$l* in this word accords with the retention of the labiovelar in **k$^{w}$ek$^{w}$lo-* 'wheel' (see above).

The relative chronology of changes implied by the varying reflexes of original labiovelars in these different environments is summarized in Table 2.5. Syncope, epenthesis, stress shift and liquid metathesis will be more extensively discussed in Chapter 5.

[13] According to Kim (1999: 165), the verb PT **yok-* 'drink' retained its original labiovelar before suffixes starting with a consonant (contrary to the delabialization in that context supported by Pinault and Hackstein and discussed above).

Table 2.5: A relative chronology of sound changes involving labiovelar stops.

| relative chronology | example | instead of | reconstruction |
|---|---|---|---|
| *nkʷt* > *nkt* | *piṅkte* | ***piṅk$_u$te* | **penkʷto-* |
| ? *kʷr*, *kʷm*, *kʷn* > *kR* | *kärnā-* | ***kurnā-* | **kʷrinḫ$_2$-* |
| syncope 2nd syllable | *laṅk$_u$tse* | ***laṅktse*, ***läṅkwatse* | **h$_1$lo̯gʰu+ti̯o* |
| epenthesis/stress shift | *kokale* | ***kokle*, ***kwakle* | **kʷekʷlo-* |
| liquid metathesis | *kulypa-* | ***kälypa-* | **kʷlʸəpa-* |

I have argued above, against Kim (1999) and in accordance with Pinault (2008) and Hackstein (2017), that there is at present no convincing example of preserved labiovelars before other consonants, apart from **kʷl*, as necessitated by TB *kokale* 'wagon' from **kʷekʷlo-* 'wheel'. The labiovelar in TB *laṅk$_u$tse* only secondarily came to be followed by **t*ˢ after syncope, as the stem ended in a vowel: **h$_1$lng*ʷʰ*u-*. Furthermore, the labiovelar in the stem PT **kʷərya-* as in TA *kuryar* 'trade, commerce' was not followed by the **r* from the moment that syllabic resonants had become vocalized: **kʷr̥i̯ḫ$_2$-* became **kʷəraya-*. The velar in the present stem TB *kärnā-*, on the other hand, was directly followed by the **r* until liquid metathesis: **kʷrinḫ$_2$-*. Whether liquid metathesis had already occurred in Proto-Tocharian is unclear, so that either PT **krəna-* or PT **kərna-* could be reconstructed (see also 5.2.2.3).

Because the Tocharian labiovelars are so complicated, perhaps in some respects more complicated than has been assumed, it is difficult to fit their developments into the relative chronology. Importantly, the development from **Kʷo* to **Ko* cannot be taken for granted on account of TB *walkwe* 'wolf (?)' and *sekwe* 'pus'. If **Kʷo* did remain labialized with delabialization occurring only in clusters and before original PIE **e* or **ē*, the latter change can still be dated before the merger of PIE **o* and PIE **ē* that occurred with palatalization (2.4), but it cannot be placed before the unrounding of PIE **o* to pre-PT **ë* that took place earlier.

#### 2.3.3.3 Secondary labiovelars

Sequences of a PIE velar plus consonantal **u̯*, i.e., of the type **Ku̯o-*, end up as labiovelars in Tocharian (Kim 1999: 158, 163, 167). This is seen in the following examples:

35. PIE **h$_1$eḱu̯o-* > **ekwo* > PT **yəkʷe* > TB *yakwe*, TA *yuk* 'horse'
49. PIE **h$_2$erǵu-* → PT **arkʷi* > TB *ārkwi*, TA *ārki* (**kʷi* > TA *ki*) 'white'
61. PIE **ḱu̯onm̥* > **kwonəm* > PT **kʷenə* > TB *kweṃ*, TA *koṃ* 'dog (obl.sg.)'
79. PIE **n̥ḱu̯o-* > **enkwo* > PT **enkʷe* > TB *eṅkwe*, TA *oṅk* 'man'

119. PIE ±*u̯erǵ-unt- → *werkwənt- > PT *wʸərkʷənt- > TB obl. *yerkwantai* /yerkʷə́ntay/, TA *wärkänt* 'wheel'[14]

As shown by the examples *h₁eḱu̯o- 'horse' and *n̥ḱu̯o- 'mortal → man', these new labiovelars did not disappear before PIE *o (cf. Ringe 1996: 42), and thus accord with the evidence from PIE *ulkʷo-, TB *walkwe* 'wolf (?)' and *sokʷo- 'sap', TB *sekwe* 'pus' with their retained labiovelars before *o.

Additional secondary labiovelars came into existence in contact with original vocalic PIE *u as well, but this development is treated in subsection 2.4.3 instead, together with a more general overview of the development of PIE *u.

### 2.3.4 The stop system

Tocharian is notorious among the Indo-European languages for its lack of distinctive manners of articulation in the stops. The Tocharian stops are simply /p t k kʷ/ with the additional affricates /tˢ/ and /c/ (±[tɕ], derived from palatalized PIE *t/dʰ). This is a clear departure from the Proto-Indo-European system, which had up to three distinct manners of articulation at each place, cf. the traditional reconstruction of the dental series as PIE */t dʰ d/. As mentioned briefly in section 2.2, the exact phonetic values of the Proto-Indo-European stops are still a matter of debate (e.g., Kümmel 2007: 299–309), but it is clear that there were three distinct series. At some point in pre-Proto-Tocharian, these distinctions were lost. The examples below show how voiceless stops *p, t, k, kʷ* result in Tocharian from PIE "voiced" and "voiced aspirated" stops (cf. Krause & Thomas 1960: 64–68; Ringe 1996: 152–154, Pinault 2008: 422–424, etc.).

1. PIE *bʰeh₂ǵʰu- > PT *pok- > TB obl. *pokai*, TA *poke* 'arm'
2. PIE *bʰreh₂tēr > PT *protʸer > TB *procer*, TA *pracar* 'brother'
3. PIE *bʰr̥ǵʰro- > PT *pərkre > TB *pärkare* /pərkə́re/, TA *pärkär* 'tall'
21. PIE *dʰugh₂tēr > PT *təkatʸer > TB *tkācer* /tkácer/, TA *ckācar* 'daughter'
29. PIE *gʷm̥-sk- > PT *kʷəmnəsk > TB *kənməsk-*, TA *kumnäsk-* 'come'
32. PIE *gʷʰn-sk- > PT *kəska- > TB *kəska-* 'scatter, strike apart'
35. PIE *h₁eḱu̯o- > PT *yəkʷe > TB *yakwe* /yə́kʷe/, TA *yuk* 'horse'
43. PIE *h₁regʷ- → *h₂rgʷont- > PT *erkent- > TB *erkent*, TA pl. *arkaś* 'black'
44. PIE *h₁rudʰro- > PT *rətre > TB *ratre* /rətre/, TA *rtär* 'red'
76. PIE *medʰu- > PT *mʸətə > TB *mit* 'honey'

[14] The development of this word is unexpected in both Tocharian A and B. In TA the labiovelar was apparently delabialized due to dissimilation with the word-initial *w-*. In TB there is an unexplained development from *ə to *e* in the first syllable; **yärkwantai* /yərkʷə́ntai/ would have been the expected counterpart of TA *wärkänt* (Kim 1999: 168; Adams 2013: 547–548).

It is difficult to establish any intermediate stages in which for instance three manners of articulation may have been first reduced to two, before finally becoming one. It is clear that not all stops simply merged into a voiceless series, however. There are a few different developments that affected PIE *$b^h$ and *$d$, and not their respective counterparts *$p$ and *$t$/$d^h$ produced at the same point of articulation. In the following subsections, we will take a look at the developments of these stops and consider what they tell us about the evolution of the Tocharian consonant system through time.

#### 2.3.4.1 Assimilation of *$mb^h$ to *$m$

There is a discrepancy in the reflex of the Proto-Indo-European labials *$p$ and *$b^h$ when they followed a nasal *$m$. Namely, a PIE *$b^h$ was lost in that position, while *$p$ was retained.[15] There are only very few examples of these clusters:

24. PIE *$ǵomb^ho$- > PT *$keme$ > TB *keme*; TA *kam* 'tooth'
104. PIE *$stemb^hH$- > PT *$s^ytʸəma$- > TB prt. *ścəma*- 'stand'
110. PIE *$temp$- 'stretch, extend' > PT *$tʸəmp$- > TB *cəmp*-, TA *cämp*- 'be able'

This change is important for the phonetic interpretation of the stop series in the earliest pre-Proto-Tocharian, as the difference in the behaviour of *$b^h$ and *$p$ in this environment should be phonetically understandable. On the basis of the standard interpretation of *$b^h$ as breathy voiced, Ringe conceptualized this development from PIE *$mb^h$ to PT *$m$ as either an extension of breathy articulation of *$b^h$ to the preceding *$m$ followed by a deletion of the stop element, or as a progressive assimilation via an intermediary stage *$mb$, i.e., *$m$ + simple voiced *[b] (Ringe 1996: 42-44).[16]

Peyrot has considered what this sound change might mean for the interpretation of the early pre-Proto-Tocharian stop system, also from the perspective of the Glottalic Theory, and concluded that the series *$b^h$ (and by extension potentially also the other voiced aspirates) would have had a weaker articulation than *$p$ (etc.) in one way or another (Peyrot 2022a: 91–92).

---

[15] One single potential etymology containing *$b$ in the cluster *$mb$ shows a development like *$b^h$ in this position: PIE *$lembH$- 'hang' > PT *$ləma$- > TB *ləma*-, TA *lämā*- 'sit' (LIV: 411). If *$mb$ indeed developed as *$mb^h$, that could imply that *$b^h$ and *$b$ were more similar to one another in some respect than to *$p$.

[16] Ringe also discusses the possibility that voiced aspirates at other places of articulation could have been lost after nasals as well, but are no good examples (Ringe 1996: 42-44). The development of *$h₁lund^h$- to PT *$lən$- suggested by Ringe does not hold, since the Tocharian stem in question is rather *$lənn$- from *$ləntn$- (Peyrot 2013: 446–447), so that the type of nasal cluster simplification seen here is really restricted to the labial place of articulation.

However, this can be understood in all three models for the Proto-Indo-European stop system given above in 2.2. From the traditional point of view, the voicedness of *$b^h$ could have facilitated full assimilation from *$mb^h$ to *$m$. In the glottalic model, *$mb^h$ is interpreted as *"mb", so that assimilation can have taken place in the same way. Kloekhorst's model, with an interpretation of *$b^h$ as *"$p$" may be interpreted with an intermediary voiced stage *[mb] before eventual full assimilation. The changes involved in all three models are relatively minor, but they should probably precede the general devoicing of Tocharian stops that applies to all stops in non-labial nasal clusters as well.

#### 2.3.4.2 Early assibilation of dental stops

The dental series shows different developments as well. PIE *$t$ and *$d^h$, but not *$d$, undergo an assibilation process when followed by *$y$ (= *$i̯$), ending up as PT *$t^s$ (Winter 1962: 20-22; Ringe 1996: 79).[17] The sequence PIE *$dy$ yielded PT *$y$ instead, with complete loss of the *$d$ (Ringe 1996: 65). The assibilation of *$ty$ and *$d^hy$ and their contrast to *$dy$ are seen in the following examples.

18. PIE *$d^hi̯H_3g^w$- > *$d^hyag^w$- > PT *$t^saka$- > TB *tsaka*- 'pierce, bite'
86. PIE *$peh_2nti̯h_2$ > *$pāntya$ > PT *pont$^s$a* > TA *pontsāṃ* 'all'
117. PIE *$u̯lh_2nti̯h_2$ > *$wlantya$ > PT **lantsa* > TB *lāntsa*; TA *lānts* 'queen'

107. PIE *$su̯id$-$i̯e$- > *$swidye$- → PT *$səya$- > TB *səya*- 'sweat'

This indicates that *$d$ had in some way changed or disappeared in the sequence *$dy$ so as to no longer be a stop at the time that assibilation took place: if there was no stop element present in *$dy$ anymore, it logically could not have become an affricate.

There is also another proposed assibilation of only word-final *-*ti* and *$d^hi$ to TAB *ṣ*, potentially via *-*si*. This change was proposed by Jasanoff (1987: 106–112), based on the development of three suffixes with this shape.

34. PIE *$h_1id^hi$ > PT *$pə$-$yəṣə$ > TB *paṣ*, TA *piṣ* 'go!'
111. PIE *-*ti* > PT *-$ṣə$ > TA -(*ä*)*ṣ* (prs.3sg. verbal ending)
112. PIE *-*ti* > PT *-$ṣə$ > TA -(*ä*)*ṣ* (ablative ending)

The first example is the irregular imperative of 'go', PIE *$h_1id^hi$ > PT *$pə$-$yəṣə$ > TA *piṣ*, TB *paṣ* 'go!' (*pə*- is the Tocharian imperative marker). Another verbal ending, 3sg. primary

[17] For *$d^hi̯$ to PT *$t^s$ this is not as firmly established due to a dearth of reliable example; only PIE *$d^hi̯H_3g^w$- to PT *$t^saka$- 'pierce, bite' supports this development; cf. Hackstein (2017: 1324), who omits mention of *$d^hi̯$ altogether.

PIE *-*ti*, might have regularly yielded TA *-(*ä*)*ṣ* in, e.g., *lkāṣ* 'sees' from PT **ləka-ti*. Jasanoff's third example is the Tocharian A ablative in -(*ä*)*ṣ*, which he compares to the Hittite ablative in -(*a*)*z* (/(a)t$^{z}$/) from **-ti*, possibly connected with the adverb **h₁éti* 'from this' (Jasanoff 1987: 109–110). This development is accepted by Ringe (1996: 79–80) and Hackstein (2017: 1325), but not by Pinault (2008: 620).

There are a few factors that cast doubt on this sound change. If the change of word final *-*ti* to TAB *ṣ* occurred as posited by Jasanoff, it appears that assibilation of *-*ti* to an intermediate *-*si* needs to have happened before palatalization, since palatalized **s* is the usual source of TAB *ṣ*. This development would then have to be separated from that of word-final *-*nti*, because that became TAB *-ñc* instead, rather than **-*nṣ* vel sim. It is furthermore noteworthy that the development of **ty* and **ti* would not be entirely parallel, since the former remained affricated TAB *ts*, while the latter would have become totally assibilated in TAB *-ṣ*. A special word-final development on the consonant is also difficult to envision, since it is the vowel *-*i* that was word-final until relatively late apocope.

Whether or not **ty*/*d*$^{h}$*y*- initially remained distinct from initial **d*- is also difficult to tell, since the respective sequences did eventually merge as TAB *t*$^{s}$- (compare the examples above and those given for **d* in the next subsection). In order to maintain this potential distinction, I will use the symbol "*ċ/ż*$^{h}$" to denote the product of assibilated **t* and **d*$^{h}$. Especially in the case of "*ċ*", the dot is intended as a neutral diacritic to avoid confusion with TAB *c* from (later) palatalized **t*/*d*$^{h}$, which resulted from a different change, and to avoid using the same symbol as either [tʂ] <č> or [tɕ] <ć> in Uralic phonological transcription. It is not meant to represent any specific phonetic value, but may be interpreted as *[ts] or *[tʃ] or *[tɕ] at the earliest stages, all typologically reasonable affricate outcomes of sequences of the type *-*tj*- (see Kümmel 2007: 258–261).

#### 2.3.4.3 The aberrant development of PIE **d*

Just as the cluster **dy* developed differently from **ty* and **d*$^{h}$*y*, the development of PIE **d* in Tocharian in general is a more complicated matter than that of **t* and **d*$^{h}$. As a result, it may also be more informative. PIE **d* essentially has four commonly accepted reflexes, namely *t*$^{s}$, *ś*, *t*, or ∅. These will each in turn be discussed in this subsection.

It is well-known that PIE **d* disappeared before **w*, **y* and **r*, as seen in the following examples (see Ringe 1996: 64-66, Hackstein 2017: 1324):

10. PIE **dor-u-*, **dr-eu̯-* > PT **orə* > TAB *or* 'wood'
11. PIE **drem-* 'run' > der. PT **rəmer* > TB *ramer* 'quickly'
12. PIE **du̯eiˌ-* 'fear' > PT **wəy-* > TB *wəy-*, TA *wäy-* 'be frightened'
14. PIE **du̯ito-* > PT **wəte* > TB *wate*, TA *wät* 'second (2$^{nd}$)'
15. PIE **du̯oh₁* > PT **wəw* > TAB *wu* 'two (2)'

106. PIE *su̯h̥₂dro- > PT *sware > TB *swāre*, TA *swār* 'sweet'
107. PIE *su̯id-i̯e- > *swidye- → PT *səya- > TB *səya-* 'sweat'
121. PIE *u̯id-u̯on- > PT *wəwe > TB *ūwe* 'wise'
126. PIE *udr- > PT *wər > TB *war*, TA *wär* 'water'
126. PIE *udr-io- > PT *wərəye- > TB der. *wriyeṣṣe*, TA *wri* 'watery'

In the first example PIE *dor-u-, the disappearance of *d must originally have taken place in particular forms like the genitive *dre-u-, with a zero-grade in the root. Thus, apparently, *doru : *dreu became *doru : *reu and then analogically *oru : *reu (on this development, see also 2.3.4.4).

In a few cases, PIE *d is reflected as PT *t, merged with non-palatalized PIE *t and *dʰ. This is found in some clusters, especially after PIE *n in the cluster PIE *nd (see Malzahn 2010: 742; Hackstein 2017: 1324). The main examples are PIE *spend- to PT *spənta 'trust' and PIE *splend-* to PT *plənta- 'be pleased'. The development of the root PIE *kedh̥₂- to *kəta- 'strew, scatter' is less clear. The nasal present (an old formation on account of cognates like Gr. σκίδνημι 'scatter', see LIV: 550) shows a metathesised variant TB *kənta-* next to renewed *kətna-*, but it is unclear whether the metathesis would have been early enough for this verb to develop like *spend- and *splend-.

62. PIE *kedh̥₂-, prs. *kədnh̥₂- or *kn̥dh̥₂- ? > PT *kəta-, *kənta- > TB *kəta-*, *kənta~kətna-*, TA *kätā-*, *känā-* 'strew'
80. PIE *neud-, prs. *nud-sk- > *nutsk > PT *nətk- 'push away'
103. PIE *spend- > PT *spənta- 'trust' (secondary *-a*)
104. PIE *splend- > PT *planta- 'be pleased' (secondary *-a*)

Outside of these clusters, PIE *d in word-initial position generally became TAB *tˢ* (Ringe 1996: Hackstein 2017: 1323–1324). When palatalized, the result was TAB *ś*. There are not many examples available. The palatalized *ś* is most clearly seen in the numeral 'ten (10)' TB *śak*, TA *śäk* from PIE *deḱm̥ (Ringe 1996: 146–147; Hackstein 2017: 1324). The reflex TAB *tˢ* is found most clearly in a few verbal roots, where palatalization was analogically removed in certain forms.

5. PIE *demh̥₂- > PT *tˢəma- > TB *tˢəma-*, TA sbj. *tˢämā-* 'grow'
6. PIE *der- 'split' > PT *tˢər- > TB *tˢər-*, TA *tˢär-* 'separate'
7. PIE *deu̯h̥₂- > PT *tˢəwa- > TB *tˢəwa-* 'attach oneself to', TA *tˢäwā-* 'fit; obey'
8. PIE *deu̯k- 'pull' > PT *tˢəwka- > TB *tˢawka-*, TA *tˢäwkā-* 'suck (out)'

It is notable that there are no secure intervocalic examples of *-d-. It is therefore difficult to say for sure what happened to it in that position (cf. already Winter 1962). The only

instance adduced by Ringe is TB *witsako* /witˢəko/ 'root', which he derives from a pre-form **wẹdəga-* from pre-Proto-Ossetic **wēdāga* (Ringe 1996: 146). However, as Bernard shows, it is not possible to derive this word from any known Iranian source (Bernard 2023: 222–227), so that its value for determining of the details of a Tocharian sound law is extremely questionable. This word is not inherited from Proto-Indo-European in any case, meaning that it would not show a development of PIE **d*. In old borrowings from Indo-Iranian, IIr. **d* is rendered as PT **t* instead: e.g., **spaldaka-* ⇒ PT **speltke* > TB *speltke*, TA *spaltäk* 'zeal, effort' (Bernard 2023: 75–77).

Unfortunately, the few potential reflexes of PIE word-internal **d* are not fully probative. One example of Indo-European origin could be in the dual of 'foot', **pod-h₁e*, which, according to Winter, regularly became pre-PT **peyə*. This was rather untransparent dual was then enlarged with the productive dual suffix *-ne* to create PT **peyəne*, attested as TA *peṃ*, TB *peyne*, *paine* (Winter 1962: 31). This would point to a *y*-reflex in intervocalic position (or ∅ with *y* arising in the resulting hiatus). The nominative singular *paiyye* may be derived from the dual via **peyə-ye*, so that it is of little help in determining what happened. With a parallel loss of **d*, it is assumed that PT **sʸəm-* 'sit' derives from PIE **sed-* 'sit' in some way; the final *-m* would be taken from *ləma-*, with which it forms a suppletive paradigm, although clearly the final *-a* was not taken over (Adams 2013: 717, Pinault *apud* Malzahn 2010: 927).[18]

#### 2.3.4.4 Interpreting the development of PIE **d* in Tocharian

The development of PIE **d* in Tocharian can be viewed in a number of different ways. In order to understand the development of PIE **d*, a coherent account is needed of all its different reflexes, including a plausible phonetic interpretation. Key points that need to be explained by such an account are:

1. The loss of PIE **d* in various clusters (and likely in intervocalic position), especially in contrast with preservation of **t* and **dʰ* in similar environments
2. The merger of PIE **d* with the outcomes of PIE **t* and **dʰ* in the cluster **-nd-*
3. The development of PIE **d* to TAB *tˢ* (with palatalized counterpart TAB *ś*)

[18] A new word-internal reflex of *ts* has been proposed for *tənts-* 'scatter', derived by Hackstein from PIE **tund-* 'strike, beat'. The *ts* would in this case have spread by analogy from the non-nasal allomorph of the root, **tud-* > **təts-* (Hackstein 2017: 1324), so that it would not be a direct reflex as such. Since no allomorph without the nasal is attested, this etymology as a whole remains rather speculative, so that this word cannot give us reliable information on the development of intervocalic **-d-*.

The difficulty lies in plausibly combining these three developments in a single phonetic interpretation while also taking into account the continued separate existence of PIE **t* in Tocharian. Ringe provides a rather straightforward reconstruction of the development from PIE **d* to TAB *t*$^{s}$ via an intermediate **"d$^{z}$"* and explains the different behaviour of these dentals in clusters with **y* by positing an earlier devoicing of **d$^{h}$* and the other voiced aspirates to **"t$^{h}$"*, etc. (Ringe 1996: 47–48, see also 2.3.4.5). However, neither the putative development from **d* to **"d$^{z}$"* nor the development from **d$^{h}$* to **"t$^{h}$"* are otherwise observable in the Tocharian material.

Furthermore, while a change from **d* to **"d$^{z}$"* to TAB *t*$^{s}$ is at first glance straightforward and economical, I do not think that this is particularly likely as a development for typological and phonetic reasons. Assibilation, or affrication in general, can be understood as a reinterpretation of the turbulent airflow that occurs right after a stop's release as a fricative or sibilant. Such turbulent airflow is generally more pronounced in voiceless and especially in aspirated stops than in voiced ones, so that one would least expect a voiced stop to become an affricate (Hall & Hamann 2006). For example, when followed by [i] or [j], two sounds that increase post-release turbulence with their restriction of the airflow through the mouth and thus often promote reinterpretation of a stop as an affricate, it appears that there is less turbulent airflow following [d] than following [t]. Based on this, [d] is only expected to assibilate when [t] does, too, not the other way around (see Hall, Hamann & Zygis 2006 on this topic).

Similarly, on "spontaneous" affrication, Kümmel (2007: 173) states that it is almost exclusively found to happen with voiceless plosives. He does, however, give the Tocharian change as **d* to **"d$^{z}$"* and also palatalized **d$^{y}$* to **"d$^{zy}$"* without further comment, directly following Ringe.[19] However, there does not seem to be a good parallel for direct affrication of just **d* and not **t* —let alone a supposed **"t$^{h}$"* from PIE **d$^{h}$*, which should be the most likely to undergo affrication—, and this lack of phonetic naturalness casts doubt on the seemingly straightforward path from **d* to TAB *t*$^{s}$ via **"d$^{z}$"* before general Tocharian devoicing.

[19] Kümmel's other example of spontaneous assibilation of **d* is found in Middle Cornish, where **d* changed to **d$^{z}$*, and eventually *z*, while **t* changed to **t$^{s}$*, and on to *s*. Here the environment in which **t* was changed, namely after **l* and **n*, seems more restricted than that of **d*, which assibilated after any vowel. However, this cannot be taken as an example of **d* being after all more prone to assibilation than **t*, since Cornish intervocalic **d* is the result of earlier **t* through lenition. At the point in time when the Cornish assibilation took place, **t* existed almost exclusively word-initially, where neither it nor **d* was affected. Between vowels, there was a contrast between geminate **tt*, voiced **d* and fricative **ð*. Singleton **t* did not occur intervocalically, a position where **d* assibilated to **d$^{z}$*/*z*, because it had turned into that self-same **d* at an earlier stage through a process of lenition.

The particular development of PIE **d* in clusters is not properly addressed by Ringe's hypothesis either, so that the lack of affrication of **d* in PIE **spend-* > PT **spənta-* 'trust' and PIE **splend-* > PT **planta-* 'be pleased' remains unexplained. On the basis of precisely the development in these verbs, Malzahn instead assumes that **d* first became a voiced fricative *[ð] before turning into TAB *t*$^{s}$ (*ś* when palatalized). An intermediate *[ð] fits here, because in a position after a nasal, fricativization of obstruents is typologically less likely to occur (Malzahn 2010: 742). A lenited intermediary *[ð] could also be connected with the disappearance of PIE **d* in the clusters **dw*, **dy* and **dr* (i.e., via *[ðw], *[ðy], *[ðr]), as an intermediate stage in that development.

If we consider the disappearance of **d* assumed for PT **peyə* 'foot' (and perhaps somehow in PT **s*$^{y}$*əm-* 'sit'), lenition via a [ð] provides a plausible path for **d* to have yielded intervocalic **y* or ∅ as well. Especially intervocalic [ð] is often further lenited and lost (e.g., Kümmel 2007: 109–113). Since the loss of **d* before **w*, **y* and **r* also took place in word-initial position, lenition of **d* to *[ð] could be a general development, not restricted to word-internal position. In this way, one could reconstruct a general development of **d* to *[ð], blocked in at least the clusters **nd* (PIE **spend-* > PT **spənt+a-*, PIE **splend-* > PT **plant+a-*) and **dsk* (PIE **nud-sk-* > PT **nətk-*).

For the development to TAB *t*$^{s}$ in initial position there are many paths that could be imagined, but none can be proven on the basis of the available data. If *[ð] is taken as a first step in the development, based on the reflexes ∅ and *-t-* after nasals, some possibilities for further developments to TAB *t*$^{s}$ are:

- [ð] > [z] > [dz] > [ts]
- or [ð] > [θ] > [tθ] > [ts]
- or [ð] > [θ] > [t$^{h}$] > [ts]
- or [ð] > [dð] > [dz] > [ts]
- or even [ð] > [j̟] > [j̟j̟] > [cɕ] > [ts][20]

The fortition that seems to be necessary in all of these sequences could be tentatively interpreted as the result of the phonologically strong onset position in which TAB *t*$^{s}$ is obtained from **d*. Again, none of these hypothetical intermediary stages are reflected in the Tocharian material, and they thus remain speculative.

Peyrot briefly considered the possibility of interpreting **d* as glottalized **ʔd* in early pre-Proto-Tocharian, but concludes that there is no direct evidence in favour of it to be found in Tocharian (Peyrot 2022a: 91–92). We may still explore if there is any indirect evidence. Disappearance of PIE **d* in the clusters **dw*, **dy* and **dr* could in principle be

[20] Compare the allophonic realization of initial /j̟/ in Spanish as [j]~[j̟]~[j̟j̟]~[ɟ] (Hualde 2005: 383–384) and the development from **ty*/*d*$^{h}$*y*, with a clear original palatal element, to TAB *t*$^{s}$ (see 2.3.4.2).

interpreted in a glottalic framework as a change from glottalic *“*ʔd*” to *“*ʔ*”. The unusual analogical development seen in **doru* : **dreu* to **oru* : *reu* ‘wood’ seen above could then be understood as a development from glottalic **ʔdoru* : **ʔdreu* to **ʔdoru* : **ʔreu* and then **ʔoru* : **ʔreu* instead, as suggested by Alexander Lubotsky (p.c., March 2025). In that way, the element that was analogically spread was a consonant, **ʔ*, rather than a lack of a consonant, ∅. The Tocharian development of PIE **d* in clusters with resonants could then be considered as a conditioned debuccalization from *“*ʔd*”. In the cluster **-nd-*, where PIE **d* behaves like **t* and **dʰ*, a conditioned deglottalization could also be assumed, thus from glottalic *“*-nʔd-*” to * “*-nd-*”,[21] whence TAB *-nt-*. Within Kloekhorst’s hypothesis that the PIE **d* was originally an unvoiced *[ʔt], the interpretation of the Tocharian behaviour of **d* could be essentially the same as just outlined for *“*ʔd*”.

The behaviour of PIE **d* in clusters can thus be understood in two different ways depending on the assumed original qualities of this consonant in Proto-Indo-European (either voiced as per the traditional reconstruction of Proto-Indo-European, or glottalized as per the Glottalic Theory). The standard word-initial reflex of **d* as TAB *tˢ*, with its palatalized counterpart TAB *śˌ* still requires further discussion. Reasoning backwards from TAB *tˢ*, an aspirated *[tʰ] is a good candidate for a direct precursor to TAB *tˢ* from a typological perspective (see Kümmel 2007: 173–175). How this would have arisen from PIE **d* is unclear, but a partially parallel development is found in Germanic. PIE **d* became PGm. **t*. This is reflected as aspirated [tʰ] in most Germanic languages, with a further development to an affricate [tˢ] in High German and Danish. However, the other, concomitant changes to the stop system found in Germanic are not paralleled in Tocharian, and a reconstruction as *[tʰ] in itself does not neatly match the other observable developments of PIE **d* in Tocharian: it is difficult to understand how such a sound would be lost in clusters with a **r*, **y* or **w*, or in intervocalic position.[22]

In the interpretations of the development of **d* just discussed, several intermediary stages have to be posited, summarized in Table 2.6. On the one hand, various seemingly disparate developments of **d* in Tocharian can be connected with the hypothesis of a single initial change: lenition. On the other hand, if a glottalic stop is taken as a precursor to a *[tʰ], such a reconstruction is more parsimonious. Preference for one scenario over the other will probably depend on the interpretation of stop qualities in the Proto-Indo-European from which Tocharian split, which cannot be reliably established on the basis of Tocharian itself due to the general lack of distinctions. The geminates of Kloekhorst’s

[21] In the glottalic framework, the *“*d*” corresponds to the **dʰ* in the traditional reconstruction.

[22] One could imagine an alternative lenited intermediate [θ] from [tʰ] rather than [ð], as the latter is based on the assumed characteristics of PIE **d* rather than on the Tocharian reflexes, so that a further alternative scenario could also involve a general reconstruction of this phoneme as **tʰ* with a conditioned lenition to *[θ] and then further to ∅ in clusters with **r*, **y*, **w*.

reconstructed system are just as absent from the Tocharian reflexes as the traditional voicing distinctions, and even an originally glottalic voiced *“*ʔd*” could have developed via a fricative like [ð] at some intermediary stage.

Table 2.6: Special developments of PIE **d* in Tocharian interpreted from the traditional reconstruction of Proto-Indo-European stops, within the Glottalic Theory, and based on Kloekhorst's model of Proto-Indo-Anatolian stops.

| Traditional | Glottalic | Kloekhorst |
|---|---|---|
| *dw* > *ð̯w* > *w* | *ʔdw* > *ʔw* > *w* | *ʔtw* > *ʔw* > *w* |
| *dy* > *ð̯y* > *y* | *ʔdy* > *ʔy* > *y* | *ʔty* > *ʔy* > *y* |
| *dr* > *ð̯r* > *r* | *ʔdr* > *ʔr* > *r* | *ʔtr* > *ʔr* > *r* |
| *nd* > *nt* | *nʔd* > *nd* > *nt* | *nʔt* > *nt* |
| *-d-* > *-ð̯-* > ∅ | *-ʔd-* > *-ʔ-* > ∅ | *-ʔt-* > *-ʔ-* > ∅ |
| *d-* > *ð̯* > (e.g.) *ϑ* > *tϑ* > *ts* | *ʔd-* > *ʔt* > *tʰ* > *tˢ* | *ʔt-* > *tʰ* > *tˢ* |

For the present purposes of establishing a workable relative chronology, the disappearance (or weakening) of PIE **d* in clusters with **r*, **w* and **y* can be placed before the assibilation of PIE **t* and **dʰ* before the glide **y*. The latter two both give TAB *tˢ* in combination with **y*, rather than disappearing like **d* (see subsection 2.3.4.2). We can interpret this as either (1) **t* and **dʰ* were from the outset more similar to one another in some way to the exclusion of **d*, so as to be affected by the glide in the same way, or (2) **d* had already disappeared in this environment or changed in some significant way so as to avoid participating in assibilation. The first explanation is difficult to substantiate without positing an otherwise unsupported change like the hypothetical development **dʰ* to **tʰ*, whereby original **t* and new **tʰ* would become united by the feature [-voice] (this is done by Ringe 1996 47–48). Such a development assumes aspiration as a feature of **dʰ* in pre-Proto-Tocharian, even though the evidence in favour of that is very slim (see the next subsection, 2.3.4.5). It is clear, however, that **d* was more liable to disappear than the other two dental stops, and that it changed in some way in several different positions. It is thus possible that **d* had already become *[ð] (or *[θ]) or *[ʔ] or even ∅ in clusters like **dy* before the assibilations of **ty* and **dʰy* took place. This change would therefore be an early one in the relative chronology.

The specific development of PIE **d* in Tocharian thus remains highly uncertain. In combination with the discussion of assibilation in the previous subsection, we can make a provisional notational difference between the two ancestors of TAB *tˢ* as pre-PT *“**ḋ*” on the one hand, from PIE **d*, and pre-PT *“**ċ/ż*ʰ” on the other, from assibilation of PIE **ty* and **dʰy*. The dot in *“**ḋ*” is intended to mark an indeterminate phonetic interpretation, similar to the dot in *“**ċ/ż*ʰ”. On balance, I think that this is a more appropriate representation than any symbol with a more definite phonetic meaning.

#### 2.3.4.5 A Tocharian version of Grassmann's Law

PIE **d* also plays an important role in discussion of another sound change, namely the Tocharian version of Grassmann's Law (Winter 1962: 24–25; Ringe 1996: 47; Hackstein 2017: 1324). Grassmann's Law is a well-known sound change from Greek and Sanskrit, whereby the first in a sequence of two successive aspirates **$C^hVC^h$* was de-aspirated due to dissimilation, yielding a sequence **$CVC^h$* instead. It is clear that two different versions of Grassmann's Law happened in Greek and Sanskrit, independently of one another, and it is thus in principle not particularly far-fetched to assume that a similar sound change happened in Tocharian at some point as well. If this type of sound change occurred in early Tocharian, it would imply that the Proto-Indo-European voiced aspirates like **$d^h$* were indeed phonetically aspirated in this branch as well. This is not obvious from any other Tocharian evidence, so that it cannot be confirmed in any other way than by reference to this Tocharian Grassmann's Law. The small set of possible examples all involve a development from PIE **$d^h$-* to TAB *$t^s$-*, as expected for PIE **d-*, rather than to TAB *t-*. De-aspiration by Grassmann's Law of, e.g., **$ǵ^h$* in a root like **"$ǵ^hed^h$-"* to **"$ǵed^h$-"* would not be visible in Tocharian, as both **$ǵ^h$* and **$ǵ$* regularly yield PT voiceless **k*.

16. PIE **$d^heg^{wh}$-* > **"$deg^{wh}$-"* > TB *tsək-*, TA *tsäk-* 'burn'
17. PIE **$d^heig^h$-* > **"$deig^h$-"* > TB *tsəyk-*, TA *tsāykā-* 'form tr.'
19. PIE ?**$d^heng^h$-* 'reach' > **"$deng^h$-"* > TB *tsənk-* 'rise'

There is also a counterexample against Grassmann's Law, namely PIE **$d^hub^hro$-* 'deep', which became TA *tpär*, TB *tapre* 'high' (20). With Grassmann's Law, the expected reflex would have been TA ***tspär*, TB ***tsapre*. Ringe does not cite this form as a counterexample, because he reconstructs it as **$d^hubro$-* instead, with a non-aspirated **b*. The reconstruction of this root was amended by Kroonen; according to him, a verbal derivative from the same root in Germanic, ON *deyfa* 'dip', from **$d^houb^h$-éie-*, while the adjective ON *djúpr* 'deep', otherwise thought to reflect PIE **b*, has a secondary *p* that arose from PIE **$b^h$* first through gemination to **pp* before a nasal in **$d^heub^h$-no-* (Kluge's Law), and was subsequently shortened after the long vowel, i.e. **$d^heub^h$-no-* > **deuppa-* > **deupa-* 'deep' (Kroonen 2011: 253).

The root for 'deep' is therefore not an example of PIE **b* as previously thought, and an exception to a Tocharian version of Grassmann's Law. On the basis of this, Peyrot casts doubt on Grassmann's Law in Tocharian, suggesting that the **i* might have played a role in the example **$d^heig^h$-* by causing assibilation in the zero-grade **$d^hig^h$-* (Peyrot 2022a: 92 fn. 10). Unfortunately, this development is not directly paralleled elsewhere either. It furthermore might require differing developments of **$d^hi$/ti* word-initially and word-finally, with a potential full assibilation to **-si* and eventual **-ṣ* taking place word-finally (see subsection 2.3.4.2). This is not impossible, but it should be borne in mind. A

development from *$d^he$ to *$t^sə$ in some way due to the *$e$ might seem attractive on the face of it based on of the examples above, but this encounters similar problems: TB *lac*, TA *läc* 'he/she went out' is derived from PIE *$h_1lud^het$ and show a development to TAB *c* instead, in line with palatalized *$t$ (see 2.4 on palatalization).

These issues make it practically impossible to properly decide what actually happened, but it seems clear that the exact phonetic properties of *$d^h$ and the other PIE voiced aspirates in early pre-Proto-Tocharian cannot be established based on such limited evidence (cf. Peyrot 2022a: 91–92). Only PT *$t^sək$- from *$d^heg^{wh}$- remains as a relatively secure example. If it does truly reflect an early Grassmann's Law that did not operate in PT *$təpre$ from *$d^hub^hro$, this change should have taken place very early. This is because it appears from other sound changes that PIE *$d$ changed its pronunciation in some way early on as well (cf. subsection 2.3.4.3), which could have untethered it from its connection to *$d^h$ as an unaspirated counterpart in a direct phonetic sense, at least within the traditional framework necessary for the operation of a Grassmann's Law. Any hypothetical Tocharian Grassmann's Law changing *$d^h$ into *$d$ by means of simple de-aspiration should therefore also have taken place at an early stage.[23]

### 2.3.4.6 Tocharian stop system developments and reconstruction

The general development of the Proto-Indo-European stops in Tocharian is a loss of distinctions with regards to features such as voicing, aspiration, glottalization or length (depending on the phonetics assumed for Proto-Indo-European). The labial and velar series each collapsed into a single stop, yielding /p k $k^w$/. However, PIE *$b^h$ developed differently from *$p$ in clusters with an *$m$ and disappeared completely. This likely happened via an intermediary stage *[mb] before stops were generally devoiced in Tocharian, although the precise situation preceding *[mb] remains unclear (see 2.3.4.1).

In the dental series, PIE *$d$ developed markedly differently from its counterparts *$t$ and *$d^h$, as discussed in some detail in 2.3.4.3 and 2.3.4.4. The disappearance of *$d$ in clusters with *$r$, *$w$ or *$y$ can be understood as a lenition (perhaps via *[ð]) and eventual elision, or as a debuccalization (to intermediary *[ʔ], with a glottalic interpretation of PIE *$d$). The intermediary steps that have to be assumed depend on the reconstruction of the original, Proto-Indo-European value of *$d$, and cannot be recovered directly from the Tocharian data. The development from PIE *$d$ to TAB $t^s$ may have happened via an aspirated *[$t^h$] in some way, but the possible path towards such a *[$t^h$] from PIE *$d$, once again, depends on the phonetic interpretation of PIE *$d$ itself. It is unattractive to simply assume *[$t^h$] was the value of *$d$ in all positions originally (without positing further

[23] Within a glottalic framework, the apparent change in *$d^heg^{wh}$- is more difficult to interpret, as it would entail an irregular secondary glottalization from *"$deg^w$-" to *"$d^ʔeg^w$-" (*"$tek^w$-" to ±*"$t^ʔek^w$-" in the framework of Kloekhorst's Indo-Anatolian stop system).

intermediate steps), as such a phonetic interpretation does not accord with its disappearance in clusters with *$r$, *$w$ and *$y$, or its merger with PIE *$t$ and *$d^h$ as TAB $t$ after the nasal *$n$.

Considering both the discussions surrounding the interpretation of the Proto-Indo-European stop system as a whole and in particular the hypothesis that Tocharian was the second branch to split off, a short note on the relevance of the Tocharian data in this perspective is in order. There are essentially three competing proposals, summarized in Table 2.7. If Tocharian is considered to be closer to either the traditional model or the "classic" glottalic model, there is no additional argument for the Tocharian second hypothesis. If Tocharian is considered to be more similar to Kloekhorst's interpretation of Proto-Anatolian and Proto-Indo-Anatolian, that might support the Tocharian second hypothesis and/or Kloekhorst's reconstruction of the Anatolian system for earlier nodes of Proto-Indo-European.

Table 2.7: Three competing interpretations of the Proto-Indo-European (or Proto-Indo-Anatolian) stops, with the corresponding Tocharian system on the right.

| Traditional | Glottalic | Kloekhorst | Tocharian |
|---|---|---|---|
| *$t$ *$d$ *$d^h$ | *$t$ *$ˀd$ *$d$ | *$t$: *$ˀt$ *$t$ | $t$ $t^s$ $t$ |

The traditional system requires a number of intermediate steps to get from the *$d$ to the Tocharian $t^s$ (see 2.3.4.4), and the traditionally assumed aspiration of *$d^h$ is not greatly supported within Tocharian, as it relies on the interpretation of a sound change with only two examples and one counterexample (see 2.3.4.5). Devoicing of *$d$ and both devoicing and de-aspiration of *$d^h$ are required to derive the Tocharian system from the traditional reconstruction of Proto-Indo-European.

The "classic" glottalic interpretation also requires a devoicing of the same two consonants, but they are instead interpreted as *"$ˀd$" and *"$d$" respectively. Some kind of shift from gottalization to affrication is also needed, perhaps with an intermediate step of aspiration (*[$t^h$]). Kloekhorst's model requires a degemination from *"$t$:" to Tocharian single $t$ and a shift from glottalization to affrication, again possibly via aspiration.

In trying to determine the preferable solution, the traditional model is, at least on paper, less economical than the others, unless a Tocharian version of Grassmann's Law is assumed, since that appears to require aspiration of the phoneme *$d^h$ (see 2.3.4.5). The "classic" glottalic model and Kloekhorst's model are on equal footing, since both require two systematic changes: devoicing and affrication in the "classic" glottalic model, and degemination and affrication in Kloekhorst's model. The position of Tocharian closer to either the Anatolian system as reconstructed by Kloekhorst or to the other Indo-

European languages therefore cannot be determined on the basis of the Tocharian stop system without additional evidence.

### 2.3.5 Summary

Before moving on through Tocharian historical phonology, it may be beneficial to briefly review the changes discussed so far. For reference, the Proto-Indo-European consonant inventory is reproduced again below. This is the system that the earliest stage of pre-Proto-Tocharian inherited, and the starting point for all subsequent developments.

Table 2.8: The Proto-Indo-European consonant inventory.

| PIE | labial | alveolar | palatal | palatovelar | velar | labiovelar | "laryngeal" |
|---|---|---|---|---|---|---|---|
| stop | *p (b) bʰ* | *t d dʰ* | | *ḱ ǵ ǵʰ* | *k g gʰ* | *kʷ gʷ gʷʰ* | |
| | | | | | | | *h₁ h₂ h₃* |
| sibilant | | *s* | | | | | |
| nasal | *m* | *n* | | | | | |
| liquid | | *l r* | | | | | |
| glide | *u̯* | | *i̯* | | | | |

In the preceding subsections, we have seen some significant changes to this system already. As in most Indo-European languages, the laryngeals disappeared from the consonant system to become integrated into the vowel system instead, creating the additional vowel phonemes **a* (from **H̥* and **h₂e*) and **ā* (from **eh₂*), and generally increasing the frequency of long vowels like **ē* (from **eh₁*) and probably also **ō* (from **oH*) in the language. Perhaps consonantal laryngeals **H* were retained in some positions until a later stage, but after vowel colouring it seems that they cannot be differentiated anymore. Syllabic resonants were vocalized and added another vowel to the system, namely **ə*, which probably became phonemic at the latest when consonantal laryngeals disappeared. The impact of these changes on the vowel system of early pre-Proto-Tocharian is illustrated in Table 2.9 below.

Table 2.9: The Proto-Indo-European vowel inventory with initial developments to the early stages of pre-Proto-Tocharian

| PIE | front | back | > pre-PT | front | central | back |
|---|---|---|---|---|---|---|
| high | *i* | *u* | | *i ī* | *ə* | *u (ū)* |
| mid | *e ē* | *o ō* | | *e ē* | | *o ō* |
| low | | | | | *a ā* | |

In the consonant system, the palatovelars as a group became plain velars, and many labiovelars also became plain velars. In contact with another consonant (except perhaps in intervocalic *-*kʷl*-) labiovelars also seems to have been delabialized. Intervocalic **Kṷo* developed to PT **kʷe*, preserving the (originally separate) labial element. Thus, while the palatovelars were eliminated from the consonant system, and the labiovelars suffered some casualties, the latter did survive in a more limited distribution and remained part of the consonant system.

With the assibilation of **t* and **dʰ* before **y* and **i*, an additional consonant phoneme (or phonemes) came into existence, provisionally written as *"*ċ*" (and perhaps also *"*żʰ*"). The assibilation likely happened after PIE **d* in clusters had changed in some way to lose the stop element of its original pronunciation, potentially becoming a fricative like *[ð] in all positions, or by debuccalization to *[ʔ] clusters with a resonant. I use the provisional symbol *"*ḋ*" to mark that some change had occurred to original PIE **d*. The consonants provisionally written as *"*ċ*/*żʰ*" and *"*ḋ*" merged at some unknown point in time, so that their reconstruction as separate entities is hypothetical at this stage of the discussion. Some justification will arise in the discussion of palatalization below (2.4.1). Finally, at some point in time, all (remaining) oppositions of voicing and/or aspiration became neutralized, so that the system ended up containing only voiceless stops. The resulting consonant system is given in the table below. The rewriting of PIE **i̯* and **u̯* as **y* and **w* is merely a notational change. Several word-final consonants and consonant clusters disappeared, but this did not affect the overall system, which is presented here in Table 2.10.

Table 2.10: The pre-Proto-Tocharian consonant inventory after loss of stop contrasts and special developments of PIE **ti̯*/*dʰi̯* and **d*.

| pre-PT | labial | dental/alveolar | palatal | velar | labiovelar | "laryngeal" |
|---|---|---|---|---|---|---|
| stop | *p* | *t* | | *k* | *kʷ* | |
| | | "*ḋ*" | | | | *H* (?) |
| affricate | | "*ċ*" | ~ "*ċ*" | | | |
| sibilant | | *s* | | | | |
| nasal | *m* | *n* | | | | |
| liquid | | *l r* | | | | |
| glide | *w* | | *y* | | | |

This consonant system represents the minimal system that must have existed at the point in time that the next great change occurred: palatalization. Palatalization further altered the profile of the Tocharian phonological system, and it is the main topic of the next section.

## 2.4 Palatalization

A pervasive opposition between palatal and non-palatal consonants is one of the most salient features of synchronic Tocharian phonology. It characterizes the general consonant system and moreover plays an important role in the morphophonology of the language (cf. Krause & Thomas 1960: 61–64). Almost every consonant /C/ had a palatal(ized) counterpart /C$^{y}$/ in Proto-Tocharian and in both Tocharian A and B. It is important for our view on Tocharian relative chronology to understand how palatalization came about, since it was a pivotal development affecting both the consonant and the vowel systems. In the following subsections we will discuss the process of palatalization and the ways in which it transformed the phonological system of pre-Proto-Tocharian.

### 2.4.1 Palatalization as it affected the consonant system

Before palatalization took place, the pre-Proto-Tocharian consonant system was probably much like in Table 2.11 below (also given in 2.3.5). Since there are no developments with regards to palatalization that depend on voicing or aspiration, the system can be reconstructed without those distinctions. On this point I differ from Ringe's relative chronology, in which general devoicing and de-aspiration is one of the latest changes (Ringe 1996: 152–154). It has to be admitted that there is no way to be certain (cf. also the considerations in section 2.6), but this is the minimal and most economical system that is required going into palatalization.

Table 2.11: The early pre-Proto-Tocharian consonant inventory; repeat of Table 2.10.

| pre-PT | labial | dental/alveolar | palatal | velar | labiovelar | "laryngeal" |
|---|---|---|---|---|---|---|
| stop | *p* | *t* | | *k* | *k*$^{w}$ | |
| | | "*ḍ*" | | | | *H*? |
| affricate | | "*ć*" | | | | |
| sibilant | | *s* | | | | |
| nasal | *m* | *n* | | | | |
| liquid | | *l r* | | | | |
| glide | *w* | | *y* | | | |

As a process, palatalization can be understood as first a coarticulation and then transference of frontness (or palatality) from a vowel to a neighbouring (in this case, preceding) consonant. Vowel mergers made the consonantal palatality phonemic. For example, an original opposition between a sequence /Ce/ with a front vowel /e/ and a sequence /Cë/ with a mid or back vowel /ë/ can be transformed into an opposition

between the sequences /C$^{y}$e/ and /Ce/ instead, with the same phonological vowel, but a contrast in palatality of the consonants. In this way, front-back distinctions in the vowel system were transferred to the consonants, and the consonant system almost doubled in size as a consequence.

Only **r* and **y* did not have a palatal counterpart, the latter for obvious reasons. The absence of **r$^{y}$* is also common cross-linguistically (Hall 2000). Perhaps pre-PT **ċ*, too, was not palatalized, as we have no reliable examples of original sequences **ty* and **d$^{h}$y* in environments with a palatalizing front vowel. The only potential instance of this is in the infinitive ending TAB *-t$^{s}$i*, perhaps from **-tyey* to **-ċey* (assibilation) to **-ċ$^{(y)}$əy* (palatalization?) and to PT **-t$^{s}$əy*. This does not show any reflex of a **ċ$^{y}$* as distinct from the development of **ċ* to TAB *t$^{s}$*. The lack of a distinct palatalized counterpart would make sense if **ċ* was an inherently palatal affricate like *[tɕ] at the time. However, as the infinitive suffix **-d$^{h}$yey* would be the only available example of a palatalized assibilated dental in any case, caution is advisable. See Ringe (1996: 79) for a discussion on the reconstruction of the infinitive suffix.

An overview of the pre-Proto-Tocharian system is given in Table 2.12 below. Palatalization of **k$^{w}$* is not distinct from a palatalized **k*, so any potential **k$^{wy}$* that may have existed at some point had merged with **k$^{y}$*, or **k$^{w}$* had already lost its labialization in palatalizing environments early on (cf. 2.3.3.2); **k$^{wy}$* is not included in Table 2.12. In combination with some of the changes discussed in subsection 2.3.3, this further reduced the number of labiovelars in pre-Proto-Tocharian.

Table 2.12: The pre-Proto-Tocharian consonant inventory after phonemic palatalization.

| pre-PT | labial | dental/alveolar | palatal | velar | labiovelar |
|---|---|---|---|---|---|
| stop | *p p$^{y}$* | *t t$^{y}$* | | *k k$^{y}$* | *k$^{w}$* |
| | | *“ḋ ḋ$^{y}$”* | | | |
| affricate | | *“ċ”* | | | |
| sibilant | | *s s$^{y}$* | | | |
| nasal | *m m$^{y}$* | *n n$^{y}$* | | | |
| liquid | | *l l$^{y}$* | | | |
| trill | | *r* | | | |
| glide | *w w$^{y}$* | | *y* | | |

From the system obtained after palatalization, the changes to Proto-Tocharian, and also onward to Tocharian A and B, are relatively minor. In general, we see that many palatalized consonants developed into true palatal consonants, although the palatalized sibilant **s$^{y}$* actually became retracted to a retroflex *[ʂ]. At some point, word-initial **“ḋ”*

(see the discussion in 2.3.4.4) merged with *"*ċ*" to yield PT *$t^s$, while the palatalized counterpart *"*ḋ$^y$*" became */ɕ/ and merged with palatalized *$k^y$.

Table 2.13: The developments of the pre-Proto-Tocharian consonant inventory to Proto-Tocharian.

| pre-PT | labial | dental | retroflex | palatal(ized) | velar | labiovelar |
|---|---|---|---|---|---|---|
| stop | *p p$^y$* | *t* | | *t$^y$* | *k* | *k$^w$* |
| affricate | | *t$^s$* < *ċ, ḋ* | | | | |
| sibilant | | *s* | *ʂ* < *s$^y$* | *ɕ* < *k$^y$, ḋ$^y$* | | |
| nasal | *m m$^y$* | *n* | | *n$^y$* | | |
| lateral | | *l* | | *l$^y$* | | |
| trill | | *r* | | | | |
| glide | *w w$^y$* | | | *y* | | |

Finally, with some minor adjustments, the Tocharian A and B systems can be obtained. The palatalized labial consonants were lost as distinctive segments, with *$w^y$ developing into TA *w* and TB *y*. The result of *$k^y$ may have been an affricate still at the Proto-Tocharian stage (Adams 2000: 19–20; Peyrot 2008: 78), but in the attested languages it had become a non-affricate palatal sibilant TAB *ś*. The shared consonant system of both Tocharian A and B is given again in the table below, this time represented in the (transcribed) spelling of both languages.

Table 2.14: The synchronic consonant inventories of Tocharian A and B.

| TAB | labial | dental/alveolar | retroflex | palatal | velar | labiovelar |
|---|---|---|---|---|---|---|
| stop | *p* | *t* | | | *k* | *kw*~*k$_u$* |
| affricate | | *ts* | | *c* [cɕ] | | |
| sibilant | | *s* | *ṣ* [ʂ] | *ś* [ɕ] | | |
| nasal | *m* | *n* | | *ñ* [ɲ] | (*ṅ* [ŋ]) | |
| lateral | | *l* | | *ly* [ʎ] | | |
| trill | | *r* | | | | |
| glide | *w* | | | *y* | | |

Until now the discussion of palatalization has centred around changes in the consonants. In the next subsections, the connected vowel developments will be considered in more detail, with a special focus on the developments of PIE **i*, **e* and **u*.

### 2.4.2 Developments of PIE **i* and **e*

To understand palatalization in Tocharian, it is important to understand the developments of PIE **i* and especially **e*. These two short front vowels are both reflected in Tocharian as PT **ə*, generally with palatalization on the preceding consonant. In word-initial position, they developed an onglide, becoming PT **yə-*.

27. PIE **gʷenh₂-* > PT **kʸəna* > TB *śana*, TA *śäṃ* 'wife'
34. PIE **h₁itr-* → ?**h₁itōr* > PT **yətar* > TB der. *ytārye*, TA *ytār* 'road, path'
35. PIE **h₁ekʷo-* > PT **yəkwe* > TB *yakwe*, TA *yuk* 'horse'
65. PIE **kʷetʷores* > PT **kʸətwerə* > TB *śtwer*, TA *śtwar* 'four (4)
73. PIE **limn̥-* > PT **lʸəmə* > TB *lyam*, TA *lyäm* 'lake'
81. PIE **-nti* > PT **-nʸtʸə* > TA *-ñc* prs.act.3pl. verbal ending
114. PIE **ueǵʰno-* > PT **wʸəkne* > TB *yakne*, TA *wkäṃ* 'way, manner'

Since the reflexes of PIE **i* and **e* are mostly the same, it is difficult to tell when exactly the two merged. Their behaviour is not entirely identical, however, so that some change must have occurred before they both became PT palatalizing **ə*. The difference is that PIE **i* appears to have had a lower rate of success in palatalizing preceding consonants. This has led to the view that **i* was not as capable of causing palatalization as **e*; e.g., Pinault (2008: 423) writes that it did not palatalize labials, velars, labiovelars and **s*.

It did, however, clearly cause palatalization in some instances (cf. PIE **limn̥-* and **-nti* in the list above), and it would be expected to palatalize more than **e*. In the linguistic typology of palatalization, the high front vowel ±[i] normally has more of a palatalizing effect than a lower front vowel ±[e], with an implicational hierarchy about the vowels that will cause palatalization. According to Bateman, "if lower front vowels trigger palatalization, so do higher front vowels" (Bateman 2011: 597) and "[i] is the best palatalization trigger in in any language, regardless of what the other triggers are" (Bateman 2007: 64). It may therefore be more natural to assume that **i* had been changed into a different vowel in those cases where it failed to palatalize. This is generally the assumption of sequences involving **i* and a labial glide or a labiovelar (see Ringe 1996: 66; Hackstein 2017: 1326).

14. PIE **duito-* > "*wuto*" > PT **wəte* > TB *wate*, TA *wät* 'second'
67. PIE **kʷi-so* > "*kuso*" > PT **kʷəse* > TB *kᵤse*, TA *kus* 'who'
107. PIE **uiso-* > "*wuso*" > PT **wəse* > TB *wase*, TA *wäs* 'poison'
123. PIE **suidie-* > "*suye-*" > PT **səya* > TB *səya-* 'sweat'

After the changes affecting **kʷi* and *wi*, **e* and **i* could theoretically be merged as palatalizing **i*, although this is not strictly necessary (cf. 4.3.1). Assumed long **ī*

developed differently and was able to palatalize a preceding *w, as seen in PIE *du̯idḱm̥t > *du̯ih₁ḱm̥t > *wīkəmt > PT *wʸəykən 'twenty' (13).

Other instances of lacking palatalization before original *i have an intervening laryngeal.[24] The chief example is TB *skiyo* 'shadow', derived from PIE *skHieh₂ (Ringe 1996: 18–19; Del Tomba 2020: 144–145), but there are a few other potential instances as well (see the discussion by Friis 2024: 372–373).

22. PIE *gnh₁-ih₁- > PT *kənəy-tər > TA opt.mid.3sg. *knitär* rather than **kñitär
96. PIE *sh₂i-nu- > PT *səyn- > TB *səyn-* 'satiate' rather than **ṣəyn-
99. PIE *skHieh₂ > PT *skəyo > TB *skiyo* 'shadow' rather than **ṣiyo

Ringe suggests that the laryngeals may have become *ə rather than the usual *a before a vocalic *i, which would explain these examples (Ringe 1996: 19). In this way, a form like *sh₂i-nu would regularly have yielded *səynə-, whence TB *səyn-*. Since *ə does not palatalize, as seen from the reflexes of syllabic resonants (2.3.2), this explanation would account for the lack of palatalization in the examples with apparent non-palatalizing *i without the need for an unnatural development whereby *e has more palatalizing potential than the higher vowel *i.

### 2.4.3 Developments of PIE *u

In Proto-Tocharian, PIE *u yielded PT *ə, just like PIE *e and *i discussed in the previous subsection. However, there are a few marked differences in the development of *u as opposed to the other two short high vowels. Most important is a consistent lack of palatalization before PT *ə from original *u. Instead, it may have caused labialization of a preceding velar to *kʷ, although there are only a few examples, none of which are certain. Such a development could constitute an additional source of PT *kʷ, besides those labiovelars that were preserved from Proto-Indo-European (cf. 2.4.3). A labial onglide developed from *u- in word-initial position, yielding *wə-, similar to how initial *i- and *e- became *yə-.

- ? PIE *g⁽ʰ⁾ul- > PT *kʷəla- > TB *kʷəla-*, TA *kʷälā-* 'fail, weaken, recede'
- PIE *udr-io- > PT *wəriye > TB der. *wriyeṣṣe* 'pertaining to water, TA *wri** 'watery'
- PIE *h₂eḱu- > PT der. *akʷətˢe > TB *akwatse* /akʷə́tˢe/ 'sharp'

[24] Pinault (2008: 241) has argued that PIE *si did not undergo palatalization either, and he is followed in this by Hackstein (2017: 1312). The Proto-Indo-European reconstructions and etymologies of these examples are not compelling due to their imperfect semantic derivations (proposed PIE *luksi- 'white, shining' > TB *laks* 'fish' and *h₃esilo- 'ash-tree' > TB *esale* , TA *asäl* 'post').

Labialization of velars due to PIE *u has been posited in preconsonantal position as well. This would provide a straightforward way to account for TB *lak$_u$tse* 'bright, light' from an earlier **luk-* + *-tio-* (Pinault 2008: 426, see also Kim 1999: 164). The difficulty is that there may have been analogical interference from the full-grade of the root, **leuk-*, which regularly became PT **l$^y$uk-*. Additionally, TB *lak$_u$tse* could also be derived from a secondary *u*-stem quasi-PIE **luk-u-ti̯o*, in which case the labiovelar should result from the syncopated sequence **ku* via **k$^w$ə* as in TB *laṅk$_u$tse* from quasi-PIE **h$_1$l̥ng$^{wh}$u-ti̯o*. Furthermore, no labiovelar appeared in PIE **leug-* → **lug-ro-*, which is posited as the ancestor of TB *lakle* 'sorrow' with assimilation of the **r* to **l* (Kim 1999: 164; perhaps it could rather be a different formation in **-lo* from the same root, i.e., **lug-lo-*). There is no labiovelar reflex, i.e., no ***lak$_u$le* (or ***lkwale*).

It is clear that intervocalic velars were not labialized by a preceding **u*, chiefly on account of TB *tkācer*, TA *ckācar* 'daughter' from PIE **d$^h$ugh$_2$tēr* via pre-PT **tukat$^y$er*. If labialization had occurred, we should find TB ***tkwācer* instead. The labiovelar in TB *sakw*, TA *suk* 'happiness' from Skt. *sukha-* 'happiness', the only example of **u* causing labialization of a following intervocalic velar, may have arisen from loan adaptation strategies rather than from genuine sound law. I think that the evidence of 'daughter' outweighs it (*pace* Kim 1999: 162 incl. fn. 53). Pinault (2008: 426) takes *lak$_u$tse* and TA *lukśanu* as from an unattested root noun **lək$^w$* (TB **lakw*, TA **luk*), but since the same type of phonological change would be involved, just in word-final position, this does not make much of a practical difference.[25]

21. PIE **d$^h$ugh$_2$tēr* > PT **təkat$^y$er* > TB *tkācer*, TA *ckācar* 'daughter'
71. PIE **leuk-* 'shine' → ?**luktio* > PT **lək$^w$tse* > TB *lak$_u$tse* 'bright, light'
74. PIE ?**luglo-* > PT **ləkle* > TB *lakle* 'sorrow'

Since PIE **u* merged with **i* and **e*, and with the **ə* that arose from the vocalization of syllabic resonants, it is not immediately obvious up until which point all of these vowels continued to exist as separate entities. However, PIE **u* did take part in various umlaut phenomena that happened after palatalization (see subsections 2.6.2 and 2.6.3), which means that the existence of PIE **u* as such is assured until a relatively late stage. It therefore needs to be kept separate from **i*, **e*, and **ə* in our interpretation of the phonological system as palatalization took place (cf. subsection 2.5.1).

[25] The development of PIE **uksō-* 'ox' is often cited in this context, since the initial *o-* in TB *okso* has been ascribed to earlier sequence **wək$^w$so*, where the **u* labialized the following velar (Ringe 1996: 127; Kim 1999: 164–165 incl. fn. 54). In my opinion, it is more likely that the change to *o-* was brought about by *o*-umlaut, as discussed in subsection 2.5.2.

### 2.4.4 Palatalization as it affected the vowel system

As the general Tocharian palatalization created new oppositions between neutral and palatalized consonant phonemes, some oppositions once inherent to the vowel system were lost. To reach the system directly preceding palatalization, it is preferable to reconstruct bottom-up. This is the same approach I took in Warries (2022). Although there are some minor differences in the current presentation, the conclusions remain the same.

#### 2.4.4.1 Syllable types and palatalization

We will start by taking a look at the types of syllables that were found in Proto-Tocharian. Each Proto-Tocharian reconstruction is followed by its Proto-Indo-European source in parentheses. Palatalization is marked in a uniform way with a superscript $^{y}$.

- */C$^{y}$i/ : PT **w$^{y}$ikən* 'twenty' (13. **dm̥ih₁ḱm̥t*), PT **w$^{y}$ike* 'place' (116. **u̯eiḱos*)
- */C$^{y}$e/ : PT **w$^{y}$ente* 'wind' (53. **h₂u̯eh₁nto-*), PT **m$^{y}$en$^{y}$e* 'moon' (77. **meh₁nē-*)
- */Ce/ : PT **keme* 'tooth' (24. **ǵombʰos*), PT **y$^{y}$ək$^{w}$e* 'horse' (35. **h₁eḱu̯os*)
- */Ca/ : PT **pat$^{y}$er* 'father' (91. **ph₂tēr*)
- */Co/ : PT **prot$^{y}$er* 'bother' (2. **bʰreh₂tēr*), PT **wəlo* 'king' (118. **uḷHōnts*)
- */C$^{y}$u/ : PT **l$^{y}$uke-* 'light up' (72. **leuko-*)
- */Cu/ : PT **ku* 'dog' (61. **ḱu̯ō*), PT **wu* '2' (15. **du̯oh₁*)
- */C$^{y}$ə/ : PT **t$^{y}$ək$^{w}$e* 'river' (108. **tek$^{w}$os*), PT **l$^{y}$əmə* 'lake' (73. **limn̥*)
- */Cə/ : PT **kənte* '100' (60. **ḱm̥tom*), PT **təpre* 'high' (20. **dʰubʰros*)

The syllable types */Ca/ and */Co/ did not originally have palatalized counterparts */C$^{y}$a/ and */C$^{y}$o/, as these were only created later due to umlaut (see 2.5.3 and 2.5.5). This means that */Ca/ and */Co/ can be projected to a stage before palatalization simply as such. It should be noted, however, that the situation surrounding these two Proto-Tocharian vowels is not at all straightforward. They will be dealt with in more detail in the next subsection below (2.4.5). PT */C$^{y}$i/ did not have a non-palatal counterpart and is therefore easy to reconstruct to a stage before palatalization took place. The difference between palatalized */C$^{y}$e/ and non-palatalized */Ce/ can be reconstructed as an earlier contrast between */Ce/ and */Cë/ respectively. Based on just internal reconstruction */C$^{y}$u/ and */Cu/ could be rendered as */Cü/ and */Cu/, but we will return to these vowels soon. The schwa */ə/ can also be reconstructed with a front (palatalizing) and back (non-palatalizing) variant, thus "front */Cə/" and "back */Cə/" respectively.

This is generally in line with the known different Proto-Indo-European sources of these syllable types. For example, the distinction between */Ce/ and */Cë/ reflects PIE **Cē* and **Co*, with front and back vowels respectively. The precursors to "front */ə/" and

"back */ə/" are more varied. The palatalizing "front */ə/" derives from PIE **e* and, partially, **i* (cf. subsection 2.5.2), while the non-palatalizing "back */ə/" emerged from original syllabic resonants **R̥* (cf. subsection 2.4.2) as well as from PIE **u* (cf. subsection 2.5.3). These latter two can sometimes be distinguished in Tocharian, since the sequences of the type PIE **Ku-* became PT **kʷə-* with labialization, while for instance **km̥-* developed into PT **kəm-*. This indicates that there used to be at least a three-way contrast, between a palatalizing short vowel*/i/ (from PIE **i* and **e*), a labializing short vowel */u/ (from PIE **u*) and a neutral short vowel */ə/ (originating mainly from the syllabic resonants).

On the basis of the umlaut phenomena involving PIE **u* (2.5.1, 2.5.2), we can conclude that at least */u/ did not merge with */i/ and */ə/ when palatalization first took place. This means that palatalization involved just the merger of the vowels */i/ and */ə/, with the former causing palatalization. */u/ remained as a separate vowel. This in turn indicates that */ə/ can be understood as a back counterpart to */i/, most naturally to be interpreted as ±*[ɨ].[26] The development outlined in Table 2.15 reflects this more detailed understanding of the mergers between the three high short vowels into PT **ə* and their relationships with palatalization.

Table 2.15: Updated intermediary overview of the origins of Proto-Tocharian schwa.

| PT | < | < | < | PIE |
|---|---|---|---|---|
| Cʸə | Cʸə | Cʸi | Ci | Ci, Ce |
| Cə | Cə | Ci | "Cɨ" | CR̥ |
| Cə (kʷə) | Cə (kʷə) | Cu | Cu | Cu |

These short *i*-vowels in Table 2.15 are not to be confused with the vowel of the syllable type */Cʸi/ in Proto-Tocharian (examples like **wʸike* 'place', cf. above), which mainly derives from the diphthong **ei*. For that sound, a reconstruction of pre-PT */ī/ or */iy/ is possible, differentiating it from the short **i* until that was reduced to **ə*. Similar to PIE **ei* yielding PT palatalizing **i*, PT */Cʸu/, which we provisionally reconstructed as */Cü/ above, goes back to PIE **eu*. A reconstruction of the type */iw/ may thus be adequate before palatalization took place. PT */Cʸu/ existed next to PT */Cu/, which may be rewritten as either */Cū/ or */Cuw/ at the pre-Proto-Tocharian stage before palatalization, parallel to the other high vowel PT */Cʸi/ as pre-PT */Cī/ or */Ciy/.

[26] A phonetically high realization of the **ə* phoneme as ±*[ɨ] instead of a "normal" undefined vowel ±*[ə] could be due to a push from the mid unrounded vowel **ë* that resulted from PIE **o*.

#### 2.4.4.2 The changing system

Based on these considerations, the vowel system right before palatalization occurred must have contained at least the following vowel phonemes: */i e a o u ə ë/. PIE **ei* and **eu* can at this stage be reconstructed as */iy/ and */iw/; the distinction between PIE **i* and **e* is not maintained in Tocharian, so that they may have merged as **i* before palatalization took place; some considerations on the timing of this merger were given above in subsection 2.4.2. With the phonemicization of palatalization, the contrasts between */Ci/ and */Cə/ and between */Ce/ and */Cë/ seem to have been given up as their distinctions were transferred to the preceding consonant. The process is illustrated in the table below with various stages represented.

Table 2.16: An overview of the vowel developments from Proto-Indo-European to Proto-Tocharian with various intermediary stages.

| PT | palatalized | pre-PT | post-PIE | PIE |
|---|---|---|---|---|
| C$^{y}$i | C$^{y}$iy | Ciy/Cī | Cei, Cī | Cei, Cih$_1$? |
| C$^{y}$e | C$^{y}$e | Ce | Cē | Cē, Ceh$_1$ |
| Ce | Ce | Cë | Co | Co, Ch$_3$e? |
| Ca | Ca | Ca | Ca | C$\underset{\circ}{\text{H}}$, Ch$_2$e |
| Co, Ca? | Co | Co | Cō | Cō, CoH |
| Co (?) | Cå | Cå | Cā | Ceh$_2$ |
| C$^{y}$u | C$^{y}$iw/C$^{y}$uw | Ciw | Ceu | Ceu |
| Cu | Cuw | Cuw/Cū | Cū | Cō#, Cuh$_1$? |
| C$^{y}$ə | C$^{y}$i | Ci | Ci, Ce | Ci, Ce |
| Cə | Ci | "Cɨ" | "Cɨ" | C$\underset{\circ}{\text{R}}$ |
| Cə (k$^{w}$ə) | Cu | Cu | Cu | Cu |

Further changes in the vowel system, leading to the creation of PT */C$^{y}$o/ and */C$^{y}$a/, were brought about by a number of umlaut phenomena. These will be discussed in the following section (2.5). Some developments involving PIE **ō* and **eh$_2$/ā* will receive more attention first, as that is a very complex aspect of Tocharian historical phonology relevant for our understanding of the relative chronology of vowel changes.

### 2.4.5 Developments of PIE **ō* and **eh$_2$/ā*

From a careful study of the table at the end of the previous subsection it will be clear that the development of PIE **ō* is split. The PIE syllable type **Cō* appears in the rows for PT */Co/, */Ca/ and */Cu/. The development of this vowel is a great problem in the relative chronology of Tocharian sound changes and it affects our interpretation of the loss of vowel length. The relation between the developments of **ō* and PIE **eh$_2$/ā* in the vowel

system especially difficult, since *$eh_2/\bar{a}$ yields PT *$o$ as well. The present subsection will further address this situation.

#### 2.4.5.1 Three reflexes of PIE *$\bar{o}$

The vowel phoneme PIE *$\bar{o}$ is considered to be the antecedent of three separate vowels PT *$u$, *$o$ and *$a$. At least the reflex PT */Cu/ can be dealt with quite easily, since it concerns a development that was restricted to word-final position. It is to be understood as a relatively early development, preceding the loss of several final consonants and clusters, and it is usually reconstructed as a change from PIE *-$\bar{o}$# to post-PIE *-$\bar{u}$#. This can be seen in the examples listed below.[27]

15. PIE *$d\underset{\text{̯}}{u}oh_1$ 'two (2)' > *$dw\bar{o}$ > *$dw\bar{u}$ > PT *$wu$ > TAB $wu$ 'two (2)'
61. PIE *$\acute{k}\underset{\text{̯}}{u}\bar{o}$ 'dog' > *$kw\bar{o}$ > *$kw\bar{u}$ > PT *$k^{(w)}u$ 'dog' > TAB $ku$ 'dog'
50. PIE *$\underset{\text{̥}}{h_2}ntb^hoh_1$ 'both' > *$antb^h\bar{o}$ > *$antb^h\bar{u}$ > PT *$ampu$- > TA $\bar{a}mpuk$ 'both'
84. PIE *$o\acute{k}t\bar{o}$ 'eight (8)' > *$okt\bar{o}$ > *$okt\bar{u}$ + $k\partial mt$ 'ten' > PT *$oktu$- > TA der. $oktuk$ 'eighty'

In some words, PIE *$\bar{o}$ has a Proto-Tocharian word-final reflex *-$o$ as well, without raising to *-$u$. This occurs in ($m\bar{o}$)$n$-stems and $nt$-stems, indicating that *-$\bar{o}$ was protected by word-final consonants at the time that *-$\bar{o}$# became *-$\bar{u}$#. When the final consonants had disappeared due to apocope, the new final *-$\bar{o}$# did not change much further, only losing its length to become PT *-$o$ (Ringe 1996: 11). The main examples are the following:

59. PIE *$\acute{k}leum\bar{o}n$ > PT *$kl^yomo$ > TB $klyomo$, TA $klyom$ 'noble'
124. PIE *$uks\bar{o}n$- > PT *$okso$ > TB $okso$, TA pl. $opsi$ 'ox'
118. PIE *$\underset{\text{̯}}{u}\underset{\text{̥}}{l}Hm\bar{o}nts$ > PT *$w\partial lo$ > TB $walo$, TA $w\ddot{a}l$ 'king'

These words clearly show a development from PIE *$\bar{o}$ to PT *$o$. A number of other examples are considered to suggest a development from PIE *$\bar{o}$ to PT *$a$, however. The reconstructions of most of these are problematic in some way. The precise origin of final -$ar$ is difficult and could be analogical in PT *$y\partial tar$ 'road' (Kortlandt 1988b: 84–85). The derivation of PT *$w^y\partial sar$ 'wheat' from Proto-Indo-European is uncertain.

[27] It is also often considered to have affected the PIE thematic ending of the first person singular *-$oH$ > *-$\bar{o}$, which may be the origin of the Proto-Tocharian first-person ending *-$u$/-$w$ (Ringe 1996: 89). The issue with this reconstruction is that it turns up in verbal classes where we would expect the athematic ending *-$m$ instead, so that this etymology is disputed (but see Malzahn 2010: 39–41). However, alternative explanations for PT *-$u$/-$w$ usually involve an irregular, unparalleled change from *-$m$, so that they in turn fail to be particularly convincing.

37. PIE *$h_1i$-tr- → ?*$h_1$itōr > PT *yətar* 'road'
120. ? PIE *ues-r → ?*u̯esōro- > PT *$w^y$əsare* > TB *ysāre*, TA *wsār* 'wheat'

An additional potential example is the preterite 3sg. of TB *yok-* 'drink', namely TB *yāś*, which may go back via *$yōg^{wh}$- to a perfect *$h_1e$-$h_1og^{wh}$-*e* (K.T. Schmidt 1997: 261; Peyrot 2013: 449). Clearly, this example would be of a different order, as the *$ō$ here should derive from a contraction without any exact parallels. Some very doubtful examples of putative PIE *$ō$ to PT **a* include the following:

23. PT **aknat$^s$a* 'foolish' from PIE *$n̥ǵneh_3nti̯h_2$ (the first and last nasals were lost due to dissimilations) may rather go back to *$n̥ǵnh_3nti̯h_2$ with a zero-grade and a development to PT **a* from a vocalized laryngeal (Friis 2021);
23. The corresponding verbal stem PT **kna-* 'know' from *$ǵneh_3$- 'know' could likewise reflect a zero-grade with a vocalized laryngeal, i.e., *$ǵnh_3$-;
30. PT **kramər* 'heaviness, weight' from PIE "*$g^wroh_2$-$mr̥$" could also reflect a zero-grade *$g^wrh_2$-$mr̥$ with vocalization of the laryngeal to **a*. This formation found in Tocharian is not paralleled in any other branch, so that its precise structure is difficult to determine, and a development via *$ō$ from *$oh_2$ is by no means assured.
54. PT **anse* 'shoulder' from PIE "**ōmsos*", with a long vowel based on the Greek reflex ὦμος 'id.'. The original form may have been PIE **$h_2$emsos* instead, with PT **a* from PIE **$h_2e$* in Tocharian (Beekes 2010: 1680; Kloekhorst 2008: 178), or perhaps **$h_3$emsos* (so reconstructed by Kroonen 2013: 25) if **$h_3e$* could become PT **a* like **$h_2e$* (cf. 2.3.1);
94. PT **puwar* 'fire' from PIE "**peh$_2$uōr*" via **ph$_2$uōr* may reflects PIE **puh$_2$r* instead (Winter 1965a: 192–193);
· PT **tano* 'grain' from putative PIE "**$d^h$oHneh$_2$*" should rather be regarded as a borrowing from Iranian (Peyrot 2018b: 257–259), and is thus uninformative.

Due to the great uncertainty surrounding these examples supposedly showing the change from PIE *$ō$ to PT **a*, I remain (see Warries 2022) sceptical about this development. There is a clear continuation of PIE *$ō$ as PT **o* in examples like PT **walo* and PT **kl$^y$omo*. The change from PIE *$ō$ to PT **a* is also surprising on account of the opposite development of earlier *$ā$ (PIE **$eh_2$*) to PT **o*. If both developments turn out to be correct, it seems like the two vowels need to have crossed paths or swapped places in some way. This is assumed by Ringe in his relative chronology. While I do not find the evidence for PIE *$ō$ to PT **a* compelling, it is commonly assumed, and in the end the interpretation rests on a number of reconstructions on which opinions differ. Therefore, in the next subsection, I will show that the assumption of such a change does not have

to affect our relative chronology and the overall interpretation of the phonological system in this dissertation, even if one does believe in it, or if better evidence is found. This would entail that after original word-final *$\bar{o}$ had become *$\bar{u}$, new word-final *$\bar{o}$ again developed separate from non-final *$\bar{o}$, with the latter eventually becoming *$a$.

#### 2.4.5.2 PIE *$\bar{o}$ to PT *$a$ and the other vowels

To understand the putative development from PIE *$\bar{o}$ to PT *$a$ within the pre-Proto-Tocharian vowel system, Ringe tentatively reconstructs an intermediate pre-PT long *$\ddot{\bar{e}}$. This would parallel the development of short PIE *$o$ to pre-PT *$ë$, and affect the relative chronology of long vowel changes. The hypothetical pre-PT long *$\ddot{\bar{e}}$ could then merge with older *$a$ (from laryngeals) after the long *$\bar{a}$ had become a rounded vowel like *$\mathring{\bar{a}}$ on its way to becoming PT *$o$. Ringe dates this in the relative chronology before the loss of vowel length (Ringe 1996: 93). This idea is summarized in Table 2.17.

Table 2.17: Potential development of PIE *$\bar{o}$ and adjacent vowels according to Ringe (1996), with a merger of *$\bar{o}$ and *$a$ upon the loss of vowel length.

| PT | < | < | PIE |
|---|---|---|---|
| e | ë | ë | o |
| a | ā | $\ddot{\bar{e}}$ | ō |
| a | a | a | a |
| å | $\mathring{\bar{a}}$ | $\mathring{\bar{a}}$ | ā |

While such a scenario is theoretically possible, I think that it is difficult in practice, since long *$\ddot{\bar{e}}$ then needs to merge with short *$a$ without interfering with short *$ë$. The latter, in turn, would have to merge with original long *$\bar{e}$. These mergers are most understandable if vowel length did not play a role anymore at the time. On the other hand, vowel length would still be required to separate *$ë$/$o$ from *$\ddot{\bar{e}}$/$\bar{o}$ in Ringe's model. We may therefore see how a similar sequence of changes can work perfectly well without a retention of vowel length.

When short PIE *$o$ changed to *$ë$, long PIE *$\bar{o}$ can take its place as the new short *$o$. Similarly, if short PIE *$e$ had merged with PIE *$i$, we can assume that long PIE *$\bar{e}$ became a new pre-PT *$e$. At this stage, vowel length has ceased to be distinctive without any problems. After palatalization had occurred, early pre-PT *$ë$ and *$e$ merged as *$e$, with the latter having caused palatalization of the preceding consonant (i.e., */Cë/ and */Ce/ became */Ce/ and */C$^{y}$e/). This then would create the possibility for non-final *$o$ to become a new *$ë$, and from that position it can have been lowered and merged with *$a$. If we reconstruct an intermediary stage *$å$ (±*[ɒ] or ±*[ɔ] as opposed to ±*[o]) for

original long PIE *$eh_2/\bar{a}$, that vowel is then continuously differentiated from PIE *$\bar{o}$ in its development via short *$o$ and *$\ddot{e}$ to eventual *$e$. This is summarized in Table 2.18.

Table 2.18: Potential developments of *$\bar{o}$ and adjacent vowels with a merger of *$\bar{o}$ and *$a$ after loss of vowel length.

| PT | < | < | < | PIE |
|---|---|---|---|---|
| e | e | e | e | ē |
| e | e | e | ë | o |
| a | ë | o | o | ō |
| a | a | a | a | a |
| å | å | å | å | ā |

This is the same type of movement around the vowel space as Ringe proposes with vowel length intact, from PIE *$\bar{o}$ to intermediate *$\bar{\ddot{e}}$, and from there down to *$a$, but it turns out that vowel length is not necessary. However, vowel length cannot be lost before *$\ddot{e}$ merges with *$\bar{e}$ to avoid merging a of *$\ddot{e}$ and *$\bar{\ddot{e}}$, while the merger of *$\ddot{e}$ and *$\bar{e}$ depends on a neutralization of vowel length.

It would thus in any case be preferable to keep long PIE *$\bar{o}$ as *$o$ initially, until short PIE *$o$ has developed through *$\ddot{e}$ and with palatalization to eventual *$e$, merging with *$e$ from *$\bar{e}$. To avoid a merger of long PIE *$\bar{o}$ and *$\bar{a}$ as *$o$, the *$\bar{a}$ can be considered as still lower *$\mathring{a}$ (*[ɒ] or *[ɔ]). This vowel only later merged as PT *$o$ with the PIE long *$\bar{o}$ that remained in (secondarily) word-final position (type TB *walo* 'king'). Some additional discussion of the PT *$o$ will be given in parts of the next subsection, 2.5.

### 2.4.6 Summary

Palatalization was one of the most far-reaching changes that took place in pre-Proto-Tocharian, with significant ramifications for the phonological system, as discussed in this subsection. Most consonant phonemes were split into a palatalized and a non-palatalized version. The schematic summary of the changes that took place is reproduced here in Table 2.19.

Table 2.19: An overview of the vowel developments with palatalization from Proto-Indo-European to Proto-Tocharian.

| PT | palatalized | pre-PT | post-PIE | PIE |
|---|---|---|---|---|
| C$^{y}$i | C$^{y}$iy | Ciy/Cī | Cei, Cī | Cei, Cih |
| C$^{y}$e | C$^{y}$e | Ce | Cē | Cē, Ceh$_1$ |
| Ce | Ce | Cë | Co | Co |
| Ca | Ca | Ca | Ca | CH, Ch$_2$e |

| PT | palatalized | pre-PT | post-PIE | PIE |
|---|---|---|---|---|
| Ca, Co# | Co | Co | Cō | Cō, CoH |
| Co | Cå | Cå | Cā | $Ceh_2$ |
| C$^{y}$u | C$^{y}$iw/C$^{y}$uw | Ciw | Ceu | Ceu |
| Cu | Cuw | Cuw/Cū | Cū | Cō#, $Cuh_1$? |
| C$^{y}$ə | C$^{y}$i | Ci | Ci, Ce | Ci, Ce |
| Cə | Ci | “Cɨ” | “Cɨ” | C$\underset{\circ}{\text{R}}$ |
| Cə | Cu | Cu | Cu | Cu |

The developments of the vowel system can be summarized as first a reduction in the number of long vowels, hand in hand with an increase in the number of vowel qualities. Since palatalization involved a merger of long PIE *$\bar{e}$ and short PIE *$o$, the length distinction of the former must have been nullified at the time of palatalization. Marginal *$\bar{\imath}$ (from *$ih_1$) and *$\bar{u}$ (from word-final *-$\bar{o}$) did not merge with their short counterparts. Since short *$i$, *$e$ and *$u$ remained distinct from one another until after the occurrence of certain types of umlaut, the original of length distinction must have been preserved among the high vowels in some form. Since *$\bar{\imath}$ and *$\bar{u}$ did coalesce with the diphthongs *$ei$ and *$eu$, they could be considered as *$iy$ and *$uw$ instead.

Table 2.20: The vowel inventories of two stages of pre-Proto-Tocharian, before the loss of phonemic vowel length on the left and after the loss of phonemic vowel length on the right.

| pre-PT | front | central | back | > | front | central | back |
|---|---|---|---|---|---|---|---|
| high | *i ī* | “*ɨ*” | *u ū* | | *i iy* | “*ɨ*” | *u uw* |
| mid | *e ē* | | *o ō* | | *e* | *ë* | *o* |
| low | | *a ā* | | | | *a* | *å* |

With palatalization, short *$i$ (having subsumed PIE *$e$ and most PIE *$i$) and *$ə$ (derived from vocalized syllabic resonants and some PIE *$i$ after labiovelars and labials) merged as pre-PT *$i$, with only the former causing palatalization. Similarly, *$e$ from PIE *$ē$ and *$ë$ from PIE *$o$ merged as pre-PT *$e$, again with only the former causing palatalization.

Table 2.21: The vowel inventories of two stages of pre-Proto-Tocharian, before phonemic palatalization on the left and after phonemic palatalization on the right.

| pre-PT | front | central | back | > | front | central | back |
|---|---|---|---|---|---|---|---|
| high | *i iy* | “*ɨ*” | *u uw* | | *i iy* | | *u uw* |
| mid | *e* | *ë* | *o* | | *e* | | *o* |
| low | | *a* | *å* | | | *a* | *å* |

The remaining changes needed to derive the Proto-Tocharian vowel system are relatively simple. Short **i* and **u* finally merged as PT **ə* so that **iy* and **uw* can now be rewritten as **i* and **u* (alternatively, **əy* and **əw*). These developments are illustrated in Table 2.22.

Table 2.22: An overview of the vowel inventories of a late stage of pre-Proto-Tocharian on the left, with Proto-Tocharian on the right, after the final changes had taken place.

| pre-PT | front | central | back | > PT | front | central | back |
|---|---|---|---|---|---|---|---|
| high | *i iy* | | *u uw* | | *i* | *ə* | *u* |
| mid | *e* | | *o* | | *e* | | *o* |
| low | | *a* | *å* | | | *a* | (*å?*) |

It is sometimes assumed that the merger of PIE **ē* and **o* did not actually take place until after Proto-Tocharian (Ringe 1990: 222–226; Adams 2013; Kümmel 2009). At various points in the next section (2.5), it will become clear why it is not necessary to keep these vowels separate until palatalization. Similarly, (pre-)PT **o* from post-PIE **ō* and (pre-)PT **o* from PIE **eh₂* (=**ā*) are often kept separate as well as PT **o* and **å*, a view which I also held in an earlier article (Warries 2022). I now think it possible that these two vowels were merged in pre-Proto-Tocharian already as **o* (cf. the table above). This will also be discussed further in the next section, specifically in 2.5.6.4.

## 2.5 Late vowel developments

Some types of syllables that existed in Proto-Tocharian were left out of consideration in the previous section about palatalization because they are the result of a different phenomenon: umlaut. This most importantly concerns the syllables of the type TB /Cʸo/ and /Cʸa/, both of which are derived from earlier */Cʸe/ due to effects of the vowel in the following syllable. It is important to discuss these changes, as they inform a number of decisions made about the interpretation of the relative chronology in the previous section, and there is no consensus.

### 2.5.1 *u*-umlaut of pre-PT **e*

It has been known for a long time that an original **u* caused umlaut of an **e* in the preceding syllable. Some scholars use a special symbol for the product of this umlaut, namely "*o̰*" with a dot underneath. This reflects the view that "the new *o*-vowel did not merge with other *o*'s in PT" (Ringe: 1996: 98). I write the potentially different *o*-vowels as "*o*" and "*å*" rather than "*o̰*" and "*o*", and the issue will be discussed throughout this section.

Furthermore, while Ringe only discusses *u*-umlaut of the **ë* that arose from PIE **o*, also (post-)PIE **ē* underwent the same development (cf. Hilmarsson 1986: 23–28; Del

Tomba 2020: 88, 91). We can thus simply say that it was the **e* that arose from the merger of **ë* (PIE **o*) and **e* (PIE **ē*) after palatalization that was targeted by *u*-umlaut. Representative examples are listed below.

10. PIE **dor-u-*, **dre-u-* 'wood' >> **eru* > PT **orə* > TA *or*, TB *or*
82. PIE **nokʷtu-* > **nektu-* > **nokt-* > TA obl. *nokte* 'last evening', der. *noktiṃ* 'in the evening'
84. PIE **oktō* 'eight (8)' > **ektū* > PT **oktə* (analogical **-ə* after **sʸəptə* 'seven (7)') > TB *okt*; TA *okät* (earlier *oktu-* visible in TA *oktuk* 'eighty' from PT **oktu-kə*)
100. PIE **sneh₁-ru* 'sinew' > **snēru* > **sʸnʸeru*[28] > PT **ṣñorə* > TB *ṣñor*
102. PIE **solh₂u-* > **selu+me*[29] > PT **solme* > TB *solme* 'complete(ly), altogether'

Hilmarsson (1986: 20-21) dated *u*-umlaut after the Proto-Tocharian period, but this cannot be correct, since most **u* that caused this sound change had disappeared by that point in the merger with original **i* and **e* as **ə* (cf. also 2.7). The *u*-umlaut of **e* indicates both that PIE **o* and **ē* were already phonologically merged when the umlaut took place, after palatalization, and conversely, that PIE **u* and **i*/*e* had not yet merged at that point in time.

### 2.5.2 *o*-umlaut of pre-PT **u*

Original PIE **u* and **eu* seems to have been turned into PT **o* when followed by an **o* in the next syllable (Hilmarsson 1986: 38–42). There are only two examples.

59. PIE **ḱleumōn-* > ?**klʸuwmo* > PT **klʸomo* > TB *klyomo*, TA *klyom* 'noble'
124. PIE **uksōn-* > *ukso* > PT **okso* > TB *okso*, TA pl. *opsi* 'ox'[30]

This development of PIE **u* is best seen in PIE **uksōn* to PT **okso* 'ox' (TB *okso*, TA pl. *opsi*). PIE **ḱleumōn* to PT **klʸomo* 'noble' (TB *klyomo*, TA *klyom*) shows a similar development to that in **ukso*, although the vowels in initial syllables of these words are

[28] With metathesis from **-ur* to **-ru*, see Del Tomba (2020: 87–94).

[29] TA *salu* 'complete(ly), altogether' is likely from *solh₂wV* with consonantal **u* as **w*, which did not cause umlaut.

[30] The vowel in **okso* is sometimes instead taken to be due to rounding between to labials with a reconstruction of pre-PT **wəkʷso* (Ringe 1996: 127; Kim 1999: 164–165 incl. fn. 54), but on account of the *o* in TA pl. *opsi*, this particular sound change cannot have been the reason for the TB *o* here (Pinault 2008: 432–433). This is because all the other examples of this sound change, namely TB *kokale* : TA *kukäl* 'chariot' from **kʷəkʷle* and TB *pokkāka* : TA *pᵤkāk* 'call!' from **pəkʷaka* show that no development to TA *o* took place in this context (cf. 2.5.6.3). Umlaut is thus the best way to account for the vocalism in TB *okso* and TA *opsi*.

not identical. The fact that we do not see this change in PIE *u̯l̥h₂ōnts, which became PT *wəlo 'king' (TB *walo*, TA *wäl*), indicates that the umlaut took place before initial *u- developed into *wə- in pre-PT *ukso. Otherwise, we would have had a development from PIE *u̯l̥h₂ōnts to PT *wolo as well, to yield TB **olo, TA **wal. This development in its entirety can thus be dated to a time before short *u merged with *i and *ə as PT *ə, a change in which the rise of initial *wə- from *u- was probably implicated. If the first *o in PT *klʸomo is the result of the same development, it must have taken place after palatalization, just like the other umlaut developments.

### 2.5.3 *o*-umlaut of pre-PT *e

The vowel *o from earlier PIE *eh₂ also appears to have caused a preceding *e to turn into *o (Hilmarsson 1986: 29ff; Ringe 1996: 163; Pinault 2008: 432). Reflexes of this umlaut are only clearly visible in Tocharian B, where there are a few examples.

89. PIE *proskeh₂ > *preskå > PT *prosko (*pråskå) > TB *prosko~proskiye*, ? TA *praski* 'fear'
95. PIE *seh₂i- → prt.ptc. pre-PT *se-soyəwə (*se-såyəwə) > TB *sosoyu*, ? TA *sasyu* 'satisfied'
97. PIE *sem- → som-eh₂m > *semåm > PT *somo (*såmå) > TB *somo*, TA *ṣom* 'one (1), some' (f.obl.sg.)

Pinault (2008: 436–438) writes that "il n'existe pas de correspondence sûre avec le tokh. A qui garandisse un Umlaut de ce type au stade tokh. commun" (437). This is because the clearest examples in Tocharian B do not have straightforward counterparts in Tocharian A. Each example listed above has its own problems.

TA *praski* 'fear' is superficially similar to the secondary Tocharian B nominative *proskiye*, but since this is a later creation the two cannot go back to Proto-Tocharian as such. TA *praski*, with its alternant plural *praskintu* shows different morphological behaviour from TB *prosko~proskiye*, leading Peyrot (2008: 103) to conclude that the two are most likely independent nominal formations. It is thus uncertain if TA *praski* shows that the *o*-umlaut visible in TB *prosko~proskiye* was restricted to Tocharian B only.

It seems like the counterpart to the preterite participle TB *sosoyu* to *soy-* 'be sated' is TA *sasyu*, with *a*-vocalism rather than *o*-vocalism (Pinault 2008: 436; Ringe 1996: 163; Malzahn 2010: 955). However, Peyrot has pointed out that TA *sasyu* is synchronically the participle to a different verb TA *säy(n)-* 'satiate; be satiated', which is the cognate of TB *səyn-* 'be satiated'. The cognate to the verb TB *soy-* 'be sated', from which TB *sosoyu* is formed, is not attested in Tocharian A. The preterite participle TA *sasyu* is parallel to the preterite participle TA *raryu* from the verb *räy-* 'give up', cognate to TB *rəyn-* 'id.' (Peyrot 2013: 833 fn. 949). This means that TB *sosoyu* and TA *sasyu* could reflect two different

formations, so that TA *sasyu* cannot give us reliable information on the status of *o*-umlaut caused by the reflex of PIE *$*eh_2$/$\bar{a}$* in Proto-Tocharian.

The feminine of the numeral 'one (1), some', which shows umlaut caused by the reflex of *$*eh_2$* in TB *somo*, pl. *somona*, is analogically modified in Tocharian A *ṣom*, pl. *ṣoman*, by at least a change of the initial sibilant from PT **s-* to TA *ṣ-* on the basis of the masculine TA *ṣom*, pl. *ṣome* (TB *ṣeme*, *ṣemi*). It is therefore uncertain whether the analogical change affected just the initial segment of an umlauted pre-TA **som*, pl. **soman*, or if the analogical modification affected the first two segments of an earlier pre-TA **sam*, pl. **saman* without umlaut (see also Pinault 2008: 437–438). Affection caused by the labial *m* may also have caused a non-umlauted PT **e* to become TA *o*, as in the development from *$*n^yemə$* to TA *ñom* 'name', beside TB *ñem* 'id.' (cf. 2.5.6.1).

There is one example that appears to show *o*-umlaut of **e* caused by an earlier PIE *$*\bar{o}$*, namely TB *orkamo*, TA *orkäm* 'darkness; dark', via pre-PT **erkmo*.[31]

42. PIE *$*h_1reg^w$- → *$*h_1rg^w$-mōn* > **erkmo* > PT *orkmo* > TB *orkamo*, TA *orkäm* 'darkness; dark'

This example would also show that the umlaut was shared between Tocharian A and B, since both have the same vowel *o*. However, it could alternatively reflect a secondary derivation from a *u*-stem adjective *$*h_1rg^{(w)}$-u-*, thus *$*h_1rg^{(w)}$-u-mōn*, where *u*-umlaut would have caused the change from **e* to **o* instead: *$*h_1rg^{(w)}$-u-mōn* > **erkumo* > **orkumo* > **orkəmo*. Such a reconstruction circumvents the need to posit *o*-umlaut of **e* caused by PIE *$*\bar{o}$* in the basis of this single example, but we would expect a labializing effect on the velar caused by **u*, thus TB ***orkwamo* vel sim. (cf. TB *laṅk$_u$tse* in 2.3.3.3).

It is conceivable that PIE *$*eh_2$/$\bar{a}$* and PIE *$*\bar{o}$* were still different when *o*-umlaut took place, and that they caused changes to different vowels, one with the reflexes TB *o* (TB *prosko*, *sosoyu*), TA *a* (*praski*, *sasyu*), and the other with TAB *o* (TB *orkamo* and TA *orkäm*). With so few examples and the complicated interpretation of TA *praski* and *sasyu* it seems impossible to further substantiate this possibility.

In conclusion, the available examples in Tocharian A do not allow us to determine with certainty whether *o*-umlaut caused by the reflex of original *$*eh_2$/$\bar{a}$* already took place in Proto-Tocharian, or if it only occurred in Tocharian B. It is furthermore possible that the resulting vowel of this umlaut was not the PT **o* that became TAB *o*, but rather PT

[31] Another similar case is that of TB *kolmo*/*a**, TA *koläm* 'basin', which has long been interpreted as the result of *o*-umlaut from pre-PT **kelmo* (Pinault 2008 432; Del Tomba 2020: 146 fn. 218; Hilmarsson 1986: 29). Huard has recently argued that the vowel should be the result of *u*-umlaut of **e* instead (Huard 2022: 411-416). This word is best excluded from the evidence due to the uncertainties involved.

**å*, which is treated further in the next subsection. Whether original PIE **ō* caused *o*-umlaut in (pre-)Proto-Tocharian, and whether the resulting vowel was different from the *o*-umlaut was different from the vowel obtained from *o*-umlaut caused by original PIE **eh₂/ā* cannot conclusively be determined.

### 2.5.4 *o*-umlaut of pre-PT **a*

Especially Tocharian B shows additional cases of umlaut where the *o*-vowel that caused the umlaut arose secondarily due to contractions (e.g., Pinault 2008: 433–436). It is thus etymologically different from both PIE **ō* and PIE **eh₂/ā*. The vowel that resulted from this umlaut must also have been different from PT **o* in phonological terms, since it always became TA *a* rather than ***o* (i.e., not TA ***ora-* 'cease' prs.[4], etc., from a PT ***oro-*). The vowel sign **å* may be used to express this. The development here is thus separate from the PT **o* from at least the products of *u*-umlaut of **e* (PIE **o* and **ē*), *o*-umlaut of **(e)u*.[32]

While the precise development of the second-syllable **-o* in the present class [4] is uncertain (Malzahn 2010: 385–389, Peyrot 2013: 592–593), it is clear that the vowel it created by umlaut was different from the vowel belonging with the two other cases of *o*-umlaut discussed above in subsections 2.5.2 and 2.5.3. The Tocharian A forms cannot reflect the original, non-umlauted vowel, since that should have become TA ***ā* rather than *a*. The examples are given in the list below.

49. TB *onolme* 'human' from **ånålme*, earlier **anålme* contracted from **ana-elme*; no TA cognate
- TB *oro-*, TA *ara-* 'cease' prs.[4] from **årå-*, earlier **arå-*
- TB *ortto-*, TA *arta-* 'praise' prs.[4] from **årttå-*, earlier **arttå-*
- TB *oso-*, TA *asa-* 'dry out' prs.[4] from **åså-*, earlier **aså-*
- TB *yoto-*, TA *yata-* 'be capable of' prs.[4] from **yåtå-*, earlier **yatå-*
- TB *plonto-*, TA *planta-* 'be glad' prs.[4] from **plåntå-*, earlier **plantå-*
- TB *kloyo-*, TA *klawa-* 'fall' prs.[4] from **klåwʸå-*, earlier **klawʸå-*

[32] Pinault (2008: 435–436) considers the word for 'elephant', TB *oṅkolmo* and TA *oṅkaläm* to have undergone the type of umlaut under discussion here, from an earlier **ankålme* (with a secondary inflection in *-o* after *okso* 'ox, cow'). This would show a Tocharian A reflex *o* in the first syllable, counter to the verbs of present [4]. Since the etymology of 'elephant' in Tocharian remains a thorny problem, and analogical influences from both the ancestor of TB *onolme* 'man' and of TB *okso* 'ox, cow' are assumed, this word is best not used for determining regular sound change. See Adams (2013: 118–119) for an overview of other etymological suggestions of TB *oṅkolmo*, TA *oṅkaläm*, with references.

The correspondences in these verbs crucially show that a vowel **å* must be reconstructed for Proto-Tocharian. The relation of this vowel to the result from PIE **eh₂/ā* is not entirely clear, see the discussion below in 2.5.6.4.

### 2.5.5 *a*-umlaut of pre-PT **e*

The final type of umlaut that we will discuss is the *a*-umlaut of pre-PT **e*. Just like the *u*-umlaut of **e* (2.5.1), this *a*-umlaut affected both **e* from PIE **o* and **e* from PIE **ē*, and it added the syllable type */Cʸa/ to the system. Its effects are mainly seen in verbal morphophonology in Tocharian B. Winter first published about *a*-umlaut of pre-PT **e* in a short aside, connecting the ablaut grades of *a*-preterites [class 1] or *a*-subjunctives [class 5] with their respective root counterparts, and offering the plural noun TB *ārwa* to *or* 'wood' from earlier **erwa* as a parallel outside of the verbal system (Winter 1962: 32–33). The *a*-umlaut allows for an interpretation of, e.g., 3sg. *tārkaṃ* /tárkan/ 's/he dismisses, emits' and 3sg. *tekäṃ* /tekən/ 's/he touches' as having the same type of ablaut grade, namely an original **e*. This original **e* was then changed to **a* in the former, due to the presence of a root-final *-a*. In the third person plural forms, such as *tarkaṃ* /tərkan/ 'they dismiss, emit' and *takäṃ* /təkən/ 'they touch', both verbs have the same *ə*-grade, which remained unaffected by umlaut (e.g., Pinault 2008: 594–596; Peyrot 2013: 69).

Table 2.23: The effects of *a*-umlaut in the *a*-subjunctive (*tārkaṃ*) next to an example from the root-subjunctive without *a*-umlaut (*tekäṃ*).

| | ablaut | example | reconstruction |
|---|---|---|---|
| no final *-a* | *e* : *ə* | /tékən/ : /təkən/ | |
| final *-a* | *a* : *ə* | /tárkan/ : /tərkan/ | < */térkan/ : /tərkan/ |

This situation is not found in Tocharian A as such. The form corresponding to TB *tārkaṃ* is TA *tarkaṣ*, with initial TA *a* reflecting a PT **e*, rather than ***tārkaṣ* with TA *ā* from PT **a* due to umlaut. In general, the Tocharian A subjunctives [5] with stem-final **-a* do not show the *a*-umlaut seen in their Tocharian B counterparts. According to Cowgill (1967: 176–177), who is followed by Ringe (1996: 160–163), Pinault (2008: 428), Hackstein (2017: 1320), etc., this is because *a*-umlaut did not happen in stressed syllables in Proto-Tocharian. The *a*-umlaut of stressed **e* only took place in Tocharian B, but in Tocharian A *a*-umlaut was restricted to unstressed syllables, from the point of view of Tocharian B stress. Cowgill's parallel outside the verbal system is the feminine form for 'four (4)' in TA *śtwar*, rather than ***śtwār* from **śtwéra*; cf. TB *śtwāra* (Cowgill 1967: 177).

The reduplication syllable in preterite participles and the negative prefix do show *a*-umlaut in Tocharian A as well as B, so that an earlier **nenákuwə* from the root **nak-* 'reproach' has become PT **nanákuwə* and thence TB *nanāku* and TA *nānku*. Similar are

pre-PT **kekámawə* from **kama-* 'bring' to PT **kakámawə* yielding TB *kakāmau* and TA *kākmu*, and pre-PT **enáywatˢe* 'unpleasant' to PT **anáywatˢe* yielding TB *anaiwatse* and TA *ānewāts* (Ringe 1996: 160). Examples such as these are the reason to assume that *a*-umlaut also occurred in Tocharian A, but that its operation was limited to unaccented syllables (Cowgill 1967). This is summarized in Table 2.24.

Table 2.24: An illustration of a-umlaut affecting unstressed syllables.

| pre-PT | *a*-umlaut | TB | TA |
|---|---|---|---|
| **nenákuwə* | *nanákuwə* | *nanāku* /nanákəw/ | *nānku* |
| **kekámawə* | *kakámawə* | *kakāmau* /kakámaw/ | *kākmu* |
| **enáywatˢtˢe* | *anáywatˢtˢe* | *anaiwatstse* /anáywatˢtˢe/ | *ānewāts* |

However, there are also some instances where Tocharian A shows traces of *a*-umlaut in syllables that are accented in Tocharian B. Jasanoff adduces the *a*-subjunctive of the stem **mənta-* 'hurt', which has a non-ablauting subjunctive showing umlaut from **ménta-* to TA *mānta*, as seen in TA *māntlune* (Jasanoff 2013: 116 fn. 37). Similarly, the sbj.3pl. TA *pālantär* from TA *pälā-* 'praise' would have been TA ***palantär* without umlaut from earlier **pélantər*.[33] These examples support the idea that *a*-umlaut of also stressed **e* actually did apply in the ancestor of Tocharian A as well, but was partly undone. Jasanoff argues that any Tocharian A *a*-subjunctives that seem to reflect **e*-grade could have been remade on the basis of the productive ablaut pattern of TA *a* : *ä* (type PT **e* : **ə*), which is seen in many subjunctives without root-final *-a* and could plausibly have replaced the ablaut of TA *ā* : *ä* that resulted from *a*-umlaut (Jasanoff 2013: 107 fn. 7).

Ringe instead suggests that the **ā* in the root in examples such as these arose in the preterite, where the accent was on the stem-final **-a* instead (Ringe 1996: 161). However, of *a*-umlaut did not only occur in verbal forms; its effects are also seen in a few nouns. The change of pre-PT **erwa* (plural of *or* 'wood') to PT **arwa*, whence TB *ārwa* /árwa/, has already been mentioned above, and it constitutes a case of *a*-umlaut of stressed pre-PT **é* in Tocharian B. The Tocharian A counterpart to this particular word is not attested, but Pinault has convincingly argued that the same plural suffix **-a* did cause *a*-umlaut in TA *wmār* 'jewel'. The cognate in Tocharian B is *wamer* 'id.' The discrepancy in the vowel TA *ā* : TB *e* (expected would be TA *a* : TB *e* or TA *ā* : TB *a*) can be understood if we consider the kind of plural that TB *wamer* forms, namely *wmera* /wméra/. If we start from such a form in pre-PT **wəméra* and apply *a*-umlaut, this should become PT **wəmára*. Such a PT **wəmára* is then the regular source of TA *wmār*, which could be reinterpreted as a

[33] TB *pālāntär* from THT 562 a2 is uninformative, and the 1sg. *pālamar* in THT 240 b6 (arch.) is not reliable, but the subjunctive *pāloymar* PK AS 5A a5 points to initial accent in this subjunctive stem.

singular when the ending was lost due to apocope (Pinault 2011: 160–164 and 171–173). The TB plural *wmera* for expected **wmāra* has been analogically changed to better fit the singular TB *wamer* again. The only way to continue the belief that accented **é* specifically was not affected is to say that the Tocharian B accentuation is innovative, and that the accent was somewhere else at the time of *a*-umlaut. Such an approach inevitably leads to a circular argumentation, since the connection between *a*-umlaut and the accent is based on the place of the accent in Tocharian B in the first place.

The lack of *a*-umlaut in TA feminine *śtwar* 'four (4)' rather than **śtwār* can probably be attributed to analogical influence from the masculine *śtwar*. The numeral for 'one (1)' also merged in Tocharian A due to analogical influence from the masculine form *ṣom* (PT **sʸeme*) on the feminine TA *ṣom*, with the initial retroflex sibilant taken over from the masculine. As a process, *a*-umlaut was not active anymore in either Tocharian language, and its results were clearly prone to being analogically remodelled. It can thus be assigned to the pre-Proto-Tocharian period, and, contrary to common belief, applied to both stressed and unstressed pre-PT **e*.

The examples of *a*-umlaut that we have seen so far affected pre-PT non-palatalizing **e*, which derives from PIE **o*. Yet there are also a few examples of palatalizing TAB *a*, and these can be derived from PIE **ē* by way of *a*-umlaut. These examples are restricted to the preterite of a small group of verbs, namely *lyaka-* to *lək-* 'see', *śawa-* to *śəw-* 'eat', *plyawa-* to *pləw-* 'complain', and *lyawa-* to *ləw-* 'rub'. The relevant forms derive via *a*-umlaut from earlier **lʸeka-*, **kʸewa-*, **plʸewa-*, and **lʸewa-* respectively. The origin of the *ē*-grade in these verbal forms has long been a contentious issue in the Tocharian scholarly literature, see for instance Peyrot (2012b: 102–109, with further references). However, the clear effects of *a*-umlaut seen here imply that PIE **o* and **ē* had phonologically merged before *a*-umlaut, with the two originally distinct vowels both being affected in the same way. This aspect of *a*-umlaut agrees with *u*-umlaut (2.5.1). The fact that original PIE **ē* did palatalize the preceding consonant before becoming **a*, in turn, confirms that *a*-umlaut, again like *u*-umlaut, took place after palatalization.

### 2.5.6 Effects of labial consonants on vowels

Aside from the various types of umlaut discussed in the preceding subsections, some effects of labial consonants on neighbouring vowels needs to be taken into account as well. These effects of labial consonants largely occurred in the separate histories of Tocharian A and B, and led to the creation of additional *o*-vowels. It is mostly Tocharian A that was affected, but some instances of are known from Tocharian B as well. The conditioning environments are different, however, and not always entirely clear. Apparent irregularities have been interpreted as the result of a retained contrast between PIE **ē* and **o* until after Proto-Tocharian, contrary to our relative chronology, so that they need to be discussed.

#### 2.5.6.1 Rounding due to *p/m in Tocharian A

Instead of turning into the expected reflex TA *a*, PT **e* next to a labial *p* or *m* has in some cases changed to TA *o*. A comparison with Tocharian B shows how the Tocharian A vowel has changed, with Tocharian B here preserving the more original situation.

Since this change only affects Tocharian A, it took place after the split of Proto-Tocharian. It is nevertheless important to address, as it has been adduced as evidence that the contrast between pre-PT **e* (PIE **ē*) and **ë* (PIE **o*) was retained for longer than assumed above in our discussion of palatalization above. This is argued by Ringe (1990: 223ff.), based on the fact that both instances of **o* with a certain Indo-European etymology go back to **ē* (viz., *ṣom* 'one (1)' and *ñom* 'name'), while those with **a* go back to PIE **o* (viz., *masäk* 'joint, link', *maku* 'nails', *pats* 'husband', *kam* 'tooth'). For this reason, Ringe (1990) set up a continued contrast between PT **e* from PIE **ē* and PT **ë* from PIE **o*. The other words are borrowings (e.g., *paräṃ* 'axe' from Iranian), or in some other way difficult to interpret.

Here we will look at the material again, to see if Ringe's continued separation of PIE **ē* and **o* as different phonemes in Proto-Tocharian is indeed warranted, in spite of the indications to the contrary (cf. especially 2.5.1 and 2.5.5). We will start in Table 2.25 with a list of the cases where TA *o* corresponds to TB *e*.

Table 2.25: Examples of TA *o* corresponding to TB *e*.

| TA | | pre-TA | TB |
|---|---|---|---|
| *opäntäs* | 'in between' | **apäṅkta* (?) | *epiṅkte* |
| *opäśśi* | 'clever, skilled' | **apästya* | *epastye* |
| *opyāc* | 'memory' | **apäyāc* | *epiyac* (OSIr. **abiyāta-*) |
| *omäl* | 'hot' | **amällä* | *emalle* |
| *omlyi* | 'heat' | **amälyā* | *emalya* |
| *ñom* | 'name' | **ñamä* | *ñem* |
| *cmol* | 'birth' | **cämal* | *camel* |
| *porat* | 'axe' | **paratä* | *peret* (OSIr. **paratu-*) |
| *ṣom* | 'one (1), some' | **ṣama* | *ṣeme* |

In the labial environments in the following words in Table 2.26, PT **e* changed to TA *a* per the usual development. It is immediately obvious from the relative length of this list that a development to TA *a* is more common than to TA *o*.

Table 2.26: Examples of TA a corresponding to TB e in a labial environment.

| TA | | pre-TA | cf. TB |
|---|---|---|---|
| *anapär* | 'before, in front of' | **anapra* | *enepre* |
| *apärkär* | 'a long time' | **a(m)pärkra* | *emparkre* |
| *apälkāt* | 'unworried' | **a(m)pälkātte* | *empalkatte* |
| *kapśañi* | 'body' | **kakśañi* | *kektseñe* |
| *kam* | 'tooth' | **kama* | *keme* |
| *pañi* | 'splendour' | **pańiya* | *peñiyo* |
| *pats* | 'husband' | **patsä* | — |
| *paräṃ* | 'glory' | **parna* | *perne* (OSIr. *farnah-*) |
| der. *parno* | 'glorious' | **parnawä* | *perne*$_u$ |
| *pare* | 'debt' | **paray* | *peri* |
| *pal* | '(right) way, law' | **pala* | *pele* |
| der. *paloṃ* | 'praise' | **palawnā* | *palauna* (analogical *a*) |
| *maku* | 'nails' | **makwa* | *mekwa* |
| *mañ* | 'moon' | **mańa* | *meñe* |
| *malañ* | 'nose' | **malañä* | *meli* (diff. pl. suffix) |
| *malke* | 'milk' | **malka* | *malkwer* (parallel deriv.) |
| *malto* | 'first' | **maltawä* | *melte* (TA is derived) |
| *masäk* | 'joint, link' | **maska* | *meske* |
| *ram* | 'witness' | **rama* | *reme* |
| *wram* | 'thing' | **wrama* | *wreme-* |
| *slam* | 'flame' | **slama* | *sleme* |

The words with a reflex TA *o* can be dividied into two groups: one where the vowel precedes the bilabial that is supposed to affect it and one where the vowel follows the bilabial. These two environments may have had different kinds of effects, so we will consider them separately. Further subdivisions in the first group can be made as in Table 2.27. The majority of cases of rounding to TA *o* are in this group, but there are also clear instances where no rounding occurred.

Table 2.27: Reflexes of PT **e* before a labial consonant in Tocharian A, classified into four categories according to the environment and the different reflexes TA *o* and *a*.

| Cam < Cem | C$^{y}$om < C$^{y}$em | #op/m < #ep/m | #ap < #emp |
|---|---|---|---|
| *kam < keme*<br>*ram < reme*<br>*wram < wreme*<br>*slam < sleme* | *ñom < n$^{y}$emə*<br>*ṣom < s$^{y}$eme* | *opäśśi < epəst$^{y}$ye*<br>*omäl < eməlle*<br>*omlyi < eməlya*<br>*opyāc < epəyac*<br>*opäntäṣ < ep$^{y}$ənkte* | *apärkär < empərkre*<br>*apälkāt < empəlkatte* |

A subdivision into four categories can be made, depending on the position of PT **e* in the word. Here it becomes clear that PT **e* became TA *o* when followed by **m* if the preceding consonant was palatalized, but not when it was neutral. This corresponds to Ringe's PT **e* from PIE **ē*, which always caused palatalization of the preceding consonant, but formulates the same correspondence in terms of a differently conditioned development. In word-intitial position, PT **e* became TA **a* if the TB cognate shows a consonant cluster *-mp-*, which was probably original (cf. Hilmarsson 1991: 192–199). In other cases, when the TB cognate has a simple *-m-* or *-p-*, the initial PT **e* followed by a labial is reflected as TA **o*. Since the words in each of the four groups belong to a different environment, the double reflex of PT **e* in Tocharian A in front of bilabials can be seen as a conditioned development, so that it does not require the existence of two separate Proto-Tocharian *e*-phonemes like "**e* vs. **ë*" or "**e* vs **æ*".

According to the material with a PT **e* with a following bilabial, there is a difference in the development of those words where the preceding consonant is non-palatalized as opposed to those where it is palatalized. The development of TA *ñom* from **nʸemə* and *ṣom* from **sʸeme* is similar to the rounding rule that produced PT **yok-* 'drink', **yop-* 'enter' (sbj. of *yəp-*) and **yom* 'get' (prt. of *yəm-*) from earlier **yekʷ-*, **yep-* and **yem-*. A change from **yeP* to **yoP* is shared with Tocharian B, while **Cʸem* to **Cʸom* is not, but they can in principle be seen as the same type of change.

In the group of words where the vowel follows the bilabial, we can see that the picture there is more difficult. The majority of examples do not get TA **o*. Of the only two examples where **e* changed to TA *o*, TA *cmol* from PT **cəmel* stands out as the only instance containing a second-syllable sequence of bilabial + **e* + consonant. In the Tocharian B counterpart *camel* /cə́mel/, the *e* is unstressed. The other example, TA *porat* from PT **peretə* is originally a loan from Old Steppe Iranian **faratu-* (Bernard 2023: 43–48), and thus does not directly reflect any inherited Proto-Indo-European vowel. Its development of TA *o* seems to be truly unexpected, next to a number of other original trisyllabic examples that have TA *a* instead.

Table 2.28: Reflexes of PT **e* following a labial consonant in Tocharian A.

| paC < peC | maC < maC | p/moC < p/meC |
|---|---|---|
| *pañi* < **penʸəyo* | *maku* < **mekwa* | *cmol* < **cəmel* |
| *paräṃ* < **perne* | *mańk* < **menke* | *porat* < **peretə* |
| *parno* < **pernewə* | *mañ* < **menʸe* | |
| *pare* < **perey* | *malke* < **melkey* | |
| *pal* < **pele* | *masäk* < **meske* | |
| *paloṃ* < **pelewna* | *malañ* < **meleñə* | |
| *pats* < **petˢə* | *malto* < **meltewə* | |

Since it is possible to come up with secondary developments to get the unexpected *o*-vowels in Tocharian A in all cases except TA *porat*, which is a borrowing from Old Steppe Iranian, there is no need to set up a continued differentiation of PIE **ē* and **o* as PT **e* and **ë* respectively to account for it. Rather, only a single PT **e* is needed. This is also in line with the evidence from both pre-Proto-Tocharian *u*-umlaut (2.5.1) and *a*-umlaut (2.5.5), which targeted the results of PIE **ē* and **o* in the same way, implying a merger between the two vowels.

In conclusion, the Proto-Tocharian vowel **e* seems to have been regularly changed to TA *a*, with secondary rounding to TA *o* in the environments *C*$^{y}$*_m*. The only unexpected TA *o* from PT **e* are found in TA *cmol* 'birth' and in TA *porat* 'axe', the latter of which is not an inherited word from Proto-Indo-European, but instead a borrowing from Old Steppe Iranian **faratu-*. These words cannot uphold a separation of PIE **ē* and **o* until after the Proto-Tocharian period in the face of the other evidence in support of their merger in pre-Proto-Tocharian.

### 2.5.6.2 Rounding due to **k*$^{w}$ in Tocharian A

Some vowels in the vicinity of labiovelars were also rounded in Tocharian A. This very clearly affected the PT vowel **ə*, which surfaces as TA *u* in words such as PT **yək*$^{w}$*e* 'horse', TA *yuk* (TB *yakwe* /yákʷe/) and PT **səkʷə* 'happiness', TA *suk* (TB *sakw* /sákʷ/). The sequence may also be written as TA *$_{u}$k* in derived and inflected forms, such as in the oblique singular *y$_{u}$kas* from *yuk* and the derived adjective *s$_{u}$kaṣi* from *suk*.

Given its effect on **ə*, it is not surprising that **k*$^{w}$ is often thought to cause rounding of PT **e* in a similar way to **p* and **m* as well. This has been proposed as the reason for the TA *o* in at least the following words:

- TA *oṅk* 'man' beside TB *eṅkwe* 'id.'
- TA *oṅkrac* 'immortal' beside TB *oṅkrocce* 'id.'
- TA *opṣäly* 'season; time for action (of a Buddha)' beside TB *ekṣalye* 'id.'
- TA *orkäm* 'darkness; dark' beside TB *orkamo* 'darkness; dark'
- TA *śorki* 'mortal fear' and *śorkmi* 'strings' beside TB *śerkw* 'cord, string'
- TA *nokte* 'at night', *noktiṃ* 'last night' beside TB *nekciye* 'at night; last night'
- TA *koṃ* 'dog obl.sg.' beside TB *kweṃ* 'id.'

The first two examples, TA *oṅk* 'man' and *oṅkrac* 'immortal' are structurally very similar, but an important difference is seen in the Tocharian B counterparts TB *eṅkwe* and TB *oṅkrocce*. Only the former, TB *eṅkwe*, has the vowel TB *e*, whereas TB *oṅkrocce* has *o*-vowels just like TA *oṅkrac* (in the first syllable). This allows for the possibility that the Proto-Tocharian form was already **oṅkrocce* rather than **eṅk*$^{w}$*rocce* or the like, and that the first *o* was caused by *o*-umlaut from the second *o*. The processes at work in TA *oṅkrac* :

TB *oṅkrocce* may thus not be truly parallel to those seen in TA *oṅk* : TB *eṅkwe*, and this word is left aside here. The same applies to TA *orkäm* next to TB *orkamo*, which is claimed to result from affection in PT **erkʷmo* by Kim (1999). However, the matching vowel TAB *o* and the lack of a labiovelar in Tocharian B indicate that this TAB *o* arose at a different time and due to a different development (see above).

The TA *o* in *oṅk* 'man' is mostly considered to be caused by the labiovelar **kʷ* (Hilmarsson 1989a: 108–109; Kim 1999: 167–168).[34] With the shift from **e* to **a* and general apocope of word-final vowels in Tocharian A, PT **enkʷe* might be expected to yield ***anku*, especially when compared to the development in a number of words with similar structure, like PT **kəntwo* 'tongue; language' to TA *käntu* (TB *kantwo*), PT **mekʷa* 'nails' to TA *maku* (TB *mekwa*) and PT **sekʷe* 'pus' to TA *saku* (TB *sekwe*; cf. Hilmarsson 1989a). These all realise the original labiovelars or clusters with **-w-* as a final *-u* in Tocharian A, while no *-u* is found in TA *oṅk* 'man'.

The development of PT **enkʷe* 'man' to TA *oṅk* shares this feature with examples like PT **yəkʷe* 'horse' to TA *yuk* (TB *yakwe*) and PT **tənkʷə* 'love' to TA *tuṅk* (TB *taṅkw*), which have no final *-u* either; and just like TA *ṅonk* they have undergone a change in vowel quality instead. TA *tuṅk* even shows affection of the vowel by **kʷ* across an intervening velar nasal *ṅ* [ŋ], whose presence also separates TA *oṅk* from the likes of TA *saku*, *maku* and *käntu*. The development can thus be considered as one from pre-TA *[aŋʷkʷ] to [oŋk], with loss of the labiovelar articulation upon vowel affection (cf. Kim 1999: 167–168). Considering the lack of affection in TA *saku* and *maku* (rather than ***sok* and ***mok*), the presence of certain clusters appears to be a prerequisite for rounding of PT **e* or pre-TA **a* to TA *o* by a following labiovelar.

A similar case is found in TA *śorkmi* 'strings', which derives from a PT stem **śerkʷ-* reflected in TB *śerkw* 'cord, string' (see Malzahn 2014: 87–90 for more details) with affection of PT **e* to TA *o*. The only element that could have caused this vowel change is the labiovelar, which was subsequently lost, like in the case of TA *oṅk* from PT **enkʷe*. It thus turns out that the effect of the labial quality could traverse an intervening *r* as well as [ŋ], but did not apply when **kʷ* immediately followed the vowel PT **e*. Only the "weaker" vowel PT **ə* was more generally affected by a following labiovelar, as in TA *yuk* 'horse' from PT **yəkwe* and TA *tuṅk* 'love' from PT **tənkwə*.

The case of TA *opṣäly* '(fit) season; time for action (of a Buddha)' and TA *nokte* 'at night' together with *noktiṃ* 'last night' is somewhat different again, since the TB cognates

[34] Pinault takes TA *oṅk* as the result of *u*-umlaut from an earlier **enku*. TB *eṅkwe* then constitutes a different, independent development, a secondary derivation from the same stem that was spared from *u*-umlaut (Pinault 2008: 452). It seems preferable to derive the word for 'man' in Tocharian A and B from the same Proto-Tocharian source, however, which for Tocharian B certainly can only be reconstructed as **enkʷe*.

*ekṣalye* and *nekciye* to not have a labiovelar attested. This may be due to spelling inconsistencies known to plague the writing of $k_u$ (e.g., Fellner 2005) or late changes to this cluster over the course of Tocharian B attestation (see Peyrot 2008 on a number of such developments), but the Tocharian A vowels can be explained differently as well. For TA *opṣäly* and TB *ekṣalye*, Adams reconstructs a PT **ekʷsʸəlye* to achieve the rounding of the first vowel in Tocharian A (Adams 2013: 80-81), but in Tocharian A there was a general change from PT **kS* to TA *pS* that could provide a *p*; cf. TA *opsi*, the plural of PT **okso* 'ox'. The *p* that resulted from this may have affected the preceding PT **e* or pre-TA **a* in the same way as the original *p* in TA *opyāc* 'memory' beside TB *epiyac*, etc., discussed above.

TA *nokte* and *noktiṃ*, meanwhile, have an additional third cognate in TA *nakcu* 'last night, at night'. This is a closer cognate to TB *nekciye* and does not show a TA *o*, i.e., not ***nokcu*. If a labiovelar had affected the vowel in TA *nokte* and *noktiṃ* from **nekʷt-*, the same effect should have been expected in TA *nakcu* if from **nekʷtʸ-*. Since this is not the case, a different explanation may be preferred: the vowel TA *o* in *nokte* and *noktiṃ* can be explained as the result of *u*-umlaut, based on an earlier *u*-stem PIE **nokʷtu-*. The resulting pre-PT **nektu-* became **noktə-* to become the base of TA *noktiṃ*, while the derived adjective **nokʷtewyo* yielded PT **nektʸəwʸe*, whence TB *nekciye* and TA *nakcu* (Pinault 1990: 184–190).

Finally, in TA *koṃ* (TB *kweṃ*), the oblique singular of *ku* 'dog', there is a clear change from PT **e* to TA *o* next to an original labiovelar. It appears that a labiovelar was not preserved in word-initial position next to pre-TA **a* (**kʷan*) and merged with the vowel. No similar process is seen in TA *śtwar* '4' (not ***śtor*) from PT **kʸətwerə*, suggesting that the effect was caused by PT **kʷ* in particular, rather than by any consonantal **w*.[35]

This concludes our overview of rounding phenomena next to labials and labiovelars in Tocharian A. While a few changes remain obscure, it should be clear that no two distinct vowels for PT **e* need to be reconstructed to account for these developments (*pace* Ringe 1990): such a distinction does not properly solve the remaining issues either. This is also clear from the limited labial affection in Tocharian B, treated in the next subsection.

#### 2.5.6.3 Rounding due to labials in Tocharian B

Just like Tocharian A, Tocharian B had its own marginal labial affection as well, specifically between two labials. There are two examples of this, although there are also two instances where it did not take place. Both of the examples belonging to this later

[35] The development of TA *por* from earlier **puwar* (TB *puwar*) does not quite belong here, since it involves the vowel PT **a* (TA *ā*) rather than **e* (TA *a*).

group are derived from the same verbal root *pəkʷ-*, so that some kind of paradigmatic levelling may have occurred. Perhaps only unstressed **ə* was affected.

64. PIE **kʷekʷlos* > "**kukʷlos*" > PT **kʷəkʷle* > TB *kokale*, TA *kukäl* 'wagon'
- PT **pə-kʷaka* > TB *pokkāka*, TA *pᵤkāk* imp. of *kaka-* 'call'
- cf. PT **pəkʷəmane* > TB *pkwamane*, not ***pokamane* or ***pkomane*, prt.ptc. of *pəkʷ-* 'cook'
- cf. PT **empəkwətte* > TB *empakwätte*, not ***ompokätte* vel sim., privative of *pəkʷ* 'cook'

In the available Tocharian A cognates, the vowel is reflected as *u* or as part of a labiovelar *kᵤ*, not as TA *o*. The word TB *troṅk*, TA *truṅk* 'cavity, hollow, inner part of the torso; cave' shows this same vowel correspondence between Tocharian B *o* and B *u*, but it does not quite share the same phonological environment, and its etymology is not certain (see Adams 2013: 341).

Some word-initial PT **e* were also rounded to TB *o*, similar to the change observed in Tocharian A (see above). A number of examples have a TB *o* in the second syllable as well, which would regularly have caused **e* to change to TB *o* due to umlaut, but four examples remain where no such explanation is applicable, namely *oṅkipṣe* 'shameless', *ompakwättäññe* 'unreliability', *ompalsko* 'meditation' *oṣṣale* 'north'.

- TB *oṅkipṣe* 'shameless' derived from TB *kwipe* 'shame'
- TB *omotruññaiṣṣe* 'southern', cf. TA *mäcriṃ* 'southern'
- TB *ompakwättäññe* derived from *empakwätte* 'unreliable'
- TB *ompalsko* 'meditation' derived from TB *palsko* 'thought'
- TB *ompostäṃ* 'afterwards' derived from TB *postäṃ* 'after'
- TB *omprotärtstse** 'related as brothers' derived from TB *procer*
- TB *oṣṣale* 'north' derived from TB *ṣale* 'mountain', cf. TA *ṣul* 'mountain' from PT **sʸwʸəle* (**wʸ* regularly changed to **y* in TB, which is lost in the initial cluster in **ṣyale* > *ṣale*, but is responsible for the gemination in *oṣṣale*)

It is very difficult to differentiate between TB *eṅkwe* 'man' from PT **enkʷe* and TB *oṅkipṣe* 'shameless' from earlier **enkʷipsʸe*, and between TB *empakwätte** /empə́kʷətte/ 'unreliable' and *ompakwättäññe* /ompə́kʷəttəńńe/ 'unreliability' from PT **empəkʷətte* by a regular sound change. Especially the latter cannot be differentiated on the basis of features like stress placement or syllable structure, but we may assume that these differences were caused by analogy and do not all reflect regular sound change. All instances do contain an intervening element between the affected vowel and the labial. The change in TB *oṣṣale* 'north' needs to have happened before PT **wʸ* had change into

TB *y*, after which no labial would have been present anymore to affect the preceding vowel. The list of examples below shows that, more commonly, TB *e* is found in this kind of environment.

- TB *eṅkwe* 'man', cf. TA *oṅk*
- TB *epastye* 'skillful', cf. TA *opäśśi*
- TB *epiṅkte* 'between', cf. TA *opäntaṣ*
- TB *epiyac* 'memory', cf. TA *opyāc*
- TB *eprer* 'sky', where the *e*- actually alternates with an *i*- as *iprer*
- TB *emalle* 'hot, warm', cf. TA *omäl* 'hot'
- TB *eprete* 'resolute, steadfast', cf. TA *pratiṃ* (without the prefix)
- TB *eplyuwai* 'swimming' derived from *pləw*- 'swim'
- TB *empreṃ* 'truth'
- TB *empakwätte* 'unreliable' derived from *pək*$^{w}$-
- TB *emparkre* 'wide(ly)', derived from TB *parkare* 'wide'
- TB *empalkatte* 'unworried', cf. TA *apälkāt*
- TB *empele* 'terrible, horrible', derived from *pele* 'way, law' → TA *empele*

The next subsection will continue with an opposite development of unrounding that seems to also have taken place in Tocharian A under some very specific circumstances.

#### 2.5.6.4 Potential unrounding in Tocharian A

One remaining thorny problem is the Tocharian A correspondences to TB *o*. In the non-first syllable, this is regularly TA *a*, while in the first syllable it is generally TA *o*.

1 TB obl. *pokai* vs. TA obl. *poke* 'arm'
59. TB *klyomo* vs. TA *klyom* 'noble'
84. TB *okt* vs. TA *okät* 'eight (8)'
85. TB m.pl. *poñc* vs. TA m.pl. *poñś* 'all' (and other forms of the paradigm)
109. TB (arch.) *toṃ* vs. TA *tos-äṃ* pronoun stem f.pl.
124. TB *okso* vs. TA pl. *opsi*, sg. *okäs** 'ox'

- TB obl. *kroścaṃ* vs. TA obl. *krośśäṃ* 'cold'
- TB *yoñiye* vs. TA *yoñi* 'path'
- TB *kroriya* 'horn' vs. TA *kror* 'crescent of the moon'
- TB *moko* vs. TA *mok* 'old'
- TB *oṅkolmo* vs. TA *oṅkaläm* 'elephant'
- TB *oṅkrocce* vs. TA *oṅkrac* 'immortal'
- TB *kos* vs. TA *kos* 'how much'

- TB *yok* vs. TA *yok* 'colour; hair'

There are just two examples with TA *a* in the first syllable instead of the *o* found in a direct Tocharian B cognate. In both cases, the vowel is a precious reflex of PIE *$*eh_2$*/$\bar{a}$. The discrepancy between the words in the two lists below is the main reason to distinguish the result of *$*eh_2$*/$\bar{a}$ from the product of *u*-umlaut as separate Proto-Tocharian vowels *$*å$* and *$*o$* or alternatively *$*o$* and *$*ǫ$* (cf. Pinault 2008: 420; Ringe 1996: 98).

2. TB *procer* vs. TA *pracar* 'brother' from PIE *$*b^hreh_2tēr$*
115. TB *ost* vs. TA *waṣt* 'house' from PIE *$*u̯eh_2stu$-*

This same correspondence is found in a number of *o*-presents [class 4], but the origins of those vowels are probably of a different nature, and they are discussed in subsection 2.5.4. It is problematic, however, that TA *pracar* 'brother' and TA *waṣt* are the only two examples of a development from PIE *$*eh_2$*/$\bar{a}$ into TA *a*, while the other examples show TA *o*. From the examples listed above, this concerns the following:

1 TB obl. *pokai* vs. TA obl. *poke* 'arm' from PIE *$*b^heh_2ǵ^hu$-*
85. TB m.pl. *poñc* vs. TA m.pl. *poñś* 'all' from PIE *$*peh_2$-nt-*
109. TB (arch.) *toṃ* vs. TA *tos-äṃ* pronoun stem f.pl. from PIE *$*teh_2$-ns*

TA *poke* 'arm' (TB *pokai*) from PIE *$*b^heh_2ǵu$-* could conceivably have its *o* in Tocharian A from a kind of *u*-umlaut (cf. the *u*-umlaut of pre-PT *$*e$* in 2.5.1), but there is no parallel for *u*-umlaut of *$*eh_2$*/$\bar{a}$. The word that is structurally most similar is PIE *$*u̯eh_2stu$-*, which has a different vowel correspondence. In TA *poke* 'arm' and *poñś* 'all', the initial labial consonant *p-* may have had a rounding effect, but the vowel of TA *tos-äṃ* cannot be easily explained in the same way. It is difficult to assume some kind of umlaut due to a suffixed element in this word because the correspondence of TB *-ṃ* (/n/) to TA *-s* is normally found only in word-final position. This development thus contradicts the evidence of TB *procer*, TA *pracar* 'brother' and TB *ost*, TA *waṣt* 'house'. It therefore indicates that a reconstruction of two different *o*-phonemes PT *$*o$* and *$*å$* from distinct Proto-Indo-European sources might not be sufficient to differentiate these varying correspondences.

Burlak and Itkin suggested effects of the initial labial on the vowel in TA *waṣt*, with a sound change from PT *$*#wo$-* to TA *#wa-* and TB *#o-* (Burlak & Itkin 2003: 31). This might work, as a dissimilation is phonetically plausible and there are no counterexamples. The explanation given by Burlak and Itkin for the TA *a* in *pracar* 'brother' is unacceptable, however: they claim that this goes back to PT *$*pərot^yer$*, so that the PT *$*o$* stood in the second syllable, where it was regularly changed to TA *a* as in TB *oṅkolmo* vs. TA *oṅkaläm*

'elephant' (Burlak & Itkin 2003: 33). This explanation relies on the assumption that all pre-Proto-Tocharian clusters of the type **CR* were broken up with epenthesis to **CəR*, for which there is simply no evidence. Furthermore, the lack of a TA *a* in TA *klyom* 'noble' (not ***klyam* from ***kəlyomo*) and *kror* 'crescent of the moon' (not ***krar* from ***kərorə*) directly contradicts such an explanation for TA *pracar*. The TAB *o* in PT **klyomo* and **krorə* is not derived from late sporadic simplifications of a diphthong **eu* to **o*, as Burlak and Itkin claim (2003: 26–27), but from umlaut (see 2.5.1 and 2.5.2 with references).

If we assume with Burlak and Itkin a dissimilation from PT **wo* to TA *wa* in TA *waṣt* 'house', perhaps a similar dissimilation can have affected PT **pro* in TA *pracar* 'brother' from PT **protʸer*. This would amount to a potential rule whereby **o* was unrounded in Tocharian A in certain labial environments. However, **po* and **mo* would clearly have remained unaffected; cf. TA obl. *poke* 'arm' (TB *pokai*) and *mok* 'old' (TB *moko*). If dissimilation occurred in **pro* one might also expect it in **po*, perhaps more strongly because the labial is adjacent to the vowel. However, this may not necessarily be the case, since we can see in the development of TA *oṅk* 'man', *śorki* 'mortal fear' and *śorkmi* 'cords' as opposed to *maku* 'nails' and *saku* 'pus', that **kʷ* had actually had a different effect on an adjacent vowel (in this case rounding) when there was an intervening resonant present. The situation for **pro* as opposed to **po* could thus have been similar.

There is an additional, though more uncertain example which may further support this sound change: TA *kroś* 'cold' is once attested as $k_u raś$ instead, which displays both an initial TA labiovelar $k_u$ and a change of the vowel to TA *a*. Itkin (2011: 251) has characterized TA $k_u raś$ as "tout simplement due à une faute de scribe". The variant *kroś* in the nominative singular is attested twice for sure, two more times potentially (damaged context), and $k_u raś$ is found only once.[36] It would be a rather strange scribal error, however, so that one may suppose that $k_u raś$ constitutes a real form, and reflects a real phonological process.

If a sound law changing PT **wo-* and **pro-* (and perhaps ?**kʷro-*) to TA *wa-* and *pra-* (and perhaps *kʷra-*) is assumed, this would eliminate the only remaining reason to reconstruct the Proto-Tocharian reflex of PIE $*eh_2$ as different from the PT **o* that arose due to *u*-umlaut of **e* and different from the second-syllable **o* in PT **wəlo*, **klʸomo*, etc. from PIE **ō*. They could furthermore be written exactly as the *o*-vowel resulting from *u*-umlaut and *o*-umlaut, all simply PT **o*. The Proto-Tocharian vowel **å* still existed, but it was the result of contraction and umlaut caused by the contracted vowel (see subsection 2.5.4 and also 5.4.1). Its correspondences are uniformly TB *o* : TA *a*. Because the uncertainties involved cannot easily be resolved, it might be prudent to provide reconstructions with PT **å* as an alternative, thus PT **wostə* (**wåstə*) 'house'.

[36] The meaning in all cases seems clear.

### 2.5.7 Summary

Two types of umlaut, namely *u*-umlaut of **e* and *a*-umlaut of **e*, applied in the same way in Tocharian A and B, and can thus be dated to pre-Proto-Tocharian as a set of sound laws (see 2.5.1 and 2.5.5). There is no difference in the application of these various types of umlaut to the Proto-Indo-European vowels **ē* and **o*, which can thus have merged as (pre-)PT **e* without any issue. The labial affection that turns some PT **e* into TA *o* rather than TA *a* appears to have operated in certain restricted environments and was not demonstrably confined to a PT **ë* or **e* as reflexes of PIE **o* or **ē* (see 2.5.6.1 and 2.5.6.2).

The Proto-Indo-European vowel **ō* seems to have caused *o*-umlaut of pre-PT **u*, although there are only two potential examples (see subsection 2.5.2). A similar *o*-umlaut of pre-PT **e* caused by a following PIE **eh₂/ā* clearly affected Tocharian B, but the evidence for this umlaut in Tocharian A is uncertain (see subsection 2.5.3). It is difficult to determine the reflex of the vowels PIE **eh₂/ā* in Proto-Tocharian was merged with the PT **o* from *u*-umlaut of **e* and *o*-umlaut of **u*, or if it became identical with the contracted vowel PT **å*. Three examples suggest that PIE **eh₂/ā* was identical with **o* and two suggest a merger with **å*. The TA *a* rather than *o* in TA *pracar* 'brother' (TB *procer*) and TA *waṣt* 'house' (TB *ost*) could be the result of a secondary, dissimilatory unrounding phenomenon that affected the sequences **pro-* and **wo-*, but it does not seem possible to come to a firm conclusion (see 2.5.6.4).

## 2.6 Information from language contact

Parts of the relative chronology can be put in a broader perspective on the basis of Tocharian words that were borrowed from early Iranian languages. The phonological characteristics of these loanwords can help determine the state of early Tocharian phonology at the time of contact with Iranian. Some Tocharian words derive from another, unknown prehistoric language, thought to be related to the language of the BMAC, and a little additional information may be gleaned from this as well.

### 2.6.1 Early Iranian languages

At a certain point, (pre-)Proto-Tocharian came into contact with an Old Iranian language, based on Old Iranian loanwords that have entered into the Tocharian lexicon. Examples include TB *perne*, TA *paräṃ* 'glory' from Old Iranian **farnah-*. The earliest Old Iranian donor language has been labelled "Old Steppe Iranian" by Chams Bernard (2023), and the contact seems to have taken place a short while before the split of Proto-Tocharian into pre-Tocharian A and pre-Tocharian B. Bernard has established regular sound correspondences between the Old Iranian donor language and the Tocharian borrowings, where, for example, each Old Steppe Iranian */a/* corresponds to a Proto-Tocharian **e*, while Old Steppe Iranian */ā/* corresponds to a Proto-Tocharian **a*. On the

basis of the regularity of these vowel correspondences, Bernard argues that the contact with Old Steppe Iranian took place over only a short, concentrated period of time, in which neither it nor (pre-)Proto-Tocharian underwent significant phonological change (Bernard 2023: 242–246).

The Old Steppe Iranian layer of borrowings in Tocharian can give us some insight into which Tocharian sound laws had already run their course. Based on the borrowing TB *yentuke* 'India(n)' from putative Old Steppe Iranian **i̯anduka-* or **i̯enduka-* (Bernard 2023: 60-62), we can tell that *u*-umlaut of **e* applied before contact with Old Steppe Iranian. Since *u*-umlaut changed earlier **e-u* into **o-u*, this word should have become ***yontuke* if it had been borrowed before *u*-umlaut. This indicates that the phonological stage of the pre-Proto-Tocharian language as it was when it came into contact with Old Steppe Iranian was already quite advanced, since *u*-umlaut of **e* was a relatively late development that took place after palatalization. Additionally, TB *eñcuwo*, TA *añcu** 'iron', borrowed from an early form of Khotanese (Peyrot, Dragoni & Bernard 2022) shows no *u*-umlaut either. The apparent *u*-umlaut in TB *mot* 'alcoholic beverage', borrowed from a descendant of Proto-Iranian **madu-* likely took place in the Old Steppe Iranian source language, and not in Tocharian itself (Bernard 2023: 103–104).

A similar argument can be made for *a*-umlaut of **e* (PIE **o*, **ē*), on the basis of the archaic attestations of Tocharian B *ainake* 'evil' as *eynāke*, borrowed from Old Steppe Iranian **a̯ināka-* (Bernard 2023: 27–29). If *a*-umlaut had taken place after the adoption of this word, no *e* in archaic TB *eynāke* would have been expected, as it should have changed into *a* as well: archaic TB ±***āynāke*. Both *u*-umlaut and *a*-umlaut occurred after palatalization, so that the contact itself can also be dated after palatalization. Indeed, the way in which certain Old Steppe Iranian sounds were adopted corroborates this; e.g., TB *waipecce* 'property, possession' from OSIr. **hwai-paϑya-* and TB *śāte* 'rich' from OSIr. **čyāta-* 'happy' (Bernard 2023: 73–75).

The situation is more nuanced in the case of the *o*-umlaut of pre-PT **u* (type **ukso* to TB *okso* 'ox'), because two borrowings from early Khotanese-Tumshuqese seem to show this sound change as well. Two further borrowings did not undergo this change, however, so that no clear conclusions can be drawn. Perhaps TB *cowo** 'stealing' and TB *koto** 'hole, pit; excrement (?)' were borrowed at a relatively earlier stage, but the exact timeline of developments remains elusive.

- TB *cowo** 'stealing' from pre-Khot. acc.sg. **dyūwu*, derived from PIr. **dab-ya* 'stealing', to be compared with LKhot. *dyūka-* 'robber' (Dragoni 2023: 118–120).
- TB *koto** 'hole, pit; excrement (?)' from PTKhot. or pre-Khot. acc.sg. **gūϑu*, whence Khot. *gūha-* 'excrement, faeces' (Dragoni 2023: 103–105).

- TB *wicuko* (not ***wicoko*) 'cheek, (jaw)bone' form pre-Khot. **wi-jwäka-*, a derivative of K °*jv-* 'chew' based on the Tocharian borrowing (Dragoni 2023: 169–170).
- TB *sumo* (not ***somo*) 'libation (?)' from LKhot. *ysūma-* 'broth' (Dragoni 2023: 196–197).

The situation for *o*-umlaut of pre-PT **a* is similar, since there are Khotanese-Tumshuqese borrowings that have undergone this sound change in Tocharian B. This may point to a relative late date for this sound change, perhaps restricted to pre-Tocharian B, after Proto-Tocharian had ceased to exist as a unitary language.

- TB *kompo** '?' perhaps via **kampo* from OKhot. acc.sg. *ggaṃphu* 'plain', or its PTKhot., pre-Khot. or OK ancestor (Dragoni 2023: 106)
- TB *kontso** '?' perhaps via **kantso* from OK *ggaṃjsā-* 'flaw', or its PTK or pre-K ancestors. (Dragoni 2023: 105)
- TB *koro* 'donkey' perhaps via **karo* from Khot. acc.sg. *kharu* 'donkey', alternatively from PTK **gōru* 'wild ass', in which case no umlaut is needed (Dragoni 2023: 106–107)

An aspect of the Tocharian phonology that remains difficult to pin down in the timeline is the general voicelessness in the stop system. Voiceless stops are used in borrowings from Old Steppe Iranian and Khotanese-Tumshuqese when the source has a voiced stop, but whether this is caused by an incorporation of the borrowings into the native sound system, or due to a later loss of the voicing opposition after the adoption of these words is not strictly visible. In both scenarios, the attested situation would show the stops as voiceless.

However, an argument in favour of general devoicing prior to contact with at least Indo-Iranian can be made as follows: Starting from contact with Old Steppe Iranian, Tocharian would have continuously been in contact with some form of Iranian until its eventual attestation. Since all the Iranian languages with which Tocharian was in contact, like Khotanese and Sogdian, also had a voicing opposition, there would have been no pressure for Tocharian to undergo a general devoicing of its stops. Rather, a retention of a voicing distinction, if still present at the time, should have been facilitated by the abundance of other languages with this distinction in the vicinity. This suggests that the devoicing of the Tocharian stops took place at least before contact with Old Steppe Iranian.

It is also clear that the sequence PIE **t̯i̯/dʰi̯* had already become pre-PT **tˢ*, since Old Steppe Iranian **t̯i̯* or **d̯i̯* became **cc* instead (e.g., OSIr. **gr̥d̯i̯a-* → TB *kercci* 'palace'; Bernard 2023: 157–159). All available information thus points to an advanced stage of

development in the Proto-Tocharian phonological system at the time of contact with Old Iranian, meaning that the specific Tocharian sound changes had developed at an even earlier stage already.

### 2.6.2 "BMAC language"

Potentially somewhat farther back in time, Tocharian was in contact with a language related to the unknown language of the BMAC culture. This hypothesis is based on a set of similar loanwords that are found both in Tocharian and in Indo-Iranian languages that would have been in more direct contact with the BMAC Culture known from archaeology (e.g., TB *kercapo* ~ Vedic *gardabhá-* 'donkey', Bernard 2023: 199–232 with extensive discussions and references)
Bernard has argued that phonological differences in the borrowings that are not easily ascribed to either Tocharian or Indo-Iranian suggest that the language encountered by the early Tocharians was not quite the same as the one encountered by the Indo-Iranians farther to the south-west, but rather a related variety (Bernard 2023: 232–241).

Although little is known directly about the BMAC language (family), the shape of the borrowings in Tocharian can give us some crucial information about the state of the Tocharian phonological system at the time of borrowing. As with the Old Steppe Iranian borrowings, it looks like Tocharian phonology had developed to a relatively advanced state already, with palatalized consonants being present (e.g., in TB *kercapo* 'donkey', TB *kroṅkśe* 'bee', TB *śerwe* ~ TA *śaru* 'hunter'; Bernard 2023: 215–228). However, no sequences liable to umlaut are found in the words that have so far been identified as borrowings from a variety of the BMAC language, so that it not possible to determine whether or not those developments had already taken place. The nature of the stop system of the BMAC language is also difficult to accurately determine, so it cannot be conclusively argued that Tocharian should have lost its stop distinction before contact with this language.

### 2.6.3 Summary

The general picture as can be gleaned from traces of interactions between Tocharian and Iranian languages like Old Steppe Iranian and early Khotanese-Tumshuqese in the later stages of Proto-Tocharian is that most major sound changes had occurred at an even earlier stage. Palatalization had certainly already taken place, and at least two types of umlaut seem to have occurred before contact with at least Old Steppe Iranian as well, namely *u*-umlaut of **e* (on account of TB *yentuke* 'India(n)') and *a*-umlaut of **e* (on account of archaic TB *eynāke* 'evil'). It cannot be ascertained whether *o*-umlaut of **u* had occurred, or if **u* and **i* had merged as **ə*. If the early Khotanese-Tumshuqese borrowings TB *cowo** 'stealing' and TB *koto** 'hole, pit; excrement' indeed show *o*-umlaut

of **u*, this sound change may have occurred relatively later. The Tocharian B *o*-umlaut of PT **a* is reflected in a number of Khotanese-Tumshuqese borrowings found in this language, and it may therefore have been a late sound change as well, although the available data is sparce and partially contradictory.

Palatalization is confirmed by borrowings from an unattested language related to the "BMAC-language" that donated loanwords to various early Indo-Iranian languages as well. Due to the limited information available about this language, not much else can be said on the basis of the current evidence.

## 2.7 Overview of Tocharian relative chronology

The developments that occurred in pre-Proto-Tocharian can be roughly split in two groups: those that took place before palatalization, and those that took place after palatalization. The relatively more significant changes belong to the first group, and the second group mostly consists of umlaut phenomena and a few additional vowel changes, as well as further shifts in the palatalized pronunciation of certain consonants, e.g., **s$^{y}$* > TAB *ṣ*, **k$^{y}$* > TAB *ś*.

We will begin our overview once again with the Proto-Indo-European phonological system. The phonemes that were affected by changes at an early stage were the laryngeals and the various velar series. A few changes are datable relative to laryngeal developments, which as a whole belong before palatalization.

Table 2.29: The Proto-Indo-European consonant inventory.

| PIE | labial | dental | palatal | palatovelar | velar | labiovelar | laryngeal |
|---|---|---|---|---|---|---|---|
| stop | *p b$^{h}$* | *t d d$^{h}$* | | *ḱ ǵ ǵ$^{h}$* | *k g g$^{h}$* | *k$^{w}$ g$^{w}$ g$^{wh}$* | |
| unknown | | | | | | | *h$_1$ h$_2$ h$_3$* |
| sibilant | | *s* | | | | | |
| nasal | *m* | *n* | | | | | |
| liquid | | *l r* | | | | | |
| glide | | | *i̯* | | | *u̯* | |

Table 2.30: The Proto-Indo-European vowel inventory.

| PIE | front | back |
|---|---|---|
| high | *i* | *u* |
| mid | *e ē* | *o ō* |
| low | (*a*) | |

Changes involving laryngeals seem to have happened particularly early on. This caused an early increase in the number of vowels in the system due to developments such as *$h_2e$ to *$a$, *$eh_2$ to *$\bar{a}$, *$ih_1$ to *$\bar{\imath}$, $\bar{o}$# > $\bar{u}$#, and the frequency of certain vowel phonemes increased as well by developments like those from *$eh_1$ to *$\bar{e}$. The vowels that arose in this way behave exactly like their more original counterparts in further developments, so that the colouring effects of the laryngeal can be considered particularly early. After laryngeal colouring, the three laryngeals can no longer be distinguished in the phonology, so that they might be merged as *$H$ and eventually lost (2.3.1). The vocalization of syllabic resonants from *$\underset{\circ}{R}$ to *"$\dot{\imath}R$" added one further vowel to the system (2.3.2), but it is difficult to date exactly. It had certainly taken place before the disappearance of laryngeals and before phonemic palatalization. An additional source of the new vowel *$\partial$ may have been the laryngeal in sequences *$Hi$ (2.4.2). Other changes either did not interact with the laryngeal developments, or occurred demonstrably later.

Table 2.31: An overview of early vowel changes from Proto-Indo-European to pre-Proto-Tocharian.

| PIE | > post-PIE | PIE | > post-PIE |
|---|---|---|---|
| $i$ | $i$ | $ih_1$? | $\bar{\imath}$ |
| $e$ | $e$ | $\bar{e}$, $eh_1$ | $\bar{e}$ |
| $u$ | $u$ | -$\bar{o}$# | -$\bar{u}$# |
| $o$, $h_3e$? | $o$ | $\bar{o}$, $oH$ | $\bar{o}$ |
| $\underset{\circ}{H}$, $h_2e$ | $a$ | $eh_2$ | $\bar{a}$ |
| $\underset{\circ}{R}$ | "$\dot{\imath}R$" | | |

Table 2.32: The Proto-Indo-European vowel inventory (left) and the new early pre-Proto-Tocharian vowel system after the first changes (right).

| PIE | front | back | | front | central | back |
|---|---|---|---|---|---|---|
| high | $i$ | $u$ | | $i$ $\bar{\imath}$ | "$\dot{\imath}$" | $u$ $\bar{u}$ |
| mid | $e$ $\bar{e}$ | $o$ $\bar{o}$ | > | $e$ $\bar{e}$ | | $o$ $\bar{o}$ |
| low | ($a$) | | | | $a$ $\bar{a}$ | |

The depalatalization of palatovelars probably took place very early, but it cannot be dated relative to any change other than the general Tocharian palatalization, which it preceded (2.3.3.1). Delabialization of labiovelars next to PIE *$e$ and *$\bar{e}$ must have happened before *$\bar{e}$ merged with *$\ddot{e}$ from PIE *$o$ as (pre-)PT *$e$, which occurred after with palatalization (2.4). PIE *$d$ changed in some uncertain way (see 2.3.4.4) at least in clusters before resonants before assibilation of PIE *$t\underset{\wedge}{\imath}$ and *$d^h\underset{\wedge}{\imath}$ took place, and in its turn, assibilation also happened before palatalization (2.3.4.2 and 2.3.4.3). The putative Tocharian version of Grassmann's Law would probably have taken place even earlier

(2.3.4.5). The details of the developments of PIE **d* in Tocharian remain unclear, however, and I have decided to provisionally use the symbol "*ḋ*" to mark an altered state of PIE **d* that is of unknown quality (see 2.3.4.4 and 2.3.4.6).

These changes happened before obstruents were generally devoiced and deaspirated. The simplification of the cluster PIE **mb$^h$* to **m* belongs here too (see 2.3.4.1). Since no distinctions in voicing or aspiration are necessary to understand the outcomes and effects of palatalization in general, it can be assumed that the loss of these features occurred before palatalization, but this cannot be confirmed. A summary is provided in Table 2.33 with an overview of the new consonant inventory in Table 2.34.

Table 2.33: An overview of early changes to the Proto-Indo-European consonants and their relative chronology.

| Sound changes | | Chronology | Treated in |
|---|---|---|---|
| A1 ? *d$^h$VD$^h$* | > *dVD$^h$* | if this did happen, before A2., D. | 2.3.4.5 |
| A2 *d* | > "*ḋ*" | before A3, B, D. | 2.3.4.3, 2.3.4.4 |
| A3 *du̯, di̯, dr* | > *w, y, r* | likely before B, D. | 2.3.4.3 |
| B *ti̯, d$^h$i̯* | > "*ċ*" ("*ż$^h$*") | before D. | 2.3.4.2 |
| C ? *mb$^h$* | > *m* | before D. | 2.3.4.1 |
| D1 *p, b$^h$* | > *p* | | 2.3.4 *passim* |
| D2 *t, d$^h$* | > *t* | | 2.3.4 *passim* |
| D3 *ḱ, ǵ, ǵ$^h$* | > *k* | | 2.3.4 *passim* |
| D4 *k, g, g$^h$* | > *k* | | 2.3.4 *passim* |
| D5 *k$^w$, g$^w$, g$^{wh}$* | > *k$^w$* | | 2.3.4 *passim* |

Table 2.34: The pre-Proto-Tocharian consonant inventory after the first consonant changes had taken place.

| pre-PT | labial | dental | palatal | velar | labiovelar |
|---|---|---|---|---|---|
| stop | *p* | *t*<br>"*ḋ*" | | *k* | *k$^w$* |
| affricate | | "*ċ*" | ~ "*ċ*" | | |
| sibilant | | *s* | | | |
| nasal | *m* | *n* | | | |
| liquid | | *l* | | | |
| trill | | *r* | | | |
| glide | *w* | | *y* | | (*w*) |

One further development that seemingly took place after the loss of laryngeals but before palatalization was a development from PIE **wi* and **k$^w$i* to **wu* and **k$^{(w)}$u*. This, in turn,

preceded a possible development of PIE **e* to **i*, as it eventually merged with original **i* that had not undergone the aforementioned centralization. The full merger of these two vowels may have happened around the time of palatalization, or followed soon thereafter (2.4.2).

Some additional vowel changes that took place before palatalization involve the loss of distinctive vowel length. This process involved an unrounding of PIE **o* (potentially including **$h_3e$*) to a pre-PT **ë*, and a shift from PIE **$eh_2$*/**ā* to **å*. With the shortening of PIE **ē* (including **$eh_1$*) to **e*, perhaps original short **e* (including **$h_1e$*) had already merged with **i*. The very marginal early **ī* ends up as equivalent to PIE **ei* and may be represented as **iy*; in parallel, marginal **ū* may be represented as **uw* and PIE **eu* as **iw*.

Table 2.35: An overview of further early changes to the vowel pre-Proto-Tocharian vowel system and their relative chronology.

| Sound changes | | Chronology | Treated in |
|---|---|---|---|
| A1 *$k^we$, $k^w\bar{e}$* | > *ke, kē* | uncertain change | 2.3.3.2 |
| A2 *{w,$k^w$}i* | > *{w,$k^w$}u* | before C1. | 2.4.2 |
| B1 *o* | > *ë* | before B2. | 2.4.4 |
| B2 *ō* | > *o* | | 2.4.5.1 |
| C1 *i, e* | > *i* | before D. | 2.4.2 |
| C2 *ei, ew* | > *iy, iw* | | 2.4.4 |
| C3 *ī, ū* | > *iy, uw* | after A2. | 2.4.4 |
| C4 *Hi* | > *əy* | | 2.4.2 |
| D *ē* | > *e* | | 2.4.4 |
| E *ā* | > *å* | | 2.4.4 |

Table 2.36: Two stages of the pre-Proto-Tocharian vowel inventory before vowel length was given up (left) and after vowel length was given up (right).

| post-PIE | front | central | back | | front | central | back |
|---|---|---|---|---|---|---|---|
| high | *i ī* | *ə* | *u ū* | | *i iy* | *"i"* | *u uw* |
| mid | *e ē* | | *o ō* | > | *e* | *ë* | *o* |
| low | | *a ā* | | | | *a* | *å* |

Palatalization involved a merger of the vowels **i* and **"i"*, and of **e* and **ë*. In the resulting system, these can be represented as new **i* and **e* respectively. Almost every consonant gained a palatalized counterpart. Only palatalized counterparts to **r* and **"ć"* cannot be established. Palatalized original labiovelars are identical to palatalized **k* (2.3.3.2), and these eventually merge with the palatalized **"ḋy"* (counterpart to **"ḋ"*) as TAB *ś*.

Developments that took place before palatalization are the shift from **o* to **ë*, the loss of vowel length, backing of certain PIE **i* before (labio)velars, assibilation and the depalatalization of palatovelars. Any change that can be dated before these developments also took place before palatalization, namely laryngeal colouring and the vocalization of syllabic resonants, the possible early delabialization of labiovelars next to PIE **e*, the change (whatever the exact nature, see 2.3.4.4) of PIE **d* to **"ḋ"* and the putative Tocharian Grassmann's Law. The phonological system resulting from the Tocharian palatalization is given in the tables below.

Table 2.37: An overview of sound changes that took place as phonological palatalization came into effect.

| Sound changes | | Chronology | Treated in |
|---|---|---|---|
| A $CV_{front}$ | $C^yV$ | before B. | 2.4 esp. 2.4.1 |
| B1 *i*, *"i"* | > *i* | | 2.4.2 |
| B2 *e*, *ë* | > *e* | | 2.4.4 |

Table 2.38: The pre-Proto-Tocharian consonant inventory after phonemic palatalization.

| pre-PT | labial | dental | palatal | velar | labiovelar |
|---|---|---|---|---|---|
| stop | *p* $p^y$ | *t* | $t^y$ | *k* $k^y$ | $k^w$ |
| | | *"ḋ"* | *"ḋy"* | | |
| affricate | | *"c"* | | | |
| sibilant | | *s* | $s^y$ | | |
| nasal | *m* $m^y$ | *n* | $n^y$ | | |
| liquid | | *l* | $l^y$ | | |
| trill | | *r* | | | |
| glide | *w* $w^y$ | | *y* | | |

Table 2.39: The pre-Proto-Tocharian vowel inventory before (left) and after (right) phonemic palatalization.

| pre-PT | front | central | back | +palatalization | front | central | back |
|---|---|---|---|---|---|---|---|
| high | *i iy* | *"i"* | *u uw* | | *i iy* | | *u uw* |
| mid | *e* | *ë* | *o* | > | *e* | | *o* |
| low | | *a* | *å* | | | *a* | *å* |

After palatalization had taken place, we see *u*-umlaut of pre-PT **e* (from PIE **ē* and **o*) and *o*-umlaut of pre-PT **u* (from PIE **u*), which were both followed by the centralization of **u* to **ə* and its concomitant merger with earlier **i* (from PIE **i*, **e*, and secondarily

from vocalized PIE syllabic resonants and earlier centralized **i*). The *a*-umlaut which affected pre-PT **e* (from PIE **ē* and **o*) also took place after palatalization had run its course. The potential changes from pre-PT **o* (PIE non-final **ō*) to PT **a* and from pre-PT **å* (PIE **eh₂/ā*) to PT **o*, are uncertain, but they would also have taken place after palatalization and in that order (2.4.5.2).

Table 2.40: An overview of late vowel changes in pre-Proto-Tocharian, like umlaut and contraction.

| Sound changes | | Chronology | Treated in |
|---|---|---|---|
| A1 *e-u* | > *o-u* | before B. | 2.5.1 |
| A2 *u-o* | > *o-o* | before B. | 2.5.2 |
| A3 *e-a* | > *a-a* | | 2.5.5 |
| B *i, u* | > *ə* | | 2.4.4 |
| C1 *a.e* | > ? *å* | | 2.5.4, 5.4.1 |
| C2 *a-å* | > *å-å* | | 2.5.4 |

Table 2.41: The late pre-Proto-Tocharian vowel inventory (left) developing into the Proto-Tocharian vowel inventory (right).

| pre-PT | front | central | back | to PT | front | central | back |
|---|---|---|---|---|---|---|---|
| high | *i i̯y* | | *u uw* | | *i* (=*əy*) | *ə* | *u* (=*əw*) |
| mid | *e* | | *o* | > | *e* | | *o* |
| low | | *a* | *å* | | | *a* | (*å*) |

The character of a few consonants was further changed as well. The palatalized **sʸ* became a retroflex TAB *ṣ* [ʂ] while palatalized **kʸ* and **“ġʸ”* became a palatal sibilant *ś* [ɕ]. In the case of **kʸ*, this development may have taken place through an intermediary stage as a palatal affricate *[cɕ] or *[tɕ], so that the full palatalization of **tʸ* to TAB *c* [tɕ] probably occurred after **kʸ* had become de-affricated.

Several final consonants and consonant clusters had been lost before this point, but the exact dating in the relative chronology is difficult. This apocope did not affect the nature of the consonant inventory at any point and has therefore been left out of consideration in this overview.

Table 2.42: An overview of the remaining changes that took place in the consonants to yield the Proto-Tocharian system.

| Sound changes | | Chronology |
|---|---|---|
| A1 *sʸ* | > *ṣ* [ʂ] | before A2. |
| A2 *kʸ*, "*ḋʸ*" | > *ś* [ɕ] | before A3. |
| A3 *tʸ* | > *c* [tɕ] | |
| B "*ḋ*", "*ċ*" | > *tˢ* | |

Table 2.43: The Proto-Tocharian consonant inventory.

| PT | labial | dental | retroflex | palatal | velar | labiovelar |
|---|---|---|---|---|---|---|
| stop | *p pʸ* | *t tʸ* | | | *k* | *kʷ* |
| affricate | | *ts* < "*ḋ*", "*ċ*" | | | | |
| sibilant | | *s* | *ṣ* < *sʸ* | *ś* < *kʸ*, "*ḋʸ*" | | |
| nasal | *m mʸ* | *n* | | *nʸ* | | |
| lateral | | *l* | | *lʸ* | | |
| trill | | *r* | | | | |
| glide | *w wʸ* | | | *y* | | |

The great majority of these changes took place before pre-Proto-Tocharian came into contact with Iranian languages in the relatively late period preceding the split of Proto-Tocharian, and perhaps even before it came into contact with a language related to the BMAC language (2.6). This means that the developments that turned Proto-Indo-European phonology into what we recognize as Tocharian phonology happened in the deeper past.

# 3 Relative chronology of Samoyed sound changes

From the previous chapter we have a model of pre-Proto-Tocharian sound developments, and to test the specific hypothesis that pre-Proto-Tocharian was in contact with early Samoyed, we require the same for pre-Proto-Samoyed. This chapter will detail the sound changes that occurred in pre-Proto-Samoyed from Proto-Uralic based on the inherited lexicon that has so far been established for this branch. The further developments to the individual Samoyed languages are left out of consideration to avoid making the matter more complicated than it needs to be for the purpose of comparing pre-Proto-Samoyed and pre-Proto-Tocharian phonology.

## 3.1 Introduction

The foundations for Proto-Samoyed reconstruction were laid down by Janhunen 1976, with the bulk of etymologies presented in the 1977 publication *Samojedischer Wortschatz* (SW). Janhunen (1981) treated the correspondences with the rest of Uralic. Further work on etymology with minor revisions to the phonological reconstruction are found in, e.g., Helimski (1986, 1992/3, 1997, 2005). The aim is to come to a considered full relative chronology that also takes into account more recent Samoyed historical studies. The numbers that accompany the cited etyma in this chapter correspond to etymological reference list 2, which contains more etymological information and references. It can be found at the end of this dissertation.

## 3.2 Samoyed and Proto-Uralic phonology

Samoyed is a relatively conservative branch of Uralic, but it still underwent a number of significant phonological changes. The Proto-Uralic consonant inventory is given in Table 3.1 below; this basic information can be found with slight variations in both recent handbook chapters like Aikio (2022a: 5) and Zhivlov (2023: 118), and in older articles as in Janhunen (1981: 249), Sammallahti (1988: 482).

Table 3.1: The Proto-Uralic consonant inventory.

| PU | labial | dental | retroflex | palatal | velar |
|---|---|---|---|---|---|
| stop | *p* | *t* | *č* | *ć* | *k* |
| fricative? | | *δ* | | *δ′* | *γ*/*x* |
| sibilant | | *s* | *š* | | |
| nasal | *m* | *n* | | *ń* | *ŋ* |
| liquid | | *l r* | | | |
| glide | *w* | | | *j* | |

A few consonants are of uncertain phonetic quality, and their place in the phonological system is tentative. PU *δ may have been a voiced fricative *[ð], which is how it turns up in Saami, although elsewhere this phoneme merged with *t, *l or *r. PU *δ′ is traditionally reconstructed as the palatalized counterpart to *δ, but whether this really was a *[ðʲ] is doubtful. It did merge with *δ in the western branches Saami, Finnic, Mordvin and Mari, and elsewhere it has a palatal reflex *l′*, *ɟ* or *j*. One alternative interpretation of this phoneme in Proto-Uralic is as a palatal lateral *l′ (Zhivlov 2023: 119). The nature of the so-called Uralic "laryngeal", represented in reconstruction by either *γ or *x, is uncertain. It disappeared in many branches, consistent with its interpretation as a weak fricative, but in Saami it became *k, and merged with original PU *k. A velar fricative *ɣ* is found in Mansi and Khanty, but its precise original nature of this phoneme is not fully understood (see Janhunen 1981: 240–246; 2007; Sammallahti 1988: 482; Aikio 2012: 227; 2022a: 7–8; Zhivlov 2023: 119–120). I will use the notation *γ for this phoneme, without the intention of taking a stance on its phonetic value.

Due to various mergers, the Samoyed consonant system was somewhat reduced. PU *δ, *δ′, *γ, *š and *s were either lost or merged with other phonemes. The development of PU *č to PS *c is a purely notational change. Furthermore, PS *s derives from PU *ć in an entirely straightforward manner. There is a possibility that PS *s was actually still palatal *ś (cf. Janhunen 1998: 462), or even retained its affricate pronunciation *ć (Zhivlov 2023: 150), but for the moment, this is of no consequence for our investigation. The Proto-Samoyed consonant system is given here in Table 3.2.

Table 3.2: The Proto-Samoyed consonant inventory.

| PU | labial | dental | retroflex | palatal | velar |
|---|---|---|---|---|---|
| stop | *p* | *t* | *c* | | *k* |
| sibilant | | | | *s* | |
| nasal | *m* | *n* | | *ń* | *ŋ* |
| liquid | | *l r* | | | |
| glide | *w* | | | *j* | |

While the consonant system became smaller, the Samoyed vowel system expanded with the innovation of a few additional distinctions. Most Proto-Samoyed vowels directly continue Proto-Uralic vowel. One important exception is PS *ü, which is almost always innovative in Proto-Samoyed. PS *a, *å and *o will be discussed more extensively in sections 3.6 and 3.7. The exact position of PS *ə̈ and *ə̂ in the vowel space is uncertain, but they are interpreted as reduced vowels (e.g., Janhunen 1998: 463). The Proto-Uralic and Proto-Samoyed vowel systems are presented in Table 3.3.

Table 3.3: The Proto-Uralic (left) and Proto-Samoyed (right) vowel inventories (Janhunen 1981: 247, Helimski 2005; Aikio 2006).

| PU | front | central | back | > PS | front | central | back |
|---|---|---|---|---|---|---|---|
| high | *i ü* | | *u* | | *i ü* | *į* | *u* |
| mid | *e* | *ę* | *o* | | *e ö* | *ę* | *o* |
| low | *ä* | *a* | | | *ä* | *a* | *å* |
| reduced | | | | | | *ə ə̂* | |

The structure of the Proto-Samoyed phonological system has been slightly revised in recent years. In particular the vowel system has been increased by two extra vowels relative to earlier reconstructions. In 1993, Helimski introduced the idea of the Proto-Samoyed front schwa **ə*[37] in addition to the back schwa **ə̂*. The only difference in the reflexes of these phonemes is in terms of the type of vowel harmony they trigger in Nganasan and, as later discovered by Salminen (2012), in Nenets as well. Etymologically, this front-back contrast is furthermore a continuation of the contrast between PU **i* and **u*, which are understood as having developed into PS **ə* and **ə̂* respectively under certain circumstances (see 3.6.1). It is uncertain whether the two were still phonetically distinct at the Proto-Samoyed level, since the only remaining observable difference between **ə* and **ə̂* in any Samoyed language is the harmonic class they belong to.

Once again based on data from Nganasan, Helimski (2005) furthermore changed the way we understand the development of PU **e* in Samoyed. It was originally thought that PU **e* had merged with **i* as PS **i* (Janhunen 1981: 225; 234; 247). Helimski argued that PU **e* was actually preserved as such, pointing to the notation of Old Nganasan <e> or <ê> in Castrén's data for Modern Nganasan *į* or *i*. The modern reflex *į* can be considered a recent outcome of a raising sound change. Old Nganasan *e* or *ê* developed into Modern Nganasan *į*, but this development is obfuscated by the secondary fronting of this *į* to *i* next to palatal sounds like *ń* (Helimski 2005).

The retention of PU **e* as PS **e* entails a change in notation for some of the other vowels: PS **e* in Janhunen's notation is now rendered as **ä*, identical to its usual Proto-Uralic source, and Janhunen's PS **ä* is written as **a* instead. This latter change renders the consultation of some sources dealing with Proto-Samoyed reconstruction somewhat confusing, since Janhunen's PS **å* was also written as **a* by some scholars (e.g., Helimski 1997; Aikio 2002; see also Aikio 2006a for an overview of the revised reconstruction). I accept Helimski's conclusions and give PS **e* as the reflex of PU **e*, and PS **ä* as the reflex of PU **ä* throughout this dissertation. See Warries (2022) for a more extensive justification and a specific address to the scepticism of Peyrot (2019a).

[37] Alternatively, simply **ə*, see Salminen (2012). I have opted to use **ə* for maximal contrast and clarity.

There are a number of changes that took place between Proto-Uralic and Proto-Samoyed that are rather straightforward on the face of it, such as a merger between PU **δ* and **r* as PS **r* (Janhunen 1981: 250; Sammallahti 1988: 485) and a simplification of clusters containing PU **k* in combination with another obstruent; the **k* in such cases is simply lost (Janhunen 1981: 251). However, most changes show an intricate interplay with other sound laws, attesting to the relative chronology of their occurrence.

In some cases, there are only one or two examples of a change or a conditioning factor, but if the etymologies themselves are reliable and there are no counterexamples, I have decided to take these at face value to come to a workable model of the relative chronology. There are also a number of words that point to irregular developments, going against more generally noticeable trends. These will be mentioned, but for the purposes of this study I have decided to take the majority development as regular, and adopt that in a working hypothesis of the relative chronology.

## 3.3 Loss of final PU **-i*

The first sound change to address is the development of second-syllable PU **-i* to PS **-ə/ə̂* and its disappearance in certain environments. An intermediate development to **-ə/ə̂* can be assumed for all PU **-i*.[38] This vowel usually disappeared in pre-Proto-Samoyed when it followed a single consonant (Janhunen 1981: 32ff.; Sammallahti 1988: 485). Following a consonant cluster, PS **-ə/ə̂* remains. This distribution is illustrated with some examples below.

| Loss of PU final **-i* in Proto-Samoyed | |
|---|---|
| 3. | PU **aŋi* > PS **aŋ* [all] 'mouth' |
| 54. | PU **kari* > PS **kar* [only Ne] 'skin, surface, bark' |
| 98. | PU **lomi* (**lumi*) > PS **jom*(-) [Ng-Sk] 'snow' |
| 114. | PU **nusi-* > PS **nə̂t-* [En-Ne-Sk-Km] 'scrape' |
| 115. | PU **nüδi* > PS **nir* '[no Mt] handle' |
| 154. | PU **sẹni* > PS **cẹn* [all] 'sinew' |
| 193. | PU **weti* > PS **wet* [all] 'water' |

| Retention of PU final **-i* in Proto-Samoyed | |
|---|---|
| 4. | PU **aŋti* > PS **aŋtə* [all] 'blade' |
| 32. | PU **ẹpti* > PS **ẹptə* [all] 'hair of the head' |
| 77. | PU **kuńći* > PS **kunsə* [Ng-Sk-Km-Mt] 'urine' |

[38] It is disputed whether this is a development at all, since one could reconstruct PU **-i* itself as **-ə* instead (Kallio 2012).

95. PU **lẹnti* > PS **lịntə* [only Ng] 'plain, valley'
177. PU **tumti-* > PS **tumtə-* [no Sk] 'know, recognize'
200. PU **woŋki* > PS **wåŋkə* [Ng-En-Ne] 'hole, cave'

There are a number of words that do not behave according to these rules, however. This is due to the fact that any root PU **CVCi* that changed to **CVC* would have continued to alternate with **CVCə* in certain forms of the paradigm. Due to these alternations, final PS **-ə/ə̂* could be analogically added or removed, sometimes in the individual Samoyed languages (cf., e.g., Janhunen 1981: 252 on PS **tə̂ŋ(ə̂)* 'summer'). For example, gen.sg **ẹptə̂n* is clearly the genitive of a stem **ẹptə̂* 'hair of the head', since ***ẹpt* would have an illegal final cluster, but gen.sg. **utə̂n* can be interpreted as the genitive of either **ut* or **utə̂* 'passage, way'. The latter is the expected regular outcome (see 3.5.2), but the former stem was easily created based on analogy with words like **nir* : gen. **nirə̈n* 'shaft', which never contained a consonant cluster and regularly lost the final **-i*. Here are a few examples where the Samoyed languages are in disagreement about the continued presence of final **-ə*:

56. PU **käri-* > PS **kärə-* [Ng-En-Ne] ~ **kär-* [Sk-Km-Mt] 'dress, put on'
160. PU **suŋi* > PS **tə̂ŋə* [Ng-En-Sk-Km] ~ **tə̂ŋ* [Ne] 'summer'
181. PU **ukti* > PS **utə* [Ng-NeF] ~ **ut* [EnF-NeT] 'passage, way'

It should be noted that loss of PU **-i* did not affect the suffix-final *-i* of the personal endings of the possessive declension and the objective conjugation PU 1sg. **-mi*, 2sg. **-ti*, which became PS **-mə* and **-rə* respectively. This makes it difficult to interpret the loss of PU **-i* as a general apocope of word-final PU **-i*, and to my knowledge, this issue has not yet been properly addressed. The adjectival suffix PU **-ji* was not affected either, as this became PS **-jə* (Aikio 2022a: 20). The issue remains to be fully resolved.

## 3.4 Changes affecting single consonants

A number of consonant changes are relatively straightforward and will be discussed first. In the following subsections, all of them will be woven into the relative chronology more or less firmly, as some of them are involved in further changes affecting the vowels.

### 3.4.1 PU **δ*, **s* and **š*

PU **δ* is reflected in Proto-Samoyed as **r*, which means that it merged with already existing PU **r* (Janhunen 1981: 250; Sammallahti 1988: 485; Zhivlov 2023: 150). Examples include the following:

58. PU **käwδi* > PS **kürə* [der. Ng-En-Ne] 'band (for tightening something)'
71. PU **kuδa* > PS **kârâ* [Sk-Km] 'morning'
93. PU **lęδi-* > PS **lęr-* [der. Ne-Sk] 'be frightened'
115. PU **nüδi* > PS **nir* '[no Mt] handle'
134. PU **piδi* > PS der. **pirə* [Ng-En-Ne-Sk] 'height'
135. PU **piδkä* > PS **pirkä* [no Ng] 'high'
166. PU **täγ/wδi* > PS **tärə* [Ng-Ne-Sk] 'full'

A number of PU **t* in suffixes also changed to **r*, chiefly in the 2sg. possessive suffix PU **-ti*, which became PS **-rə*, and in one of the Proto-Uralic causative suffixes PU **-tA*, which developed into PS **-rA* (Janhunen 2002: 82; Aikio 2022a: 20; Zhivlov 2023: 150). It could be assumed that this happened via an intermediate stage **-δV* (Zhivlov 2023: 150), in which case this development could be taken together with the general rhotacization of PU **δ*.

With that assumption, the change from PU **δ* to PS **r* can receive a tentative place in the relative chronology. If the intervocalic environment after the second syllable is an important trigger for the change from PU **t* to PS **r*, as seems to be the case since postconsonantal **t* is preserved as such (e.g., Zhivlov 2023: 150), it can be tentatively dated before the loss of second-syllable **-i*. Forms like PU **tuliti* 'your fire' may have developed via **tulâδâ* to **tulârâ* to PS **tujrâ* (cf. Fi. *tulesi*, also from PU **tuliti*). If PU **t* in this position first merged with **δ* before the loss of second-syllable **-i*, the development may also have taken place via an intermediary stage of the type pre-PS **tulδâ* or **tujδâ*.

The development as a whole has to be dated before a notable change of two other consonant phonemes, namely those of PU **s* and **š*. At some point in the development of Samoyed, both **s* and **š* became PS **t* (Janhunen 1981: 150; Zhivlov 2023: 150). This can be seen in the examples below.

| PU **s* > PS **t* |
|---|
| 69. PU **kosi-* (**kusi-*) > PS **kot-* [all] 'cough' |
| 70. PU **koska* > PS **kåtå* [Ng-En-Ne] 'grandmother' |
| 78. PU **kupsa-* > PS **kâptå-* [all] 'extinguish' |
| 91. PU **läsi-* > PS **jät-* [Ng-En-Ne-Sk] 'near' |
| 100. PU **lupsa* > PS **jâptå* [all] 'dew' |
| 103. PU **męksa* > PS **mįtâ* [no Mt] 'liver' |
| 114. PU **nusi-* > PS **nât-* [En-Ne-Sk-Km] 'scrape' |
| 140. PU **poski* > PS **påtâ* [all] 'cheek' |
| 148. PU **sala-* > PS **tålä-* [all] 'steal' |
| 149. PU **sarka* > PS **tårkå* [En-Ne] 'branch' |
| 153. PU **sęksa* > PS **tįtå-jŋ* [no Ng] 'Siberian pine' |

156. PU **soski-* > PS der. **tåtu-* [Sk-Km] 'chew'
158. PU **suδ′a* > PS **tâjå* [only Mt] 'finger'
159. PU **suksi* > PS **tutâ* [Ng-En] 'ski'
161. PU **süli* > PS **tij* [Sk; der. Ng-En-Ne-Mt] 'fathom'

| PU **š* > PS **t* |
|---|

48. PU **kajši-*; noun *kajšo* > PS **kåjtâ-*; noun **kåjto* [Ng-En-Ne-Sk] 'be sick; sickness'
163. PU **šęra* > PS **tįrå-* [En-Ne-Sk?; der. Mt] 'dry (itr.)'
162. PU **šelki-* > PS **tej-* ~ ?**teə-* [no Km] 'fly'

The **t* from PU **s* never developed further to **r* in any suffix; cf. the third person singular possessive suffix PS **-tA* from PU **-sA*. The completion of the development from **s*/**š* to PS **t* can thus be placed after the change from **δ* to **r*, although the initial merger of the two sibilants and developments through intermediary stages (such as **[θ]*) cannot be further pinned down at present.

### 3.4.2 PU **l*, **δ′* and **j*

The changes to PU **l*, **δ′* and **j* are best treated together, since they share a reflex PS **j* in many environments. PU **δ′* is relatively straightforward, since it is always reflected as PS **j* (Janhunen 1981: 250; Sammallahti 1988: 485; Zhivlov 2023: 150), as illustrated by the following list.

12. PU **ćoδ′a-* > PS **såjå-* [Ng-Ne] 'wage war'
17. PU **ćüδ′i* > PS **sijə* [En-Sk-Km-Mt] 'charcoal'
23. PU **δ′ęmi* > PS **jęm* [Sk-Km] 'bird cherry'
24. PU **δ′ümä* > PS **jimä* [all] 'glue'
45. PU **kaδ′a-* > PS **kåjä-* [all] 'leave'
62. PU **koδ′ka* > PS **kåjkâ* [Ng-En-Ne-Sk] 'idol; ghost'
106. PU **muδ′a* > PS **mâjå* [Ng-Sk-Km-Mt] 'earth, land, terrain'
122. PU **oδ′a* > PS **åjå* [no Ng] 'raw'
142. PU **puδ′a-* > PS **pâjä-* [no Km] 'chop wood'
158. PU **suδ′a* > PS [only Mt] **tâjå* 'finger'

PU **l*, on the other hand, underwent a split development in Samoyed. It is either preserved as **l* or changed into **j* (Janhunen 1981: 250). Word-initially the general development was into **j*, except when followed by PU **ę* (Aikio 2014b: 86).

| PU *l- > PS *j- | |
|---|---|
| 87. | PU *lapta > PS *jåptå [no Km] ‘flat, thin’ |
| 88. | PU *läćä- > PS *jäsä- [En-Ne-Sk] ‘cover the tent’ |
| 89. | PU *lämi > PS der. *jämVjə [Ng-Ne] ‘broth’ |
| 90. | PU *lämpi ‘warm’ > PS *jämpə [Ng-En-Ne-Sk] ‘clothes’ |
| 91. | PU *läsi- > PS *jät- [Ng-En-Ne-Sk] ‘near’ |
| 98. | PU *lomi (*lumi) > PS *jom(-) [Ng-Ne-Sk] ‘snow’ |
| 100. | PU *lupsa > PS *jâptå [all] ‘dew’ |

| PU *l- = PS *l- before PU *ę | |
|---|---|
| 93. | PU *lęδi- > PS *lęr-, der. *lęrö- [der. Ne-Sk] ‘be afraid’ |
| 94. | PU *lęmpi > PS *lį̈mpâ [Ng-En-Ne-Sk] ‘swamp, bog’ |
| 95. | PU *lęnti > PS *lį̈ntâ [Ng] ‘lowland’ |
| 97. | PU *lęwi > PS *lę [all] ‘bone’ |

| PU *l > PS *l- ~ *j- | |
|---|---|
| 96. | PU *lepći > PS *lępsâ [Ng-En] ~ *jępsâ [Ne-Sk-Km-Mt] ‘cradle’ |

In one instance, initial *l shifted to *j before *ę in only part of Samoyed, namely in PU *lępći ‘cradle’, PS *lępsâ. This has a reflex with initial *l-* in EnF *l'ičᴜ* and Ng. *labsə*, but initial *j-* in NeT *jebc°*, PSk. **ťopsə*, Km. *ďepsü*, Mt. *čĕpsë*, as if from PS **jępsâ* instead (Aikio 2014b: 86). Salminen reconstructs this as PS **l'ępsâ*, with a unique palatal **l'*, and projects this back to a PU **l'ępćâ* with the same phoneme, arguing that the developments of an initial sequence PU **lę* with both *l* and *j* in the Samoyed languages would otherwise be impossible (Salminen 2023: 383–384). Rather than reconstructing an additional Proto-Uralic phoneme, perhaps in this case assimilation occurred with the following **ć* or **s* (when it was still palatal **ś*) to cause a secondary palatalization. An intermediary stage in the development from **l* to **j* may have been a palatal lateral **l'* in pre-Proto-Samoyed. However, the problem of the correspondence *l* and *j* within Samoyed is also found in different words of non-Uralic origin, and the situation is not at all straightforward (see Kaheinen 2023: 88).

The word-internal developments are different. Before consonants, and potentially also word-finally after the first syllable, the reflex is **j* (Janhunen 1981: 250). Between vowels the development could go further, depending on the type of stem: in *i*-stems **l* changed to **j* and even disappeared in intervocalic position, while in *A*-stems it was retained as PS **l* (ibid.). After the second syllable, **l* was also retained, as seen in in PU **aćkal* ‘step (n.)’ > PS **asâl-* ‘step (v.)’ and certain suffixes like the verb suffix PU **-li-* > PS

*-*l*(*ə*)- (Aikio 2012: 244). The list of examples below shows how, normally, PU **j* and PU **l* merged in *i*-stems.

| PU **Vji* and **Vli* > PS **Vj* (no vowel sequence) | |
|---|---|
| 13. | PU **ćoji-* > PS **soj-* [En-Ne] 'be heard, be audible' |
| 132. | PU **peli-* > **pẹj-* [Sk; der. in rest] 'fear' |
| 161. | PU **süli* > PS **tij* [Sk; der. in Ng-En-Ne-Mt] 'fathom' |
| 170. | PU **toli-* (**tuli-*)> PS **toj-* [all] 'come' |
| 175. | PU **tuli* > PS **tuj* [all] 'fire' |
| 182. | PU **ulki* > PS **uj* [no Km] 'pole' |

| PU **Vj*(*k*)*i* and **Vl*(*k*)*i* > PS **Və* (vowel sequence) | |
|---|---|
| 18. | PU **ćülki-kći-* > PS **siə-sə-* [En-Sk-Km] 'spit' |
| 47. | PU **kaji* > PS **kåə̂* [Sk] 'slender object' |
| 51. | PU **kali-* > PS **kåə̂-* [all] 'die' |
| 74. | PU **kulki-*[39] > PS **kuə̂* [Ne-Sk] 'float downstream' |
| 143. | PU ?**puji* > PS **puə̂* [all] 'back, behind' |
| 180. | PU ?**uji-* > PS **uə̂-* [Ne–Sk–Mt] 'swim' |
| 188. | ? PU **wali-* > PS **wåə̂-* [Ng-En-Ne-Mt] 'speak' |

| PU **Vli* > PS **Vj* and **Və* | |
|---|---|
| 198. | PU **woli-* 'be' > PS **åj-*, **åə̂-* [paradigmatic alternation in En-Ne; all] 'be' |

| PU **Vl*(*k*)*i* > PS **Vj* or **Vj* (ambiguous cases) | |
|---|---|
| 55. | PU **käli* > PS ?**käj* [Mt] or ?**käə* [Ng-Sk]; [Ne ambiguous] 'tongue' |
| 162. | PU **šelki-* > PS **tej-* or ?**teə̈-* [potentially in EnT; rest ambiguous] 'fly' |

| PU **Vli* > PS **V* (exceptional example) | |
|---|---|
| 183. | PU **üli-* > PS **i-* [no Mt] 'on; space above' |

The variability in the reflexes of **Vj* and **Və* most likely result from paradigmatic alternations between, e.g., nom.sg. **tuj* : gen.sg. **tujə̂n*, which regularly became **tuə̂n*

[39] Connected with PS **kəj-* 'go' rather than **kuə̂-* by Aikio (2002: 26). One could consider a paradigmatic split, with **kəj-* from forms like the aorist pre-PS **kuj-ŋå-* with the vowel change to **kə̂j-ŋå-* (NeT *xəya*), cf. 3.6.1. However, the particularly common palatalization of suffixes attached to this stem in Tundra Nenets, like in fut.ptc. *xæw°ntya*, ipfv.vn. *xæbya*, and ipfv.ptc. *xænya* next to *xæna* (Salminen 1997: 66, 83–84), might suggest that it originally was front vocalic PS **käj-* rather than the back vocalic **kə̂j-* that would be expected from PU **kulki-*.

'fire', and could be levelled in either direction, i.e., either **tuj* : **tujân* or **tuâ* : **tuân* (Aikio 2012: 246–247). The verb 'to be' still shows such an alternation in Tundra Nenets: the present participle NeT *ŋæda* regularly derives from PS **åjtå*, with the regular consonant stem PU ±**wol-* 'be' suffixed with **-(n)ta*, while the imperative and connegative NeT *ŋaq* can be derived from PS **åâk*, from PU ±**wolik* (cf. SW: 16–17 with a reconstruction of **åâ-* ~ **åâ-j-* ~ **åj-* for Nenets). A similar distribution of allomorphs is found in Forest Enets, with *ŋa-* in the aorist, imperative, jussive and connegative and *ɛ-* elsewhere (Khanina & Shluinsky 2023: 827). The imperative/connegative PS **åâk* is also reflected in Nganasan with the (suppletive) connegative *ŋuəʔ* (Wagner-Nagy 2019: 272). Similar to the preservation of PU **l* after the second syllable, suffixal PU **-ji* was able to preserve its **j* in the suffix PU **-ji* > PS **-jə*.

There are no examples containing PU **δ′* that show that this consonant was ever lost in the same way as many PU **l* and **j* in intervocalic position before **-ə*, but this may simply be a coincidence due the general lack of words that show this phoneme in the relevant environment. The clearest instance of PU **Vδ′i* reflected in Samoyed seem to be PS **sijə* from PU **ćüδ′i* 'charcoal' (17). While this could suggest that PU **δ′* only became **j* after **Vjə* from both **Vji* and **Vli* had developed into vowel sequences, PS **sijə* could also represent a development from earlier **sij* (type PU **tuli* > PS **tuj*) with an analogical **-ə* added at the end after the sound law removing **j* in intervocalic position had ceased to operate (cf. section 3.3).

### 3.4.3 PU **w* and **γ*

Proto-Uralic **γ* (alternatively written as **x*) is lost in Samoyed (Janhunen 1981: 245–246; Sammallahti 1988: 485; Zhivlov 2023: 150). Generally, the sequence PU **Vγi* yields a simple vowel in Samoyed. The only known instance of **Vki* behaves in the same way, so that **Vki* may be routed through **Vγi*. The examples are given together here.

| PU **Vγ/ki* > PS **V* (no vowel sequence) |
|---|
| 39. PU **jęγi-* > PS der. **ęr-* [Ne-Sk-Mt] 'drink', **ękâl-* [En-Ne] 'gulp' |
| 64. PU **koki-* > PS **ko-* [no Ng] 'see' |
| 102. PU **meγi-* > PS **mi-* [all] 'give, sell' |
| 130. PU **peγi-* > PS **pi-* [no Mt] 'cook' |
| 169. PU **toγi-* > PS **tå-* ~ **tâ* [all?] 'bring' |
| 192. PU **weγi-* > PS **i-* [Sk-Km] 'take, grab' |
| 195. PU **wiγi-* > PS **ü-* [Sk; der. Ne] 'take, carry' |

| PU **Vγi* > PS **V* or **Və* (ambiguous example) |
|---|
| 120. PU **ńoγi-* (**ńuγi-*) > PS **ńo-* or perhaps **ńoâ-* [Ne-Sk-Mt] 'pursue' |

| PU *Vγi > PS *Və (vowel sequence) |
|---|
| 157. PU *suγi- > PS *suə̂- [Sk; der. Ng] 'row' |

Only PU *suγi- 'row' unambiguously yielded a vowel sequence *uə̂ on account of Ng. *tuobsa* 'oar' (PS ±*tuə̂-psåń); SkTym *tū-* and SkKet *tū-* are suggestive of a vowel sequence, but not probative, and this verbal root is not attested in the other Samoyed languages. Since this is the only example of a sequence *Vγi being reflected as a vowel sequence in Samoyed, alternative explanations may also be imagined. For instance, this verb may have been influenced by the semantically close verbs PS *uə̂- 'swim' and PS *kuə̂- 'drift downstream'.

The reflexes of PU *Vwi are normally also PS *V, same as for PU *Vγi. There are only a few ambiguous cases, as can be seen in the examples below.

| PU *Vwi > PS *V |
|---|
| 97. PU *lę̨wi > PS *lę̨ [all; der. Ng] 'bone' |
| 126. PU *pawi > PS *pa [all; der. Ng] 'tree' |
| 151. PU *sewi- > PS der. *tima [all] 'tooth' |
| 152. PU *sewi- > PS der. *titü [Ng-En-Sk] 'bait' |
| 167. PU *täwiw > PS ?*tä/ew [Sk-Km-Mt; ders. Ng, En, Ne] 'lung' |
| 174. PU *towi (*tuwi) > PS *to [all; der. Ng] 'lake' |

| PU *Vwi > PS *V or *Və (ambiguous example) |
|---|
| 16. PU *ćowi (*ćuwi) > PS der. *so-j or perhaps ?*soə̂-j [no Ng] 'throat' |

| PU *Vwi > PS *Və (uncertain example) |
|---|
| 155. PU ?*si/äwi > PS tiə̈ [Ng-En-Ne-Sk] 'year ring (of tree)' |

| PU *Vwi > PS *Vw or *Və (uncertain ambiguous example) |
|---|
| 92. PU ?*lewi- > ? PS *jew- or *jeə̈- [only Sk] 'shoot' |

On account of PU *sewi- 'eat' to PS *ti- (in derivations), the development of *eγi to PS *i appears to have been shared with *ewi. It seems that word-final *-w was preserved in Proto-Samoyed (Gusev 2008), so that perhaps loss of the *-w- in these roots should have applied in intervocalic position, before loss of final *-i. However, the potential Selkup reflex of putative PU *lewi- 'shoot', Sk. *t'ǖ- 'shoot' (Aikio 2002: 53 fn. 5; Salminen 2023: 392) could imply preservation of the *-w- at least in this root: PS *je- would not yield the front rounded vowel Sk. *ǖ. None of the other roots with the structure *CVwi show effects of the *-w-, so that it is difficult to determine what vowel effects, if any, should have been

expected if the **-w-* was preserved until a late stage. A vowel sequence **jeə-* might also yield Sk. *ćīī-*, however, as it is expected to merge with the regular development from PS **iə* to Sk. *īī* (PS **i* and **e* merge in Selkup). Loss of intervocalic **-w-* might also be dated before the loss of intervocalic **j* in **Vjə*, since such sequences yield a vowel sequence in Proto-Samoyed, which is not atteetd with certainty for PU **Vwi*. This distribution may be accidental, however, owing to the low number of examples.

The reflex of *A*-stem sequences like PU **uwa* are often given as uniformly resulting in **uə̂* in Proto-Samoyed, with a preservation of the original disyllabic structure, but not of the second-syllable vowel. In some words, a reconstruction of **uå* is now preferred: for instance, PS **puə̂* 'back' (143) gives Ng. *huo* while PS **puə̂l-* 'blow' yields Ng. der. *hüoltə-*, with no other way to account for the difference; thus Ng. *hüoltə-* may be derived from PS **puål-* instead (Kaheinen 2023: 68–73; Zhivlov 2023). Compare the split development of PU second-syllable *-a* discussed in subsection 3.7.1 below.

82. PU **kuwakka* > PS **kuə̂kå-nå* [only Ng] 'long ago'
150. PU **puwa-* > PS **puå-* [Sk; der. all] 'blow', der. **puål-* 'blow'
201. PU **(w)uwa* > PS **wuə̂-* [only Ng] 'current'

The similar sequence apparently **üwä* ends up as PS **iə* 'belt', with the regular change from **ü* to **i* (on which see 3.6.1) and a prothetic consonant. The disappearance of the **-w-* in this position before PU **A* might be later than before PU **i*, on account of the completely reliable preservation of the original disyllabic structure in these examples, as opposed to the apparent reduction to monosyllables in all cases of **Vwi*.

### 3.4.4 Summary

The simple consonant changes are now briefly summarized. PU **δ* merged with **r* most likely before loss of second-syllable **-ə*. The development of PU **s* and **š* took place after the change of PU **δ* and some suffixal **t* to **r*. PU intervocalic **-l-* before **-ə* merged with PU **-δ'-* and PU **-j-* as **-j-*, and was eventually lost in this position. In word-final position **-j* from all sources was preserved, so that the intervocalic elision of **-j-* cannot be dated before apocope of final **-ə*. Intervocalic PU **-w-*,**-γ-* and **-k-* disappeared in original *ə*-stems most likely before both loss of intervocalic **-j-* in the same environment, and before the apocope of **-ə* after open syllables. This is summarized in Table 3.4.

Table 3.4: An overview of the changes discussed in this section and their relative chronology as far as can be determined.

| Sound changes | | Chronology | Treated in |
|---|---|---|---|
| A *-i* | > *-ə* | a notational change | 3.3 |
| B1 *CVCVtV* | > *CVCVδV* | before B2, C1. | 3.4.1 |
| B2 *δ* | > *r* | before E. | 3.4.1 |
| C1 *CVCə* | > *CVC* | before C5. | 3.3 |
| C2 *Vγ/kə* | > *V* | **eγi* > **i*; | 3.4.3 |
| C3 *Vwə* | > *V(w)* | | 3.4.3 |
| C4 *uwa* | > *uə̂/uå* | | 3.4.3 |
| C5 *Vjə* | > *Və* | | 3.4.2 |
| D1 *Vli* | > *Vjə* | before C5 | 3.4.2 |
| D2 *δ´* | > *j* | possibly with D1 before C5. | 3.4.2 |
| E *s/š* | > *t* | | 3.4.1 |

## 3.5 Changes affecting consonant clusters

A number of Proto-Uralic consonant clusters were reduced in Samoyed, and these reductions are relevant for the relative chronology of a number of vowel developments. Aside from the specific change of **lC* to **jC*, already discussed above (3.4.2) together with the general development of **l*, the remaining changes each involve the loss of a segment.

### 3.5.1 PU **rk* and **lk*

The sequences PU **rki* and **lki* were reduced to pre-PS **rə* and **lə*. This change could well have occurred before **lC* had become **jC*, allowing for an abstraction to **Rkə* > **Rə* (Aikio 2002: 26).

10. PU **ćęrki* > PS **sęr* [Ng-En-Ne-Sk; der. Km-Mt] 'ice; white; salt
18. PU **ćülki-kći-* > PS **siə-sə-* [En-Sk-Km] 'spit'
74. PU **kulki-* > ? PS **kuə̂-* [ders. Ne-Sk] 'float downstream'
146. PU **purki* > PS **pur* [der. Ne-Sk?] 'smoke'
162. PU **šelki-* > PS **tej-* (or **teə-*?) [all] 'fly'
182. PU **ulki* > PS **uj* [no Km] 'pole'

Most of these words behave in Samoyed exactly as though the **k* was never there. In particular, apocope in PU **ulki* > ** uli* > PS **uj* pole', PU **šelki-* > **šeli-* > PS **tej-* (or **teə-*?) 'fly', and PU **ćęrki* > **ćęri* > PS **sęr* 'ice; white; salt' should have occurred when the **k* had already disappeared. We can thus date this change before the development **CVCə* > **CVC* in the relative chronology.

The cluster **rk* was not affected before **-A*, as evident from PU **sarka*, which became PS **tårkå* 'form, bifurcation, branch' (149). There are, however, some words that possibly show a loss of **lk* even before **-A*, namely PU **tulka* > PS **tuå* 'feather' (176), PU **ülkä* > PS ?**ij* 'boy' (184), and PU **pälkä* > PS ?**pea* 'thumb' (128) (Aikio 2020: 95). The developments of these words are difficult to understand, especially next to PU **wolka* > PS **wajk* 'shoulder' (199) with the more expected change from **lk* to **jk*. It is also problematic that the **lk*-cluster would have completely disappeared from PU **tulka*, since in *a*-stems the expected intermediary of **j* should not have been lost (cf. 46. PU **kaja* > PS **kåjå* 'sun'). This would force us to set up a sound development that avoids an intermediary step **j* in some way, such as a speculative **tulka* > **tulγa* > **tujγa* > **tuwå* > **tuå*. An intermediate stage like **tuwå* would also seem to be necessary to account for the lack of reduction in the sequence **u-a* > PS **ə̂-å*, which is otherwise only known to be blocked by an intervening **-w-* (the development of **u* is discussed in subsection 3.6.1). This remains highly speculative, but the development of **-lk-* needs to be resolved if this etymology of PS **tuå* is to be accepted.

PU **δ′k* is not lost in a similar way in PU **koδ′ka* > PS **kåjkə̂* 'idol; ghost' (62), so that the change of **lk* should perhaps be dated before the merger of **l* and **δ′*, if this example is reliable. If a cluster **lγ* is posited as an intermediary stage in the development of PU **tulka* > PS **tuå*, the different reflex PS **ij* of PU **ülkä* may be accounted for by positing a change of **lγ* or later **jγ* to either **j* or **w* depending on the surrounding vowels: **j* in a front environment, and **w* in a back environment. This remains difficult, however: while pre-stage **tuwa* would yield PS **tuå*, an earlier **üjä* to **ijä* should be expected to yield Sk. **ićə* rather than **ī* if the final vowel was not lost in some way.

All of these changes remain very tentative, but it should be clear that if (part of) these etymologies are correct, some chain of developments needs to be set up that somehow does not interfere with either **Vja*, **Vδ′a*, **Vla*, or, for that matter, **Vka*, and that also blocks the development of **u-a* to **ə̂-å*. An intermediate **lγ* with further changes specific to that cluster is the best option I currently see short of positing an entirely new phoneme such as **ɬ* or the like.

### 3.5.2 PU **kC*, **Ck* and geminates

Original **k* in consonant clusters was not preserved in several other clusters either. Not only did it disappear in the sequences **rki*, **lki* and possibly even **lkA*, it was eliminated from all clusters of the type **kC* and **Ck* where *C* was an obstruent (Janhunen 1981: 251). This development is seen in a good number of examples, given in the lists below.

| PU **Ck* > PS **C* (loss of **k*) | |
|---|---|
| 1. | PU **aćkal* > PS **asVl-* [Sk-Km] 'step' |
| 22. | PU **čočki* (**čučki*) > PS **cocə̂* [Sk] 'log' |

57. PU **kätki-* > PS **kätə* [Ng-Ne-Mt] 'wrap up, swaddle'
61. PU **koćki-* > PS **kåsə̂-* [no Mt] 'dry (itr.)'
70. PU **koska* > PS **kåtå* [Ng-Ne-En; Sk?, Mt?] 'grandmother'
105. PU **moćki-* > PS **måsə̂-* [no Ng] 'wash'
140. PU **poski* > PS **påtə̂* [all?] 'cheek'
156. PU **soski* > PS der. **tåtu-* [Sk-Km] 'chew'

PU **kC* > PS **C* (loss of **k*)

28. PU **ękta-* > PS **i̯tå-* [all] 'hang'
103. PU **męksa* > PS **mi̯tə̂* [no Mt]'liver'
131. PU **peksä-* > PS **pet-* [Ng-Ne-Sk-Km] 'beat'
141. PU **pučki* > PS **pucə̂* [all] 'hollow stalk'
153. PU **sęksa* > PS **ti̯tå-jŋ* [no Ng] 'Siberian pine'
159. PU **suksi* > PS **tutə̂* [Ng-En] 'ski'
173. PU **totki* > PS der. **tåtu* [Sk] 'tench'
181. PU **ukti* > PS **ut(ə̂)* [Ng-En-Ne] 'passage, way'

Janhunen interprets these developments in terms of assimilation and subsequent degemination (on which see below); e.g., **ks*, **kt*, **sk* > **tt* > **t* and **ck* > **cc* > **c* (Janhunen 1981: 251). This may indeed have been an intermediate step before the cluster was fully lost, but whether it really occurred cannot be properly verified. Contrary to the development of **rki* and **lki*, final **-ə* generally did not disappear due to apocope after these original consonant clusters, even if those stems had only a single intervocalic consonant in Proto-Samoyed after losing the **k*. This indicates that this specific type of *k*-loss is to be dated after the general development of stems of the type **CVCə* to **CVC* (see 3.3).

Geminate stops show a similar interaction with apocope. There are not many examples of Proto-Uralic geminate consonants, and only a very few are known to be reflected in Samoyed, listed here.

30. PU **ęna-ęppi* > PS [Ng-En-Ne-Km] **i̯nə̂pə̂*
60. PU **kęčči-* > PS **kęcə̂-* [Sk] 'bad smell, stench'
113. PU **nokki* > PS der. **nåku* [Sk] 'neck'

If the development of PU **kęčči-* to PS **kęcə̂-* represents a regular development, the simplification occurred after *ə*-apocope. On account of PU **nokki* to PS der. **nåku* 'it was also after the disappearance of simple **-k-* in stems of the type **CVki* like PU **koki-* > **ko-*. Fittingly, this latter change was already dated before the former in the discussion above, but the available evidence remains slim.

Thus, if the loss of **k* in the clusters described above went through an intermediary stage of assimilation to geminates (Janhunen 1981: 251), that change should be dated before degemination itself. Such a middle step cannot be dated relative to any other changes, however, as it cannot be confirmed in the first place.

### 3.5.3 PU pre-consonantal glides

The final type of consonant cluster to discuss is the type starting with the labial glide **w*. This glide was lost in Samoyed in preconsonantal environment, but seemingly only at a relatively late date. The three examples are provided here.

38. PU **jäwji* > PS **jüjə* [Ng-Ne] 'lichen (on trees)'
58. PU **käwδi* > PS **kürə* [ders. Ng-Ne] 'band (for tightening something)'
166. PU ?**täwδi* or **täγδi* > PS *tärə* [Ng-En-Ne-Sk] 'full'

Final **-ə* was apparently shielded from apocope by the original cluster **wC*, so that the simplification seems to have happened after the loss of **-ə* after open syllables (Aikio 2002: 33). On account of the development from PU **jäwji* to PS **jüjə* 'lichen (on trees)' rather than to ***jüə*, this cluster simplification also seems to have happened after the loss of intervocalic **j* in sequences of the type **Vjə* (cf. 3.4.2).

Clusters of the type **jC* seem to have behaved differently, as both examples show a loss of the final PU **-i*.

136. ? PU **pijri* > ? PS **pijr* [Ne] 'inner bark of birch'
186. PU **wajŋi* > PS **wajŋ* [all] 'spirit, soul, breath'

Some vowel changes seem to have been involved with the disappearance of preconsonantal **w*, most clearly a change from **äwC* to **üC*. The lack of a vowel-change in PS **tärə* 'full' (not ***türə*) has been explained by Zhivlov with a reconstruction as PU **täγδi* rather than **täwδi* (Zhivlov 2023, cf. Aikio 2002: 31–34). This would imply the continued presence of PU **γ* in the pre-Proto-Samoyed consonant system until after apocope, since otherwise PU **täγδi* would be expected to yield PS ***tär* instead. Naturally, since it is only a single example, the final **-ə* may have been added secondarily in Proto-Samoyed as well, so that we must remain cautious.

### 3.5.4 Summary

To summarize this section, a number of Proto-Uralic consonant clusters were simplified in Samoyed. These simplifications can be dated relative to a number of other developments, in particular the loss of **-ə* after open syllables. The reduction of PU **rki*

and **lki* happened before **CVCə* changed to **CVC*, but every other cluster simplification happened after **CVCə* had become **CVC*. For structural reasons, we may suppose that **lC* had not changed to **jC* yet, since it patterned with the resonant **r* in the sequence **Rki*; there are no original sequences PU **jki* to strictly corroborate this, however. For the relative chronology, we may conclude:

1) **rki*, **lki* > **ri*, **li*
2a) **l* > **j*; 2b) **CVCə* > **CVC*
3a) *Vjə* > *Və*; 3b) **kC*/*Ck* > **C* or **CC*; 3c) **CC* > **C*
4) **wC* > **C*

A development **γC* > **C* might be added on the basis of only PS **tärə* 'full' if from PU ?**täγδi* rather than ?**täwδi*. Table 3.5 gives a summary.

Table 3.5: An overview of the changes discussed in this section and their relative chronology as far as can be determined.

| Sound changes | | Chronology | Treated in |
|---|---|---|---|
| A *-Rkə* | > *-Rə* | before B1, B4, C1 | 3.5.1 |
| B1 *CVCə* | > *CVC* | before B2, C1, C2 | 3.3 |
| B2 *w*/*γC* | > *C* | | 3.5.3 |
| B3 *Vγ*/*kə* | > *V* | before C2, C1? | 3.4.3 |
| B4 *Vjə* | > *Və* | | 3.4.2 |
| C1 *kC*/*Ck* | > *C* or *CC* | if via *CC*, before C2 | 3.5.2 |
| C2 *CC* | > *C* | | 3.5.2 |
| D1 *l*/*δ´* | > *j* | before B4 | 3.4.2 |

## 3.6 Changes affecting non-low vowels

In this section we will go through the changes that affected the non-low vowels **u*, **i*, **ü*, **ę*, and **o*. They are all very important for the development of the pre-Proto-Samoyed vowel system. A more thorough discussion focussing on PU **a* is given in the next subsection.

### 3.6.1 PU **u*, **i* and **ü*

The Proto-Uralic high vowels **i* and **u* both show a double reflex in Proto-Samoyed. They are either preserved as such, or reduced to "front" **ə̈* and "back" **ə̂* respectively. We will start with a look at PU **u*, which changed into PS **ə̂* before a following PU **-a* (Janhunen 1981: 223-224).

41. PU **juka* > PS **jə̂kå* [no Ng] ‘small river’
42. PU **jupta-* > PS **jə̂ptə̂-* [Sk; der. Ng] ‘say, speak’
71. PU **kuδa* > PS **kə̂rə̂* [Sk-Km] ‘morning’
75. PU **kuma-* > PS **kə̂mə̂-* [no Km] ‘fall over’
76. PU **kuńa-* > PS **kə̂ńə̂-* [all] ‘close eyes’
78. PU **kupsa-* > PS **kə̂ptå-* [all] ‘extinguish’
79. PU **kura-* > PS **kə̂rå-* [Ng-En-Ne-Km; der. Sk] ‘bend; crooked, bent’
80. PU **kura* > PS der. **kə̂ru* [En-Ne-Mt] ‘knife’
100. PU **lupsa* > PS **jə̂ptå* [all] ‘dew’
106. PU **muδ′a* > PS **mə̂jå* [Ng-Sk-Km-Mt] ‘earth, land, terrain’
107. PU **muja-* > PS **mə̂jə̂-* [Ne] ‘become happy’
108. PU **muka* > PS **mə̂kå* [all; der. Km] ‘back’
109. PU **muna* > PS **mə̂nå* [Ng-En-Sk; der. Km] ‘egg’
110. PU **mura* > PS der. **mə̂rVŋkå* [Ng-En-Ne] ‘cloudberry’
142. PU **puδ′a-* > PS **pə̂jä-* [Ng; der. En-Ne-Sk-Mt] ‘chop wood’
144. PU **puna-* > PS **pə̂n-* [Ng-Sk-Km; der. En-Ne] ‘braid, weave’
158. PU **suδ′a* > PS **tə̂jå* [only Mt] ‘finger’
178. PU **tunta-* > PS **tə̂ntå-* [En-Sk; der. Ne] ‘teach, accustom to’

In some examples, like in PU **kuma-* > PS **kə̂mə̂-* ‘fall over’, the PU **-a* was changed into PS **-ə̂* (see also subsection 3.7.1), which must have happened later. The exceptions, in which PU **u* was not reduced before a second-syllable **-a*, all contained an intervocalic **w* in Proto-Uralic. The rule can thus be formulated as “**u-a* > **ə̂-å/ə̂* except in the sequence **uwa*” (Zhivlov 2023: 152). Whether this is due to a true retention of **u* in this environment or a reversal of the change from **ə̂w* to **uw* is difficult to tell.

82. PU **kuwakka* > PS **kuə̂kå-nå* [only Ng] ‘long ago’
147. PU **puwa-* > PS **puå-* [Sk; der. all] ‘blow’, der. **puål-* ‘blow’
201. PU **(w)uwa* > PS **wuə̂-* [only Ng] ‘current’

There are a few examples where **u* was reduced, even though the conditioning environment as described so far did not apply.

81. PU **kurki* > PS der. **kə̂rö* [no Ng] ‘crane’
114. PU **nusi-* > PS **nə̂t-* [En-Ne-Sk-Km] ‘scrape’
160. PU **suŋi* > PS **tə̂ŋ* [all] ‘summer’

For some of these aberrant cases, paradigmatic alternations could be assumed. For example, PU **nusi-* ‘scratch’ would originally have formed a verbal noun based on the

consonant stem **nus-*, thus **nus-ma*, and an aorist stem **nus-ŋa-*, from which an expected allomorph **nâs-* may have been levelled to other forms of the paradigm, so that the expected imperative **nutâ-k* changed to PS **nâtâ-k*. PU **suŋi* 'summer' may have undergone the vowel change in certain case forms like the old locative **suŋna* > **tâŋnå* 'in summer', or in other, similar formations. The same might have applied for PU **kurki* 'crane', but it is more difficult to imagine that a locative pre-PS **kur(k)na*, or an ablative pre-PS **kur(k)ta*, or the poss.3sg. pre-PS **kur(k)sa* would have been common enough to spread via analogy. The suffix **-ö* (or **-V w*, see Salminen 2012: 341, Kaiheinen 2023: 64) could theoretically have had some influence, depending on its age in this word and the precise origins of the suffix.

For completeness, the remaining words in which PU **u* did not change to PS **â* are listed below. It is indeed clear that all these examples have a second-syllable PU **-i* and not **-a* (cf. Janhunen 1981: 233).

43. PU **juri-* > PS **ju/ürâ-* [En-Ne-Sk-Km] 'get lost; forget'
77. PU **kuńći* > PS **kunsâ* [Ng-Sk-Km-Mt] 'urine'
141. PU **pučki* > PS **pucâ* [all] 'hollow stalk'
146. PU **purki* > PS **pur* [ders. Ne-Sk?] 'smoke'
159. PU **suksi* > PS **tutâ* [Ng-En] 'ski'
157. PU **suγi-* > PS **tuâ-* [Sk; der. Ng] 'row'
175. PU **tuli* > PS **tuj* [all] 'fire'
177. PU **tumti-* > PS **tumtâ-* [no Sk] 'know'
181. PU **ukti* > PS **utâ* [Ng-En-Ne] 'passage, way'
182. PU **ulki* > PS **uj* [no Km] 'pole'

PU **i* similarly changed to PS **ə* in some cases, although the circumstances are not as clear due to a smaller number of examples (Janhunen 1981: 225, 237). The majority of reliable examples do show a second-syllable **-ä*, however.

| PU **i-ä* yielding a PS **ə* |
|---|
| 11. PU **ćilmä* > PS **səjmä* [all] 'eye' |
| 34. PU **ipsä-* > PS **əptä-* [Ne] 'smell' |
| 36. PU **itä-* > PS der. ?**əti/u-* [all] 'appear' |
| 104. PU **minä* > PS **mən* [all] '1sg.' |
| 168. PU **tinä* > PS **tən* [no Ne] '2sg.' |

| PU **i-i* yielding a PS **ə* |
|---|
| 33. PU **imi-* > PS der. **əmmä(-)* [Ne] 'suckle; breast' |
| 35. PU **ipsi* > PS **əptə* [Ng-En-Ne-Sk] 'smell' |

Examples where PU **i* is continued as PS **i* are even more scarce (Janhunen 1981: 234). Moreover, they contain some instances where PU **i* is apparently preserved unchanged before **-ä*, namely in PS **ńimsä* 'breast' and **pirkä* 'high'. In both cases we are dealing with derivations from original PU *i*-stems, which are preserved regularly in PS **ńim-* 'suck' and **pir* 'high'. These most likely exerted analogical pressure to change the vowel in otherwise expected PS ***ńəmsä* and ***pərkä*.

| PU **i-i* yielding a PS **i* |
|---|
| 112. PU **nimi* > PS **nim* [all] 'name' |
| 119. PU **ńimi-* > PS **ńim-* [Sk; ders. all] 'suck' |
| 134. PU **piδi* > PS der. **pirə* [Ng-En-Ne-Sk] 'height' |

| PU **i-ä* yielding a PS **i* |
|---|
| 119. PU **ńimćä* > PS **ńimsä* [Sk] 'breast' |
| 135. PU **piδkä* > PS **pirkä* [no Ng] 'high' |

The remaining Proto-Uralic high vowel **ü* changed to PS **i*, except in PU **ćüńćä* > PS **sünsə* 'breast' and PU **künti* > PS **küntə* 'smoke, vapour'. In both cases, the **ü* precedes a cluster with a nasal, which could potentially be significant.

| PU **ü* > PS **i* (majority development) |
|---|
| 17. PU **ćüδ′i* > PS **sijə* [En-Sk-Mt-Km] 'coal' |
| 18. PU **ćülki-kći-* > PS **siə-sə-* [En-Sk-Km] 'spit' |
| 24. PU **δ′ümä* > PS **jimä* [all] 'lime' |
| 115. PU **nüδi* > PS **nir* [no Mt] 'handle' |
| 161. PU **süli* > PS **tij* [no Km] 'fathom' |
| 179. PU **tütki-* > PS **titə-* 'open' [Sk-Km] |
| 183. PU **üli-* > PS **i-* [no Mt?] 'space above' |
| 185. PU **üwä* > PS **(n\|j)iə* [all] 'belt' |

| PU **ü* = PS **ü* (exceptional cases) |
|---|
| 19. PU **ćüńćä* > PS **sünsə* [Ng-En-Ne?-Mt] 'breast' |
| 85. PU **künti* > PS **küntə* [all] 'smoke, vapour' |

Pystynen (2017) has suggested that the neighbouring **k* may have been of some influence in **küntə*, with the additional etymologies PU **kütki-* 'tie' to PS **küt-* 'tie' and PU **ńüktä-*

'pull' to PS **n$^{l}$üt$^{l}$-*[40] 'pull', but neither of these etymologies seem to be compelling. As far as I can tell, Pystynen's PS **küt-* refers to **ket$^{l}$-* in SW, which is only reflected as a verb in Km. *šüʔ-/šöʔ-* 'tie, sew'. However, this may also derive from PS **såjtə-* 'sew' (SW 135; see KSz 1081), and the derivatives meaning 'thread', under SW **ketâj*, could plausibly be derived from the root PS **kätə-* 'wrap up' instead; in any case, only Kamas allows for PS **ü*. As for PS **n$^{l}$üt$^{l}$-* 'pull', the long vowel of NeT *nyúcy°* requires a different Proto-Samoyed reconstruction, such as PS **n$^{l}$üət$^{l}$-* or **n$^{l}$üjt$^{l}$-*, which cannot regularly be provided by PU **ńüktä-*.

There are two etymologies where PU **ü* may have developed into PS **ə*, similar to **i* and **u*, but neither involves a following vowel **-A*:

84. PU **külki* > ? PS **käj*, der. **käjwä* [Ng-En-Ne-Sk] 'side'
86. PU **küńil* > PS **käńäli* [all] 'tear'

After the development from **l* to **j*, the **ü* in both words would have followed **k* and preceded a palatal, which in combination may have had a particular effect. Otherwise, the reason for this development remains, at present, unknown. Analogical influence from the verb **kuńa-* 'close eyes', PS **kâńâ-* has been assumed to explain the vowel of PS **käńälį* 'tear' (Pystynen 2014), but since the verb is back vocalic while the noun is front vocalic (Ng. *kəəlï-ðï* 'his/her tear' rather than ***kəəlï-ðu*), this may be difficult.

On account of PU **ð'ümä* > PS **jümä* 'lime' it seems that the change from **ü* to **i* happened after the development from **i-ä* to **ə-ä* discussed above (Sammallahti 1990: 8), although, as Pystynen cautions (2014, and p.c. 2023), the **j* could potentially have had some assimilatory effect as well.

Since additional PS **ü* were created with the developments of PU **käwδi* > PS **kürə* 'rope' and PU **jäwji* > PS **jüjə* 'lichen (on trees)' (see subsection 3.5.3 above), the change from **ü* to **i* can be dated before the simplification of clusters with an initial *w*-component.

Word-initial PU **wi-* also yielded PS **(w)ü-* in at least two examples. The development of PU **wiγi-* 'drag', is separate from that of PU **weγi-* 'take', which did not develop a rounded front vowel, but seeing as the initial **w-* of the latter word was lost, it is difficult to connect this to the development of PU **eγi* to PS **i* in the relative chronology.

195. PU **wiγi-* > PS **ü-* [Ne-Sk] 'drag'
194. PU ±**wijit(t)i* > PS **wüət* [all] 'ten (10)'

[40] **n$^{l}$* is used when the attestations do not allow differentiation of PS **n* and **ń*, and **t$^{l}$* is used when the attestations do not allow for a differentiation of PS **t* and **c* (Janhunen 1977).

### 3.6.2 PU **ę* and PS **ę* and **į*

The Proto-Uralic back unrounded vowel **ę* underwent a phonemic split into two distinct Proto-Samoyed vowels: PS **ę* and **į*. This can be seen in the following lists of examples.

| PU **ę* > PS **į* (before original **-a*) | |
|---|---|
| 31. | PU **ęńa* > PS **įńâ* [Ne-Sk] 'tame' |
| 103. | PU **męksa* > PS **mįtâ* [no Mt] 'liver' |
| 153. | PU **sęksa* > PS **tįtå-jŋ* [no Ng] 'Siberian pine' |
| 163. | PU **šęra* > PS **tįrå-* [En-Ne-Sk-Mt] 'dry out (itr.)' |

| PU **ę* > PS **ę* (before original **-i*) | |
|---|---|
| 10. | PU **ćęrki* > PS **sęr* [all] 'ice; white; salt' |
| 23. | PU **δ´ęmi* > PS **jęm* [Sk-Km] 'bird cherry' |
| 26. | PU **ęćiw-* > PS **ęso-* [En-Ne] 'camp' |
| 32. | PU **ępti* > PS **ęptâ* [all] 'hair of the head' |
| 39. | PU **jęyi-* 'drink' > PS der. **ęr-* 'drink' [Ne-Sk-Mt], der. **ękâl-* 'id.' [En-Ne] |
| 59. | PU **kęčči-* > PS **kęcâ* [Sk] 'spoiled' |
| 96. | PU **lępći* > PS **lępsâ* [Ng-En] ~ **jępsâ* [Ne-Sk-Km-Mt] 'cradle' |
| 117. | PU **ńęli* > PS **ńęj* [Ne-Sk-Km-Mt] 'arrow' |
| 118. | PU **ńęri* 'wetness' > PS **ńęr* 'cartilage' [Ng-En-Ne-Sk] |
| 133. | PU **pęni-* > PS **pęn-* [all] 'put' |
| 154. | PU **sęni* > PS **cęn* (irregular **c-*) [all] 'sinew' |

This change is difficult to date, since it did not interact very strongly with other sound laws. Janhunen (1981: 233–234) and Sammallahti (1988: 484) inferred that the syllable structure influenced the reflex of PU **ę*, with a reflex as PS **į* being found only in open syllables. On the other hand, Aikio (Aikio 2015: 34) and recently also Zhivlov (2023: 151) take the vowel in the second syllable as the decisive conditioning factor: a sequence PU **ę-a* yielded a PS **į* while PU **ę-i* yielded PS **ę*. This accounts neatly for the majority of examples. In the relative chronology this should have happened before any PU **-a* had changed into **-â*.

The consistency of these developments is greater than in the case of the development from **u* and **i* to **â* and **ä* before **-A*, even though here, too, some paradigmatic alternations might be expected. For example, the aorist stem of pre-PS **pęnâ-* or already **pęn-* 'put' should have been **pęnŋå-*, whence perhaps regular ***pįnŋå-*. There are only two examples where PU **ę* changed to **į* before **-i*, both with a very similar phonological environment (Aikio 2015: 34; Zhivlov 2023: 151).

94. PU **lęmpi* > PS **lį̈mpə̑* [Ng-En-Ne-Sk] ‘pond, swamp’
95. PU **lęnti* > PS **lį̈ntə̑* [Ng] ‘lowland’

This is interpreted as an additional condition for raising to **į̈* before a nasal cluster (Aikio 2015: 34; Zhivlov 2023: 151). If not only coda nasals in a cluster but also word-final coda nasals should in principle have effected this phonological change, it could be dated before PU **δ́ęmi* ‘bird cherry’, **pęni-* ‘put’ and **sęni* ‘sinew’ developed into **jęm*, **pęn-* and **cęn* respectively. Paradigmatic alternations between otherwise expected **jęm* : gen.sg. **jęmə̑n*, etc., might also have eliminated this change by analogy if it did apply in these words, so that this is not a hard *terminus ante quem*, but it is simpler to avoid including them in the scope of the sound law by preservation of the final **-ə̑* at that point in time.

Zhivlov describes the raising of **ę* to **į̈* as taking place only after **l-*, which works for the two clearest and most straightforward examples (**lį̈mpə̑* and **lį̈ntə̑*; Zhivlov 2023). This would disconnect the development from other sound laws like apocope. However, this could complicate the development of PU **joŋsi* > PS **(j)į̈ntə̑* ‘bow’, this likely happened via an intermediary **jęŋsi*, which then developed further with the rule of raising before a nasal. If that development is restricted to **l_NC*, it implies a direct shift from **o* to **į̈* in **joŋsi* ‘bow’. Since **o* is two steps removed from **į̈*, and the environment **NC* applies in this word, it seems to me more economical to generalize pre-nasal raising in such a way as to include also **(j)į̈ntə̑*, i.e., as **_NC*.

Zhivlov (2023: 152) has proposed an additional, minor source of PS **į̈* from PU **u* in the specific environment **k_j* on account of two potential examples:

72. PU **kuji-* > ? PS der. **kį̈jtV* [Sk-Mt] ‘lie’
83. PU ±**kuwi-* > PS **kį̈j* [Ng-Km-Mt] ‘moon’[41]

The origin of the **j* in PS **kį̈j* ‘moon’ is uncertain, and while Mt. *kistə* ‘lie’ can go back to PS **kij-tə-*, the Sk. **kučal-* ‘lie down’, only attested in Ket Selkup, points to a vowel **u* (Helimski 1997: 280–281). Due to the great uncertainties involved, it will not be possible to draw any conclusions on the basis of this potential sound change.

[41] The original internal consonant of this word is difficult to accurately reconstruct in the absence of a Saami cognate. PU ±**kuwi* should have regularly become PS ***ku*; the **-j* in the Samoyed form might thus also be a derivational suffix of the same kind as seen in PS **soj* from PU **ćowi* ‘throat’ (Aikio 2020: 145).

### 3.6.3 PU **o*

In a number of words, Proto-Samoyed **o* is a puzzling entity from the Finno-Ugric perspective. The problem is that in a number of words a **o* in Samoyed corresponds to a **u* in the Finno-Ugric languages. This was first discussed by Janhunen (1981: 231), who explained it as a change in all other Uralic branches except Samoyed, which would then preserve the archaic situation. The conditioning would be PU **oCi* > FU **uCi*, as there are plenty of examples of PS **uCCə̂* = FU **uCCi*, such PU **kuńći* 'urine' > PS **kunsə̂*, FU **kuńći* and PU **pučki* > PS **pucə̂*, FU **pučki*. There are also two words where FU **oCi* corresponds to PS **o*, however.

| FU **u* ~ PS **o* | |
|---|---|
| 18. | FU **ćuwi* ~ PS der. **so-j* (or ?**soə̂-j*) [no Ng] 'throat' |
| 22. | FU **čučki* > PS **cocə̂* [Sk] 'log' |
| 69. | FU **kusi-* ~ PS **kot*(-) [all] 'cough (noun and verb)' |
| 98. | FU **lumi* ~ PS **jom-* [Ng-Ne-Sk] 'snow' |
| 120. | FU **ńuγi-* ~ PS **ńo-* [Ne-Sk-Mt] 'pursue' |
| 139. | FU **puri* ~ PS **por-* [Mt; der. Ng-En-Ne-Sk] 'bite' |
| 170. | FU **tuli-* ~ PS **toj-* [all] 'come' |
| 174. | FU **tɯwi* ~ PS **to* [all; der. Ng] 'lake' |
| 197. | FU **wuδ'i* ~ PS **oj-* [der. Sk] 'again, anew' |

| FU **o* ~ PS **o* | |
|---|---|
| 13. | FU **ćoji* ~ PS **soj-* [En-Ne] 'be heard, be audible' |
| 64. | FU **koki* ~ PS **ko-* [no Ng] 'look' |

The sound change posited by Janhunen has also been inverted, with the Finno-Ugric languages retaining an original **u* that changed to **o* in Samoyed in *i*-stems with an open syllable. This is the position taken by Zhivlov, who does add the exceptional position before **l* as retaining **u* (Zhivlov 2023: 152). This seems to be mainly on account of the example PU/FU **tuli* > PS **tuj* 'fire' (not ***toj*), but it yields an incorrect result for FU **tuli-* ~ PS **toj-* (~**tu-*) 'come'.

In either direction, the sound laws as suggested should have led to paradigmatic alternations between the vowel stems and the consonant stems: if an original **pori-* : **porma*, for example, developed with a sound law changing **CoCi* into **CuCi* (Janhunen 1981: 231), the result should have been **puri-* : **porma*. Conversely, **puri-* : **purma* should have regularly become **pori-* : **purma* with a change from **CuCi* to **CoCi* (Zhivlov 2023: 152). Such alternations should have been resolved in either Finno-Ugric or Samoyed in a

remarkably consistent way to result in the attested distribution between FU **u* and PS **o*.

There are two etymologies that favour Janhunen's interpretation. First, the opposition in Samoyed between PS **tuj* 'fire' and **toj-* 'come' as opposed to homophony in FU **tuli* 'fire' and **tuli-* 'come' is easier to understand if an earlier opposition of **tuli* 'fire' vs. **toli-* 'come' merged in Finno-Ugric, rather than if earlier **tuli* 'fire' and **tuli-* 'come' split in Samoyed. Second, the assumption that PS **ńo-* 'pursue' and FU **ńuɣi-* 'id.' derive from **ńoɣi-* rather than **ńuɣi-* allows for the derivation of PU **ńoma* 'hare' from this verb (Janhunen 2007: 221). Only **ńoɣ-ma* could yield PU **ńoma*, not ***ńuɣ-ma*.

At this point it does not seem possible to decide exactly how the differentiation of original **o* to PS **o* and FU **u* took place. However, we can see that the great majority of PU **o* became PS **å*, but that this did not happen to any PS **o* that corresponds to FU **u*.[42]

12. PU **ćoδ′a-* > PS **såjå-* [ders. Ng-Ne] 'war'
14. PU **ćojma* > PS **såjmå* [Ng-Sk] 'sound'
62. PU **koδ′ka* > PS **kåjkâ* [Ng-En-Ne-Sk] 'spirit, god, idol'
65. PU **kompa* > PS **kåmpå* [Ng-En-Ne-Sk] 'wave'
66. PU **konta* > PS **kåntâ-* [Ng-En-Ne-Mt] 'hunt; kill; sacrificial animal'
70. PU **koska* > PS **kåtå* [Ng-En-Ne-Sk?] 'grandmother'
68. PU **korpi* > PS **kårpâ* [En-Ne] 'blaze'
61. PU **koćki-* > PS **kåsâ-* [no Mt] 'dry (itr.), be dry'
105. PU **moćki-* > PS **måsâ-* [no Ng] 'wash'
117. PU **ńoma* > PS **ńåmå* [Ng-En-Ne-Sk] 'hare'
122. PU **oδ′a* > PS **åjå* [no Ng] 'meat, body'
123. PU **ojwa* > PS **åjwå* [Ng-En-Ne-Mt] 'head'
138. PU **ponči* 'tail' > PS **påncâ* [all] 'hem, lower edge'
156. PU **soski-* > PS der. **tåtu-* [Sk-Km] 'chew'
172. PU **tora-* > PS **tåro* [Ng-En-Ne-Mt] 'fight'
173. PU **totki* > PS der. **tåtu* [Sk] 'crucian carp'
169. PU **toɣi* > PS **tå-* ~ **tâ-* [all] 'bring'
196. PU **woča* > PS **wåc* [Ng-En-Ne-Sk] 'enclosure'
198. PU **woli-* > PS **åâ-*, **åj-* [all] 'be'
200. PU **woŋki* > PS **wåŋkâ* [Ng-En-Ne] 'hole'

[42] Notice the near complete absence of **oCi* in the following list, only **toɣi-*, ?**woli-* and **ćoji-* are reconstructed with this structure and have reflexes in Samoyed. Especially **toɣi-* > PS **tå-* (next to also **tâ-*) is puzzling besides PS **ńo-* 'pursue' from **ńoɣi-* and PS **tuâ-* from **suɣi-*: three different vowel reflexes apparently need to come from an original binary opposition.

This change, too, has a few exceptions, characterized by a retention of **o* "after **k* in an open syllable", according Zhivlov (2023: 151–152); but also **ćoma* 'good' and **ćoji-* 'be heard' retain the **o* despite not having initial **k* (see Aikio 2020: 135; Salminen 2023: 380–385). It cannot be said that **k* and **ć* fully patterned together in this regard, since PU **ćoδ'a* 'war' did become PS **såjå-* rather than ***sojå*. If Zhivlov's conditioning is otherwise correct, the developments of PU **koćki-* > PS **kåsə̂-* 'dry (itr.)' and PU **koska* > PS **kåtå* indicate that this change happened before the simplification of their internal clusters. It should also be borne in mind that, although the verb **ko-* 'look' looks like a clear open syllable in its citation form, it would practically always have been suffixed in some way, creating forms with both open and closed syllables; e.g., in the connegative **ko-k*, and the aorist stem **ko-ŋå-*. Any resulting alternations between **ko-* and ***kå-* expected to have resulted from regular sound change could have been levelled in favour of **ko-*. Similar alternations might have existed for PU **ćoji-*, PS **soj-* 'be heard, be audible', since the derivation PU **ćojma*, PS **såjmå* 'sound' does show a development from PU **o* to PS **å*.

13. PU **ćoji-* > PS **soj-* [En-Ne] 'be heard'
15. PU **ćoma* > PS **somå* [Ng-En-Ne-Sk] 'good'
63. PU **kojra* > **kora* > PS **korå* [no Mt] 'male reindeer'
64. PU **koki-* > PS **ko-* [no Ng] 'look'
67. PU **kopa* > PS **kopå* [all] 'skin'

It seems most economical to say that, assuming a shape PU **CoCi* for the type PS **por-* ~ FU **puri* (etc.) at the time, pre-PS **CoCi* did not change to **CåCi* while **CoCCi* and **CoC(C)a* did get a pre-PS **å*, and that PS **tå-/tə̂-* from PU **toγi* is an irregular exception. The sound law then took place before loss of final **-ə* and before simplification of clusters in PU **moćki-* > PS **måsə̂-* (etc.).

Not only **o*, but also PU **a* underwent a split development, namely into PS **a* and **å*, thus partly merging with the **o* that developed into PS **å*. This change is discussed in more detail in its own subsection directly after this one, but the basic premise is necessary for the understanding of a final change involving PU **o*. Namely, there are two examples in which PU **o* is reflected as PS **a* rather than **å*, in the specific environment PU **olC* (Zhivlov 2023: 152). The **l* in this position is later changed to **j*, but on account of the different development in PU **ojwa* > PS **åjwå* 'head' and PU **ćojma* > PS **såjmå* 'sound', the shift from **olC* to **alC* can be dated before **l* had merged with **j*. The examples below show how syllable-final **l* and **j* both became PS **j*.

14. PU **ćojma* > PS **såjmå* [Ng-Sk] 'sound, noise'
49. PU **kajwa* > PS **kajwå* [no Mt] 'spade'

123. PU **ojwa* > PS **åjwå* [no Ng] 'head'
165. PU **talwa-* > PS **tåjwå-* [no Sk] 'reach'
171. PU **tolwa* > PS **tajwå* [En-Ne-Mt] 'wedge'
199. PU **wolka* > PS **wajk* [Sk; der. in all] 'shoulder'

This development did not affect PU **alC* on account of PU **talwa-* > PS **tåjwå-* 'reach', which shows the general development of PU **a* to **å* instead (discussed in more detail in the next subsection). The only way to get PU **talwa-* to PS **tåjwå-* and PU **tolwa* to PS **tajwå*, assuming that both words underwent regular developments, is to date **a* to **å* before **o* to **å* and directly change **tolwa* to **talwa* in between these developments. With the reverse chronology, **tolwa* would first become **tålwå*, which leaves no room at all for it to cross with **talwa-* before that becomes **tålwå-* as well. This entails a relative chronology of 1) **a* > **å*; 2) **olC* > **alC*; 3a) **o* > **å*; 3b) **lC* > **jC*.

### 3.6.4 Summary

The relative chronology of the changes discussed in this section is summarized in Table 3.6. All vowel changes discussed happened before any PU **-a* became PS **-ə̂*, which leads us neatly to our next subsection.

Table 3.6: An overview of the changes discussed in this section and their relative chronology as far as can be determined.

| Sound changes | | Chronology | Treated in |
|---|---|---|---|
| A1 *u/i-A* | > *ə̂/ə̈-A* | before A2? | 3.6.1 |
| A2 *ü* | > *i* | before C1 | 3.6.1 |
| C1 *äwC* | > *ü* | before C2 | 3.6.1 |
| C2 *w/γC* | > *C* | | 3.4.3 |
| C3 *CC,kC,Ck* | > *C* | | 3.5.2 |
| D1 *oC(C)a* | > *åC(C)å* | before E2 | 3.6.3 |
| D2 *oCCi* | > *åCCə̂* | before C3 | 3.6.3 |
| D3 *ę-a* | > *į-å* | before E1? | 3.6.2 |
| E1 *CVCə* | > *CVC* | before C1, C2, C3 | 3.3 |
| E2 *-A* | > *-ə̂* | | 3.7.1 |

## 3.7 Changes affecting PU **a*

There are a number pre-Proto-Samoyed developments that are sometimes considered to be irregular. Most prominent among these is the split development of both first-syllable and second-syllable PU **a*.

### 3.7.1 PU second-syllable *-*a* to *-*å* and *ə̂*

We will begin our discussion with the second-syllable *-*a*, which underwent a reduction to PS *-*ə̂* in some cases (Janhunen 1981: 226–230). This happened in at least the following etyma:

2. PU **aŋa*- 'open' > PS **ńaŋə̂*- [Ng; der. Sk] 'take off (clothes)
7. PU **ćala*- > PS **sålə̂*- [Ng-Ne-Mt] 'lighten, flash'
29. PU **ęla* > PS **įlə̂* [no Mt] 'below (relational noun)'
31. PU **ęńa* > PS **įńə̂* [Ne-Sk] 'tame'
42. PU **jupta*- > PS **jə̂ptə̂*- [Ng-Sk] 'speak'
52. PU **kama* > PS **kamə̂* [Ne-Sk] 'scale'
62. PU **koδ'ka* > PS **kåjkə̂* [Ng-En-Ne-Sk] 'spirit, god; idol'
66. PU **konta* > PS **kåntə̂* [Ng-En-Ne; der. Mt] 'hunt, kill; sacrificial animal'
75. PU **kuma*- > PS **kə̂mə̂*- [no Km] 'fall'
76. PU **kuńa*- > PS **kə̂ńə̂*- [all] 'close eyes'
103. PU **męksa* > PS **mįtə̂* [no Mt] 'liver'
107. PU **muja*- > PS **mə̂jə̂*- [only Ne] 'smile'
125. PU **pata* > PS **patə̂*- [no Mt] 'put something in a pot'
164. PU **taka*- > PS **takə̂*- [no Mt] 'behind (relational noun)'

In a few instances, the PU *-*a* was lost. This probably happened via an intermediate stage pre-PS *-*ə̂*, which could be removed by analogy.

144. PU **puna*- > PS **pân*- [no Mt] 'weave'
191. PU **watka*- > ? PS **wåt*- [only Ne] 'debark (a tree)'
196. PU **woča* > PS **wåc* [Ng-En-Ne-Sk] 'enclosure'
199. PU **wolka* > PS **wajk* [Sk; der. in all] 'shoulder'

In a few rare instances PU *-*ä* was also affected by a similar reduction. In the pronouns, this could be easily understood as an irregular reduction. (The change in PU **üwä* > PS **iə* 'belt' (185) may be connected with the disappearance of the *-*w*- and the creation of a vowel sequence, and could thus be a separate phenomenon.)

19. PU **ćüńćä* > PS **sünsə̈* [Ng-En-Ne-Mt] 'breast'
104. PU **minä* > PS **mən*- [all] '1sg'
131. PU **peksä*- > PS **pet*- [Ng-Ne-Sk-Km] 'hit, soften a hide'
168. PU **tinä* > PS **tən*- [no Ne] '2sg'

Janhunen explained some of these by reference to the development of the sequences **-åj|äj* to PNe. **-ə*. This is seen in the oblique plural of, e.g., NeT *xoba* < **kopå* 'skin' obl.pl. *xob°* < **kopåj*. The Nenets form for 'eye' NeT *sæw°*, NeF *xäm°* < PNe. **sæmə* can also be explained by this, with an earlier paradigmatic alternation between PNe. **səjmä* : *səjmə*, the latter from PS **səjmäj* (Janhunen 1981: 228–229). However, PS **-Aj* does not seem to have become **-ə* in the other Samoyed languages, meaning that this type of development is possible only for Nenets; cf., e.g., Ng. acc.pl. *kubuj* < PS **kopåj* 'skin' and Ng. acc.pl. *śejmij* < PS **səjmäj* 'eye'. For the development of PU **-A* to **-ə* in Proto-Samoyed itself, another explanation has to be found.

To give a more complete picture, the list below contains examples of PU second-syllable **-a* becoming PS **-å*. With 29 examples as opposed to 19 in the list above, this is the majority development.

6. PU **ćaδa-* > PS **sårå-* [Ng-En-Ne-Sk] 'rain'
14. PU **ćojma* > PS **såjmå* [Ng-Sk] 'sound'
15. PU **ćoma* > PS **somå* [En-Ne-Sk; der. Ng.] 'good'
20. PU **čaŋa-* > PS **cåŋå-* [Ne-Km] 'rub, wear out (tr.)'
28. PU **ękta-* > PS **įtå-* [all] 'hang up'
41. PU **juka* > PS **jôkå* [no Ng.] 'river'
46. PU **kaja* > PS **kåjå* [all] 'sun'
49. PU **kajwa* > PS **kajwå* [Ng-En-Ne-Km] 'spade'
63. PU **kojra* > PS **korå* [no Mt.] 'male, male reindeer'
65. PU **kompa* > PS **kåmpå* [Ng-En-Ne-Sk] 'wave'
70. PU **koska* > PS **kåtå* [Ng-En-Ne-?Sk-Mt*] 'grandmother'
78. PU **kupsa-* > PS **kôptå-* [all] 'extinguish, put out'
79. PU **kura* > PS **kôrå* [Ng-En-Ne-Sk] 'bend'
87. PU **lapta* > PS **jåptå* [no Km] 'thin'
100. PU **lupsa* > PS **jôptå* [all] 'dew'
106. PU **muδ´a* > PS **môjå* [Ng-Sk-Km-Mt] 'earth, land, terrain'
108. PU **muka* > PS **môkå* [all] 'back'
109. PU **muna* > PS **mônå* [Ng-En-Sk-Km] 'egg'
121. PU **ńoma* > PS **ńåmå* [Ng-En-Ne-Sk] 'hare'
122. PU **oδ´a* > PS **åjå* [no Ng] 'meat, body'
123. PU **ojwa* > PS **åjwå* [Ng-En-Ne-Mt] 'head'
149. PU **sarka* > PS **tårkå* [En-Ne] 'fork, bifurcation, branch'
158. PU **suδ´a* > PS **tôjå* [only Mt] 'finger'
163. PU **šęra-* > PS **tįrå-* [En-Ne-?Sk; der. Mt] 'dry out (intr.)'
165. PU **talwa-* > PS **tåjwå-* [no Sk] 'reach'
171. PU **tolwa* > PS **tajwå* [En-Ne-Mt] 'nail, wedge'

178. PU **tunta-* > PS **tânt̊å-* [En-Ne-Sk] 'teach'
190. PU **wara* > PS **wårå* [Ng-Sk-?Km] 'mountain, ridge'

It is clear from the examples that the change to **-ə̂* in the earlier list looks late: both the changes of **ę* to **į* before **-a* (3.6.2) and of **u* to **ə̂* before **-a* (3.6.1) occurred in all instances, regardless of whether the **-a* in question ended up as PS **-å* or PS **-ə̂*. The resulting **-ə̂* is furthermore generally preserved after open syllables (5 cases of preservation vs. 2 of loss, which could be analogical), so that we may suppose that it was still **-a* at the time when **-ə* from original PU **-i* was removed in that environment (3.3). It can thus be dated as a late development in the relative chronology of Samoyed sound changes.

### 3.7.2 PU first-syllable **a* to **å* and **a*

There is another vowel change connected to the split of the second-syllable PU **-a* into **-ə̂* and **-å*, namely the development of first-syllable PU **a* into **a* or **å* (Aikio 2002: 50; Zhivlov 2014). The examples with PS **a* are here given first.

1. PU **aćkal* > PS **asVl-* [Sk-Km] 'step'
2. PU **aŋa-* 'open' > PS **ńaŋə̂-* [Ne-Sk] 'take off (clothes)'
3. PU **aŋi* > PS **aŋ* [all] 'mouth'
4. PU **aŋti* > PS **aŋtə̂* [all] 'blade, point'
37. PU **jasi* > PS **jat(ə̂)* [Sk] 'chilly weather'
52. PU **kama* > PS **kamə̂* [Ne-Sk] 'scale'
54. PU **kari* > PS **kar* [Ne] 'skin, surface, bark'
125. PU **pata* > PS **patə̂-* [no Mt] 'put something in a pot, in water'
126. PU **pawi* > PS **pa* [all] 'tree'
164. PU **taka-* > PS **takə̂-* [no Mt] 'behind (relational noun)'
186. PU **wajŋi* > PS **wajŋ* [all] 'breath'

The distribution appears to be a development of PU **a-a* > **a-(ə̂)* next to PU **a-a* > PS **å-å* (Zhivlov 2023: 151). The following examples show a development of PU **a-a* to PS **å-å*.

6. PU **ćaδa-* > PS **sårå-* [all] 'rain'
8. PU **ćaŋka-* > PS **såŋkå-* [Sk] 'stick into', der. *såŋkə̂* [En-Ne] 'penis'
20. PU **čaŋa-* > PS **cåŋå-* [Ne-Km] 'rub, wear out (tr.)'
44. PU **kaća-* 'give (as a gift)' > PS der. **kåso* (or ?**kåsåw*) [En-Ne-Sk] 'gift'
46. PU **kaja* > PS **kåjå* [all] 'sun'
53. PU **kanta-* > PS **kåntå-* [all] 'carry'

87. PU **lapta* > PS **jåptå* [no Km] 'thin'
111. PU **nataw* > PS **nåto* (or ?**nåtåw*) [Ne-Km] 'brother-in-law'
149. PU **sarka* > PS **tårkå* [En-Ne] 'fork, bifurcation, branch'
165. PU **talwa-* > PS **tåjwå-* [no Sk] 'reach'
189. PU **wančaw* > PS **wånco* (or ?**wåncåw*) [all] 'root'
190. PU **wara* > PS **wårå* [Ng-Sk-?Km] 'mountain, ridge'

Exceptions to the general distribution are PU **watka-* > **wåt-* 'debark (a tree)' rather than either ***wat(ə̂)* or ***wåtå-* and PU **ćala-* > PS **sålə̂-* 'lighten, flash' rather than expected ***sålä-*. The development of PU **kani-* > PS **kån-* rather than PS ***kan-* can be understood as a result of analogical influence from the causative PU **kanta-* > PS **kåntå-* (Zhivlov 2023: 160). Additionally, the aorist stem could have been a potential factor, since that should probably have regularly developed into PS **kånŋå-*. Similarly, PS **kåjtə̂-* 'be sick' could have been influenced by the derived noun PU **kajšaw* > PS **kåjto* 'sickness', where the second-syllable vowel may have facilitated rounding of the **a* to **å*. It is noteworthy that PU **kaji* and **kali-* developed into PS **kåə̂(-)* rather than ***kaə̂(-)*.

If interpreted as a development from pre-PS **å-ə̂* > PS **a-(ə̂)*, this change has to be dated before PU **o* had become **å*, since original PU **o* is never affected; cf. PU **woča* > PS **wåc* not ***wac*, PU **ponči* > PS **påncə̂* not ***pancə̂*, etc. A true exception seems to be PS **kajwå* 'spade' from PU **kajwa*, where the tautosyllabic palatal may have had some unexpected effect (Aikio 2002: 40).

The causation is somewhat circular here, as it makes about as much sense to say that second-syllable **-a* changed to **-ə̂* in those **a-a*-stems where the first **a* did not become PS **å*. A reverse development could also be posited as being the actual sound change, with PU **a-a* developing to both pre-PS **a-ə̂* and **a-å*, with the latter changing further to PS **å-å*. The same exceptions would of course still apply.

Since there is no known conditioning factor due to which PU second-syllable **-a* become PS **-å* and which become PS **-ə̂*, it seems expedient to set up two different *a*-phonemes at the pre-Proto-Samoyed level already, before one of them merged with PS **-ə̂* from PU **-i*. If a rounding from PU **a* to PS **å* is the sound change at hand, it makes sense to reconstruct one of these **a*'s as pre-PS **-å*, as opposed to pre-PS **-a* that ends up as PS **-ə̂*. A difference in development between, e.g., PU **taka-* and PU **čaŋa-* can then be written as developments from **taka-* to **takə̂* vs. **čaŋå-* to **čåŋå-*. Such a solution is posited for Proto-Uralic by Zhivlov, and may have reflexes in other Uralic branches as well (Zhivlov 2014). At the same time, it has to remain tentative. It would certainly make it easier to understand the Samoyed changes, which otherwise constitute an unmotivated split development.

### 3.7.3 PU second-syllable *-*a* to *-*ä*

Aside from the specific shift of **olC* to **ajC* to the exclusion of **ojC* to **åjC* discussed at the end of the previous subsection, there is another vowel change that seems to only have been caused by **l* and **δ′*, but not by **j*. This is the fronting of PU **-a* to **-ä* (Janhunen 1981: 221–222). It is noteworthy that the internal **a* still changed to **å* in these cases (cf. 3.7.2), so that it may have happened in a sequence from PU **kala* to *kålå* and PS **kålä* 'fish'. The examples and counterexamples are listed below.

| PU **-a* > PS **-ä* (special development in this environment) | |
|---|---|
| 45. | PU **kaδ′a-* > PS **kåjä-* [all] 'leave' |
| 50. | PU **kala* > PS **kålä* [all] 'fish' |
| 124. | PU **pala-* > PS **pålä-* [Ne-Sk] 'swallow' |
| 142. | PU **puδ′a-* > PS **pə̂jä-* [Ng; ders. En-Ne-Sk-Mt] 'chop wood' |
| 148. | PU **sala-* > PS **tålä-* [Ng-En-Ne-Sk; der. Km-Mt] 'steal' |
| 187. | PU **wala* > PS **wålä* [Ng-En-?Ne] 'song' |

| PU **-a* > PS **-å* (standard development) | |
|---|---|
| 12. | PU **ćoδ′a-* > PS **såjå-* [ders. Ng-Ne] 'war' |
| 106. | PU **muδ′a* > PS **mə̂jå* [Ng-Sk-Km-Mt] 'earth, land, terrain' |
| 122. | PU **oδ′a* > PS **åjå* [no Ng] 'meat, body' |

PU **ćala-* to PS **sålə̂-* 'lighten, flash' does not show this development, in addition to its unexpected vocalism **å-ə̂* (expected would be either ***sålä-* or ***salə̂-*). The one available example indicating that **j* did not have the same effect is PU **kaja*, which became PS **kåjå* 'sun', not ***kåjä*. In a sequence **oδ′a*, the stem-final **-a* was not affected in this way, and developed as expected; e.g., PU **oδ′a* 'wet, raw' > PS **åjå* 'meat, body', PU **ćoδ′a* > PS **såjå-* 'war'. The sequence PU **uδ′a* shows a contradictory development in PU **puδ′a-* to PS **pə̂jä-* 'chop wood' next to PU **muδ′a* to PS **mə̂jå* 'earth', so that perhaps some levelling of earlier paradigmatic alternations was involved.

## 3.8 Information from language contact: Yukaghir

There are various theories on the relationship between Yukaghir and Uralic, or Yukaghir and Samoyed. According to some scholars, Yukaghir is a distant relative of the Uralic languages (e.g., Collinder 1940; Piispanen 2013; Nikolaeva 2020). However, it would be prudent to wait until our knowledge of historical changes in Yukaghir and pre-Proto-Yukaghir reconstruction has advanced further, as this is still an underresearched topic (cf. Zhivlov 2022). An alternative account of especially lexical similarities between Yukaghir and Samoyed in particular has been to consider an early exchange of

borrowings (Rédei 1999; J. Häkkinen 2012; Aikio 2014d). The phonological nature of these proposed borrowings may give us some potential insights into Samoyed sound changes, similar to how Iranian borrowings could inform us on early Tocharian developments. I base the following on Aikio's (2014d) evaluation of the most likely loanword material.

Unfortunately, the proposed Yukaghir borrowings paint a somewhat confusing picture, with a number of contradicting correspondences. This is especially true as regards the sound **l*. On the one hand, there are instances of a correspondence of PU **l* to pre-PY **l*, as in PU **mälki* (would-be-PS **mäə*/*mäj*) ⇒ PYuk. *mel* 'breast', and PU **lańća* (would-be-PS **jåńćå*) ⇒ PYuk. **lančə* 'calm, quiet'. On the other hand, the correspondence between PS **lâmto* 'low' (no PU etymology) and PYuk. **lamtə* 'low' indicates that this word at least entered Samoyed after **l-* had already changed to **j-*; otherwise, this word should have taken part in that change and become PS ***jâmto* (cf. PU **lupsa* > PS **jâptå* 'dew'). At such a point in time, PU **lańća* should already have become pre-PS **jańća*, making borrowing as pre-PYuk. ±**lančə* difficult. These correspondences could be saved by positing that Samoyed **l* had already become palatalized (i.e., PU **lupsa* had changed into **Ĺupså* and PU **lańća* into **Ĺåńćå*), but that this **Ĺ* was borrowed as pre-PYuk. **l*. Borrowings in the other direction (such as PS **lâmto*) could then have been rendered with pre-PS **l* without danger of this being later palatalized to **Ĺ* and eventual PS **j* (as in potential pre-PS **Ĺupså* to PS **jâptå*).

Further interesting examples are PU **ćoδ'ka* (would-be-PS **såjkå* 'goldeneye'), potentially borrowed as PYuk. *so*/*aľqə* 'loon' (both are diving birds), where Yukaghir could show an earlier stage of the development of PU **δ'* before it became PS **j*; conversely, in PU **puδ'a*, which became PS **pâjä-* 'strike, hit; chop wood' and was potentially borrowed as PYuk. **paj-* 'strike, hit', this development appears to have been completed already. PYuk. **paj-* may thus point to a relatively later stage of Samoyed at the time of borrowing than PYuk. *so*/*aľqə*, if both Yukaghir words are correctly etymologized

The comparison between PS **sâppâ-* 'hit, chop' and PYuk. **sapa-* 'hit' (possibly pre-PYuk. **sappa-* vel sim; cf. Zhivlov 2022 on a reconstruction of Proto-Yukaghir voiceless stops as pre-Proto-Yukaghir geminates) indicates a stage of Samoyed after geminate simplification had run its course (3.5.2); that is to say, a rather advanced stage of late pre-Proto-Samoyed. This word could also be of onomatopoeic origin, however, and very similar words are found in other Uralic languages, such as SaaN *čuohppat* 'chop, cut' and Hu. *csap-* 'strike'. The Samoyed word cannot regularly be related, since it should then be PS ***sapâ-* with different vocalism and a singleton stop, and Saami *č-* normally corresponds to Hu. *sz-* [s] rather than to *cs-* [tʃ].[43] PYuk. **weg-* 'lead', if compared to PU

[43] TB *t*$^{s}$*apa-* 'hit' is also very similar, perhaps especially to Yukaghir *sapa-* 'hit', but it might equally be sound-symbolic.

**wiγi-* > PS **ü-* 'take (somewhere), lead', on the other hand, would again suggest an early stage of Samoyed, before intervocalic **-γ-* had been lost, and before **ü* had become *i* or **wi* changed to **ü*.

These issues may be understood in two ways: either the period of contact between pre-Proto-Samoyed and pre-Proto-Yukaghir was long, allowing for continued borrowing as sound changes occurred on both sides, or some of these etymologies are mistaken and should be left aside. It may also be the case that both statements are true. At the current state of knowledge as regards pre-Proto-Yukaghir sound developments, it is impossible for me to draw a well-founded conclusion fit as an instrument to test further hypotheses, and more research in that area is definitely needed.

## 3.9 Overview of Samoyed relative chronology

This final section will give an overview of how the changes altered the Samoyed phonological system through time, providing us with a relative chronology that can be compared with the Tocharian phonology of the pervious chapter. The starting point for Samoyed was the Proto-Uralic system, repeated here in Table 3.7 and Table 3.8.

Table 3.7: The Proto-Uralic consonant inventory.

| PU | labial | dental | retroflex | palatal | velar |
|---|---|---|---|---|---|
| stop | *p* | *t* | *č* | *ć* | *k* |
| fricative? | | *δ* | | *δ′* | *γ*/*x* |
| sibilant | | *s* | *š* | | |
| nasal | *m* | *n* | | *ń* | *ŋ* |
| liquid | | *l r* | | | |
| glide | *w* | | | *j* | |

Table 3.8: The Proto-Uralic vowel inventory.

| PU | front | central | back |
|---|---|---|---|
| high | *i ü* | | *u* |
| mid | *e* | *ę* | *o* |
| low | *ä* | *a* | |

An early change seems to be that PU **a* became differentiated into pre-PS **a* and **å*, although the exact mechanism remains elusive (3.7.2). More difficult to date is the split of PU **ę* into pre-PS **ę* and **į* (3.6.2). This is because the precise conditioning is quite complicated, and only potentially interacts with a number of other shifts. It definitely pre-dated the reduction of some PU second-syllable **-a* to PS **-ə̂*, but that itself appears

to be a late development (3.7.1). Another conditioning for a change from PU **ę* to **į* concerns a pre-nasal environment, where it may have applied at a relatively early stage, before apocope of PU **-i* (3.3). An intermediate pre-Proto-Samoyed phonological system that resulted from these changes is provided in Table 3.9 and Table 3.10.

Table 3.9: A representation of a possible intermediate, pre-Proto-Samoyed consonant inventory.

| pre-PS 1 | labial | dental | retroflex | palatal | velar |
|---|---|---|---|---|---|
| stop | *p* | *t* | *č* | *ć* | *k* |
| fricative | | *s* > *ϑ?* | (*š*) | | (*γ*) |
| nasal | *m* | *n* | | *ń* | *ŋ* |
| liquid | | *l r* | | (*ĺ?*) | |
| glide | *w* | | | *j* | |

Table 3.10: Potential early changes in the pre-Proto-Samoyed vowel system.

| PU | front | central | back | | front | central | back |
|---|---|---|---|---|---|---|---|
| high | *i ü* | | *u* | | *i ü* | *į* | *u* |
| mid | *e* | *ę* | *o* | > | *e* | *ę* | *o* |
| low | *ä* | *a* | | | *ä* | *a* | *å* |

The shift from PU **s* and **š* to PS **t* is difficult to date, but it probably happened after PU **δ* had become **r*, or at least after PU **-t-* after a second-syllable vowel had merged with PU **δ* (3.4.1). The change from PU **s* and **š* to PS **t* possibly took place via an intermediate stage like **ϑ*. The complete disappearance of PU **γ* from the system may have been after apocope, but this is only supported by a single example: PU **täγri* 'full' > PS **tärə* (3.5.3). PU **i* and **u* became split to create two additional vowels **ə* and **ə̂*, potentially before most PU **ü* had become PS **i*, since the latter was apparently not affected (3.6.1). The simplification of certain consonant clusters and geminates seems to have happened at a late date (3.5.2). It is entirely uncertain when the marginal Proto-Samoyed vowel **ö* arose, since it is not known to regularly reflect any Proto-Uralic vowel. The Proto-Samoyed phonological system is repeated in Table 3.11 and Table 3.12.

Table 3.11: The Proto-Samoyed consonant inventory.

| PS | labial | dental | retroflex | palatal | velar |
|---|---|---|---|---|---|
| stop | *p* | *t* | *c* | | *k* |
| sibilant | | | | *s* | |
| nasal | *m* | *n* | | *ń* | *ŋ* |
| liquid | | *l r* | | | |
| glide | *w* | | | *j* | |

Table 3.12: Developments from a possible pre-Proto-Samoyed vowel system (left) to the Proto-Samoyed vowel system (right).

| pre-PS | front | central | back | PS | front | central | back |
|---|---|---|---|---|---|---|---|
| high | *i ü* | *ɨ* | *u* | | *i ü* | *ɨ* | *u* |
| mid | *e* | *ę* | *o* | > | *e ö* | *ę* | *o* |
| low | *ä* | *a* | *å* | | *ä* | *a* | *å* |
| lax? | | | | | | *ə ə̂* | |

The relative dating of some changes remains impossible due to their lack of interaction, and the two stages of the vowel system presented here could also be inverted, as illustrated in Table 3.13.

Table 3.13: Two alternative sequences of vowel developments from Proto-Uralic to Proto-Samoyed.

| | PU | | | pre-PS | | | PS | | |
|---|---|---|---|---|---|---|---|---|---|
| high | *i ü* | | *u* | *i ü* | *ɨ* | *u* | *i ü* | *ɨ* | *u* |
| mid | *e* | *ę* | *o* | *e* | *ę* | *o* | *e ö* | *ę* | *o* |
| low | *ä* | *a* | | *ä* | *a* | *å* | *ä* | *a* | *å* |
| lax? | | | | | | | | *ə ə̂* | |
| | PU | | | pre-PS | | | PS | | |
| high | *i ü* | | *u* | *i ü* | | *u* | *i ü* | *ɨ* | *u* |
| mid | *e* | *ę* | *o* | *e* | *ę* | *o* | *e ö* | *ę* | *o* |
| low | *ä* | *a* | | *ä* | *a* | *å* | *ä* | *a* | *å* |
| lax? | | | | | *ə ə̂* | | | *ə ə̂* | |

Future research into Samoyed etymology and sound change will surely still improve our understanding, but hopefully the present scheme can provide a starting point for more detailed comparison with other languages with which early Samoyed has come into contact, such as (pre-)Proto-Yukaghir. I furthermore hope that the particular difficulties that plague Samoyed historical phonology are made more explicit with this overview.

# 4 Phonological comparison

In this chapter, the conclusions from the previous two chapters will be considered together to determine whether the developments that took place between Proto-Indo-European and Proto-Tocharian, discussed in chapter 2, can be understood as the result of substrate influence from early Uralic or pre-Proto-Samoyed, of which the sound developments are treated in detail in chapter 3.

## 4.1 Introduction

The consonant system will be treated first in subsection 4.2; it constitutes the most obvious and long-standing point of comparison between Tocharian and Uralic (e.g., Ivanov 1985, Kallio 2001, Bednarczuk 2015, Peyrot 2019a, although with differing details). The vowel system of pre-Proto-Samoyed was first compared to that of pre-Proto-Tocharian by Peyrot (2019a), and I emended some of his conclusions in an article dedicated specifically to the comparison of pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems (Warries 2022). The comparison provided here in subsection 4.3 is basically in line with my conclusions in that article, which may be consulted for some additional considerations. I argue that pre-Proto-Tocharian vowel system as it would have looked before palatalization took place is particularly similar to the early pre-Proto-Samoyed vowel system, with the important caveat that it remains difficult to confidently date some of the important Samoyed vowel changes.

In order to evaluate whether the phonological systems of pre-Proto-Tocharian and pre-Proto-Samoyed are specifically more similar to one another than to those of other languages spoken in the area, subsections 4.2.2 and 4.3.2 will also briefly highlight the stop systems and vowel systems respectively of Proto-Turkic, Proto-Yeniseian and Proto-Yukaghir. These are the other known language groups that may have been in the region, in addition to early Uralic or early Samoyed. Proto-Yeniseian and Proto-Yukaghir were also involved in the comparison and prehistoric scenario advanced by Peyrot (2019a), with special reference to similarities to pre-Proto-Tocharian and pre-Proto-Samoyed in the vowel systems.

The final subsection of this chapter, subsection 4.4, will provide an overview of the similarities between the pre-Proto-Tocharian and pre-Proto-Samoyed phonological systems, and an interpretation of their potential significance. It is concluded that substrate influence from pre-Proto-Samoyed or a closely related language could explain the more unusual and drastic developments that took place in the development from Proto-Indo-European to Tocharian.

It furthermore turns out that Old Turkic phonology is clearly different from pre-Proto-Tocharian phonology, and so is Proto-Yeniseian in its stop system. The most recent

reconstruction of the Proto-Yeniseian vowel system is not particularly close to that of pre-Proto-Tocharian either, but ongoing research may still change this picture. On the other hand, the consonant system of especially pre-Proto-Yukaghir as recently reconstructed by Zhivlov (2022) is very similar to the pre-Proto-Tocharian and pre-Proto-Samoyed systems. The reconstruction of the pre-Proto-Yukaghir vowel system is not so clear, however. Early Uralic, and perhaps particularly pre-Proto-Samoyed thus remains the most securely reconstructed and therefore most viable comparison for pre-Proto-Tocharian phonology at this point in time.

## 4.2 Comparison of the consonant systems

The Tocharian lack of the Proto-Indo-European stop contrasts traditionally interpreted as voicing and aspiration is a strong indication of substrate influence, since it is quite a unique development within the Indo-European language family (Peyrot 2019a: 78) and also a relatively unusual feature of languages around the world (Kallio 2001: 226). There are no noticeable compensatory processes that shifted a voicing or aspiration contrast onto adjacent segments. Vowel quality and vowel length remained entirely unaffected. This means that a very meaningful double phonological contrast was given up entirely (on this point particularly Peyrot 2019a: 73).

For these reasons, the stop system has often been the main point of comparison with Uralic. The pervasive palatalization found for almost every Tocharian consonant has also been considered as a potential result of substrate influence (Ivanov 1985, Kallio 2001, Bednarczuk 2015), so this aspect will be addressed as well. The comparison with Proto-Uralic and pre-Proto-Samoyed will be considered first (subsection 4.2.1), and then the stop systems of other ancient languages of the area will be treated as well (subsection 4.2.2), in order to see whether the comparison with Uralic or Samoyed is specific enough to support the contact hypothesis.

### 4.2.1 The pre-Proto-Tocharian and pre-Proto-Samoyed stop systems

The Proto-Uralic and (pre-)Proto-Samoyed stop systems consisted primarily of a single series of voiceless stops, providing a good match for the stop system adopted by early Tocharian when the Proto-Indo-European distinctions were abandoned. This is regarded as the strongest indication of Uralic influence on Tocharian (Kallio 2001: 225–226; Bednarczuk 2015: 56, Peyrot 2019a: 78). The difference between Proto-Uralic and Proto-Samoyed is not significant in this regard, so that it is difficult to make any meaningful distinction between these different stages. Table 4.1 shows a comparison of the stop systems. It should be noted that, if the Tocharian stop system is derived from a PIE system without voicing distinctions, as in Kloekhorst's reconstruction (see 2.2, and 2.3.4.4 for a discussion), the transformation of the Tocharian stop system seems less significant.

Table 4.1: A summary of the possible transformation of the (post-)Proto-Indo-European consonant system into an early pre-Proto-Tocharian consonant system under the influence of pre-Proto-Samoyed.

| post-PIE | > pre-PT | cf. pre-PS | PU |
|---|---|---|---|
| *p*, *b*$^{h}$ | *p* | *p* | *p* |
| *t*, *d*$^{h}$ | *t* | *t* | *t* |
| *ty*, *d*$^{h}$*y* | *ċ* | *ć*, *č* | *ć*, *č* |
| *ḱ*, *ǵ*, *ǵ*$^{h}$<br>*k*, *g*, *g*$^{h}$ | } *k* | *k* | *k* |
| *k*$^{w}$, *g*$^{w}$, *g*$^{wh}$ | *k*$^{(w)}$ | – | – |

The development of the clusters **ty* and **d*$^{h}$*y* into a monophonemic affricate can be connected with either of the Uralic affricates palatal **ć* or retroflex **č*. Especially the palatal affricate provides an adequate match, since the original cluster **ty*/*d*$^{h}$*y* contained a palatal glide. Here, pre-PT *ċ* is used rather than *t*$^{s}$ to differentiate the result of **ty*/*d*$^{h}$*y* from that of word-initial PIE **d*, as these may have remained phonologically distinct until after palatalization (cf. subsections 2.3.4.2 and 2.3.4.3). Changes from palatal affricates like **ć* [tɕ] to [ts] are attested in various languages (see Kümmel 2007: 197–200 for examples), so that a phoneme like PU **ć* may also be an adequate intermediate stage for the development from **ty*/*d*$^{h}$*y* to PT **t*$^{s}$.

The palatal affricate PU **ć* eventually became a fricative in the Samoyed languages, but it may have retained its affrication in Proto-Samoyed (Zhivlov 2018; Zhivlov 2023: 150). The time depth of this change is, in any case, unknown, so that it is possible that pre-Proto-Samoyed still had an affricate **ć* at the point in time that it would have come into contact with pre-Proto-Tocharian. It is uncertain if retroflex PU **č*, PS **c*, which definitely retained its affricate character in Proto-Samoyed, can be compared to pre-PT **ċ*, PT **t*$^{s}$ (cf. Peyrot 2019a: 79–80). Within Uralic, a development from PU **č* to *c* [ts] is at least known from Saami (e.g., Korhonen 1988: 273).

The Tocharian development of PIE **d* is problematic, and it is not certain precisely how it changed through time (see 2.3.4.4). Due to the great uncertainties about the development of this phoneme in Tocharian, I think that it cannot be included in a comparison with the early Samoyed or Uralic stop system.

The most straightforward comparison between the pre-Proto-Tocharian and pre-Proto-Samoyed or Proto-Uralic stop systems does not yet contain the palatalization that characterized almost the entirety of the Proto-Tocharian consonant system (cf. Peyrot 2019a: 79 fn. 6). While pre-Proto-Samoyed and Proto-Uralic contained some palatalized consonants, namely **ć*/*ś*, **ń*, and potential **δ́*/*ĺ*, the prevalence of palatalization is not quite comparable to the full system of palatalized consonants in Proto-Tocharian.

Nevertheless, the palatalization in Tocharian has been compared to Uralic (Ivanov 1985: 413; Kallio 2001: 227–228; Bednarczuk 2015: 56); and similarly, palatalizations in Balto-Slavic and in Indo-Iranian have been connected with Uralic influence as well (Thomason & Kaufman 1988: 247–249; Kallio 2001: 228).

Kallio rightfully remarks that it is unknown whether Proto-Uralic perhaps had phonetic palatalization of consonants before front vowels (Kallio 2001: 228). The main difficulty here lies in the fact that reconstructions primarily concern phonological systems and phonological change, which may lag behind phonetic developments. It is thus entirely possible, perhaps even probable, that pre-Proto-Tocharian consonants were non-phonemically palatalized for a period of time before palatalization was phonemicized (2.4).

It is equally possible, but more difficult to determine, that automatic palatalization of consonants before front vowels was a phonetic feature in early pre-Proto-Samoyed. For example, it is possible that pre-PS **ki* was phonetically realized as palatalized *[kʲi], in opposition to pre-PS **kį* as non-palatalied *[ki]. Phonetic palatalization would have enhanced the contrast between these two phonemically distinct sequences in a natural way, and later in Samoyed the sequences PS **ki*, **ke*, **kä* were independently fully palatalized and assibilated in most Samoyed languages (Janhunen 1998: 467, bearing in mind that **e* in Janhunen's notation is **ä* in the modern notation).

Thus, while it is perhaps plausible that a phonetic palatalization contrast existed in early Samoyed, and that this could have been transferred to Tocharian through contact to become phonologized later, there is no way to establish the presence of this phonetic feature in the source language. Phonological reconstruction alone does not provide us with direct access to such phonetic interactions between the consonants and vowels. Further considerations of the vowel system will be addressed in subsection 4.3.1; first we will turn to the stop systems of other prehistoric languages of the region.

### 4.2.2 Stop systems of other languages in the region

In this subsection, the general character of the stop systems of Proto-Turkic, Proto-Yeniseian and (pre-)Proto-Yukaghir will be briefly summarized on the basis of the available literature, and compared to the consonant systems of early Tocharian and early Uralic/Samoyed that were discussed more extensively above.

Proto-Turkic had an opposition between strong/fortis stops and weak/lenis stops, as shown in Table 4.2. According to Róna-Tas, both series may originally have been voiceless, with an aspiration contrast (Róna-Tas 1998: 71). In Old Turkic, the weak consonants were voiced and perhaps fricative (Erdal 2004: 61ff.) This does not correspond to the Tocharian single stop series, which is precisely the most conspicuously non-Indo-European feature of the Tocharian phonological system. Substrate influence

from Proto-Turkic therefore clearly cannot explain the loss of Proto-Indo-European stop contrasts.

Table 4.2: The Old Turkic inventory of stops and related obstruents (Erdal 2004: 62).

| OTrk. | labial | alveolar | palatal | velar |
|---|---|---|---|---|
| voiceless | *p* | *t* | *č* | *k* |
| voiced | *b* | *d* | *y* | *g* |

The reconstruction of the Proto-Yeniseian consonant system is an ongoing effort, but both older reconstructions (Starostin 1982) and recent reconstruction (Vajda 2022) agree that Proto-Yeniseian had a phonological voicing opposition in its stops; see Table 4.3. This voicing opposition in Yeniseian does not accord with the pre-Proto-Tocharian development of general voicelessness.

Table 4.3: The Proto-Yeniseian stop inventory as reconstructed by Vajda (2022); the retroflex and labiovelar stops in parentheses are not reconstructed by Starostin (1982).

| PYen. | labial | alveolar | post-alveolar | retroflex | palatal | labiovelar | velar | uvular |
|---|---|---|---|---|---|---|---|---|
| voiceless | *p* | *t* | *č* | (*ʈ*) | *c* | (*kʷ*) | *k* | *q* |
| voiced | *b* | *d* | *ǰ* | (*ɖ*) | *ɟ* | (*gʷ*) | *g* | *ɢ* |

The Proto-Yukaghir consonant system can also be reconstructed with an opposition between voiceless and voiced stops. Many voiced stops probably developed from earlier consonant clusters like **-nt-* for **-d-*, etc. (Zhivlov 2022; the reconstruction by Nikolaeva (2006: 66) different in the details, interpreting Proto-Yukaghir voiced stops as fricatives instead).

Table 4.4: Proto-Yukaghir stops and affricates (Zhivlov 2022: 47).

| PYuk. | labial | dental | palatal | velar |
|---|---|---|---|---|
| voiceless | *p* | *t c* | *č* | *k* |
| voiced | *b* | *d ʒ* | *ǯ* | *g* |

Zhivlov has argued that pre-Proto-Yukaghir can be reconstructed entirely without this voicing opposition, and that Proto-Yukaghir voiceless stops can be traced back to earlier geminates and specific changes in consonant clusters (Zhivlov 2022: 53–57). This entails a consonant system for pre-Proto-Yukaghir as in the table below; this is very similar to both the pre-Proto-Tocharian and pre-Proto-Samoyed systems.

Table 4.5: A reconstruction of the pre-Proto-Yukaghir consonant system (Zhivlov 2022: 57).

| pre-PYuk. | labial | dental | palatal | velar |
|---|---|---|---|---|
| stop | *p* | *t* | | *k* |
| affricate | | *c* | *č* | |
| nasal | *m* | *n* | *ń* | *ŋ* |
| liquid | | *l r* | *l´* | |
| glide | *w* | | *j* | |

Out of the three early languages of the region, Proto-Turkic, Proto-Yeniseian and (pre-)Proto-Yukaghir, only pre-Proto-Yukaghir may have had the same lack of voicing oppositions as found in Tocharian and Uralic/Samoyed. This means that, aside from early Uralic and Samoyed, early Yukaghir represents another potential source for the lack of voicing oppositions in the Tocharian stop system.

## 4.3 Comparison of the vowel systems

The Tocharian vowel system was first compared to potential outside sources by Peyrot (2019a), who argued for a significant similarity with the Ket (Yeniseian), Yukaghir and pre-Proto-Samoyed vowel systems. These are all marked by a lack of phonemic vowel length and the presence of two central vowels, in addition to the five cardinal vowels. Peyrot reconstructed an identical seven-vowel system for pre-Proto-Tocharian, allowing for a very neat comparison (Peyrot 2019a: 80–89). Chapter 3 shows that the Samoyed situation was more complicated, however.

### 4.3.1 The pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems

The comparison of the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems that Peyrot (2019a) made is very striking. However, it needed to be further informed by a detailed relative chronology of both pre-Proto-Tocharian and pre-Proto-Samoyed. I attempted to further work out the details in my article on the comparison of the pre-Proto-Tocharian and pre-Proto-Samoyed vowels systems, Warries (2022), and the stages of the vowel system as discussed for Tocharian chapter 2 and for Samoyed in chapter 3 derive from this endeavour. The diachronic developments of the pre-Proto-Tocharian and pre-Proto-Samoyed vowel inventories are here repeated in a condensed form in Table 4.6 and Table 4.7 below.

The first of the two tables, Table 4.6, contains the post-PIE vowel system, after laryngeal developments had taken place, and following on this the systems required before palatalization took place ("pre-pal.") and after palatalization took place ("post-

pal."), and finally Proto-Tocharian itself. The second of the two tables, Table 4.7, contains the development of the Samoyed vowel system. The exact chronology of developments is not as clear in this case, so that the relative order of the two intermediate pre-Proto-Samoyed stages is uncertain. Proto-Uralic and Proto-Samoyed vowel systems are given on the far left and the far right of the table respectively.

Table 4.6: An overview of the development of the pre-Proto-Tocharian vowel system.

| post-PIE | | | pre-PT pre-pal. | | | pre-PT post-pal. | | | PT | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *ī* | | (*ū*) | *i̯y* | | *uw* | *ʸi̯y* | | *⁽ʸ⁾uw* | *i* | | *u* |
| *i* | *R̥* | *u* | *i* | *"i̯"* | *u* | *⁽ʸ⁾i* | | *u* | | *ə* | |
| *e ē* | | *o ō* | *e* | *ë* | *o* | *⁽ʸ⁾e* | | *o* | *e* | | *o* |
| | *a ā* | | | *a* | *å* | | *a* | *å* | | *a* | *å* |

Table 4.7: An overview of the development of the pre-Proto-Samoyed vowel system.

| PU | | | pre-PS (uncertain relative order) | | | | | | PS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *i ü* | | *u* | *i ü* | *i̯* | *u* | *i ü* | | *u* | *i ü* | *i̯* | *u* |
| *e* | *e̯* | *o* | *e* | *e̯* | *o* | *e* | *e̯* | *o* | *e ö* | *e̯* | *o* |
| *ä* | *a* | | *ä* | *a* | *å* | *ä* | *a* | *å* | *ä* | *a* | *å* |
| | | | | | | | *ə ə̂* | | | *ə ə̂* | |

Since the comparison of the consonant systems pointed to a stage of pre-Proto-Tocharian without pervasive phonological palatalization, it follows that we should ideally also compare the pre-Proto-Tocharian vowel system as it looked before palatalization with pre-Proto-Samoyed. The vowel system represented here as "pre-PT pre-pal." is the minimal vowel system necessary before palatalization (cf. 2.4 and Warries 2022). Diphthongs like pre-PT **ey* and **ew* (PT **ʸi* and **ʸu*) are left out of the table to conserve space. The pre-Proto-Tocharian vowel inventory before palatalization is nearly identical to that of pre-Proto-Samoyed, apart from the three-way contrast between the low vowels **ä*, **a* and **å* and the presence of a rounded high front vowel **ü*. Here it should be borne in mind that Proto-Uralic and early Samoyed had sequences of a vowel and a glide **j* or **w*, which may correspond to the pre-Proto-Tocharian diphthongs written here as **i̯y* and **uw*, as well as the type **ey* and **ew* (cf. 2.4 and Warries 2022: 198-200).

While the pre-PS **ü* really has no convincing pre-Proto-Tocharian counterpart, a plausible counterpart to pre-PS **ä* can be found. The two pre-Proto-Tocharian front vowels **i* and **e* as they are represented in Table 4.8 derive from three Proto-Indo-European sources: PIE **i* and **e* can be theoretically merged as pre-PT **i* (after some **i*

had changed into *$i̇$, see 2.4.2), while PIE *$\bar{e}$ ends up as a short vowel like pre-PT *$e$. It is not strictly necessary to keep these vowels distinct for palatalization to take place normally (again, see 2.4.2 for the differences between etymological PIE *$i$ and *$e$ with regards to palatalization), but it is also possible that the merger of the short front vowels happened later. If PIE *$i$ and *$e$ were kept separate, we may imagine that PIE *$\bar{e}$ initially became a vowel like pre-PT *$æ$, differentiated from short PIE/pre-PT *$e$ by vowel quality. This would then entail the earlier presence of a more direct counterpart to the third front vowel pre-PS *$ä$ in pre-Proto-Tocharian as well. This is illustrated in the Table 4.8, where the full comparison between the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems is provided with the corresponding vowels side by side.

Table 4.8: A summary of the possible transformation of the (post-)Proto-Indo-European vowel system into an early pre-Proto-Tocharian vowel system under the influence of pre-Proto-Samoyed.

| post-PIE | > pre-PT | pre-PS |
|---|---|---|
| $i$ | $i$ | $i$ |
| – | – | $ü$ |
| $e$ | $e$ | $e$ |
| $a$ | $a$ | $a$ |
| $o$ | $ë$ | $ę$ |
| $u$ | $u$ | $u$ |
| $\underset{\circ}{R}$ | $i(R)$ | $i̯$ |
| $\bar{e}$ | $æ$ | $ä$ |
| $\bar{a}$ | $å$ | $å$ |
| $\bar{o}$ | $o$ | $o$ |
| $\bar{\imath}$ | $i̇y$ | $ij$ |
| $(\bar{u})$ | $uw$ | $uw$ |

Pre-Proto-Tocharian innovated a system without Proto-Indo-European vowel length, and with various new vowel oppositions. The types of oppositions, especially with the inclusion of central vowels, is very similar to the Proto-Uralic and early pre-Proto-Samoyed systems. Here the pre-Proto-Samoyed system is an especially close match due to the contrast between high *$i̯$ and *$ę$ that is not reconstructed for Proto-Uralic itself. Similarly, the presence of three low vowels *$ä$, *$a$ and *$å$ is a pre-Proto-Samoyed innovation that can be matched quite well with pre-Proto-Tocharian if PIE *$i$ and *$e$ merged later.

### 4.3.2 Vowel systems of other languages in the region

In this subsection, the vowel systems of Proto-Turkic, Proto-Yeniseian and (pre-)Proto-Yukaghir will be briefly summarized, and compared to those of early Tocharian and Uralic/Samoyed that were discussed more extensively in the previous subsection.

The Proto-Turkic and Old Turkic vowel system is markedly different from that of (pre-)Proto-Tocharian, with a length distinction and a very systematic, symmetrical oppositions between front and back vowels, and rounded and unrounded vowels (Róna-Tas 1998: 69–70; Erdal 2004: 45). This is shown in Table 4.9. The loss of vowel length with the creation of more vowel contrasts in pre-Proto-Tocharian thus does not look like the result of influence from Proto-Turkic.

Table 4.9: The Proto-Turkic vowel inventory (Róna-Tas 1998: 69–70).

| PTrk. | front | round | back | round |
|---|---|---|---|---|
| high | *i ī* | *ü ǖ* | *ï ī* | *u ū* |
| mid | | *ö ō* | | *o ō* |
| low | *ä ǟ* | | *a ā* | |

The exact nature of the Proto-Yeniseian vowel system is more contentious. The inventory reconstructed by Starostin (1982: 185) is very similar to the pre-Proto-Tocharian and the pre-Proto-Samoyed system in both its lack of opposition in vowel length, and in its relatively high number of quality contrasts; cf. Table 4.10.

Table 4.10: Starostin's reconstruction of the Proto-Yeniseian vowel system (1982: 185).

| PYen. | front | central | back |
|---|---|---|---|
| high | *i* | *ɨ* | *u* |
| mid | *e* | *ə* | *o* |
| low | *ä* | *a* | *ɔ* |

On the other hand, Vajda's most recent reconstruction is significantly different, with a reduction of the vowel system from nine vowels to five vowels (Vajda 2022: 266). The additional contrasts in the later Ket branch of Yeniseian are derived by Vajda from earlier combinations with the glides **j* and **w* (Vajda 2022: 262-273); cf. Table 4.11. It is this Proto-Ket vowel system that provides the close match to the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems in Peyrot's (2019a) comparison, but the time depth is unclear.

Table 4.11: Vajda's reconstructions of the Proto-Yeniseian (left) and Proto-Ket (right) vowel systems (2022: 262, 266).

| PYen. | front | central | back | > PKet | front | central | back |
|---|---|---|---|---|---|---|---|
| high | *i ij iw* | | *u uj uw* | | *i* | *ɯ* | *u* |
| mid | *e ej ew* | | *o oj ow* | | *e* | *ɤ* | *o* |
| low | | *a aj aw* | | | | *a* | |

The Proto-Yukaghir vowel system is reconstructed by Nikolaeva as in Table 4.12, but many details of the developments still need to be worked out (see Maslova 2008: 258), and the reconstruction will likely change upon further research. Long vowels in Yukaghir are usually dependent on syllable structure, or derive from contractions of earlier diphthongs (Nikolaeva 2006: 64–65, Zhivlov 2022: 55–56). The few long vowels that do not transparently derive via such processes may also be secondary, so that Nikolaeva hypothesises that the pre-Proto-Yukaghir vowel system originally lacked distinctive vowel length (Nikolaeva 2006: 65). This would entail a further similarity between the phonological systems of (pre-)Proto-Yukaghir, pre-Proto-Tocharian and pre-Proto-Samoyed, in line with the comparison by Peyrot (2019a). It should be borne in mind, however, that the reconstruction of the Proto-Yukaghir vowel system is still insecure, not to mention the pre-Proto-Yukaghir vowel system.

Table 4.12: A reconstruction of the Proto-Yukaghir vowel system (Nikolaeva 2006: 57).

| PYuk. | front | central | back |
|---|---|---|---|
| high | *i* (*ü*) | *y* | *u* |
| mid | *e* | *ö* | *o* |
| low | | *a* | |

Between Proto-Turkic, Proto-Yeniseian and Proto-Yukaghir, Proto-Turkic provides the worst match with the pre-Proto-Tocharian vowel system. Especially the presence of distinctive vowel length does not accord well with the fact that vowel length was lost in pre-Proto-Tocharian. Proto-Yeniseian (especially Proto-Ket) and Proto-Yukaghir show a closer correspondence with both pre-Proto-Tocharian and pre-Proto-Samoyed. Uncertainties concerning the time depth of the Proto-Ket system, derived from a more dissimilar Proto-Yeniseian system according to Vajda (2022), and concerning the general reconstruction of Yukaghir vowels, render a proper comparison with pre-Proto-Tocharian rather insecure.

## 4.4 Overview and conclusions

The pre-Proto-Tocharian phonological system, both in the stops and the vowels, probably went through a stage that bears a great likeness to pre-Proto-Samoyed. The Tocharian developments are striking and, in some cases, entirely unique among the known Indo-European languages and thus suggestive of substrate influence. As illustrated throughout this chapter the Proto-Uralic and early Samoyed phonological systems provide appropriate matches.

Tocharian more closely resembles this language group than it does most of the other known old languages of the area. It is clear that both Proto-Turkic and Proto-Yeniseian constitute an inadequate comparison for the pre-Proto-Tocharian phonological system, especially with regards to voicing oppositions in the case of Proto-Yeniseian. The earlier stages of pre-Proto-Yukaghir have not yet been reconstructed in detail, so that any comparison remains unreliable for the present. However, at least the consonant system seems to have been similar to that of (pre-)Proto-Samoyed, so that it could provide an adequate comparison for pre-Proto-Tocharian as well. There may have been early contact between Proto-Samoyed and pre-Proto-Yukaghir with both lexical exchange (Aikio 2014, J. Häkkinen 2012) and perhaps also structural convergence (suggested by Peyrot 2019a). It should be borne in mind that the time depth of Proto-Yukaghir is perhaps at most around 2000 years (Maslova 2003: 2), which is younger than Proto-Tocharian estimated around 2500 to 3000 years ago, and thus too late for a comparison with pre-Proto-Tocharian. The earlier stages of especially the vowel systems are uncertain, and the time depth that needs to be reached for a comparison with pre-Proto-Tocharian is a great obstacle.

Based on current knowledge, early Uralic, and possibly specifically pre-Proto-Samoyed, provides a good model for the changes observed in the Tocharian phonological system. The Table 4.13 and Table 4.14 below repeat the most adequate comparisons of the consonant systems and the vowel systems.

Table 4.13: A summary of the possible transformation of the (post-)Proto-Indo-European consonant system into an early pre-Proto-Tocharian consonant system under the influence of pre-Proto-Samoyed.

| PIE | > pre-PT | pre-PS | ~ PU |
|---|---|---|---|
| *p, bʰ* | *p* | *p* | *p* |
| *t, dʰ* | *t* | *t* | *t* |
| *ty, dʰy* | *ċ* | *ć, č* | *ć, č* |
| *ḱ, ǵ, ǵʰ* *k, g, gʰ* | *k* | *k* | *k* |
| *kʷ, gʷ, gʷʰ* | *k*$^{(w)}$ | – | – |

Table 4.14: A summary of the possible transformation of the (post-)Proto-Indo-European vowel system into an early pre-Proto-Tocharian vowel system under the influence of pre-Proto-Samoyed.

| post-PIE | > pre-PT | cf. pre-PS < | PU |
|---|---|---|---|
| *i* | *i* | *i* | *i* (*ü*) |
| – | – | *ü* | *ü* |
| *e* | *e* | *e* | *e* |
| *a* | *a* | *a* | *a* |
| *o* | *ë* | *ę* | *ę* |
| *u* | *u* | *u* | *u* |
| *R̥* | *i̇*(*R*) | *i̯* | *ę* |
| *ē* | *æ* | *ä* | *ä* |
| *ā* | *å* | *å* | *a* (*o*) |
| *ō* | *o* | *o* | *o* |
| *ī* | *i̇y* | *i̯j* | *i̯j* |
| (*ū*) | *uw* | *uw* | *uw* |

In terms of the phonological comparison, both the consonant and the vowel system of pre-Proto-Samoyed may be more closely matched to a pre-Proto-Tocharian system without pervasive phonemic palatalization. Tocharian palatalization has been considered as another aspect of Tocharian phonology influenced by Uralic substrate influence (Ivanov 1985: 413; Kallio 2001: 227–228; Bednarczuk 2015: 56), but in a phonological sense, this is difficult to substantiate beyond a general plausibility. Phonetic, automatic palatalization before front vowels may already have been a feature at the time in early Samoyed, so that it may still have been contact with a Uralic language that set Tocharian on its path toward palatalization. Unfortunately, the nature of phonological comparison makes it difficult to properly test this particular aspect of the contact hypothesis.

# 5 The Tocharian accent

In this chapter, I will consider the prehistory of the Tocharian accent in various relevant categories in order to compare it with Uralic/Samoyed. I will start from epenthesis and syncope as developments that can inform our view on regular, category-independent phonological developments in early Tocharian, and measure existing theories on the accent against that background.

## 5.1 Accentuation in Tocharian

One aspect of the phonology that has been entirely left aside until now is the stress placement or accentuation. Proto-Uralic and Proto-Samoyed had regular word-initial stress (e.g., Janhunen 1982: 27; Sammallahti 1988: 480), while Proto-Indo-European had a mobile accent (e.g., Beekes 2011: 153–158). The Tocharian accent can only be determined in Tocharian B, and there it falls most generally on the second syllable, unless the second syllable is also the (Proto-Tocharian) final syllable (see below for more details). While Tocharian A provides no direct evidence for stress placement in its spelling, the situation of Tocharian B is normally reconstructed for Proto-Tocharian as well (e.g., Pinault 2008: 416). This means that the Proto-Tocharian stress placement would be different from Uralic or Samoyed stress placement, and that Tocharian would thus not show any sign of Uralic influence in this respect (Kallio 2001: 230–231).

However, the eventual Tocharian stress on the second syllable constitutes an innovation vis-à-vis the Proto-Indo-European mobile accent. For a comparison with Uralic/Samoyed, we need to look at a stage of pre-Proto-Tocharian that predates Tocharian B accentuation, and from which it can be explained. The precise history of the placement of the accent in Tocharian is not agreed upon by scholars, and various models for its diachronic development exist. These models are not always systematically applied to all categories, so that it is difficult to get a complete overview. For example, Ringe (1987) is only concerned with the accentuation of nouns. Consequently, there are various views on the development of the accent in pre-Proto-Tocharian, several of which do assume an earlier stage of initial stress with later movement to the second syllable under various conditions (see below). Such scenarios amount to a closer match between pre-Proto-Tocharian and early Uralic or Samoyed in this regard. The remainder of the present section will introduce some general characteristics of the Tocharian B accent, how it is reflected in the spelling and its general behaviour, before we attempt to reconstruct a coherent account of the Tocharian accent in the following sections.

### 5.1.1 Basics of the Tocharian B accent

Since Tocharian has not survived until the modern day and is only known from ancient textual evidence, certain aspects of the language cannot be exactly recovered. The placement of stress is among those features that are difficult to access through written sources alone. However, at least classical Tocharian B provides some direct information on this aspect of the language in the way its vowels are spelled. Crucial in this regard are the phonemes TB /ə/ and /a/, as both have two allophones that are differentiated in the orthography. The written vowel ⟨ä⟩ is used to represent unaccented /ə/, while ⟨a⟩ can represent accented /ə́/. At the same time, ⟨a⟩ may also stand for an unaccented /a/, whereas the accented counterpart is written as ⟨ā⟩. Thus, accented /ə́/ and unaccented /a/ are written with the same sign ⟨a⟩. See Peyrot (2008: 33–41), also on the chronology of these orthographical practices within the manuscripts. Table 5.1 presents an overview of this situation.

Table 5.1: An overview of the spelling of stressed and unstressed /ə/ and /a/ in classical and late Tocharian B.

| phoneme | /ə/ | /ə́/ | /a/ | /á/ |
|---|---|---|---|---|
| spelling | ⟨ä⟩ | ⟨a⟩ | ⟨a⟩ | ⟨ā⟩ |

In archaic texts there is a more variable spelling, with ⟨ā⟩ and ⟨a⟩ representing both /á/ and /a/ interchangeably, while ⟨a⟩ also still sometimes stands for /ə́/ as well. This means that the accent in such texts cannot be established (Peyrot 2008: 31–39).

Paradigmatic alternations between ⟨a⟩ and ⟨ä⟩ on the one hand, and between ⟨ā⟩ and ⟨a⟩ on the other, can be understood as the result of shifts in the position of the accent. Such alternations can be seen in, e.g., TB *tāka* /táka/ 'he/she was' : *takāre* /takáre/ 'they were', *yaltse* /yə́ltˢe/ 'thousand' : *yältsenma* /yəltˢénma/ 'thousands'. As can be observed in these examples, the accent usually falls on the second syllable of the word, unless that is also the final syllable. Thus, if a word has a basic disyllabic form and inflected trisyllabic forms, the accent moves from the first to the second syllable in inflection. In trisyllabic words, no comparable movement is observed; e.g., *takāre-ne* /takárene/ 'they were to him/her' from *takāre* /takáre/ without movement beside *takā-ne* /takáne/ 'he/she was to him/her' from *tāka* /táka/ with movement.

Unaccented /ə/ is syncopated in open syllables, as in *camel* /cə́mel/ 'birth' : *cmela* /cméla/ births' (e.g., Marggraf 1970; Ringe 1987: 254–255). When a Proto-Tocharian final schwa was lost, the preceding syllable may still be accented, even if it is now the second syllable in a disyllabic word; e.g., obl.sg. *säsweṃ* /səswén/ to *saswe* /sə́swe/ 'lord', originally PT **səswenə* (Krause & Thomas 1960: 43–44).

In verse, even otherwise accented /ə́/ may be syncopated, often with accentual movement to the preceding syllable (on this topic in more detail, see Winter 1990, Pronk 2009). The accentuation of verse forms is clearly a secondary phenomenon, and will not be discussed further. The relationship between the graphemes for long ⟨ī⟩ and short ⟨i⟩ and between ⟨ū⟩ and ⟨u⟩ do not consistently indicate the place of the accent. Furthermore, the spelling of the phonemes /e/, /o/, /ay/ and /aw/ give no indication of accentuation at all, so that our understanding of the Tocharian B accent is essentially based on the behaviour of /ə/ and /a/ (Marggraf 1970: 9–10; Winter 1990: 371).

In Tocharian A there are no comparable alternations in the spelling to give any indication of the position of the accent. The graphemes ⟨ā⟩, ⟨a⟩ and ⟨ä⟩ each consistently represent separate Tocharian A sounds. For synchronic Tocharian A there is thus no direct way to determine the position of the accent. However, due to certain sound laws that operated in the prehistory of Tocharian A, it can be surmised that the position of the accent was on the first syllable, seeing as non-first syllable vowels undergo various processes of weakening and deletion. The position of the Tocharian A accent as can be inferred from sound changes will be elaborated in subsection 5.5.1.

This brief introduction has provided an overview of the Tocharian accent and how it is spelled, and now the following sections will be devoted to various topics that are closely associated with accentuation. Frequent reference will be made to other works concerned with the Tocharian accent and its development, in particular Marggraf (1970), Winter (especially 1993, 1994a), Pronk (2009), Malzahn (2010 *passim*), Peyrot (2013 *passim*), and Jasanoff (2015).

### 5.1.2 Problems with the Tocharian B accent

As seen in the examples provided in the previous subsection, the place of the accent often shifts due to inflection within a paradigm; e.g., TB *saswe* /sə́swe/ 'lord', obl.sg. TB *säswen* /səswén/. These shifts are mostly regular and predictable if the pre-TB syllable structure is taken into account: TB *säswen* is from earlier trisyllabic **səswénə*. Marggraf analysed this by setting up an underlying morphological form as, for example, |səswé-| for TB *saswe* 'lord'. The nominative singular TB *saswe* can then be synchronically analysed as the result of accent retraction from the final syllable to the preceding syllable, schematically |səswé| + # → /sə́swe/. The oblique TB *säswemṃ*, meanwhile, with its additional ending, can be understood as |səswé| + |nə| → /səswén/. The final schwa that was lost would thus still have its effects in the phonology of the accent.

Similarly, when a pronoun suffix is added to a verbal form like TB *kameṃ* 'they came', the position of the accent shifts to yield TB *kmeṃ-ne* 'they came to him/her/it', i.e., |kəmén| + # → /kə́men/ (TB *kameṃ*), but |kəmén| + |ne| → /kménne/ (TB *kmeṃ-ne*). As a result, the stem within a paradigm can generally be regarded as stable on the morphological level despite these accent changes, and any intraparadigmatic changes

observed can be synchronically ascribed to the retraction rule (Marggraf 1970: 15–17). The issue is that there is no concrete evidence that this synchronic morphophonological rule reflects the diachronic development of the accent. The accentual movement between **sáswe* and **səswénə* could theoretically be caused by a historical accentual retraction from **səswé* to **sáswe* beside a lack of retraction in **səswénə*, or from a forward shift in **səswénə* from an earlier **sáswenə* beside a lack of shift in **sáswe*. There is no immediate way to decide between these two possibilities, but our final choice should fall on the analysis that can best accommodate exceptions.

Aside from this general pattern, there are a number of categories that do not show second-syllable accent where it should otherwise be permissible (and thus, sychronically, expected) on account of the presence of a third syllable, but rather have their accent fixed on the first syllable. These categories are seen as aberrant from the point of view of general Tocharian B accentuation and thus require an explanation (e.g., Marggraf 1970: 24–31, Winter 1993, Peyrot 2013: 507–515, Jasanoff 2015). Persistent initial accent is found in many subjunctives of classes [1] and [5], in the causative present-subjunctive [class 9] and causative preterite [2], as well as most sigmatic forms of the preterite class [3] paradigm. Middle forms of thematic verbs belonging to present class [2] and [8] also have their accent fixed on the first syllable. It also concerns the secondary case forms of many disyllabic nouns. Most diachronic accounts of the accent revolve around scenarios designed to account for these categories, as will be discussed in more detail in sections 5.3 and 5.4.

Before delving into the accentually problematic verbal categories, I will first treat the insertion of epenthetic vowels and the loss of original vowels through syncope in order to create a framework against which we can test diachronic theories of the Tocharian accent. Epenthesis and syncope will be treated from a diachronic perspective in section 5.2, while the rest of the present subsection provides an introduction to the interaction between these two phenomena and the Tocharian B accent.

In various verbal and non-verbal categories in Tocharian B we find an epenthetic vowel /ə/, which is invariably accented wherever it occurs, i.e., /ə́/. However, since before epenthesis occurred, no /ə/ was present to be accented at all, our model of the Tocharian accent should explain both how epenthetic /ə/ arose, and how it came to be accented. For this reason, I argue that a diachronic understanding of Tocharian epenthesis should be taken into account in any theory on the development of the Tocharian accent.

The accent also interacted with syncope, which affected several forms cited above already, such as the plural *cmela* /cméla/ from *camel* /cə́mel/ 'birth'. In general, in Tocharian B, any schwa in an open syllable is syncopated unless it is accented. This is the reason for the syncope in *cmela* /cméla/ from earlier **cəmela*. A number of original PIE **e* disappeared without a trace in Tocharian even in the second syllable, where they should have been accented under normal Tocharian B accentuation rules, and thus

immune to syncope. This implies that some kind of early syncope took place before the establishment of Tocharian B accentuation as we know it. This early syncope provides another avenue for our further understanding of the diachronic development of the Tochairan accent, and will be treated in the next section along with epenthesis.

## 5.2 Epenthesis and syncope

Epenthesis and syncope both affect the Tocharian vowel schwa /ə/, which can derive from five distinct sources. It regularly developed from PIE **i*, **e* and **u*, and from vocalization of syllabic resonants (cf. subsections 2.3.2, 2.4.2 and 2.4.3 for more details), as well as from epenthesis in certain consonant clusters. The latter is our primary concern here. On the other hand, some word-internal etymological PIE **e* vowels are not continued as a Tocharian schwa, which means that they were lost due to syncope. Both developments are important for our diachronic understanding of the position and movement of the Tocharian accent.

### 5.2.1 General remarks on epenthesis

Both Tocharian languages show epenthesis in a number of clusters. In Tocharian B the epenthetic vowel is written ⟨a⟩, representing an accented schwa /ə́/, while in Tocharian A it is ⟨ä⟩. Not all instances of epenthesis in Tocharian A and B are necessarily part of the same phenomenon. For instance, although the comparison of TA *āṣtär* with TB *astare* /astə́re/ 'pure' or TA *pärkär* with TB *pärkare* /pərkə́re/ 'long' give the impression that **astəre* and **pərkəre*, with schwa, have to be reconstructed for Proto-Tocharian, this is not necessarily the case. In Tocharian A, a schwa is also found in very similar environments in words like *tpär* 'high', *rtär* 'red', even though the Tocharian B cognates *tapre* /tə́pre/ and *ratre* /rə́tre/ show no epenthetic schwa at all. The insertion of *ä* in Tocharian A is a synchronically automatic phenomenon. Since no schwa is found in Tocharian B, a Proto-Tocharian reconstruction for such words as ***təpəre*, ***rətəre* is unmotivated, especially since such forms should yield TB ***tpare*, ***rtare* with regular Tocharian B accent on the second syllable. The correct reconstructions are thus PT **təpre* 'high' and **rətre* 'red'.

The schwa found in TA *āṣtär* 'pure' and *pärkär* 'long' can simply be conditioned by word-final position of the cluster after apocope (cf. the m.obl.sg. *āṣträṃ* and *pärkreṃ* without epenthesis in the same cluster), making it ambiguous whether to reconstruct PT **astre* and **pərkre* or **astəre* and **pərkəre*. TB *astare* and *pärkare* could theoretically be derived from **astre* and **pərkre* with a different epenthesis rule than the one that operated also in TA *tpär* and *rtär*. The epenthesis in TB *astare* and *pärkare* might be dated either before Proto-Tocharian or after Proto-Tocharian in pre-Tocharian B. In either case, this epenthesis rule is of great interest to the development of the Tocharian

B accent. After all, the resulting epenthetic vowel ended up being accented in TB /astə́re/, /pərkə́re/.

In Tocharian B, epenthesis of this kind has been analysed in different ways. According to Ringe (1987), the epenthesis occurred in a cluster *-CCR-*, or in *-CR-* when preceded by the Proto-Indo-European accent; e.g., PIE **bʰr̥ǵʰrós* > TB *pärkare* /pərkə́re/ 'long', PIE **kʷékʷlos* 'circle, wheel' > TB *kokale* /kokə́le/ 'cart'. Either presence or lack of epenthesis would be the only reflex of the original accentuation otherwise preserved to a late stage of Tocharian. The only two examples with epenthesis due to an original Proto-Indo-European initial accent are TB *kokale* 'cart' from PIE **kʷékʷlos* 'circle, wheel' and TB *ṣpane* 'sleep' from PIE **su̯épnos* 'id.'.[44] Its explanatory power is thus very low, and Ringe does not attempt to apply these rules outside of the noun.

Epenthesis was disconnected from the placement of the Proto-Indo-European accent by Pronk (2009: 101–103), who argued that epenthesis occurred in any cluster *-CR-* that was not preceded by just a schwa in the syllable before it; in other words, any sequence *-CCR-* or *-*$V_{[\text{non-}ə]}$*CR-* underwent epenthesis. Looking at some of the same examples as above, in accordance with this rule it occurred in (pre-)PT **astre* as it developed to TB *astare* /astə́re/ 'pure' and also in (pre-)PT **wəntre* to TB *wäntare* /wəntə́re/ 'thing' (both representing the condition *-CCR-*), but not in PT **təpre* to TB *tapre* /tə́pre/ 'high' (representing the type *-əCR-*).

A very similar rule concerning consonant clusters and "heavy" or "full" vowels (i.e., non-*ə*) was also posited by Kim to account for the distribution of the allomorphs *-ná-* and *-ána* in Tocharian B class [6] nasal presents. Kim explicitly regards also */i/ and */u/ as light vowels like */ə/, and thus not causing epenthesis (Kim 2007a). However, these epenthesis rules as formulated by Pronk and Kim do not work entirely as expected. There are a number of examples of *-CR-* that are consistently attested without epenthesis even after a heavy vowel, e.g., *etre* 'hero' rather than ***etare*, *prākre* 'firm, hard, solid' rather than ***prakare*, *ṣotri* 'sign' rather than ***ṣotari*, etc. For Pronk (2009), the only example treated in his article that would actually suggest influence from a heavy vowel is *waṣamo* 'friend', and with a diphthong *aiśamo* 'wise'. In both cases, the palatalized consonant before the schwa indicates that there may have been a palatalizing vowel there originally, so that probably no epenthesis would be required at all.[45]

---

[44] TB *ṣkas* /ṣkə́s/ 'six (6)' from PIE **s(u)eḱs* 'id.' was also derived in this way by Ringe (1987: 258–259, 263), but he later, correctly, pointed out that this form was abstracted from the ordinal TB *ṣkaste* (Ringe 1996: 76). TA *ṣäk* is the regular reflex of PIE **s(u)eḱs*.

[45] Pronk (2009: 82–83) suggests that the palatalization in these forms is secondary, based on the absence of palatalization in TB *orkamo* 'dark' with the same suffix. However, the basis for the derivatives may have been different. For instance, TB *ayk-* 'know' forms a thematic present, which shows the allomorphs |ayśə-| and |ayke-|, and if the former was used as the basis for *aiśamo* 'wise', the schwa may have been introduced in that way; cf. also *ṣlyamo** 'flying' and *alāṣmo* 'sick', both

For Kim (2007a) the conditioned epenthesis in *-CR-* after a heavy vowel is relevant for /rapə́na-/ 'dig; plough', /walə́na-/ 'cover', and /tˢapə́na-/ 'crush' with a heavy vowel, and for /laykə́na-/ 'wash', /kautə́na-/ 'chop' and /kraupə́na-/ 'gather' with a diphthong. On the other hand, no epenthesis occurred in /tˢakná-/* 'bite' (attested *tsāknaṃ* with accent retraction, in prose), while it did in /ritə́na-/ 'seek', both counter to Kim's rule. It thus seems that a further specification is desirable, and it will be necessary to consider the development of the *na*-presents [class 6] in more detail.

### 5.2.2 Epenthesis in *na*-presents

The development of the nasal suffix of the *na*-presents [class 6] is very relevant for our conception of epenthesis, seeing as the variant *-ə́na-* ⟨ana⟩ next to *-ná-* ⟨nā⟩ can best be derived by such a mechanism.[46] As mentioned at the end of the previous section, Kim (2007a) has explained the distribution of the suffix allomorphs based on a combination of vowel weight and consonant clusters. Kim determined that a root containing either a heavy vowel or a consonant cluster received an epenthetic vowel, and thus showed the allomorph *-ə́na-*. Earlier, Marggraf (1970: 17–18) also gave the consonant cluster as a conditioning, although he did not go into as much detail. Kim mentions as the remaining counterexamples *tələ́na-* 'bear' and *kərsná-* 'cut off', leaving aside *tˢakna-* 'pierce, bite' and *pərsná-* 'sprinkle' as only appearing in metrical texts (Kim 2007a: 80–81 incl. fn. 30). Based on those counterexamples, as well as *ritə́na-* 'seek', *kəltˢná-* 'oppress' and *wərná-* 'turn (intr.) (?)' (apparently from a root **wərta-*), Malzahn (2010: 414–416) rejects Kim's proposed rule. Lack of epenthesis in *tˢakna-* 'bite' and *pərsna-* 'sprinkle' is found also in prose, Malzahn points out, so that these counterexamples cannot be dismissed so easily. Malzahn instead derives the suffix variant *-ə́na-* from an original locus in *tələ́na-* restored from *təllá-* from **təlna-* 'bear' (Malzahn 2010: 416–417). However, Malzahn's preferred account leaves the significant correlation of root shape and epenthesis as noted by Kim and Marggraf unexplained. I will argue that only one of the conditions posited by Kim is correct: roots ending in a final cluster induced epenthesis of the type *-CCn-* > *-CCən-*. The root vowel is of no influence, however. Apparent exceptions become understandable if

from the palatalized root allomorph. Furthermore, a structure |-əmo|, with a stable schwa, may need to be set up for this suffix at some point in time, on account of *mlamo* from *məl-* 'melt': *|məlmo| should have yielded simply **malmo* by regular sound law due to liquid metathesis changing *CRə́C* to *Cə́RC*.

[46] Although Pronk (2009: 85–86) does entertain the possibility that the schwa arose as a reflex of an original syllabic resonant in the third person plural, e.g., **kr̥s-n̥h₂-enti* > **kərsənan* and analogical spread in some verbs. The eventual distribution, as we will see, speaks against such a scenario.

we consider how analogy would logically have applied after the effects of some other Tocharian B sound laws involving metathesis, deletion and assimilation.

#### 5.2.2.1 The suffix variant *-ná-*

The majority of *na*-presents that have the suffix allomorph *-ná-* ⟨-nā-⟩ are indeed based on roots that originally ended in a single consonant. The most straightforward examples are given here in Table 5.2.

Table 5.2: Tocharian B nasal presents without epenthesis.

| Root | Meaning | Diagnostic examples | Analysis |
|---|---|---|---|
| *klətˢa-* | 'oppress' | prs.ger. *kälsnāle*[(a)] THT 497 a4 | \|klətˢ‹n›á-\| |
| *kʷəsa-* | 'lament' | prs.mid.3sg. *kwäsnāträ* THT 88 b1 | \|kʷəs‹n›á-\| |
| *təla-* | 'bear' | prs.act.3pl. *tallaṃ* THT 44 b2 | \|təl‹n›á-\| |
| | | prs.ger. *tällālle* PK AS 6A b2 | |
| | | prs.mid.3sg-1sg. *tlanatär-ñ* THT 17 a2 | \|təl‹án›a-\| |
| | | prs.mid.3pl-1sg. *tlanantär-ñ* IOL Toch 169 b3 | |
| *trəwka-* | 'allot to' | prs.ger. *truknālle* THT 497 b5 | \|trəwk‹n›á-\| |
| *pəka-* | 'intend' | prs.mid.2sg. *päknāstar* THT 45 a1 | \|pək‹n›á-\| |
| | | prs.mid.3sg. *päknāstär* PK AS 17K a3 | |
| | | prs.mid.3pl. *päknāskenträ* THT 273 a2 | |
| | | ipf.mid.3sg. *päknāṣṣiträ* PK NS 76 a5 | |
| *pərsa-* | 'sprinkle' | ipf.act.3pl. *pärsnoṃ* THT 17 a4 | \|pərs‹n›á-\| |
| | | prs.ger. *pärsnālle* PK AS 8C b7 | |
| *pəla-* | 'praise' | prs.mid.1sg. *pällāmar* THT 596 a4 | \|pəl‹n›á-\| |
| | | prs.mid.3sg. *pällāträ* THT 16 a6 | |
| | | prs.mid.3pl. *pällānträ* IOL Toch 1020 a2 | |
| *prəwka-* | 'leap' | prs.act.3sg. *pruknaṃ* PK AS 7M a6 | \|prəwk‹n›á-\| |
| | | prs.mid.3pl. *pruknānträ* IOL Toch 1 a4 | |
| | | ipf.act.3pl. *pruknoyeṃ* PK AS 15B b4 | |
| *məwsa-* | 'rise (itr.)' | prs.mid.3sg. *musnāträ* THT 119 a2 | \|məws‹n›á-\| |
| | | prs.mid.3pl. *musnāntär* IOL Toch 1 a5 | |
| *yəka-* | 'be careless' | prs.mid.3sg. *yäknāstär* THT 305 b6 | \|yək‹n›á-\| |
| | | sbj.mid.3pl. *yäknāntär* THT 13907 b7 | |
| *səwka-* | 'bring' | prs.act.3sg. *suknaṃ* PK AS 15C b4 | \|səwk‹n›á-\| |
| | | prs.act.3pl. *suknaṃ* IOL Toch 14 b2 | |
| | | prs.ptc. *suknāmane* PK AS 17C a2 | |

| Root | Meaning | Diagnostic examples | Analysis |
|---|---|---|---|
| *skaya-* | 'strive' | prs.act.2sg. *skaināt* PK NS 167 b2<br>prs.act.3sg. *skainaṃ* PK AS 7M a2<br>prs.act.3pl. *skainaṃ* THT 286 a5<br>prs.ptc. *skaināmane* THT 95 a3<br>prs.ger. *skainālle* THT 192 a2 | \|skay‹n›á-\| |
| *skəra-* | 'scold' | prs.act.3sg. *skarraṃ* IOL Toch 124 a1<br>prs.ptc. *skärrāmane* THT 85 b6 | \|skər‹n›á-\| |
| *tˢaka-* | 'bite' | prs.act.3sg. *tsāknaṃ* IOL Toch 4 a6<br>ipf.act.3pl-2sg. *tsaknoyeñ-c* THT 231 a4 | \|tˢak‹n›a-\| |

(a) *kälsnāle* with a simplification of expected *-ltsn-* to *-lsn-*. The subjunctive act.2sg. *klātsāt* (THT 91 a2) shows that the original stem is *klətˢa-* rather than ***kəltˢa-* (see Peyrot 2013: 746).

This category also includes the *na*-subjunctives *päknā-* 'intend' and *yäknā-* 'be careless' (with their derived *nask*-presents [class 10]), and assimilated forms with earlier **-ln-* and **-rn-* in *tällā* from *təla-*[47] 'bear', *pällā-* from *pəla-* 'praise', and *skärrā-* from *skəra-* 'scold'. The present *pärsnā-* of *pərsa-* 'sprinkle; spray' can be included here if derived from **prəs‹n›a-* before liquid metathesis, just like *käl(t)snā-* from **klətˢ‹n›a-* 'oppress', cf. TA prt.act.3pl. *prasar*; in this way it is not a counterexample.[48]

Other *na*-presents with the suffix allomorph *-ná-* were subject to analogical changes after regular sound laws obscured the shape of the root. These are given in Table 5.3 and discussed below.

[47] Middle forms of this verb have the nasal suffix allomorph *-ə́na-* rather than *-ná-*, but those are probably created secondarily; cf. Malzahn (2010: 417).

[48] The attested finite forms with a nasal suffix prs.act.3pl. *parsnān* in THT 121 a6 (arch.) ipf.act.3pl. *pärsnoṃ* in THT 17 a4 are both attested in verse, and could thus represent syncopated forms. However, the gerund *pärsnālle* is found in prose in PK AS 8C b7.

Table 5.3: Tocharian B nasal presents with special developments and no epenthesis.

| Root | Meaning | Diagnostic examples | Analysis |
|---|---|---|---|
| *kəta-* | 'strew' | prs.act.3sg. *katnau* THT 205 a3 | \|kət‹n›a-\| |
| | | prs.act.3sg. *katnaṃ* THT 17 b2 | |
| | | prs.ger. *kätnālle* PK AS 3B a6 | |
| | | prs.act.3pl. *käṃtaṃ* (arch.) THT 545 a5 | \|kə‹n›ta-\| |
| | | prs.ger. *käntalle* THT 1535.a a5 | |
| *kəlpa-* | 'obtain' | sbj.act.2sg. *källāt* THT 286 a3 | \|kəl(p)‹n›a-\| |
| | | sbj.act.3sg. *kallaṃ* THT 27 a4 | |
| | | sbj.act.1pl. *källām* THT 17 a4 | |
| | | sbj.act.3pl. *kallaṃ* IOL Toch 253 a1 | |
| | | opt.act.3sg. *kalloy* THT 42 b5 | |
| | | sbj.ger. *källālle* THT 597 a1 | |
| | | vn. *källalñe* THT 181 a5 | |
| | | inf. *källātsi* THT 127 b5 | |
| *kərsta-* | 'cut off' | prs.act.2sg. *kärsnāt* THT 1573.a a3 | \|kərs(t)‹n›a-\| |
| | | prs.act.3sg. *karsnaṃ* PK AS 4A b1 | |
| | | prs.act.3sg-3pl. *kärsnān-m*= THT 2 b3 | |
| | | prs.act.3pl. *karsnaṃ* PK AS 7K a1 | |
| | | prs.mid.3sg. *kärsnātär* THT 18 a5 | |
| | | ipf.act.3pl.-2sg. *kärsnoyeñ-c* THT 231 a4 | |
| | | prs.ptc. *kärsnāmane* PK AS 1B a1 | |
| *kləwpa-* | 'rub' | prs.mid.3sg. *klupnāträ*[(a)] THT 334 b6 | \|kləwp‹n›a-\| |
| *wərta-* (?) | 'turn (itr.) ?' | prs.ptc. *wärnāmane* PK AS 17A b6 | \|wər(t)‹n›a-\| |

(a) Written as though *kluṣnāträ*, but see Adams (1983: 611; 2013: 244) for the interpretation of this form as *klupnāträ*; cf. also Malzahn (2010: 628)

The two roots *kəta-* and *klupa-* have only a single internal consonant, so that the allomorph *-ná-* is entirely expected. However, these consonants should have interacted with the *-n-* of the suffix by sound law. Regular forms were apparently restored without added epenthesis, after regular metathesis and assimilation had affected the integrity of the root: *kəta-* 'strew' shows both regular *käntā-* and renewed *kätnā-*. Similarly, *klupnā-* 'rub' has to be restored from expected ***klunmā-* from **kləwp‹n›a-*, with the Tocharian B change of **-pn-* to *-nm-*. The regular development of this cluster is seen for example in *yänmask-*, the present [class 10] to *yəp-* 'enter', and in *sänmetse* 'in a trance-like state' from **səpn-*, a root meaning 'sleep'. These phonological changes are important for our understanding of a number of *na*-presents with the allomorph *-ə́na-*, as will be discussed further below.

Next, *kärsnā-* from **krəst‹n›a-* 'cut off' and *wärnā-* from **wərt‹n›a-* 'turn (intr.) (?)' are two examples with *-ná-* that do have a stem with a consonant cluster, in both cases *-Ct-*, to which the *n-* of the nasal suffix is attached directly. It has to be assumed that the root-final *-t* was assimilated to the nasal suffix and deleted before epenthesis occurred; the deletion of the *-t* is necessary in any case to arrive at the attested forms. The nasal present *kärsnā-* 'cut off' thus derives regularly from **krəstna-* by first a change from **krəstna-* to **krəsna-* with deletion of the **-t*, bleeding the epenthesis rule, after which liquid metathesis yielded **kərsna-*, in its turn counterfeeding epenthesis.[49]

The simplification of *-rtn-* seen in *wärnā-* apparently also bled epenthesis and counterfed the assimilation seen in *skärrā-* < **skərna-* 'scold' (note, however, that *wärnā-* is a hapax legomenon in a prs.ptc. *wärnāmane*, and the *-rt* is not actually attested as such, only inferred based on the lack of assimilation of *-rn-* > *-rr-*; see Peyrot 2013: 822 with fn. 854).

A final striking form from a root with a consonant cluster is the *na*-subjunctive *kəllá-* (a former *na*-present; Peyrot 2013: 737 fn. 135) from *kəlp‹n›a-* 'obtain': apparently also the **-p-* was lost in the cluster **-lpn-*, after which *-ln-* underwent the same regular assimilation as also seen in *pällā-* 'honour, praise' and *tällā-* 'carry'. As it is probably safe to say that there are no analogical processes by which a stem *kəllá-* could be secondarily derived from a root *kəlpa-*, the development that led to this form must be regular by sound law. The fact that we see no epenthesis indicates that the sound law by which **-p-* was deleted in this context preceded epenthesis, since the cluster *-lpn-* should definitely count as heavy.

#### 5.2.2.2 The suffix variant *-ə́na-*

Turning now to those *na*-presents that have the suffix variant *-ə́na-* ⟨-ana-⟩, I will first give the examples that end in a consonant cluster and can have developed regularly by simple epenthesis in a cluster *-CCn-* here in Table 5.4.

Table 5.4: Tocharian B nasal presents with clusters and regular epenthesis.

| Root | Meaning | Diagnostic examples | Analysis |
|---|---|---|---|
| *kətka-* | 'cross' | prs.act.3sg. *kätkanaṃ* THT 29 b4<br>prs.ptc. *kätkanamāne* (arch.) SI B 119(2) b6 | \|kətk‹ə́n›a-\| |
| *kərka-* | 'steal' | ipf.mid.3sg. *kärkänoyträ* (arch.) THT 1312 b1<br>prs.ptc. *kärknāmane* PK AS 15D a3 (verse) | \|kərk‹ə́n›a-\| |

[49] See Kim (2007a: 87–97) for a more detailed account of the development. In short, an earlier shape of the root as |krVsta-| (with ablaut) can be inferred from the sbj.act.3sg. *krāstäṃ*, and according to Peyrot (2013: 742 fn. 171) also from the lack of palatalization in the preterite.

| Root | Meaning | Diagnostic examples | Analysis |
|---|---|---|---|
| *kərsa-* | 'know' | prs.act.2sg. *kärsänatä* (arch.) PK AS 12B b6<br>prs.act.3sg. *kärsanaṃ* THT 523 b6<br>prs.mid.3sg. *kärsanatär* THT 543 b2<br>prs.ger. *kärsanalle* PK AS 7B b4 | \|kərs‹ə́n›a-\| |
| *tərka-* | 'let go' | prs.act.1sg-2sg. *tärkanau-c* IOL Toch 31 b4<br>prs.act.2sg. *tärkanat* PK Bois B17 a1<br>prs.act.1pl. *tärkanam* THT 107 b2<br>prs.act.2pl. *tärkanacer* THT 108 a7<br>prs.ger. *tärkanalle* THT 337 b3 | \|tərk‹ə́n›a-\| |
| *nətka-* | 'push away' | prs.ger. *nätknallona* THT 8 b1 (verse syncope) | \|nətk‹ə́n›a-\| |
| *pləska-* | 'think' | prs.act.3sg. *pälskanaṃ* PK NS 58 a4<br>ipf.act.1sg. *pälskanoym* THT 496 a3<br>prs.ptc. *pälskanamane* THT 1681 a1 | \|pləsk‹ə́n›a-\| |
| *mansa-*[50] | 'be sad' | prs.mid.3sg. *mantsanatär* IOL Toch 1 a1<br>prs.ptc. *mantsanamane* PK NS 407 a1 | \|mant$^{s}$‹ə́n›a-\| |
| *yəksa-* | 'embrace' | prs.mid.3sg-3sg. *yäksanatär-ne* THT 354 a2 | \|yəks‹ə́n›a-\| |
| *latka* | 'cut off' | ipf.act.3pl-2sg. *latkanoyeñ-c* THT 231 a5 | \|latk‹ə́n›a-\| |
| *walt$^{s}$a-* | 'grind' | prs.act.3sg-pl. *waltsanan-me* PK AS 19.3 a1<br>prs.ger. *waltsanalle* PK AS 3A b3 | \|walt$^{s}$‹ə́n›a-\| |
| *stawkka-* | 'swell' | prs.mid.3sg.-pl. *staukkanatär-me* IOL Toch 306 b5 | \|stawkk‹ə́n›a-\| |
| *t$^{s}$arka-* | 'torment' | ipf.act.3sg. *tsarkanoy* PK NS 163 b6<br>ipf.act.3pl-2sg. *tsarkanoyeñ-c* THT 231 a5<br>prs.ger. *tsarkanalle* IOL Toch 213 a5 | \|t$^{s}$ark‹ə́n›a-\| |

Evidently, the clusters **-rkn-*, **-rsn-*, **-skn-*, **-ksn-*, **-tkn-*, **-nsn-* **-lt$^{s}$n-* and **-kkn-* avoided simplification, all undergoing epenthesis instead.

The verbs with a root ending in final *-tk* merit a brief elaboration because that type usually shows an *ṅk*-present [class 7] instead; *kətka-* 'cross' and *nətka-* 'push away' actually both have a *ṅk*-present [7] attested beside the *na*-present [6] listed in the table, and the Tocharian A cognate of *latka-* 'cut off' has a *ṅk*-present [7] as well. If these verbs had a *na*-present [6] **kətkna-*, **nətkna-*, **latkna-* at the time of epenthesis, they can of course be considered regular from that point onward, even if they were created analogically at an earlier stage.

[50] A secondary nasal present next to a thematic present, see Peyrot (2013: 791 fn. 588, 589).

The rest of the *na*-presents belonging to this group with the suffix variant *-ə́na-* require a more detailed explanation. They are provided in Table 5.5 and treated underneath.

Table 5.5: Tocharian B nasal presents with analogical epenthesis.

| Root | Meaning | Diagnostic examples | Analysis |
|---|---|---|---|
| *kawta-* | 'split' | ipf.act.3pl-2sg. *kautanoñ-c* THT 231 a5 | \|kawt‹ə́n›a-\| |
| *krawpa-* | 'gather' | prs.mid.3sg. *kraupana(t)rä* THT 1209 a2<br>prs.ger. *kraupanallona* THT 1209 b2–3 | \|krawp‹ə́n›a-\| |
| *rapa-* | 'dig; plough' | prs.act.3sg. *rapanaṃ* IOL Toch 246 a1<br>prs.mid.3sg. *rapanträ* THT 3998 a1 | \|rap‹ə́n›a-\| |
| *rəyta-* | 'seek' | prs.act.3sg. *(r)itanaṃ*[(a)]<br>prs.mid.3sg-2sg, *ritanatär-ś*[(a)] | \|rəyt‹ə́n›a-\| |
| *wala-* | 'cover' | prs.mid.1sg. *walanamar* IOL Toch 785 b1<br>prs.ger. *walanalle* W 14 b2 | \|wal‹ə́n›a-\| |
| *tˢapa-* | 'crush' | prs.ger. *tsapanale* PK AS 8B a2<br>prs.ger.f.pl. *tsapanallona* PK AS 19.1 a5 | \|tˢap‹ə́n›a-\| |

(a) Both in unpublished texts, see Malzahn (2010: 824)

First, I will consider some roots ending in a single consonant that ended up with an epenthetic vowel. This category includes the roots on which Kim (2007a) based his conclusion that heavy vowels caused epenthesis in the following cluster.

Two examples have a *-t-*: *rəyta-* 'seek' with the nasal present stem *ritana-* and *kawta-* 'chop' with the nasal present stem *kautana-* (cf. Table 5.5). Without epenthesis, which is definitely not expected in the former according to Kim's rule whereby only heavy vowels cause epenthesis in *-Cn-*, the cluster **-tn-* should have regularly metathesised to **-nt-*, yielding ***rintā-* from **rəyt‹n›a-* and ***kauntā-* from **kawt‹n›a-*. Such forms must have been restored to having a nasal suffix rather than a nasal infix, just as happened to the stems of *kəta-* 'strew': *kätnā-* is found next to *käntā-* (cf. Table 5.3). The only difference is that the other suffix allomorph **-ə́na-* was used instead of *-ná-* in these cases.

Even more extreme changes would have affected the roots ending in *-p*, namely **krawp‹n›a-* to **krawnma-* remade as *kraupana-* 'gather', **rap‹n›a-* to **ranma-* remade as *rapana-* 'dig; plough', and **tˢap‹n›a-* to **tˢanma-* remade as *tˢapana-* 'crush'. These analogical restorations are obviously well-motivated, and appropriate models would have been readily available from the type of nasal presents as listed in Table 5.4 above.

One further root with final *-l*, namely *wala-* 'cover', should have developed as **walná-* > **wallá-* by regular sound law, cf. **pəlna-* > *pällā-* 'honour, praise' in the previous

subsection. Since the attested stem is *walana-* /walə́na-/, a renewal must have taken place.

Thus, roots ending in single stop **-t* or **-p*, as well as those ending in a single **-l* are liable not to exhibit the forms that would be expected by sound law. Regular developments would have made either the root or the nasal suffix less recognisable, which meant that they could be reshaped by analogy. As a consequence, the variation between the suffixes *-ná-* and *-ə́na-* found in these roots has no real bearing on our rules for epenthesis; evidently both *-ná-* and *-ə́na-* were available for use in restoration. Since all of Kim's examples to suppose that a heavy vowel promoted epenthesis belong to this category (viz., *rapa-*, *wala-*, *tˢapa-*, *kaupa-*, *kauta-*), that rule can be left aside. The exceptional status of *tˢaka-* 'bite' with a full vowel but no epenthesis can be resolved in this way, and the remaining example of epenthesis with neither a consonant cluster or a heavy vowel, *rəyta-* 'seek', also belongs to the group where analogical developments are likely.

A few remaining roots ending in a cluster, which I argue is the sole requirement for epenthesis, require a little more explaining, but they do not change the general picture. They are listed in Table 5.6, and explained below.

Table 5.6: Tocharian B nasal presents with remaining special situations.

| Root | Meaning | Diagnostic examples | Analysis |
|---|---|---|---|
| *alpa-* | 'stroke' | prs.act.3pl. *alpanaṃ* IOL Toch 1 b2 | \|alp‹ə́n›a-\| |
| *kanta-* | 'rub off' | ipf.mid.3sg. *kantanoy*(*tr*)*ä* THT 429 a3 | \|kant‹ə́n›a-\| |
| *layka-* | 'wash' | prs.mid.3sg. *laikanaträ* IOL Toch 262 b4 | \|layk‹ə́n›a-\| |
| *wənta-* | 'cover' | prs.act.3pl-2sg. *wäntanañ-c* (arch.) THT 567 b3 | \|wənt‹ə́n›a-\| |
| *wərpa-* | 'enjoy' | prs.mid.1sg. *wärpanamar* IOL Toch 183 b1<br>prs.mid.3sg. *wärpanaträ* THT 42 b3<br>prs.mid.3pl. *wärpanantär* PK AS 7N b4<br>prs.ptc. *wärpanamane* THT 558 a1<br>prs.ger. *wärpanalle* THT 1220 b1 | \|wərp‹ə́n›a-\| |
| | | prs.mid.2sg. *wärpnātar* (arch.) THT 128 b2<br>prs.mid.3sg. *wärpnāträ* THT 521 b1 | \|wərp‹n›á-\| |
| *tˢalta-* | 'chew' | prs.ptc. *tsaltanamane* IOL Toch 269 b4 | \|tˢalt‹ə́n›a-\| |

The *na*-presents of *alpa-* 'stroke', *kanta-* 'rub off', *wənta-* 'cover', *wərpa-* 'enjoy' and *tˢəlta-* 'chew' are all special in some way. Based on the development of **kəlpna-* to *källā-* 'obtain' (discussed at the end of the previous subsection), which must be phonologically regular, also **alpna-* to ***alla-* is expected. We find *alpana-* 'stroke' instead, so that this root was

evidently renewed.[51] The motivation for such a renewal is obvious. A development of **wərpna-* to **wərna-* (or **wərra-*), whence analogically both /wərpə́na-/ and /wərpná-/ 'enjoy' can also be assumed.

The stem *laikana-* /laykə́na-/ from **layk‹n›a-* 'wash' does not neatly fit. This example could theoretically indicate that *-ay-* caused epenthesis (as assumed by Kim 2007a), and one could also think that the sequence **-ykn-* counted for the condition *-CCn-*, but I can see no clear further indication in support of either solution. The present *laikana-* was likely remade from earlier **laynka-*, with an old nasal infix (see Peyrot 2013: 539), so that it is probably not a reliable witness.

It is furthermore very likely that **wəntna-* 'cover' would have lost its root-final stop, becoming **wənna-*, fit for restoration to yield attested *wəntəna-*. The nasal present *tˢaltana-* 'chew' may be recent according to Peyrot (2013: 841), so either it was formed directly with the suffix *-ə́na-* after that had been established by epenthesis in other forms, or it is renewed from otherwise likely ***tˢalna* from **tˢaltna-*, cf. *kärsnā-* from **krəstna-* 'cut off'. The implication of this analysis of the suffix of the *na*-present [6] will be discussed in the next subsections.

#### 5.2.2.3 Conclusions and Tocharian A nasal presents

Based on the previous two subsections, the Tocharian *na*-presents [6] can be divided into three groups:

i) roots without an original root-internal cluster or with a cluster that was reduced early on (like **kəlpna-* > **kəlna-* > TB /kəlla-/ 'obtain') received the more basal allomorph *-na-*;
ii) roots with an original cluster that was not reduced due to contact with the nasal of the suffix generated a new allomorph **-əna-* from **-na-* by epenthesis;
iii) roots with a **-t-*, **-p-* or **-l-* that did not undergo epenthesis underwent assimilation and/or metathesis, and could be secondarily restored and redistributed to either of the other two categories.

The developments in the nasal presents attest to a relative chronology of five sound changes as follows:

---

[51] If one assumes with Kim (2007a) that the difference in the developments of pre-TB **alpna-* and **kəlpna-* is due to the *a*-vowel causing epenthesis in **alpna-* but not in **kəlpna-*, two separate epenthesis rules would be required. First epenthesis after an *a*-vowel to yield **alpəna-* beside **kəlpna*, followed by deletion of the **-p-*, and then second epenthesis in consonant clusters. This is not sufficiently supported by the evidence.

1) *-lpn-* > *-ln-* (and perhaps *-rpn-* > *-rn-*);
2) *-ln-*, *-rn-* > *-ll-*, *-rr-*;
3) *-rtn-*, *-stn-* > *-rn-*, *-sn-*;
4) epenthesis *-CCn-* > *-CCən-*;
5) liquid metathesis *CRəC* > *CəRC*.

The relative order of (2) before (5) was also pointed out by Malzahn (2010: 417) on the basis of *kəln-* 'resound/howl (?)' from earlier **klən-*. Changes (2) and (3) can be parallel to (4) and (5) if clusters of the type **Ctn* were immune to epenthesis anyway, perhaps because *t* and *n* are homorganic. The placement of the changes from **-tn-* to *-nt-* and from **-pn-* to *-nm-* in the relative chronology is not certain, but the latter is only found in Tocharian B. Metathesis of **-tn-* to **-nt-* is found in Tocharian A only in *länt-* 'go' from **lətn-*, and seems to have been undone or secondarily obscured elsewhere (Peyrot 2013: 446–447).

Liquid assimilation of *-ln-* > *-ll-* happened in Tocharian A as well, cf. TA prs.mid.3pl. *pällānträ*, TB prs.mid.3pl. *pällānträ* from **pəlna-* 'praise'. Based on the correspondence TA *tālo*, TB *tallā$_u$* 'miserable', Ringe considers it a shared development and traces it back to Proto-Tocharian (1996: 165). TA *tālo*, TB *tallā$_u$* could be derived from PT **talláwə* with regular simplification of the geminate in A. However, in most other instances of this sound law in Tocharian B, Tocharian A does show a geminate *-ll-*, which should then be from restored **-ln-*, e.g., in TA *pällānträ* (TB *pällāntär*) from the nasal present **pəlna-* of **pəla-* 'praise', TA *källāṣ* (TB *källāṣṣä-ṃ*) from the present **kəlna-* of **kəla-* 'bring'. Only TA *wläṣtär* 'die' (attested once, no cognate in TB) has a singleton *-l-* from presumed PT **-ll-*, next to a geminate in the more frequently attested *wälläṣtär* (restored via **-ln-* according to Ringe 1996: 165).

However, Peyrot (2013: 755 fn. 275, 281) points out that the geminate in TB *tallā$_u$* can be interpreted as an analogical introduction from the *na*-present instead, since the *a*-grade of *tallā$_u$* does not match the *ə*-grade of the present stem *tällā-* /təllá-/ itself, so that the adjective does not appear to be derived from the *na*-present directly. Thus, this geminate is possibly specific to Tocharian B. TA prs.mid.3sg. *wläṣtär* 'die' (A 102 a5) and the present-derived infinitive TA *wlässi* (A 65 a2) thus carry the full weight for dating the sequence of the sound change of **-ln-* to **-ll-* to Proto-Tocharian.[52] General degemination in Tocharian A clearly affected the **-ll-* of PT **allekə* 'other', which became TA *ālak*, TB *allek*, but since the geminate here derives from **-l̥-* instead of **-ln-*, it does not directly

[52] Non-simplified prs.mid.3sg. *wälläṣtär* is found in A 64 b6, A 65 a3 and a4, and *wälläṣträ* in A 117 a5 and in A 125 and THT 1425.e and A 117 a5. If analysed in isolation, *wläṣtär* and *wlässi* look like *s*-presents [class 8]; cf. Peyrot (2013: 117).

give us any insights into the Tocharian A development of *-*ll*- from *-*ln*-, which could also be posited after Proto-Tocharian.

The first change in the relative chronology, loss of *-*p*- in *-*lpn*-, may be independent, since TA *kälpnā*- shows no assimilation. On the other hand, a secondary restoration of the root -*p*- is also possible. The contrast between deletion of the stop in TA *kārna*- of the root *kārpā*- 'descend' vs. retention in TA *wärpnā*- from *wärpā*- 'enjoy' points to this type of analogical restoration in the latter.

Ringe takes the deletion of pre-nasal stops as a general sound law in (late) Proto-Tocharian, but notes that TA *āñcäm*, TB *āñme* 'soul; self' may suggest a (partially) independent development. On account of the sequence of sound changes whereby (2) occurs between (1) and (3), not all stop deletions in (complex) clusters with resonants can have taken place at the same time in any case.

The characteristic Tocharian A change from *-*st*- to -*ṣt*- has occurred in the TA cognate *kärṣnā*- < **kärṣtnā* to TB *kərsná*- < **kərstna*- 'cut off', so that either the deletion of the *-*t*- in *-*Ctn*- occurred independently in Tocharian A and B, or *-*t*- was briefly reintroduced in pre-TA. In theory -*ṣ*- may also have been introduced on the basis of forms like the prt.ptc. *kärṣto* (Ringe 1996: 164). It is thus easiest to assume independent loss of *-*t*- in the cluster *-*stn*- in Tocharian A and B, but analogical developments may have reshaped the evidence. This means that it is also difficult to tell if the subsequent changes of epenthesis and liquid metathesis are independent, since they involve the schwa, on which Tocharian A is not particularly informative.

Kim, in any case, does not consider the epenthesis found in the nasal presents to be of Proto-Tocharian date. The main argument that Kim (2007a: 85–86) advances is that those nasal presents with a heavy vowel in the root show Tocharian A vowel weakening of the *-*ā* in the suffix; viz., TA *kotna*- to *kota*- 'split' rather than **kotnā* (TB *kauta*-), TA *kropna*- to *kropa*- 'gather' rather than **kropnā*- (TB *kraupa*-), TA *skena*- to *ske*- 'strive' rather than **skenā*- (TB *skaya*-), and TA *kārna*- 'to *kārpā*- 'descend' rather than **kārnā*- (TB *karpa*-). However, this reasoning is based on the idea that heavy vowels induced epenthesis even in clusters of the type *-*Cn*-, which I argue is unwarranted. Since there seem to be no examples attested of Tocharian A nasal presents with both a heavy vowel in the root and an original root-final cluster that was definitely not simplified, the vowel weakening found in the nasal suffix cannot be adduced as a secure reason to date epenthesis after Proto-Tocharian. In my opinion, no regular vowel epenthesis is expected in *kotna*- to *kota*- 'split', *kropna*- to *kropa*- 'gather', and *skena*- to *ske*- 'strive' anyway, and the epenthetic vowel attested in the Tocharian B counterparts of the former two is, I argue, analogical (cf. Table 5.5 and accompanying text above).

To briefly wrap up our discussion of the *na*-presents, their development can be understood along similar lines as laid out by Kim (2007a), who considered both root-final clusters and root-internal heavy vowels as causes for epenthesis. It turns out only

the former rule is necessary. Marggraf (1970: 18), giving only two representative examples (viz., |təllá-| and |tərkə́na-|), had already arrived at this conditioning, even though he did not provide an explanation for the exceptions. I argue that the exceptions can best be understood if we consider that a number of forms simply do not represent the regular outcome by sound change, were analogically renewed with good reason, and are thus of little use for the determination of the correct sound laws. The remaining, reliable forms follow a clear pattern that also informs us of the relative chronology of a number of other sound changes. Unfortunately, a dating of these sound changes relative to Proto-Tocharian remains uncertain because the Tocharian A material does not provide us with the much needed, reliable examples.

#### 5.2.2.4 Generalization of the epenthesis rules

An epenthesis rule with full vowels causing epenthesis, as suggested by Kim (2007a) for the *na*-presents, not only failed for certain examples within that category, like the present *tsakna-* of *tsana-* 'bite' but also for words in other categories like *prākre* 'firm'. It was thus not generalisable. On the basis of the distribution of the allomorphs *-ná-* and *-ə́na-* in the nasal presents, I concluded that epenthesis visible in Tocharian B occurred regularly in the sequence *-CCn-*. On account of examples like *astare* 'pure' and *pärkare* 'tall', the rule of epenthesis in *-CCn-* established on the basis of the nasal presents can at least be extended to clusters of the type *-CCr-*. The lack of epenthesis in words like *prākre* 'firm' is thus in line with the epenthetic developments in the *na*-presents. I am not aware of any counterexamples. I think that this generalisability constitutes a further advantage over Kim's analysis, which operates in the isolation of a single morphological category.

It should also be noted that epenthesis is only observable between the first and (pre-epenthesis) second syllable, and is never found after an originally second syllable; we find, e.g., *emparkre* /empə́rkre/ 'widely' derived from *pärkare* /pərkə́re/ 'wide' rather than ***empärkare* /empərkə́re/, and *eṅkatkre* /enkə́tkre/ '± deeply' from *kätkare* /kətkə́re/ 'deep' rather than ***eṅkätkare* /enkətkə́re/. This will be briefly addressed again below in 5.4.2.

Epenthesis cannot be generalized to all resonant clusters *CCR*. The clusters *-CCl-* and *-CCm* are too uncommon, while in clusters of the type *CCw* there are examples with heavy clusters like *entwe*[53] 'then, thereupon' and *kantwo*[54] 'tongue', where no epenthesis is found. A comparison of *wäntare* vs. *entwe* suggests that *w* at least did not behave quite like *r*.

[53] Attested in prose in IOL Toch 4 b2; IOL Toch 131 b3; PK NS 30b2.

[54] Attested in prose in THT 88 a1, and as the per.sg. *kantwasa* in prose in THT 408 a6; a syncopated form of ***käntawo* would furthermore have yielded ***käntwo*.

Pronk (2009: 101) extends the rule for epenthesis to the initial position, in order to explain forms like *pūwarsa* 'fire per.sg.', and also *tuwe* 'you' (~ *twe*) and *ltuweṣ*, where a cluster with *-w* would be broken up. I am not sure that this works. For instance, there is no epenthesis in *śtwer* (**śtuwer*), so that epenthesis of the type under discussion in *ltuweṣ* is unlikely (earlier **śətwer* and **lətweṣ* would be parallel in structure). It is also unclear why specifically the *tw-* in *tuwe* would need to be broken up, while words like *twere* 'door' and *tweye* 'dust' seem to have been perfectly acceptable, as well as *entwe*, and *kantwo*. Pronk (2009: 88 fn. 28) further remarks that it is difficult to determine whether PIE **RH*, as in **puHr* 'fire' could have become pre-TB **əRa* directly without an intermediate stage **Ra*.

Based on *puwar* /páwar/, it seems indeed possible to posit a direct development PIE **puH̥r* > **pu.ar* = *puwar*. On the other hand, for *swāre* 'sweet' obl. *swareṃ* < **su̯H̥d-ro-*, we cannot easily assume earlier ***suware* on account of the accent, which otherwise unexpectedly lands on the third syllable in the oblique *swareṃ* /swarén/ from **swarénə* rather than **səwaránə* (exp. ***səwárenə* with regular accentuation). This would then have to be analogical, from an otherwise perhaps unparalleled stable accent on the apparent first syllable of a (newly) disyllabic *re*-adjective: *swáre* : ***swáren* vs. *tápre* : *təprén* next to obviously trisyllabic *pərkáre* : *pərkáren*. The development of PIE **luHs* 'animal; louse' to TB obl.sg. *luwa* (Del Tomba 2020: 131–133) may provide a parallel to *puwar*.

Remaining cases that apparently show epenthetic vowels not originating in a cluster of the type *-CCn/r-* are found in prs.ptc. *yokamane* of *yok-* 'drink', prs.ptc. *nesamane* of *nes-* 'be', prs.ptc. *pkwamane* and prs.mid.1sg. *pkwamar* of *pəkʷ-* 'rely, trust', *kokale* 'chariot' and *ṣpane* 'sleep'. The first four should all go back to a cluster of the type *-Km-*, where *K* is an obstruent. Perhaps such clusters were subjected to a special epenthesis as well. Words like TB *solme* 'complete(ly)' (not ***solame*) show that no general epenthesis occurred in any cluster *Cm* if the first consonant was a resonant. It is also possible that |yokə-|, |pəkʷə-| and |nesə-| were extracted as the basis for the respective forms from other places of the paradigm, such as prs.3sg. *yokäṃ* /yókən/ analysable as |yokə-n|, prs.3pl. *pkwantär* /pkʷántər/ analysable as |pəkʷə-ntər|, and prs.3sg. *nesäṃ* /nésən/ analysable as |nesə-n|.

TB *kokale* 'chariot' probably goes back to PT **kʷəkʷle* (cf. TA *kukäl*), where the rather unique cluster *-kʷl-* may have been broken up[55], perhaps because it functioned as "*kwl* = *CCR*" for the purposes of epenthesis.[56] No *-kw-* is visible anymore in Tocharian B *kokale*,

[55] TB *pikul* /péykʷəl/ 'year', pl. *pikwala* /pəykʷála/ would adhere to this as well. The syncopated verse form *pik$_u$la* does not have any bearing on epenthesis rules.

[56] Kim (1999) discusses a supposed de-labialization in a similar cluster in *lakle* 'sorrow' from hypothetical **ləkʷle*, even earlier **lukle*. However, the **kʷ* in this case would be of a different

but it must still have been there in pre-Tocharian B in order to change the **ə* of **kʷəkʷle* to a TB *o* (cf. 2.5.6.3 on this development).

The expected outcome of PIE **sy̯epnos* would have been TB ***ṣanme* /ṣə́nme/, with assimilation and metathesis, so that attested TB *ṣpane* 'sleep' needs to have been modified or recreated in some way. The related adjective f.nom.sg. *sänmetsa* /sənmétˢa/ 'calm, tranquil' does show this expected development. Tocharian A *ṣpäṃ* 'sleep' aligns with TB *ṣpane* /ṣpə́ne/, so that the change needs to have happened in Proto-Tocharian already. One could perhaps think of an irregular development like pre-PT **sʸwʸəpne* > **ṣpəne* due to some kind of misattribution of the position of the labial stop element to the labial **wʸ* in the initial cluster, of which no trace remains, but such an irregular change would necessarily remain speculative. A morphological explanation departing from an old genitive **suepns* is offered by Pronk (2009: 107–109), although this cannot be confirmed either. Ringe (1987) explained the epenthesis in TB *ṣpane* as a reflex of the Proto-Indo-European accent on the first syllable in PIE **sy̯epnos*, but since the Proto-Indo-European accent cannot be shown to have an effect in any other development, this solution is highly questionable. TB *ṣpane* alone cannot prove a continued Proto-Indo-European accentuation.

### 5.2.3 Epenthesis in *s*-preterites

A few other clusters were broken up by epenthetic vowels as well. In the *s*-preterite [class 3], sigmatic forms like TB prt.act.3sg. *otkasa* /otkə́sa/ to *wotk-* 'separate', prt.act.3sg. *yonmasa* /yonmə́sa/ to *yənm-* 'get', prt.act.3sg. *mlautkasa* /mlawtkə́sa/ to *mləwtk-* 'take off (tr.)' and prt.mid.1sg. *sätkasamai* /sətkə́samay/ to *sətk-* 'spread (tr.)' (Winter 1993: 200–201; Malzahn 2010: 192) have an accented schwa between the root and the preterite suffix *-sa*. To explain this situation, Winter posited an earlier suffix form **-ə́sa-* for this sigmatic ending, which would have been preserved only after the four roots mentioned due to their complicated consonant clusters (Winter 1993: 200). However, as Peyrot points out, this reconstruction is hardly supported by the evidence: epenthesis in only those forms of the type *otkasa* accounts for the same distribution, and since epenthesis is needed to get a schwa before the **-s-* from PIE **-s-* anyway, it is more economical to assume that there never was a schwa in those cases where it is not actually attested (Peyrot 2013: 509).

The supposed general presence of a schwa before the suffix **-sa-* is also used to explain the existence of Tocharian A forms like *yāmsāt* < **yaməsatə*, in which vowel weakening would have been expected to yield ***yāmsat* when departing from earlier

---

nature, from affection caused by the preceding **u*, which is not actually visible in this word, and may not have occurred at all. For example, *tkācer* rather than ***tkwācer* from **tukatʸer* provides strong evidence against a progressive labialization rule.

**yamsatə* (Malzahn 2010: 192). This argument is not compelling, since the vowel of the ending would have been preserved in other forms with the same suffix, and could then easily be generalized to all verbs after vowel weakening stopped being operational. The epenthesis before the sigmatic suffix could therefore also be considered specific to Tocharian B, although this cannot be ascertained. Finally, the rule for this epenthesis needs to be restricted to clusters of three consonants *-CCs-* where the *-CC-* could not function as a syllable coda. This is because similarly long clusters were retained in examples like TB prt.mid.3sg. *eṅksate*, prt.act.3sg.-pl. *ṣerpsa-me* from PT **enksate*, **sʸerpsa-me* (Winter 1993: 200). The accentuation of the *s*-preterite is treated further below in 5.3.3.

### 5.2.4 Epenthesis in *Cst* and *Csk*

One further environment for epenthetic vowels is actually agreed upon by both Tocharian A and B: an earlier cluster **Cst* is broken up to become TA *-Cäst-* and TB *-Cəst-*. This is reflected in, e.g., TA *ṣkäṣt*, TB *ṣkaste* from **ṣəkəste* < PIE **s(u̯)eksto* '6th' (Ringe 1990: 192–194; 1996: 71). However, whether the *ä* in Tocharian A is to be historically related to the *ə* of Tocharian B is, once again, not immediately clear, since it would have arisen automatically in the synchronic phonology of Tocharian A anyway.

I can think of one structural argument to date this particular epenthesis before syncope, so that it would have been a squarely pre-Proto-Tocharian sound law. In Tocharian B, the cluster *-kṣt-* is allowed, as it appears in, e.g., the prs.act.2sg. *prekṣt* to *prək-* 'ask', from earlier **preksʸətə*. Since *-kṣt-* and *-kst-* are phonologically so similar, assuming epenthesis in the latter, but not in the former, is, in my opinion, rather unattractive. I think it is therefore better if we assume that **-kst-* became **-kəst-* before syncope applied, so that **-kṣt-*, only derived from earlier **-kṣət-*, did not yet exist to become ***-kəṣt-* by epenthesis. Syncope of unstressed /ə/ in open syllables then only came into effect after **Cst* had become **Cəst*, creating **prekṣtə* from earlier **preksʸətə*. The epenthetic /ə/ in *-kəst-* was not deleted by this new syncope rule, because it stood in a closed syllable.

A similar distribution is found in other clusters of this type, e.g., prt.act.2sg. *orasta** (attested as arch. *orästa*; Malzahn 2010: 525–526) to the root *ar-* 'abandon' vs. prs.mid.2sg. *erṣtar* from the *s*-present *er-* 'evoke'. Epenthesis before the cluster **-sk-* appears to be very similar, seen in a number of *sk*-presents like prs.3pl. *yamaskeṃ* /yamə́sken/ from the root *yam-* 'do'. The epenthesis of /ə/ in the clusters *Cst* and *Csk* can be considered (pre-)Proto-Tocharian, and it likely took place before likewise (pre-)Proto-Tocharian syncope of unstressed /ə/ in open syllables. The following subsection is devoted to a more detailed account of this syncope.

### 5.2.5 Syncope

At the end of the previous subsection, syncope was already briefly considered, and this subsection will provide an elaboration on the topic. We can observe in both Tocharian A and B a tendency to delete **ə* in an open syllable. In Tocharian B this deletion applies only to unaccented **ə*, while in Tocharian A it applies to all **ə* in open syllables. In general, Tocharian A is uninformative with regards to the historical development of **ə* in early Tocharian, as the evidence would mostly have been secondarily lost or altered.

Tocharian B thus retains a more original situation: accented /ə́/ remained in classical Tocharian B prose even in open syllables, so that also in (pre-)Proto-Tocharian /ə́/ should have escaped deletion. As a result, the most straightforward assumption is that any etymological /ə/ that had disappeared in (non-metrical) Tocharian B was unaccented at an earlier stage. This mostly concerns the thematic suffix PIE **e* > PT **ʸə* in thematic present [2] forms like mid.3sg. *āśtär* from **akʸətər* (*ak-* 'lead'), mid.2sg. *aiśtar* from **aykʸətar* (*ayk-* 'know'), and act.2pl. *ṣamcer* from **sʸəmʸətʸer* (*ṣəm-* 'sit'). These forms can therefore be accented as pre-PT **ákʸətər*, **áykʸətar*, **sʸámʸətʸer* respectively at the time of syncope. This is Marggraf's original hypothesis (1970: 21). The palatalizing **ʸə* of the *s*-present [class 8] was another victim of this syncope rule, as we have already seen in the previous section: forms like prs.act.2sg. *tˢəmsʸətə* (*tˢəm-* 'increase') and **preksʸətə* (*prək-* 'ask') became PT **tˢəmsʸtə* and **preksʸtə* to eventually yield TB *tsamṣt*, *prekṣt*. These forms thus reflect pre-PT **tˢámsʸətə*, **préksʸətə*. Winter proposed that an originally accented /ə́/ would have lost the accent and be syncopated before dentals[57], as in **akʸátər* (Winter 1993), but based on the general retention of /ə́/ in open syllables in Tocharian B, this seems unlikely. Instead, the more straightforward assumption is that any schwa that disappeared was simply unaccented at the time of this syncope.[58]

It has been claimed that a schwa in the third syllable was lost especially early, based on the development of the sequence **-skʸət-*. For example, there is no trace of

[57] The prevalence of deletion before dental obstruents, set up by Winter as the general sound law, is only apparently the case, since *p* and *k* are not really well-represented in derivational or inflectional suffixes, and thus do not often appear in the right place in the word. However, TB *eṃṣke* 'up to, until; even; while' must surely have lost an original palatalizing schwa before **k* due to syncope, from earlier **ensʸəke*, since the **s* could not have been palatalized in any other way (cf. Adams 2013: 78).

[58] Pronk adds to Winter's rule for syncope of /ə́/ that it did not apply in the first syllable, on account of forms like *yasa* 'gold', *yasar* 'blood', *wate* 'second', etc. (Pronk 2009: 75). On the whole it is not entirely clear to me whether Winter would not reconstruct *yəsár* at the time depth that he reconstructs /ə́/-deletion before dentals, so that instead the modification could be that pretonic /ə/ was somehow safe from syncope regardless of its accentuation. This would seem to contradict Winter's analysis of the copula, however, since *skʸəté* is required to develop to *s*(*k*)*te* in his model (cf. later in this subsection).

palatalization in the prs.act.2sg *yamast* or prs.mid.2sg. TB *yaməstar* 'do' from **yaməskʸətə* and **yaməskʸətar* respectively. The lack of palatalization of the *-s-* is understandable if the ending had become **-skʸt-* > **-st-* already before regressive assimilation occurred to change **-skʸ-* to **-sʸkʸ-* (eventual TB *-ṣṣ-*) (Couvreur 1947: 63; Peyrot 2013: 510). However, this assimilation does appear to have happened in corresponding forms like the prs.mid.2sg TB *pāṣtar* from the verb *pask-* 'guard'. One may therefore wonder if **paskʸətar* remained as such at the time that **yaməskʸətar* was reduced to **yaməskʸtar* > **yaməstar*. Such a supposed lack of syncope in **paskʸətar* would perhaps be readily understandable by sound law if the accentuation was **paskʸátar*, next to **yamáskʸətar*, with regular second-syllable accent. Essentially that is Winter's explanation for the discrepancy, although he formulates it in terms of a synchronic underlying representation: *|skʸát| would yield |ṣt| and *|skʸət| would yield |st|, according to Winter (1993: 201–203).

Krause (1952: 76), on the other hand, concluded that *paskʸətar* did undergo syncope to **paskʸtar* with subsequent development to **pastar*, but that the alternation of root-final *-sk* : *-s* was restored to more regular *-sk* : *-ṣ* at a later stage. After all, *|pas-| could not have been interpreted as the expected palatalized stem variant of *|pask-|. Integrity of the alternation between non-palatal *-sk* and palatal *-ṣ* would understandably have been more important when it was root-final than when it was in the suffixal *-sk-*. Compare the inflection of *ayk-* 'know' in the same present class [2] as *pask-*, with |ayś(ə)-| next to |ayke-| showing a clear alternation between palatal and non-palatal final consonant. The very realistic possibility of analogical change means that we do not need to assume either a difference in accentuation between forms like **paskʸátar* next to **ákʸətar*, nor an earlier syncope of /ə/ in the latter but not in the former.

Krause's approach is surely correct, in spite of Winter's objections, and can be supported with evidence from the *sk*-present of monosyllabic roots like *ay-* 'give', *weñ-* 'say'. The prs.act.2sg. *aist* 'you give' clearly shows the same lack of palatalization as seen in the type *yamast*, but it must derive from trisyllabic **ayskʸətə* rather than from quadrisyllabic ***ayəskʸətə*. Malzahn (2010: 448) does assume ***ayəskʸətə* as a valid reconstruction to uphold Winter's analysis, but this cannot work within the same framework due to the accent movement exhibited by prs.act.2sg.-3sg. *aista-ne* /aistə́ne/. We can thus conclude that a schwa in the non-first syllable was lost before the development from **-skʸ-* to PT **-sʸsʸ-*, no matter if it was in the second or the third syllable. Syncope of schwa in general thus applied to non-first syllables.

### 5.2.6 The accent through the lens of syncope and epenthesis

The accent was already featured in the previous subsections most prominently in relation to syncope. I claimed that, because a schwa in the second or third syllable was lost at some point, those syllables must have been unaccented at the time. This is based

on the idea that the stressed syllable in a word, i.e., the most prominent syllable, will be the least liable to syncope. Only if the stress moves away from such a syllable can it be deleted, as probably happened in the development of PT **yásar* to pre-TA **yäśár* and TA *ysār* 'blood' or more clearly in the plural TB *ysāra* /ysára/ from **yəsara*. Thus, for syncope to work as attested, the accent should have been on the first syllable of the word at the time, irrespective of the number of following syllables, or on the last, because it was either the second syllable in trisyllabic words or the third syllable in quadrisyllabic words that was lost. Final accent, however, is generally disallowed in Tocharian B, so that it is not an immediately obvious option for our reconstruction. The retention of the first-syllable vowel in *yasar* 'blood' and words like it would then furthermore be surprising; i.e., if we should assume a starting point **yəsár*, why not syncope to PT **ysár* and TB ***ysār*? A reconstruction with initial accent is thus to be preferred.

General initial accent is contrary to the accent system as it is attested in Tocharian B, however, so that the question is how and at what point the system changed. That the accentual system changed in some way after syncope seems likely, however, as already pointed out by Marggraf (1970: 22) based on the paradigm of the thematic present. An absolute dating of the syncope is not possible, but we can be sure that it happened after palatalization, and most likely before epenthesis in clusters of the type *Cst* (see 5.2.4). We also see from examples like TB *ṣkaste* /ṣkə́ste/ from **s(u̯)eksto* 'sixth ($6^{th}$)' that the new, epenthetic vowel actually ended up bearing the accent in Tocharian B. Since it could not have borne the accent before it existed, an accent shift is implied.

But what about the other types of epenthesis? As it turns out, also the epenthetic vowels in the clusters *-CCr*/*n-* and complex *-CCs-* were accented in Tocharian B; e.g., TB *astare* /astə́re/ 'pure', *otkasa* /otkə́sa/ 'separate, decide (prt.3sg.)'. This situation can most straightforwardly be accounted for by assuming that an accent shift occurred in Tocharian B from the first to the second syllable, unless the second syllable was also the final syllable. Such an accent shift provides us with the general pattern as it is attested in Tocharian B, outside of the exceptional classes to be further discussed below. If we could establish whether epenthesis in *-CCR-* was shared between Tocharian A and B, we would also know whether the accent shift was originally a shared Proto-Tocharian development. Unfortunately, as discussed in 5.5, Tocharian A does not provide the necessary answers for a definite conclusion.

The only word I am aware of that shows divergent developments in Tocharian A and B connected with epenthesis or lack thereof is TB *āñme*, TA *āñcäm* 'soul; self'. The Tocharian B form *āñme* /áñme/ cannot regularly have developed from PT ***añcəme* with epenthesis in the cluster **-ñcm-* and accentuation as TB ***añcame* /añcə́me/. Tocharian B lost the **-c-* between the two nasals, which would only have been possible if there was no intervening vowel, and Tocharian A retains the cluster, apparently due to resolution of the word-final cluster after apocope. The development of PT **añcme* to TB *āñme* is the

only example that I am aware of that implies that we should date the epenthesis in *-CCR-* after Proto-Tocharian, and as specific to Tocharian B. However, since there are no parallels to this particular type of cluster, it is unknown if it would have behaved like other clusters of the types **CCr* and **CCn*.

The placement of Proto-Tocharian accent therefore remains rather ambiguous on the basis of Tocharian evidence alone: either it was still on the first syllable, as implied for an early stage by the syncope of second- and third-syllable schwa discussed above, or it had already shifted to the second syllable as in Tocharian B. At least a relative chronology can be established as follows:

1) epenthesis in clusters *Cst*;
2) syncope of unaccented /ə/ in open syllables (second and third syllable);
3) epenthesis of /ə/ in several complicated consonant clusters of the type *CCR* and *CCs* where the *CC* could not form a coda;
4) shift of the accent from the first to the second syllable in trisyllabic (or longer) words.

As far as I can see, there is at least one reason to assume accentuation of the type found in Tocharian B for the earliest stages of Tocharian A, and that is the accentuation of loanwords from Old Steppe Iranian (on which see 5.4.6). Another, more indirect reason is the number of analogical changes relating to the accent that took place in Tocharian B, discussed in 5.4.

### 5.2.7 Intermediary summary

So far, we have discussed a number of changes involving epenthesis and syncope that happened in Tocharian at different points in time, pulling together information from various publications on the matter. At least the epenthesis of a schwa in clusters of the type **Cst* can be considered pre-Proto-Tocharian, so that, e.g., PIE **s*(*u̯*)*éksto* '6$^{th}$' became first PT **sʸákəste* (cf. Ringe 1996). Afterwards, there was a sound law by which original schwa could be deleted in an open second or third syllable. Based on the deletion of a vowel in these positions, they should be reconstructed unstressed at the time: e.g., **ákʸətər* > PT **ákʸtər* (Marggraf 1970: 22). Perhaps during, or perhaps after the Proto-Tocharian period, at least in pre-TB, further epenthesis occurred in clusters of the type *CCR* (Kim 2007a, Pronk 2009). I have argued that the epenthesis in *CR*-clusters that has previously also been assumed after a heavy vowel in the preceding syllable (ibid.) is not sufficiently supported by the data. Nasal presents that appear to show such a change are, in my view, likely analogically remade. After the different types of epenthesis had occurred, the accent moved from the first to the second syllable, which yielded the general Tocharian B accent pattern. In the following two sections 5.3 and 5.4, the

Tocharian B categories that show aberrant or otherwise interesting and potentially informative accentuation will be treated in more detail.

## 5.3 The accent in finite verbs

In this subsection I will go over the different verbal categories that any coherent account of the Tocharian B accent needs to cover, and evaluate the explanations given for their aberrant accentuation based on how well they correspond to the sound changes for epenthesis and syncope as discussed in the previous subsection. First, 5.3.1 provides an introduction to the accent in the verb to frame the problematic categories to be discussed in more detail in the following subsections.

### 5.3.1 Mobile and fixed accent in the Tocharian B verb

In accordance with the general accentuation rules of Tocharian B, many verbs display accent movement when additional syllables are added at the end through suffixation. This can happen within a single paradigm, between the active and the middle paradigms, or when pronoun suffixes are added. Table 5.7 illustrates accent movement with the present active of *ak-* 'lead' and *ay-* 'give', and with the preterite active of *taka-* 'be'.

Table 5.7: Three verbal paradigms with accent movement in the surface forms.

| | *ak-* 'lead' prs.[2] act. | | *ay-* 'give' prs.[9] | | *taka-* 'be' prt.[1] | |
|---|---|---|---|---|---|---|
| | base | suffixed | base | suffixed | base | suffixed |
| 1sg. | *ākau** | *akau-ne** | *aiskau* | *aiskau-ne** | *takāwa* | *takāwa-ne** |
| 2sg | *āst* | *aśta-ne** | *aist** | *aista-ne* | *takāsta* | *takāsta-ne** |
| 3sg. | *āśäṃ* | *aśan-ne** | *aiṣṣäṃ* | *aiṣṣan-ne** | *tāka* | *takā-ne* |
| 1pl. | *akeṃ** | *akem-ne* | *aiskem* | *aiskem-ne** | *takām** | *takām-ne** |
| 2pl. | *āścer* | ? | *aiścer* | ? | *takās* | *takās-ne** |
| 3pl. | *ākeṃ* | *aken-ne* | *aiskeṃ* | *aisken-ne** | *takāre* | *takāre-ne** |

The number of forms that show accent movement is different depending on the category. In the present active, it can be seen everywhere except in the first-person plural, which always has its accent on the second syllable. It does not look like a trisyllabic form, but the ending used to be **-mə* with a final schwa that was eventually lost, so that *akem** /akém/ derives from trisyllabic **akémə* and *aiskem* /ayském/ from **ayskémə*. In the preterite [1] there is only a single form that shows accent movement, namely the third person singular. In all other cases the accent is already on the second syllable because the endings added an additional syllable. The 1pl. *takām** /takám/ and 2pl. *takās* /takás/ derive from earlier trisyllabic **takámə* and **takásə*, which lost their final schwa.

There are a few verbal classes that have an unexpectedly stable first-syllable accent in (part of) their paradigm. Thematic presents show accentual movement to a degree, but not in all forms. In Table 5.8 they are represented by middle of the simple thematic present [class 2] *ák-*. Many subjunctives of classes [1] and [5] have a stable initial accent as well, as shown with the other paradigms in Table 5.8.

Table 5.8: The accentuation of the thematic present middle and initially-accented subjunctives [1] and [5].

| | *ak-* 'lead' prs.[2] mid. | | *prek-* 'ask' sbj.[1] | | *taka-* 'be' sbj.[5] | |
|---|---|---|---|---|---|---|
| | base | suffixed | base | suffixed | base | suffixed |
| 1sg. | *akemar** | *akemar-ne** | *preku* | *préku-ne** | *tākau* | *tākau-ne** |
| 2sg. | *āśtar** | *āśtar-ne** | *prekt** | ? | *tākat* | *tākat-ne** |
| 3sg. | *āśtär* | *āśtär-ne** | *prekäṃ** | *prekäṃ-ne* | *tākaṃ* | *tākaṃ-ne** |
| 1pl. | *akemtär** | *akemtär-ne** | *parkäm** | *parkäm-ne** | *tākam* | *tākam-ne** |
| 2pl. | ? *āśtär** | ? | *parkcer* | ? | *tākacer* | *tākacer-ne* |
| 3pl. | *akentär** | *akentär-ne* | *parkäṃ* | *parkän-ne** | *tākaṃ* | *tākaṃ-ne* |

The present [2] middle shows accent movement within the paradigm (*akemar* /akémar/ vs. *āśtär* /áśtər/), but not when pronoun suffixes are added (*āśtär* /áśtər/ vs. *āśtär-ne** /áśtərne/). The same is seen in the present [8] middle, and this will be treated in 5.3.2. Many subjunctives [1] and [5] have a fixed initial accent, which does not move at all. Even (originally) trisyllabic forms have the accent on the first syllable, so that we have 2sg. *tākat* /tákat/ rather than ***takát* from **takatə*, 1pl. *tākam* /tákam/ rather than ***takám* from **takamə*, and 2pl. *tākacer* /tákacer/ rather than ***takácer*. See 5.3.4 for more details.

The preterites of class [3] show the normal accentuation in most forms of the active, except for the third person singular, which has a consistent initial accent. This initial accent is also found in the entire middle, and illustrated in Table 5.9. The active third person singular and the middle of the preterite [3] all share a suffix *-sa-*, which is absent in the regularly accented forms. The accentuation of the *s*-preterite [3] is discussed in 5.3.3. Finally, causatives with a present-subjunctive of class [9] and a preterite of class [2] are accented on the first syllable as well, even though they have many forms of three syllables where the accent would normally be on the second syllable; see Table 5.9, and further in 5.3.5.

Table 5.9: The accentuation of the preterite [3] and the causative present-subjunctive [9] and preterite [2]. The non-3sg. forms of the preterite [3] do not show special accentuation.

| | *prek-* 'ask' prt.[3] | | *wətk-* 'command' (causative) | |
|---|---|---|---|---|
| | active (3sg. + suffix) | middle | prs-sbj.[9] | prt.[2] (3sg. + suffix) |
| 1sg. | *prekuwa* | *parksamai** | *watkäskau* | *yātkawa* |
| | /prekə́wa/ | /pə́rksamay/ | /wətkəskaw/ | /yátkawa/ |
| 2sg | *prekasta* | *parksatai** | *watkäst* | *yātkasta* |
| | /prekə́sta/ | /pə́rksatay/ | /wətkəst/ | /yátkasta/ |
| 3sg. | *preksa, preksa-ne* | *parksate* | *watkäṣṣäṃ* | *yātka, yātka-ne** |
| | /préksa, préksane/ | /pə́rksate/ | /wətkəṣṣən/ | /yátka, yátkane/ |
| 1pl. | *prekam** | *parksamte** | *watkäskem** | *yātkam* |
| 2pl. | *prekas* | *parksat** | *watkäścer-* | *yātkas* |
| 3pl. | *prekar* | *parksante* | *watkäskeṃ** | *yātkare* |

The discussion around the accentuation of the verbs is centred on the types shown in these last two tables above. In the order that they will be treated in the following subsections, the forms that deviate from the usual Tocharian B accent rules are:

- The middle of the simple thematic present [class 2] and the *s*-present [class 8]; see 5.3.2.
- The preterite [3] forms with the suffix *-sa-*; see 5.3.3.
- The subjunctives [1] and [5] with fixed initial accent; see 5.3.4.
- The causative present-subjunctive [9] and preterite [2]; see 5.3.5.

Scholars have approached the accentuation of the verbal categories illustrated above in various ways. The remainder of this subsection provides a short overview of the different types of explanations that have been adduced to account for the fixed initial accent, which will play a role in the discussions of the next three subsections on the verb.

It has been proposed that certain aspects of Tocharian B accentuation directly continue Proto-Indo-European accentuation. This features prominently in Ringe's explanation of epenthesis and the accent in nouns (Ringe 1987), and it is also integral to Eyþórsson's explanation for the initial accent found in subjunctive [5] (Eyþórsson 1993). In both cases the lack of other, corroborating reflexes of the Proto-Indo-European accent makes its applicability questionable (cf. Malzahn 2010: 301; Jasanoff 2015: 87, 89 incl. fn. 7).

Winter (1993) instead departs from a general underlying second-syllable accentuation and proposes an analysis for the initially-accented present [2], present [8] and preterite [3] forms as originally, and still underlyingly, containing an accented /ə́/ in the second syllable that was lost, leaving the accent stranded on the first syllable. This

explanation is followed by Malzahn (2010: 191–195), but it is rejected by Peyrot (2013: 509), as it is "phonologically implausible, synchronically as well as diachronically" and furthermore incompatible with Marggraf's diachronic interpretations of the accent (Marggraf 1970: 21).

Another way to get the accent on the initial syllable is to interpret it as originally being a non-initial syllable. This is possible in originally reduplicated categories where the reduplication syllable was lost. For example, a form like prs.-sbj.act.1sg. *watkäskau* /wətkəskau/ 'I command' could be derived from **wəwətkəskau*, with regular second-syllable accentuation, by a syncope of the *ə* of the reduplication syllable and initial degemination. Schematically, earlier **CəCV́CVCV* yields **CV́CVCV* with synchronically aberrant initial accent. Winter applies this de-reduplication to subjunctives [1] and [5] and to the causative present-subjunctive [9] and preterite [2] (Winter 1994b). Malzahn basically follows this explanation, but with analogical dereduplication in those subjunctives that do not have initial acccent (Malzahn 2010: 310–316). Peyrot rejects the possibility of earlier reduplication in the subjunctives [1] and [5] (Peyrot 2013: 401ff.), but accepts it as a reason for the initial accent in the causative present-subjunctive [9] and especially the accompanying preterite [2] (Peyrot 2013: 487–491). Jasanoff does not consider dereduplication to be the cause of the initial accent in any category (Jasanoff 2015).

Jasanoff (2015) has instead developed a theory whereby an accent shift took place from the first syllable to the second syllable, but only if the second syllable did not contain a schwa following a full vowel in the first syllable. He calls this the "*yāmu*-rule", based on a reconstruction of the preterite participle *yāmu* of *yam-* 'do' as **yáməwə*. In words with this type of structure, Jasanoff claims, the progressive accent shift that led to the second-syllable accent of Tocharian B was inhibited because the second-syllable *ə* could not take the accent from preceding full vowel.[59] Jasanoff applies this "*yāmu*-rule" in various ways to explain the accentuation of the subjunctives [1] and [5], and of the causative present-subjunctive [9] and preterite [2]. Jasanoff's "*yāmu*-rule" is incompatible with existence of accented epenthetic vowels discussed in 5.2, and he can only account for them by assigning a special status to the epenthetic schwa (Jasanoff 2015: 95), which is otherwise unmotivated.

As will become clear, Jasanoff's "*yāmu*-rule" of limited accent shift in any case does not have any clear advantage over a general accent shift, including shift from a full vowel

[59] Malzahn (2010: 6) had pointed out set of words that appear to be accentuated according to this principle, which Jasanoff refers to; Malzahn called this phenomenon the "*pātär*-rule" instead, based on the interpretation of the example obl.sg. *pātär* as **pátərə*, with accent on the first syllable of originally three. Notably, however, the derived adjective *patarye* /patárye/ 'paternal' is not stressed on the initial syllable, calling into question the general applicability of the "*pātär*-rule" even in forms of the word *pācar*.

to a schwa, and it fails to account for a number of forms that do clearly show such a change. Precisely a general accent shift from the first to the second (but non-final) syllable is key to Peyrot's (2013: 510–515) explanation of the accentuation of the subjunctives [1] and [5]. Peyrot's account is more explicitly reliant on secondary regularization through analogy than, for example, Winter's scenarios with deleted /ə/ and reduplication in the subjunctive, but it has distinct advantages in being neither circular nor contradicted by forms that fundamentally should not be able to exist within its phonological framework. It is in accordance with Marggraf's diachronic interpretation of the accent (Marggraf 1970: 21) and also conforms to the epenthesis and syncope developments discussed in the previous section (5.2). In the following subsections, I will illustrate in more detail how the aberrant verbal categories can be understood with the general accent shift required by the syncope of second-syllable vowels and the accent movement to epenthetic vowels.

### 5.3.2 The accent in the thematic present [2] and *s*-present [8]

In disyllabic middle forms of the thematic present [class 2], the *s*-present [class 8], the accent never moves to the second syllable, even when pronoun suffixes are added. Examples include TB prs.2sg.mid.-1sg. *taṣtar-ñ* /tə́ṣtarñ/ of *təs-* 'put' [class 2], TB prs.3sg.mid.-pl. *śauśtär-me* /śáwśtərme/ of *śewk-* 'call, invite' [class 2], and prs.3sg.mid.-2sg. *wākṣtär-ś* /wákṣtərś/[60] of *wak-* 'differ' [class 8]. These forms thus seem to have a fixed initial accent.

Winter (1993) explained these forms by positing an underlying /ə/ which was syncopated before dentals, with an accent shift to the left; e.g., **tasʸə́tar-nʸə* > **tásʸtar-nʸə* = *taṣtar-ñ*. After this development, Winter claims that no further accent movement was possible, so that the forms of this type never came to adhere to the general Tocharian B accentuation on the second syllable. The same change is applied by Winter to all forms of this type, also non-suffixed ones like prs.3sg.mid. *āśtär* of *ak-* 'lead' [class 2] and prs.3sg.mid. *nakṣtär* of *nək-* 'destroy; mid. be destroyed' [class 8]. I already touched upon this briefly in subsection 5.2.5 on syncope, and rather interpreted the disappearance of the **ə* in these types of words as evidence that they were unaccented when syncope took place, as per Marggraf (1970: 21).

A remaining issue is thus the fixation of the accent on the first syllable, which Winter attempts to explain, but Marggraf does not. With an accent shift after syncope, one might expect forms like ***täṣtār-ñ* /təṣtárñ/ and ***śauśtar-me* /śawśtárme/ instead. However, the middle endings never seem to be accented at all, in any stem type, so that the accent of middle verbs falls on either the root or the suffix, and not on the ending; cf. the types present [3] *tsmetär* /tˢmétər/, present [6] *kärsnātär* /kərsnátər/, present [9] *eṅkastär*

[60] *-ś* here is a variant of the 2sg. pronoun suffix *-c*.

/eṅkástər/, none of which could ever have the accent on the ending. The lack of movement in *taṣtar-ñ* and similar forms can thus be interpreted as an analogical morphophonological rule disallowing accent movement to middle endings.

The sole exception is in the irregular form *star-* 'be' used with pronoun suffixes. However, this consists of only a single syllable, and thus has an exceptional status. It derives quite regularly from, e.g., earlier **skʸətər-ne* with the 3sg. pronoun suffix. With accent movement from the first to the second syllable, this developed from **skʸátər-ne* to **skʸətár-ne*, whence **s(k)tár-ne* developed due to syncope of the unstressed open syllable. No analogical placement on the root syllable was possible here because the root syllable was lost entirely.

The ban on accented endings must be applied only to the middle endings, as there are examples of accented endings of the second person singular active. There, prs.act.2sg.-3sg. *aista-ne* and sbj.act.2sg.-3sg. *aita-ne* of *ay-* 'give', prs.act.2sg.-1sg. *nesta-ñ* of *nes-* 'be' and perhaps prs.act.2sg.-1sg. *prekṣta-ñ* of *prək-* 'ask' are found (Malzahn 2010: 30–31 incl. fn. 6). Malzahn ascribes these forms to an informal style (Malzahn 2010: 31), but there do not seem to be any examples without this accented second-person singular ending before a pronoun suffix outside of verse.[61] There is thus no particular reason to assume that they are secondary creations instead of normal prose forms; and they are fully regular with Marggraf's accent shift from the first to the second syllable.

On the other hand, Winter's hypothesis involving shift away from a */ə/ before an obstruent does not accord with the existence of *aista-ne* and *prekṣta-ñ*, since these forms should be derived from **ayskʸətə-ne* and **preksʸətə-nʸə* respectively. Since an application of regular syncope of earlier/underlying second-syllable schwa in combination with accent fixation on the first syllable cannot account an example like *aista-ne*, I think that it cannot be accepted as either a regular development or as a regular synchronic morphophonological analysis.

These forms like *aista-ne* furthermore cannot be understood in terms of Jasanoff's "*yāmu*-rule", whereby no shift should occur from a full vowel to a schwa, although it could theoretically account for *wākṣtär-ś* and *śauśtär-me* from **wákṣətər* and **śéwśətər*. Jasanoff's "*yāmu*-rule" additionally falls short as an explanation for forms like *taṣtar-ñ* /təṣtarñ/, with a full vowel in the ending. In a similar vein to *aista-ne* but a different category, the prt.act.3sg.-2sg. *yaśa-c* /yaśəc/ of *yok-* 'drink' also shows accent shift from *a* to *ə*, and can hardly be explained away, since ***yāś-c* next to unsuffixed *yāś** would have been phonologically entirely unproblematic.

[61] The examples adduced by Malzahn all occur in verse: IOL Toch 285 a3 *w(e)nt-meścä* 'you will say to us/them', THT 273 b3 *yust-me* 'you turn towards us(/them)', *west-meśca* 'you say to us(/them)'. These could thus be syncopated forms of **wenta-me-ścä*, **yusta-me*, **westa-me-śca*.

I thus conclude that Marggraf's general accent shift after syncope of unstressed /ə/ in combination with a morphophonological ban on accented middle endings provides the best explanation for the accentual behaviour of forms of the presents [2] and [8] like *taṣtar-ñ* and like *wākṣtär-ś*.

### 5.3.3 The accent in the *s*-preterite [3]

The preterite of class [3] is known as the sigmatic or *s*-preterite due to its characteristic suffix *-sa-*. This suffix is not present in all forms, however, only in the third person singular active and in the middle. The remaining active forms do not contain *-sa-* and are therefore "asigmatic"; see Table 5.10. In general, on the *s*-preterite, see Krause & Thomas (1960: 247–251) and Malzahn (2010: 190–214).

Table 5.10: An example paradigm of the preterite [3].

| | prt.[3] active | prt.[3] middle |
|---|---|---|
| 1sg. | *prekuwa* /prekə́wa/[(a)] | *parksamai** /pə́rksamay/ |
| 2sg | *prekasta** /prekə́sta/ | *parksatai** /pə́rksatay/ |
| 3sg. | *preksa* /préksa/, 3sg.-3sg. *preksa-ne* /préksane/ | *parksate** /pə́rksate/ |
| 1pl. | *prekam** /prekə́m/ | *parksamte** /pə́rksamte/ |
| 2pl. | *prekas** /prekə́s/[(b)] | *parksat** /pə́rksat/ |
| 3pl. | *prekar* /prekə́r/ | *parksante* /pə́rksante/ |

(a) Two types of 1sg. forms of the prt.[3] are attested, e.g., *nekwa* beside *prekuwa*, of which the type in *-uwa* is older (Schmidt 1985: 433). Malzahn reports that *-uwa* is found mostly in late, informal style documents, but also notes that this does not preclude it from being an archaism (Malzahn 2010: 40, 193). If derived from quasi-PIE **-u-$h_2e$* (Peyrot 2013: 418, 505), a disyllabic sequence **-əwa* is expected. With the exception of prs.act.1sg. *tsaukwa* from *$t^s$əwk-* 'suckle' in the late text THT 415 a3, all other attestations of prt.act.3sg. *-wa* that I could trace are found in verse, where the lack of the *-u-*could result from verse syncope (see 5.1.2): *nekwa* in THT 246 a5, *pelykwa* in THT 46 a2, *yātwa* in THT 365 b6, *yonwa* in THT 365 b5 and PK AS 6E b2, *lyautwa* in PK AS 13E a2 and a6. At least *prekuwa* in THT 1115 a5 and *pleṅkuwa* in HWB 74(4) a11, Kucha 191 a4 and PK DA M 507.39 and .43 a8 are found in prose.

(b) There is only one attestation of 2pl. forms of the prt.[3], namely *maitas* /maytə́s/ from *məyt-* 'set out (itr.)', since *lautso* /láwtso/ to *ləwt-* 'expel' is to be interpreted as an imperative instead (Peyrot 2013: 511 fn. 94). On the basis of *maitas*, *prekas** is the expected form.

While the asigmatic forms are generally accented according to the usual Tocharian B accent rules (Peyrot 2013: 100), the sigmatic forms of most verbs have a consistent first-syllable accent. Forms that need an explanation are those of the type *preksa-ne* /préksane/ rather than ***preksā-ne* /preksáne/, and the entire middle with *parksamai*

/pə́rksamay/ rather than *pärksāmai* /pərksámay/, etc. Winter (1993: 199–201) applied the same rule for the deletion of second syllable /ə́/ here as for the thematic presents discussed in the previous subsection, interpreting the sigmatic suffix as *-ə́sa-* rather than *-sa*. Thus, *preksa-ne* would derive from **prekə́sa-ne*, and *parksamai* from **prəkə́samai*. Winter is followed in this by Malzahn (2010: 191–195). Jasanoff departs from a similar **prékəsa-ne* with an unaccented schwa instead which would have blocked the movement of the accent according to his "*yāmu*-rule" (Jasanoff 2015: 95).

The theory of the disappearing /ə́/ already had difficulties accounting for the variety of forms in the thematic presents [class 2] and the *s*-presents [class 8] in the previous subsection, and in the sigmatic preterite there are problems as well. For example, there is a small group of verbs that do have their accent on the sigmatic suffix *-sa-*. All examples are middle forms.

- *tässāte* /təssáte/, prt.mid.3sg. of *təs-* 'put' (THT 109 a5, THT 626 a2)
- *tässānte* /təssánte/, prt.mid.3pl. of *təs-* 'put' (PK NS 39 a8, THT 349 b5)
- *wässāte* /wəssáte/, prt.mid.3sg. of *wəs-* 'wear' (THT 107 b4)
- *kälnsāte* /kəlnsáte/, prt.mid.3sg. of *kəln-* 'resound' (IOL Toch 19 b3)

Winter knew only of *tässāte* and *wässāte*, which he explained with a special, earlier deletion of /ə́/ in the sequence **-səs-*, this time with a rightward accent shift instead of the leftward shift in all the other verbs. This solution is *ad hoc*, and on account of their isolation within the system, these forms should rather be interpreted as archaisms (Peyrot 2013: 513). Furthermore, a special deletion of /ə́/ between two **s* cannot account for the suffixed active form of *təs-* 'put' *tessa-ne* /téssane/ (in PK DA M 507.37 and 36 a54–55). This form has initial accent, rather than ***tessā-ne* /tessáne/ as would be predicted by Winter's rules, and thus shows that the attested situation with these verbs is messier than can be captured by Winter's phonological rules. The forms *tässāte* and *wässāte* are regular within the framework of progressive accent shift, since they were never ***təsásate* and ***wəsásate*; *tessa-ne* is like the standard prt.act.3sg. type *preksa-ne* and should receive the same explanation.

Another, more basic issue with Winter's underlying /ə́/ in the sigmatic preterite is that this vowel is ahistorical (cf. also 5.2.3 above). Unlike the thematic suffixes, which contained a palatalizing **ə* in the relative forms that derived from a real PIE vowel **e*, the schwa of the theoretical sigmatic suffix *-əsa* has no clear origin: this suffix originally started with PIE **-s-*. It would therefore have to be inserted at some point by epenthesis. In a few verbs with particularly complex consonant clusters, there is indeed a /ə́/ before the sigmatic suffix, which are interpreted as proof that the forms without /ə́/ are the result of syncope (Winter 1993: 201):

- *mlautkasa* /mlawtkə́sa/, prt.act.3sg. of *mləwtk-* 'take off' (PK NS 44 b4)
- *otkasa-me* /otkə́sa/, prt.act.3sg.-pl. of *wətk-* 'separate' (THT 366 b5)
- *yonmasa* /yonmə́sa/, prt.act.3sg. of *yənm-* 'get' (PK AS 1B b2, THT 29 a3)
- *sätkasamai* /sətksə́samay/, prt.mid.1sg. of *sətk-* 'spread (tr.)' (IOL Toch 4 b5)

However, since the /ə́/ is never seen outside of forms of this type, where it can easily be understood as secondary, its earlier presence in verbs with simpler root structures remains conjectural (Peyrot 2013: 509). The basis for its reconstruction by Winter (1993: 199–201; also Malzahn 2010: 192–193) in **prekə́sa*, **prəkə́samay*, etc., is the fixed initial accent of these forms in TB *preksa*, *parksamai*, which it also purports to explain.

If we try to understand the phonological developments involved, it would furthermore be strange that one suffix **-sV* should develop an epenthetic /ə́/ while another suffix **-sV* did not. In the other verbal category with a suffix of the structure **-sV*, the *s*-present [class 8], there is no fixed initial accent (cf. *naksentär* /naksə́ntər/ 'they blame' rather than ***näksentär* /nə́ksentər/, *tsämsen-ne* /tˢəmsə́nne/ 'they increase' rather than ***tsamsen-ne* /tˢə́msen-ne/, etc.). There are no *s*-presents with a final cluster like in *otkasa-me*, but the absence of the supposed effects of an earlier /ə́/ in this class always fixing the accent on the root, as compared to the *s*-preterite [3], speaks against Winter's solution as the result of a sound change. The various issues in the preterite [3] add to the problems that Winter's deletion of /ə́/ had in the presents [2] and [8] discussed in the previous subsection.

A prehistoric accent shift from the first to the second syllable accounts elegantly for both the type of *tässāte* and the type of *otkasa*, since they show the accent precisely on the second syllable out of three. Other forms of the paradigm, including *preksa*, *preksa-ne*, *parksante*, and middle-only *eṅksate* with their stable initial accent require a different explanation. In the case of *preksa* and *preksa-ne*, it is easiest to assume that the accent of disyllabic *preksa* was generalized (Peyrot 2013: 512). The initial accent of the middle type *parksante* /pə́rksate/ is then due to influence from the active (ibid.). Middle-only preterites [3] like *eṅksate* were subsequently regularized with initial accent as well. The few verbs like *tässāte* that do show second-syllable accent in the middle evidently managed to escape the effects of this analogy. Of course, this interpretation requires more analogy, i.e., irregular change, than Winter's, but we have seen that the purely phonological account falls short.

One might ask why initial accent was not generalized in other categories like the active of the thematic present [class 2], where initially accented disyllabic forms like 1sg. *ākau*, 2sg. *āśt**, 3sg. *āśäṃ* and 3pl. *akeṃ* routinely move their accent to the second syllable when pronoun suffixes are added. First, these forms are not as isolated within their paradigm as *preksa*, which is the only sigmatic form in the active. This isolation may have increased the relative strength of the base *preksa*. Second, it should be borne in mind

that, unlike sound change, analogical change does not apply regularly, so that it may have affected different categories in different ways. Its effects may even have been different even for verbs in the same category.

Now we briefly turn to the accentuation of the asigmatic forms in the paradigm. At first sight, it looks like there is nothing particularly strange going on with the accent there, but if we consider the regular sound developments, we can see that more analogy has been at work. To see this, it will be illustrative to consider the effects of epenthesis on the shape of the verb forms in this category of the sigmatic preterite. It is likely that different *s*-preterites developed in different ways according to the rules of epenthesis, depending on their structure. For example, it is not expected that the cluster *-*ks*- should be broken up (cf. the prt.act.3pl. *preksa* /préksa/ rather than **prekasa* /prekə́sa/, *laks* /lə́ks/ 'fish', etc.), so that an earlier 2pl. **prek-sə* should have regularly yielded TB **preks*. It is not clear for every root-final consonant cluster how epenthesis would have operated, but Table 5.11 illustrates some possibilities with the root *awn*- act. 'hit' ending in a nasal, *prek*- 'ask' ending in a single stop, *ṣərp*- 'indicate' ending in a cluster of a resonant and a stop, and *wotk*- 'separate' ending in a cluster of two stops.

Table 5.11: The expected phonological developments in the active preterite [3] paradigms of *awn*- act. 'hit', *prək*- 'ask', *sərp*- 'indicate' and *wotk*- 'separate'.

| *awn*- act. 'hit' prt. | | epenthesis | shift | apocope | attested type |
|---|---|---|---|---|---|
| 1sg. | *áwn-əwa* | *áwn-əwa* | *awn-áwa* | *awnə́wa* | = *aunuwa** |
| 2sg | *áwn-sta* | *áwn-ə-sta* | *awn-á-sta* | *awnásta* | = *aunasta* |
| 3sg. | *áwn-sa* | *áwn-t-sa* | *áwn-t-sa* | *áwntsa* | = *auntsa* |
| 1pl. | *áwn-mə* | *áwn-mə* | *áwn-mə* | *áwnm* | ≠ *aunam** |
| 2pl. | *áwn-sə* | *áwn-t-sə* | *áwn-t-sə* | *áwnts* | ≠ *aunas** |
| 3pl. | *áwn-rə* | ? *áwn-rə* | *áwn-rə* | *áwnər* | ≠ *aunar* |
| *prək*- 'ask' prt. | | epenthesis | shift | apocope | attested type |
| 1sg. | *prék-əwa* | *prék-əwa* | *prek-áwa* | *prekə́wa* | = *prekuwa* |
| 2sg | *prék-sta* | *prék-ə-sta* | *prek-á-sta* | *prekə́sta* | = *prekasta** |
| 3sg. | *prék-sa* | *prék-sa* | *prék-sa* | *préksa* | = *preksa* |
| 1pl. | *prék-mə* | ? *prék-ə-mə* | *prek-á-mə* | *prekə́m* | = *prekam** |
| 2pl. | *prék-sə* | *prék-sə* | *prék-sə* | *préks* | ≠ *prekas** |
| 3pl. | *prek-rə* | *prék-rə* | *prék-rə* | *prékər* | ≠ *prekar* |
| *ṣərp*- 'indicate' prt. | | epenthesis | shift | apocope | attested type |
| 1sg. | *ṣérp-əwa* | *ṣérp-əwa* | *ṣerp-áwa* | *ṣerpə́wa* | = *ṣerpuwa** |
| 2sg | *ṣérp-sta* | *ṣérp-ə-sta* | *ṣerp-á-sta* | *ṣerpásta* | = *ṣerpasta** |
| 3sg. | *ṣérp-sa* | *ṣérp-sa* | *ṣérp-sa* | *ṣérpsa* | = *ṣerpsa* |

| | | | | | |
|---|---|---|---|---|---|
| 1pl. | *ṣérp-mə* | *ṣérp-ə-mə* | *ṣerp-á-mə* | *ṣerpám* | = *ṣerpam** |
| 2pl. | *ṣérp-sə* | *ṣérp-sə* | *ṣérp-sə* | *ṣerps* | ≠ *ṣerpas** |
| 3pl. | *ṣérp-rə* | *ṣérp-ə-rə* | *ṣerp-á-rə* | *ṣerpár* | = *ṣerpar** |
| *wotk-* 'separate' prt. | | epenthesis | shift | apocope | attested type |
| 1sg. | *ótk-əwa* | *ótk-əwa* | *otk-áwa* | *otkáwa* | = *otkuwa** |
| 2sg | *ótk-sta* | *ótk-ə-sta* | *otk-á-sta* | *otkásta* | = *otkasta** |
| 3sg. | *ótk-sa* | *ótk-ə-sa* | *otk-á-sa* | *otkásа* | = *otkasa* |
| 1pl. | *ótk-mə* | *ótk-ə-mə* | *otk-á-mə* | *otkám* | = *otkam** |
| 2pl. | *ótk-sə* | *ótk-ə-sə* | *otk-á-sə* | *otkás* | = *otkas** |
| 3pl. | *ótk-rə* | *ótk-ə-rə* | *otk-á-rə* | *otkár* | = *otkar** |

Since the regular development of epenthetic vowels would have been different depending on the root, the accentuation of resulting pre-TB forms must also have been different originally. The paradigm of *wotk-* would probably have generated all the expected forms by sound law, but *awn-* likely only obtained the singular forms by regular sound developments. The only form that would have had an epenthetic vowel with basically every type of root is the second person singular, since the suffix-initial cluster of the ending *-sta* itself would always develop with epenthesis in the cluster *-Cst-* (see 5.2.4). The accentual type of *wotk-*, with a second-syllable accented schwa seems to have been regularized in the asigmatic forms of other preterite [3] verbs as well.

Analogical pressures would have come from many sides, partly from the first and second person singular, partly from (relatively rarer) paradigms of the type exhibited by *wotk-*, and perhaps partly as the result of the accentuation type exhibited by the preterite [1]. An extension of root shape |prekə́-| found regularly at least in the 1sg. /prekə́wa/ 2sg. /prekə́sta/*, and perhaps the 1pl. /prekə́m/* in certain verbs, to the other forms of the paradigm could have been facilitated by a comparison with the preterite [1] root shape |taká-| in the way illustrated in Table 5.12.

Table 5.12: Morphological structure of the preterite [1] and preterite [3] forms, illustrating the room for analogical influence from the former on the latter.

| | prt.[3] act. | analysis | result | prt.[1] act. | analysis |
|---|---|---|---|---|---|
| 1sg. | **prekə́wa* | \|prekə́-\| is as \|taká-\| | *prekə́wa* | *takáwa* | \|taká-\| |
| 2sg | **prekə́sta* | \|prekə́-\| is as \|taká-\| | *prekə́sta* | *takásta* | \|taká-\| |
| 1pl. | **prekə́m* | \|prekə́-\| is as \|taká-\| | *prekə́m* | *takám* | \|taká-\| |
| 2pl. | **préks* | \|prék-\| >> \|prekə́-\| | *prekə́s* | *takás* | \|taká-\| |
| 3pl. | **prékər* | \|prékə-\| >> \|prekə́-\| | *prekə́r* | *takáre* | \|taká-\| |

Jasanoff, with a limited accent shift according to his "*yāmu*-rule", has a rather different account of the accentuation of the preterite [3]. For the sigmatic forms like *préksa*, Jasanoff departs from **prékəsa*, where the "*yāmu*-rule" would prevent accent shift (Jasanoff 2015: 95). This is similar to Winter's **prekásа*, but with a different mechanism for obtaining the attested accentuation. The problems associated with the *ə* in **prekəsa*, as discussed above in the context of Winter's scenario, remain.

As for the asigmatic part of the paradigm, forms like 2sg. **prékəsta* and 1pl. **prekəmə* should have remained as such, since the accent could not have shifted from the **e* to the **ə* according to Jasanoff's "*yāmu*-rule". By way of a solution, Jasanoff considers the *ə* in these forms an "accent-attracting Fremdvokal [= Tocharian <ä>, /ə/]", which would have behaved differently from other *ə* (Jasanoff 2015: 95). The distinction between the *ə* that does not attract the accent (as allegedly in the sigmatic forms) and the *ə* that does attract the accent (as in these asigmatic forms) is arbitrary, so that its use as an explanation for the accent is circular. Additionally, the *ə* in forms like *aista-ne* /aystə́ne/ and *yaśa-c* /yaśə́c/ is not an epenthetic schwa but did end up accented, in contradiction with both Jasanoff's "*yāmu*-rule" and his assumed "accent-attracting Fremdvokal".

The accentuation of the sigmatic preterite [class 3] can thus be most adequately explained with a combination of regular accent shift to the second out of three (or more) syllables and an analogical morphophonological regularization of the accentuation of different stem types. To better appreciate the motivation for the analogical changes treated in this subsection, it may be taken into consideration that the preterite [3] paradigm obtained by sound law with regular accent shift could have had at least four different morphological stems when the placement of the accent was no longer phonologically automatic: asigmatic |prék-| in the act.2pl. and the act.3pl., |prekə́-| in the act.1sg, act.2sg. and perhaps act.1pl., next to sigmatic |prék-sa-| for the act.3sg. and |prək-sá-| for the middle. It is not terribly surprising that such complexity was reduced by levelling to just two stem variants: one sigmatic, ablauting |pré/ə́ksa-|, and one asigmatic, |prekə́-|. Both were in opposition to a subjunctive stem |prék(ə)-|, which is treated as part of the next subsection.

### 5.3.4 The accent in the subjunctives [1] and [5]

The verbs in the subjunctive classes [1] and [5] show a tendency for fixed word-initial accent, regardless of the number of syllables. Not all verbs in these subjunctive classes have a fixed initial accent, however. Table 5.13 illustrates this with two example paradigms of different types of *a*-subjunctives [class 5]: *taka-*, the suppletive subjunctive of the verb *nes-* 'be', with consistent *a*-grade and *ləma-*, the suppletive subjunctive of the verb *səm-* 'sit', with ablaut. The ablauting subjunctives [5] like *ləma-* have *a*-grade in the active singular and *ə*-grade in the active plural and the entire middle, and in derived forms such as the gerundive and infinitive. In general, on the subjunctives [1] and [5], see

Krause & Thomas (1960: 222–224, 226–230), Malzahn 2010: (274–316); Peyrot (2013: 494–560).

Table 5.13: Two example paradigms of the subjunctive [5].

| | *taka-* 'be' sbj.[5] | | *ləma-* 'sit' sbj.[5] | |
|---|---|---|---|---|
| | base | suffixed | base | suffixed |
| 1sg. | *tākau* | *tākau-ne** | *lāmau** | *lāmau-ne** |
| 2sg | *tākat* | *tākat-ne** | *lāmat* | *lāmat-ne** |
| 3sg. | *tākaṃ* | *tākaṃ-ne** | *lāmaṃ* | *lāmaṃ-ne** |
| 1pl. | *tākam* | *tākam-ne** | *lamam* | *lamam-ne** |
| 2pl. | *tākacer* | *tākacer-ne** | *lamacer** | *lamacer-ne** |
| 3pl. | *tākaṃ* | *tākaṃ-ne* | *lamaṃ* | *lamaṃ-ne** |

Despite this variation, some generalizations are possible. The remainder of this section treats the subjunctives of class [5] first, and then the subjunctives of class [1], grouping the examples according to a few different observable patterns. The general patterns are mentioned by both Malzahn and Peyrot, but the underlying data is not explicitly provided.

**Subjunctives [5] with initial accent**

Most verbs that do not show any ablaut in the subjunctive [5] are accented on the first syllable (Malzahn 2010: 283–284). This is true for all verbs with a subjunctive [5] that have a *ə*-grade in the present, as shown in Table 5.14. These are all paired with an *a*-preterite [class 1], with which they are identical in terms of stem structure except for the accentuation; e.g., *pəla-* 'praise' has a preterite [1] stem |palá-| beside its subjunctive [5] stem |pála-|.

Table 5.14: Subjunctives [5] with consistent *a*-grade in the subjunctive, but not in the present, with diagnostic subjunctive forms showing their first-syllable accent.

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *tresk-* | 'chew' | sbj.ger. *trāska*(*lye*) IOL Toch 157 a3 | \|tráska-\| |
| *nəytt-* | 'collapse' | inf. *naittatsi* THT 270 a3 | \|náytta-\| |
| *pəla-* | 'praise' | sbj.mid.1sg. *pālamar* THT 240 b6<br>opt.mid.1sg. *pāloymar* PK AS 5A a5<br>sbj.vn. *pālalñe* THT 14 a6 | \|pála-\| |
| *pəya-* | 'trumpet' | sbj.vn. *pāyalyñe* PK AS 15D b2 | \|páya-\| |

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *pəyk-* | 'paint, write' | sbj.mid.3sg.-3sg. *paiykatär-ne* IOL Toch 213 a2<br>sbj.vn.per. *paikalñesa* PK NS 27 a3<br>inf. *paikatsi* THT 357 a6 | \|páyka-\| |
| *mənt-* | 'hurt; stir' | sbj.mid.3sg.-3sg. *māntatär-ne* THT 331 a1<br>sbj.vn. *māntalñe* THT 171 a4<br>inf. *māntatsi* THT 596 b5 | \|mánta-\| |
| *məyw-* | 'tremble' | sbj.vn. *maiwalñe* PK AS 2C a2<br>inf. *maiwatsi* PK AS 13I a7 | \|máywa-\| |
| *layka-* | 'wash' | sbj.vn. *laikalñe* THT 271 b3<br>inf. *laikatsi* IOL Toch 723 b2 | \|láyka-\| |
| *ləwp-* | 'rub' | sbj.vn. *lauwalñe* (*-p-* > *-w-*) PK AS 2C a5<br>inf. *laupatsi* W 34 b3 | \|láwpa-\| |
| *skəra-* | 'scold' | sbj.ger. *skāralñe* THT 122 a7<br>inf. *skāratsi* THT 266 b3 | \|skára-\| |

Roots that contain a basic internal *a*-vowel are also overwhelmingly accented on the first syllable in the subjunctive [5] (Malzahn 2010: 285), as shown in Table 5.15. As with the subjunctives [5] in Table 5.14 above, the only difference between the subjunctive [5] and preterite [1] stem of these verbs is in the accent; e.g., *karpa-* 'descend' has a preterite [1] stem |karpá-| beside its subjunctive [5] stem |kárpa-|.

Several subjunctives [5] that are expected to belong to this group are not attested in forms that are diagnostic for the placement of the accent, as far as I can tell on the basis of the forms collected and cited by Malzahn (2010). They are *alpa-* 'stroke', *auswa-* '± cry, lament', *klapa-* '± touch', *trappa-* 'trip', *naska-* 'spin?', *patka* 'give up', *pasa* '± speak', *pauta-* 'flatter', *yanka-* 'be deluded', *latka-* 'cut off', *waka-* 'split', *skaya-* 'strive', *śampa-* 'be conceited', *spawa-* 'diminish', *swara-* 'please', and *waska-* 'stir'.

Table 5.15: Subjunctives [5] with a basic *a*-vowel in the root and consistent *a*-grade in the subjunctive, with diagnostic forms showing their first-syllable accent in the subjunctive.

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *aksa-* | 'wake up (itr.)' | sbj.vn. *āksalñe* PK AS 4A b5 | \|áksa-\| |
| *ara-* | 'cease | sbj.vn. *āralñe* PK NS 414 a3<br>inf. *āratsi* THT 282 a1 | \|ára-\| |
| *art(t)a-* | 'love, praise' | sbj.mid.3pl. *ārtantär* PK AS 13D a9<br>opt.mid.1sg. *ārttoymar* PK AS 5C a2<br>sbj.ger.m.pl. *ārttalyi* THT 1119 b1<br>sbj.vn. *ārttalñe* THT 107 a4 | \|árt(t)a-\| |

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *kaka-* | 'call' | sbj.mid.3sg. *kākatär* THT 331 b2<br>opt.mid.1sg. *kākoyma(r)* IOL Toch 82 a2<br>sbj.vn. *kākalñe* THT 543 a1<br>inf. *kākatsi* THT 241 a1 | \|káka-\| |
| *kanta-* | 'rub' | opt.mid.3sg. *kāntoytär* THT 19 b6<br>sbj.vn.abl. *kāntal(ñ)emeṃ* THT 532 a2<br>inf. *kāntatsi* THT 324 a1 | \|kánta-\| |
| *kant$^{s}$a-* | 'sharpen' | inf. *kāntsasi* THT 432 b4 | \|kánt$^{s}$a-\| |
| *kara-* | 'gather' | inf.-all. *kāratsi-śco* THT 23 a7 | \|kára-\| |
| *karpa-* | 'descend' | sbj.act.2sg. *kārpat* THT 384 b4<br>sbj.vn.gen.sg. *kārpalñents(e)* PK AS 17H b1<br>inf. *kārpatsi* PK AS 17G b6 | \|kárpa-\| |
| *kawa-* | 'desire' | sbj.vn.per. *kāwalyñesa* THT 102 a3<br>inf. *kāwatsi* PK AS 13F b6 | \|káwa-\| |
| *kawta-* | 'split, cleave' | sbj.mid.3sg. *kautatr=* THT 61 b3<br>sbj.vn. *kautalñe* THT 5 a3<br>inf. *kautatsi* THT PK AS 6L a2 | \|káwta-\| |
| *krawp-* | 'gather' | sbj.mid.1sg. *kraupamar* PK AS 6A b1<br>sbj.mid.3sg. *kraupatär* PK AS 17D a4<br>sbj.ger. *kraupalle* THT 33 a5<br>sbj.vn. *kraupalñe* THT 155 a2<br>inf. *kraupatsi* THT 1573.a b1 | \|kráwpa-\| |
| *klanka-* | 'drive' | inf. *klāṅkatsi*$^{(a)}$ see Malzahn 2010: 617–618 | \|klánka-\| |
| *klaya-* | 'fall' | sbj.act.2sg. *klāyat* IOL Toch a44 a5<br>sbj.ger.f.pl. *klāyallona* THT 1374.g a3<br>sbj.vn. *klāyalñe* PK NS 53 b4 | \|kláya-\| |
| *klayksa-* | 'dry up' | inf. *klaiksatsi* THT 11 b3 | \|kláyksa-\| |
| *klawtka-* | 'return' | sbj.ger. *klautkalle* THT 1681 b4 | \|kláwtka-\| |
| *taka-* | 'be, become' | sbj.act.2sg. *tākat* THT 22 b2<br>sbj.act.1pl. *tākam* THT 77 a1<br>sbj.act.2pl. *tākacer* THT 107 b3<br>opt.act.2sg. *tākoyt* THT 5 b8<br>opt.act.2pl. *tākoycer* THT 338 a6 | \|táka-\| |
| *nawta-* | 'disappear' | sbj.act.3pl.-3sg. *nautan-ne* PK AS 7G a1<br>sbj.ger. *nautalle* THT 268 b1<br>sbj.vn. *nautalñe* THT 30 a3 | \|náwta-\| |

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *planta-* | 'rejoice' | sbj.act.1pl. *plāntam* PK AS 17C a5<br>inf. *plāntatsi* THT 1235 a1 | \|plánta-\| |
| *mansa-*<br>*+ mens-* | 'be sorrowful' | sbj.mid.3pl. *māntsanträ* PK AS 15J b3<br>sbj.vn. *māṃtsalyñe* THT 88 a4 | \|mántˢa-\| |
| *mawka-* | 'refrain from' | sbj.vn. *maukalyñ(e)* THT 1270 b1 | \|máwka-\| |
| *mrawska-* | 'feel disgust' | sbj.ger. *mrauskalle* PK AS 7L a4<br>vn. *mrauskalñe* THT 3 b2 | \|mráwska-\| |
| *yata-* | 'be able' | opt.mid.3sg.-1sg. *yāto-ñ* IOL Toch 269 a4<br>sbj.ger. *yātalle* THT 127 a4<br>sbj.vn. *yātalñe* THT 600 a3<br>inf.all. *yātatsiś* THT 30 a3 | \|yáta-\| |
| *yawkk-* | 'use' | sbj.mid.3pl. *yaukkanträ* THT 36 b3 | \|yáwkka-\| |
| *raka-* | 'extend' ? | opt.mid.3pl.-1sg. *rākoyentär-ñ* THT 271 a1 | \|ráka-\| |
| *rapa-* | 'dig; plough' | inf. *rāpatsi*$^{(b)}$ IOL Toch 246 a2 | \|rápa-\| |
| *layta-* | 'fall, pass' | sbj.vn. *laitalñe* THT 5 a4 | \|láyta-\| |
| *wapa-* | 'weave' | inf. *wāpatsi* IOL Toch 7 b4 | \|wápa-\| |
| *ak-* | 'lead' | sbj.ger. *wāyalle* PK NS 79.1 a5<br>inf. *wāyatsi* IOL Toch 151 a3 | \|wáya-\| |
| *waywa-* | 'be wet' | sbj.vn. *waiwalñe* PK AS 2C a2 | \|wáywa-\| |
| *wala-* | 'cover' | opt.mid.3sg. *wāloyträ* PK AS 5C b2<br>sbj.vn. *wālalñe* THT 534 a4 | \|wála-\| |
| *waltsa-* | 'crush' | inf. *wāltsasi* PK DA M 507.23 a16 | \|wáltˢa-\| |
| *wrata-* | 'form' | sbj.vn. *wrātalñe* IOL Toch 133 b2 | \|wráta-\| |
| *wlawa-* | 'control' | sbj.mid.3sg. *wlāwatärä* IOL Toch 78 b1<br>opt.mid.1sg. *wlāwoymar* IOL Toch 307 a5<br>sbj.vn *wlāwalñe* IOL Toch 785 b3 | \|wláwa-\| |
| *samp-* | 'take away' | sbj.ger. *sāmpal(l)e* IOL Toch 603 b4<br>inf. *sāmpatsi* IOL Toch 214 a5 | \|sámpa-\| |
| *səwa-*<br>*+ swasa-* | 'rain' | sbj.ger. *swāsallye* IOL Toch 7 a1<br>vn. *swāsalñe* THT 1539.d a3 | \|swása-\| |
| *skawa-* | 'kiss' ? | inf. *skāwa(tsi)* THT 83 a3 | \|skáwa-\| |
| *spartta-* | 'turn' | sbj.vn. *spārttalñe* THT 149 a2<br>inf. *spārtatsi* THT 249.a b1 | \|spártta-\| |
| *tˢarwa-* | 'be comforted' | opt.mid.3sg. *tsārwoyträ* SI B 119(1) a1 | \|tˢárwa-\| |
| *tˢalta-* | chew' | sbj.ger. *tsāltalye* IOL Toch 248 b1 | \|tˢálta-\| |

(a) Cited by Thomas (1952: 50 fn. 3) without a reference, so that the existence of this form cannot be properly verified (see Malzahn 2010: 617; Peyrot 2013: 742 fn. 712).
(b) The supposed verbal noun *rapālñe*, with conflicting medial accent, was incorrectly identified, see Peyrot (2018b: 265–266).

Peyrot has provided an interpretation for the fixation of the accent on the first syllable within Marggraf's framework of a progressive accent shift from the first to the second syllable (Peyrot 2013: 510), which accords with the rules of epenthesis and syncope discussed in the previous section (5.2). Due to the shape of the person-number endings, subjunctive [5] would have regularly developed with initially accented 1sg., 3sg. and 3pl. forms next to medially accented 2sg., 1pl. and 2pl. forms. In the corresponding preterite [1], all forms except the 3sg. would have had a consistent second-syllable accent due to the addition of an extra syllable with the preterite endings (Peyrot 2013: 511–512). Table 5.16 displays the implied historical change in combination with a diachronic progressive accent shift from the first to the second syllable in trisyllabic words (Peyrot 2013: 510),

Table 5.16: An illustration of the analogical change that caused the first-syllable accent of the subjunctive [5], according to Peyrot (2013: 510–512).

| | sbj.[5] | | regularized | prt.[1] | |
|---|---|---|---|---|---|
| | initial | shift | >> \|táka-\| | initial | shift > \|taká-\| |
| 1sg. | **táka-w* | **táka-w* | *tákaw* | **táka-wa* | *taká-wa* |
| 2sg | **táka-tə* | **taká-tə* | >> *tákat* | **táka-sta* | *taká-sta* |
| 3sg. | **táka-n* | **táka-n* | *tákan* | **táka-∅* | *táka*, *taká-ne* |
| 1pl. | **táka-mə* | **taká-mə* | >> *tákam* | **táka-mə* | *taká-mə* |
| 2pl. | **táka-cer* | **taká-cer* | >> *tákacer* | **táka-sə* | *taká-sə* |
| 3pl. | **táka-n* | **táka-n* | *tákan* | **táka-re* | *taká-re* |

Of these forms of the paradigm, the subjunctive 1pl. **takámə* would have been identical to the preterite 1pl **takámə*. To repair this ambiguity, the subjunctive took over the initial accent of 1sg. **tákaw*, 3sg. **tákan* and 3pl. *tákan*. As the accent provided a new way to morphologically differentiate the preterite and subjunctive stems, the other forms followed suit (Peyrot 2013: 511–512).

The remaining subjunctives [5] with their accent on the first syllable are listed in Table 5.17. Around half of these demonstrably have ablaut of *a*-grade in the active singular and *ə*-grade in the other forms. It is possible that more of them showed this alternation, but that this is not visible in all cases due to a lack of attested diagnostic active singular forms. The subjunctives for which ablaut is confirmed by active singular

forms are marked in the table below with superscript "a." behind the morphophonological analysis of the subjunctive stem.

Table 5.17: Subjunctives [5] with *ə*-roots or ablauting roots and stable first-syllable accent.

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *kəta-* | 'strew' | opt.mid.3sg. *katoytr*= THT 42 b5<br>inf. *katasi* PK DA M 507.27 a3 | \|kə́ta-\|$^{(a)}$ |
| *kətka-* | 'cross' | sbj.act.2sg. *kātkat* THT 1103 a4<br>sbj.ger.m.pl. *katkalyi* THT 539 a3<br>sbj.vn. *katkalñe* PK NS 171 a1<br>inf. *katkatsi* IOL Toch 387 a2 | \|kátka-\|$^{(a)}$<br>\|kə́tka-\| |
| *kərka-* | 'steal' | inf. *karkats*(*i*) THT 14 b8 | \|kə́rka-\| |
| *kərsa-* | 'know' | sbj.vn. *karsalñe* THT 183 b2<br>inf. *karsatsi* PK AS 16.2 b5 | \|kə́rsa-\|$^{(a)}$ |
| *krəst-* | 'cut off' | opt.mid.1sg. *karstoymar* IOL Toch 307 b1<br>opt.mid.3sg. *karstoytär* THT 271 b1<br>inf. *karstatsi* PK AS 4B a3 | \|krə́sta-\|$^{(a)}$ |
| *kəla-* | 'lead' | sbj.mid.2sg. *kalatar* PK AS 18A b3<br>sbj.mid.3sg. *kalatär* PK AS 18.18 b2<br>inf. *kalatsi* THT 77 a4 | \|kə́la-\|$^{(a)}$ |
| *kləntˢ-* | 'sleep' | sbj.vn. *klantsalñe* PK AS 2C b4 | \|klə́ntˢa-\| |
| *kləwtka-* | 'turn' | inf. *klutkatsi* THT 1446 a3 | \|klə́wtka-\| |
| *kʷəsa-* | 'lament' | sbj.vn. → der.adj. *kwasalñeṣṣe* THT 86 a2 | \|kwə́sa-\| |
| *tərka-* | 'dismiss' | sbj.act.1pl. *tarkam* PK AS 16.2 a5<br>sbj.act.2pl. *tarkacer* THT 28 a1<br>sbj.ger.f.pl. *tarkall*(*e*) THT 518 a2<br>sbj.vn. *tarkalñe* PK AS 2C a5<br>inf. *tarkatsi* THT 42 a7 | \|tə́rka-\|$^{(a)}$ |
| *nəwa-* | 'cry' | sbj.vn. *nūwalñe* THT 1304 a3 | \|nə́wa-\| |
| *pənna-* | 'stretch' | inf..all. *pannatsiś* THT 9 b3 | \|pə́nna-\|$^{(a)}$ |
| *pəlka-* | 'see' | sbj.act.1pl. *palkam* PK AS 16.8 a3<br>sbj.vn. *palkalñe* THT 41 b7<br>inf. *palkatsi* IOL Toch 213 a1 | \|pə́lka-\|$^{(a)}$ |
| *məka-* | 'run' | inf. *makatsi* THT 10 a4 | \|mə́ka-\| |
| *mərtka-* | 'shave' | sbj.vn. *markalñ·* (*-rtk-* > *-rk-*) THT 309 a4<br>inf. *markasi* PK DA M 507.35 a46 | \|mə́rtka-\| |
| *mərsa-* | 'forget' | inf. *marsatsi* THT 556 a2 | \|mə́rsa-\|$^{(a)}$ |

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *məlka-* | 'put on' | sbj.vn. *malkalñe* IOL Toch 49 a1 | \|mə́lka-\| |
| *pləska-* | 'consider' | opt.mid.3sg. *palskoyträ* THT 300 a3<br>sbj.vn. *palskalñe* IOL Toch 748 a1<br>inf. *palskatsi* IOL Toch 96 b5 | \|pləska-\|$^{(a)}$ |
| *pəwtka-* | 'divide' | sbj.vn. *putkalñe* THT 172 b2<br>inf. *putkatsi* PK AS 6E b6 | \|pə́wtka-\|$^{(a)}$ |
| *məyta-* | 'set out (itr.)' | sbj.ger.loc. *mītalyene* PK AS 5B b4 | \|mə́yta-\| |
| *yəksa-* | 'embrace' | inf. *yaksatsi* THT 1262 a3 | \|yə́ksa-\| |
| *yəwka-* | 'overcome' | sbj.vn. *yukalñe* THT 512 a4<br>inf. *yukatsy* S 1 a6 | \|yə́wka-\| |
| *rənka-* | 'ascend' | sbj.ger. *raṅkal(l)e* THT 355 a6 | \|rə́nka-\| |
| *rəss-* | 'tear' | sbj.vn. *rassalñe* PK AS 7C b2 | \|rə́ssa-\| |
| *rəynka-* | '?' | inf. *riṅkatsi* W 39 b1 | \|rə́ynka-\| |
| *rəyta-* | 'seek' | sbj.ger. *ritale* PK AS 8C a6<br>sbj.vn. *rītalñe* SI P 2 b1<br>inf. *ritatsi* THT 1574 b1 | \|rə́yta-\| |
| *rəwtka-* | '(re)move' | sbj.vn. *rutkalyñ*= THT 290 a5<br>inf.all. *rutkatsiś* THT 537 a2 | \|rə́wtka-\|$^{(a)}$ |
| *ləma-* | 'sit' | sbj.act.2sg. *lāmat* THT 567 a1<br>sbj.ger.per. *lamallesa* IOL Toch 247 a6<br>inf. *lamatsi* IOL Toch 247 a3 | \|láma-\|$^{(a)}$<br>\|lə́ma-\| |
| *ləwka-* | '?' | inf. *lukatsi* PK DA M 507.42 and .40 b5 | \|lə́wka-\| |
| *wəta-* | 'fight' | sbj.vn. *watal(yñe)* THT 1107 a5 | \|wə́ta-\| |
| *wərpa-* | 'enjoy' | sbj.mid.1sg. *warpamar* PK AS 15E b3<br>sbj.mid.2sg. *warpatar* THT 1104 b3<br>sbj.mid.3sg. *warpatär* THT 1104 a2<br>opt.mid.1sg. *warpoymar* THT 20 a5<br>opt.mid.3sg. *warpoyträ* THT 1543.d a5<br>sbj.ger. *warpalle* THT 107 b10<br>sbj.vn. *warpalñe* THT 156 b4<br>inf. *warpatsi* PK AS 7C a2 | \|wə́rpa-\| |
| *səlk-* | 'pull out' | sbj.mid.1sg. *salkamar* THT 296 b5<br>inf. *salkatsi* THT 1323.a a4 | \|sə́lka-\|$^{(a)}$ |
| *stəma-* | 'stand' | sbj.vn. *stamalñe* THT 182 a5<br>inf. *stamatsi* PK AS 8C b3 | \|stə́ma-\|$^{(a)}$ |

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *srəwka-* | ‘die’ | sbj.ger. *srukalle* THT 309 a4<br>vn. *srukalñe* THT 108 b4 | \|srə́wka-\|(a) |
| *tˢənka-* | ‘(a)rise’ | sbj.ger.m.pl. *tsaṅkalyi* THT 608 a3<br>sbj.vn. *tsaṅkalñe* THT 197 a3<br>inf. *tsaṅkatsi* THT 561 a2 | \|tˢə́nka-\|(a) |

(a) These roots have active singular subjunctives form attested, showing ablaut.

The initial accent of the majority of the verbs in Table 5.17 can be interpreted as an extension of the type exhibited in the subjunctive of *táka-*, illustrated above in Table 5.16 after Peyrot (2013: 511–512). The *a*-grade 2sg. may have been adapted first due to the closer correspondence of the *a*-grade singular forms to the *táka*-type (Peyrot 2013: 514), creating an association between the full grade and initial accent (ibid.). This is illustrated in Table 5.18.

Table 5.18: Analogical extension of the first-syllable accent first to the full-grade forms and then the general paradigm of ablauting subjunctives [5].

| | *kətka-* ‘strew’ | | sbj.[5] | *taka-* ‘be’ | *kətka-* ‘cross’ |
|---|---|---|---|---|---|
| | shifted | | >> \|ká/ətka-\| | sbj.[5] \|táka-\| | prt.[1] \|śətká-\| |
| 1sg. | **kátka-w* | **kátkaw* | *kátkaw** | *tákaw* | *śətkáwa* |
| 2sg | **katká-tə* | >> **kátkat* | *kátkat* | *tákat* | *śətkásta* |
| 3sg. | **kátka-n* | **kátkan* | *kátkan* | *tákan* | *śə́tka* (*śətká-ne**) |
| 1pl. | **kətká-mə* | **kətkám* | >> *kátkam** | *tákam* | *śətkám** |
| 2pl. | **kəká-cer* | **kəkácer* | >> *kátkacer** | *tákacer* | *śətkás** |
| 3pl. | **kátka-n* | **kátkan* | *kətkan** | *tákan* | *śətkáre* |

A differentiation in the placement of the accent between the subjunctive and the preterite also seems to have been a potent force for the regularization of the accent in these categories, since even middle-only verbs with otherwise identical subjunctive and preterite stems ended up using this strategy, despite all forms being (originally) trisyllabic; see Table 5.18. This mechanism is thus perhaps formulated too cautiously by Peyrot (2013: 511–512). It came to apply to an entire category of verbs with otherwise identical preterite and subjunctive stems

Table 5.19: Analogical differentiation of the subjunctive [5] and preterite [1] stems, illustrated with middle-only verb *wərpa-* 'enjoy'.

| | sbj.[5] | | secondary | prt.[1] | regular |
|---|---|---|---|---|---|
| | initial | shifted | >> \|wárpa-\| | initial | > \|wərpá-\| |
| 1sg. | **wárpamar* | **wərpámar* | >> *wárpamar* | **wárpamay* | *wərpámay* |
| 2sg | **wárpatar* | **wərpátar* | >> *wárpatar* | **wárpatay* | *wərpátay* |
| 3sg. | **wárpatər* | **wərpátər* | >> *wárpatər* | **wárpate* | *wərpáte* |
| 1pl. | **wárpamtər* | **wərpámtər* | >> *wárpamtər** | **wárpamte* | *wərpámte* |
| 2pl. | **wárpatər* | **wərpátər* | >> *wárpatər** | **wárpatə* | *wərpát** |
| 3pl. | **wárpantər* | **wərpántər* | >> *wárpantər** | **wárpante* | *wərpánte* |

One remaining ablauting subjunctive [5] appears to have had a variable accent, illustrated in Table 5.20. However, the evidence for the initially accented full grade form consists of only a single attestation, *kāskat*, which could in principle be a misspelling for *kaskāt** with an accent on the second syllable. If not, this verb might conserve an intermediate stage as per the mechanism of change illustrated in Table 5.18.

Table 5.20: Accentuation of the subjunctive [5] of *kəska-* 'scatter', with diagnostic attested forms.

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *kəska-* | 'scatter' | sbj.act.2sg. *kāskat* | \|káska-\| |
| | | sbj.mid.3sg. *käskātär* | \|kəská-\| |
| | | sbj.vn. *käskälläñe* | |

**Subjunctives [5] with non-initial accent**

The rest of the subjunctives of class [5] whose accent can be determined are only found with a *ə*-grade (no ablaut) and are accented on the second syllable (Peyrot 2013: 514), as shown in Table 5.21.

Table 5.21: Subjunctives (and present-subjunctives) [5] with ablaut or with a *ə*-vowel in the root and first-syllable accent, with diagnostic attested forms.

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *kərra-* | 'scold' | sbj.mid.3sg.-pl. *kärrātär-me* THT 1575.b a4 | \|kərrá-\| |
| *kəlska-* | 'go down' | sbj.vn. *kälskālñe* THT 1597.b a2 | \|kəlská-\| |
| *kʷəla-* | 'fail' | sbj.mid.3sg.-pl. *$k_u$lātär-me* THT 21 b5 | \|kʷəlá-\| |
| | | sbj.vn. *$k_u$lālñe* IOL Toch 211 b4 | |
| | | inf. *$k_u$lātsi* THT 222 b3 | |

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *krəmpa-* | 'be hindered' | sbj.vn. *krämpālñe* THT 127 a1 | \|krəmpá-\| |
| *tətta-* | 'put' | sbj.mid.3pl. *tättāntär* THT 27 a6<br>sbj.vn. *tättālñe* THT 3 b2 | \|təttá-\| |
| *təyna-* | 'defile oneself' | sbj.mid.3pl. *tinānträ* THT 408 b6 | \|təyná-\| |
| *trəywa-* | 'be mixed' | sbj.mid.3sg. *triwātär* IOL Toch 269 a2<br>sbj.vn. *triwālñe* IOL Toch 227 a3<br>inf. *triwātsi* THT 26 b4 | \|trəywá-\| |
| *trəwka-* | 'allot to' | sbj.ger. *trukāle* THT 433 a4 | \|trəwká-\| |
| *pərka-* | 'become clear' | sbj.mid.3sg. *pärkāträ* THT 1321 b4<br>sbj.vn. *pärkālñe* PK AS 5C b6 | \|pərká-\| |
| *prəwtka-* | 'be shut' | sbj.mid.3sg. *prutkātär* THT 596 a1<br>sbj.mid.3pl. *prutkāntä(r)* THT 569 a3<br>opt.mid.3sg. *prutkoytr=* SI P 2 b6<br>sbj.vn. *prutkālñe* THT 155 b6 | \|prəwtká-\| |
| *plənka-* | 'be sold' | inf.per. *pläṅkā(ts)is(a)* PK AS 8A b9 | \|plənká-\| |
| *mənka-* | 'lack' | sbj.mid.3sg. *mäṅkātär* THT 1466.b a3 | \|mənká-\| |
| *mətˢtˢa-* | 'starve' | sbj.mid.3sg. *mätsātär* PK LC 20 a2 | \|mətˢtˢá-\| |
| *məwska-* | 'get lost' | sbj.vn. *muskālñe* IOL Toch 239 a1<br>inf.all. *muskātsiś* PK AS 5C a4 | \|məwská-\| |
| *mləwta-* | 'pluck?' | sbj.ger. *mlutālle* W 3 a4 | \|mləwtá-\| |
| *mləwtka-* | 'escape' | sbj.mid.1sg. *mlutkāmar* THT 375 a5 | \|mləwtká-\| |
| *rəma-* | 'bend' | opt.mid.3sg. *rmoytär* THT 81 a4 | \|rəmá-\| |
| *rəytta-* | 'be attached' | sbj.mid.3sg. *rittāträ* THT 108 a7<br>sbj.vn.abl. *rittālñemeṃ* THT 3090 a2 | \|rəyttá-\| |
| *ləka-* | 'see' | sbj.act.2sg. *lkāt*, 3sg.-3sg. *lkā-ne*<br>sbj.act.1pl.-2sg. *lkām-c*, 2pl. *lkācer*<br>sbj.mid.3sg. *lkātär*, 3pl. *lkāntär*<br>sbj.ger. *lkālye*, vn. *lkālyñe*<br>inf. *lkātsi* | \|ləká-\| |
| *ləypa-* | 'remain' | sbj.mid.3sg. *lipāträ* IOL Toch 306 a5<br>inf. *lipātsi* THT 497 b6 | \|ləypá-\| |
| *wətka-* | 'be separated' | opt.mid.1pl. *wätkomträ* PK AS 17E b5 | \|wətká-\| |
| *wəra-* | 'smell' | inf. *wrātsi*$^{(a)}$ PK AS 7H b3 | \|wərá-\| |
| *wəyka-* | 'dwindle' | sbj.mid.3sg. *wikātär* PK AS 17J b5<br>sbj.vn. *wikālñe* THT 1354.a a3<br>inf. *wikātsi* THT 127 b6 | \|wəyká-\| |

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *śəwa-* | 'eat' | sbj.act.3pl.-1sg. *śwā-ñ* THT 83 a6<br>sbj.ger. *śwālle* THT 335 b4<br>inf. *śwātsi* THT 335 b4 | \|śəwá-\| |
| *sətka-* | 'spread (itr.)' | sbj.vn. → der.adj. *sätkālñecci* THT 531 a4 | \|sətká-\| |
| *səya-* | 'sweat' | sbj.ger. *syālle* W 5 b2<br>sbj.vn. *syālñe* W 3 b4 | \|səyá-\| |
| *spənta-* | 'trust' | sbj.vn. *späntālñe* PK AS 17A b3 | \|spəntá-\| |
| *spərka-* | 'disappear' | opt.mid.3sg. *spärkoytär* PK AS 5D a1<br>sbj.vn. *spärkālñe* THT 409 b5 | \|spərká-\| |
| *tˢəma-* | 'grow' | sbj.mid.3pl. *tsmāntär* THT 64 a8<br>sbj.vn. *tsmālñe* THT 537 b4<br>inf.-all. *tsmātsi-śco* THT 1509 b1 | \|tˢəmá-\| |
| *tˢəra-* | 'be separated' | sbj.ger. *tsrālle* THT 405 a5<br>sbj.vn. *tsrālñe* THT 200 b2 | \|tˢərá-\| |
| *tˢəlpa-* | 'pass away' | sbj.vn. *tsälpālñe* THT 30 a1<br>inf. *tsälpātsi* THT 1121 b3 | \|tˢsəlpá-\| |

(a) A hapax legomenon, see Peyrot (2013: 820 fn. 849).

A number of the verbs in this category differentiate the finite subjunctive and preterite in another way than with the accent, as pointed out by Peyrot (2013: 513–514). The preterites of *ləka-* 'see' and *śəwa-* 'eat' are |ĺaká-| and |śawá-| respectively, with *a*-grade next to subjunctives with stable *ə*-grade, and *tətta-* 'put' combines with the preterite [3] *təs-*. Of the remainder that have attested finite forms in both the subjunctive and preterite, there are ten that belong to a pattern with middle forms in the subjunctive and active forms in the preterite. They are *kʷəla-* 'fail', *trəywa-* 'be mixed', *pərka-* 'become clear', *prəwtka-* 'be shut', *mənka-* 'lack', *rəytta-* 'be attached', *ləypa-* 'remain', *wəyka-* 'dwindle', *spərka-* 'disappear', *tˢəma-* 'grow' and *tˢəlpa-* 'pass away'. Only two, *kərra-* 'scold' and *rəma-* 'bend' have middle subjunctive forms attested next to middle preterite forms. The others have either no finite subjunctive forms or no finite preterite forms attested to make sure how they should be categorized. The present, where attested, belongs to the middle-only *e*-presents [class 3] (Jasanoff 2015: 91), except for *trəwka-* 'allot to', which has a *na*-present [class 6], and *mləwta-* '±pluck', which has an *sk*-present [class 9].

The accentuation of the subjunctives of the following roots with a subjunctive [5] are unclear to me on the basis of the forms cited by Malzahn (2010): *pərsa-* 'sprinkle', *pərska-* 'be afraid', *klətˢa-* 'oppress', *təksa-* '± destroy', *təwka-* 'hide oneself', *trəyka-* 'go astray', *nətka-* 'push away', *nəwka-* 'swallow', *prənka-* 'restrain oneself, *yəya-* 'go along', *ləwa-* 'send', *wərka*/*warka-* 'shear'?, *səyka-* 'step', and *tˢəyka-* 'form'.

None of the subjunctive [5] forms should have been exempt from progressive accent shift under to Jasanoff's "*yāmu*-rule", since they have a stem-final *-a*. One might assume, therefore, that in this category there would be no functional difference between this theoretical limited accent shift and the general accent shift assumed by Peyrot (2013) and Marggraf (1970). However, Jasanoff's account of this category is actually quite different. Rather than an analogical interplay between observable types of paradigms, as with Peyrot's scenario in Table 5.16, Jasanoff argues that verbs with a subjunctive [5] and accompanying preterite [1] represent two etymologically distinct verbal classes that merged by a process of mutual assimilation: one former class would represent PIE laryngeal-final roots, which regularly developed final *-a*, while the other would consist of roots that had no laryngeal and thus originally no final *-a*. The *e*-grade of the original *a*-less roots was taken from that class, while the *-a* was generalized in the other direction. This scenario is illustrated in Table 5.22.

Table 5.22: An illustration of the development of the subjunctive [5] in Jasanoff's scenario.

| | *initial* | "yāmu-rule" | analogy 1 | analogy 2 | analogy 3 |
|---|---|---|---|---|---|
| act.sg. | *kə́rsa-* | *kərsá-* | >> *kérsa-* | *kérsa-* | *kérsa-* |
| act.pl. | *kə́rsa-* | *kərsá-* | *kərsá-* | *kərsá-* | >> *kə́rsa-* |
| act.sg. | *wéykə-* | *wéykə-* | *wéykə-* | >> *wéyka-* | *wéyka-* |
| act.pl. | *wə́ykə-* | *wəykə́-* | *wəykə́* | >> *wəykə́-* | >> *wə́yka-* |

Thus, verbs of the type **kərsa-* 'know' were supposedly changed to **kersa-* under influence of the type **weykə-* 'dwindle', which in turn became **weyka-*. In original **weykə-*, the accent would be regularly on the first syllable according to Jasanoff's "*yāmu*-rule", and this accentuation was then retained when **wéykə-* became **wéyka-*, and extended to **kérsa* (expected ***kersá-* with Jasanoff's *yāmu*-rule) (Jasanoff 2015: 92; also Jasanoff 2013). This development is not apparent from the Tocharian data and relies on the unverifiable timing of the extension of final *-a* and accent shift. It additionally requires at least as far-reaching analogical changes as in Peyrot's scenario, and does not seem to provide any extra explanatory benefit. Peyrot's scenario is thus to be preferred over Jasanoff's.

**Subjunctives [1] with and without first-syllable accent**

In the root subjunctive [class 1] there are also verbs with and without first-syllable accent. This variation is illustrated in Table 5.23 with the initially accented forms *prek-* 'ask' and the variably accented forms of *ay-* 'ask'. Not all person-number forms are well-attested.

Table 5.23: Two example paradigms of the subjunctive [1].

| | *prek-* 'ask' sbj.[1] | | *ay-* 'give' sbj.[1] | |
|---|---|---|---|---|
| | base | suffixed | base | suffixed |
| 1sg. | *preku* | *preku-ne** | *āyu* | *ayu-ne* |
| 2sg | *prekt** | *? prekt-ne** | *ait* | *aita-ne* |
| 3sg. | *prekäṃ* | *prekäṃ-ne** | *aiṃ* | *? aiṃ-ne** |
| 1pl. | *parkäm* | *parkäm-ne** | *aim* | *?* |
| 2pl. | *parkcer* | *? parkcer-ne** | *aicer** | *?* |
| 3pl. | *parkäṃ* | *parkäṃ-ne** | *aiṃ* | *? aiṃ-ne** |

The subjunctive [1] is usually coupled with a sigmatic preterite [class 3], which was discussed in the previous subsection (5.3.3). Table 5.24 and Table 5.25 give an overview of the verbs with a subjunctive [1] for which the accentuation can be established. The verbs in Table 5.24 have consistent initial accent, while those in Table 5.25 do not.

To my knowledge, no diagnostic forms are attested for the remaining verbs in this subjunctive class, namely *al-* 'keep away', *ewk-* 'grant', *kəl-* 'bear', *kəw-* 'pour', extinguish', *kaw-* 'kill', *pəwtk-* 'close', *plak-* 'agree (act.); ask permission (mid.)', *plətk-* 'emerge', *y-* 'go', *rək-* 'extend (tr.)', *rənk-* 'lift'?, *rəyn-* 'leave, give up', *rəw-* 'open', *lənk-* 'hang up', *wotk-* 'separate, decide', *wəs-* 'wear', *səyn-* 'rest on', *səyn-* 'be satiated, depressed', *smәy-* 'smile', *tˢənk-* 'flay'.

Table 5.24: Subjunctives [1] with first-syllable accent.

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *ar-* | 'abandon' | sbj.act.3sg.-2sg. *oräñ-c* THT 46 a8 | \|órə-\| |
| *er-* | 'cause' | sbj.vn. *erälyñe* THT 1363.d a3 | \|érə-\| |
| *kaw-* | 'kill' | sbj.vn. *kāwälñe* THT 540 a2 | \|káwə-\| |
| *kłənk-* | 'doubt' | sbj.vn. *klaṅkälyñe* PK AS 6D b3 | \|klánkə-\| |
| *tək-* | 'touch' | sbj.act.3sg.-pl. *tekä(ṃ)-me* IOL Toch 23 b2<br>sbj.vn. *takälñe* IOL Toch 998 b3 | \|tékə-\|<br>\|tákə-\| |
| *tənk-* | 'stop' | sbj.act.3sg.-2sg. *teṅkäñ-c* THT 1113 a2-3 | \|ténkə-\| |
| *nak-* | 'blame' | sbj.vn. *nākälyñe* SI B 117 b2 | \|nákə-\| |
| *pyak-* | 'smash' | opt.act.1sg.-pl. *pyāśim-me* IOL Toch 82 a3<br>opt.act.3sg.-3sg. *pyāśi-ne* IOL 127 a4<br>sbj.ger. *pyākäle* PS BI 13 b2 | \|pyákə-\| |

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *prək-* | 'ask' | sbj.act.3sg.-3sg. *prekäṃ-ne* PK AS 8C b6<br>sbj.act.3pl.-pl. *parkän-me* THT 7 a2<br>opt.act.2sg. *parśit* THT 291.a a2 | \|prékə-\|<br>\|prə́kə-\| |
| *pləw-* | 'float' | sbj.act.3sg. *plyewä-ñ* PK AS 17C a5<br>sbj.vn. *pluwälyñe* IOL Toch 30 b5 | \|pléwə-\|<br>\|pləwə-\| |
| *yəp-* | 'enter' | sbj.act.3sg.-3sg. *yopäṃ-ne* THT 41 b3<br>opt.act.2sg. *yapit* THT 246 a2<br>sbj.vn. *yapälñe* PK NS 54 a5 | \|yópə-\|<br>\|yə́pə-\| |
| *tˢəwk-* | 'suckle' | sbj.ger.abl. *tsukälemeṃ* THT 139.d a3 | \|tˢə́wkə-\| |

Table 5.25: Subjunctives [1] without first-syllable accent.

| Root | Meaning | Examples of diagnostic forms | Analysis |
|---|---|---|---|
| *ay-* | 'give' | sbj.act.1sg.-3sg. *ayu-ne* THT 25 a7 | \|ayə́-\| |
| *awn-* | 'hit; begin' | sbj.mid.3pl. *aunantär* THT 27 a7 | \|awnə́-\| |
| *eṅk-* | 'take' | sbj.mid.3pl. *eṅkanträ* THT 37 a7<br>sbj.vn. *eṅkalñe* IOL Toch 133 b4 | \|enkə́-\| |
| *trənk-* | 'lament' | sbj.vn. *träṅkalyñe* THT 89 b4 | \|trənkə́-\| |
| *nes-*[(a)] | 'be' | prs-sbj.ger. *nesalle* | \|nesə́-\| |
| *pəkʷ-* | 'trust' | sbj.vn. *pkwalñe* THT 428 b8 | \|pəkʷə́-\| |
| *yam-* | 'do' | sbj.mid.1pl. *yamamtär* THT 404 b8<br>sbj.mid.3pl. *yamantär* THT 21 a2<br>sbj.ger. *yamalle* THT 575 b6<br>sbj.vn. *yamalñe* THT 77 a4 | \|yamə́-\| |
| *yok-* | 'drink' | sbj.vn. (*yo*)*kalyñeme*(*ṃ*) THT 1103 b1 | \|yokə́-\| |
| *tˢərk-* | 'burn' | sbj.ger. *tsärkalle* IOL Toch 305 b5 | \|tˢərkə́-\| |

(a) The finite subjunctive of this verb is provided by suppletive *taka-*, but *nesalle* functions both as the present and the subjunctive gerundive (see, e.g., Malzahn 2010: 691).

The distribution is quite balanced, with 12 subjunctive [1] verbs with stable initial accent (Table 5.24), and 9 without (Table 5.25). In all cases where the subjunctive [1] has initial accent, the attested asigmatic preterite [3] stem is identical to the subjunctive stem except for the accentuation; e.g., *prək-* 'ask' has prt.[3] |prekə́-| and sbj.[1] |prékə-|. Only *klənk-* 'doubt' and *tənk-* 'stop' have no preterite attested. In the group of subjunctives [1] with non-initial accent, only *awn-* 'hit; begin' has both active and middle forms in all tenses. Since *enk-* 'take' is a middle-only verb, the preterite stem is always sigmatic |énksa-|, and thus differentiated from the subjunctive. The preterite stem of *ay-* 'give' is

suppletive *wəsa-* and *yok-* 'drink' has an irregular present-preterite stem *yaśə-*/*yake-*; *yam-* 'do' has an *a*-preterite [1] *yamáṣṣa-*, based on the present stem, and *trənk-* 'lament' also seems to have had an *a*-preterite [1] *trənca-*. The preterites of *pəkʷ-* 'trust' and *tˢərk-* 'burn' are not attested.

None of the forms of the subjunctive paradigm overlap with forms in the preterite paradigm in all respects except accentuation, in the way that sbj.[5] 1pl. *tákam* would be the same as the prt.[1] 1pl. *takám* if it did not have initial accent, cf. Table 5.26. Nevertheless, the subjunctive type with initial stress expanded to this category, either to differentiate the stems of the subjunctives [1] and the (asigmatic) preterite [3] themselves despite a lack of direct competition—similar to the case of middle-only *a*-subjunctives [class 5] like *wərpa-* 'enjoy', where none of the middle endings caused a merger—or due to pressure from the great number of *a*-subjunctives with initial accent. Both factors may have been active at the same time, as illustrated in Table 5.26.

Table 5.26: Analogical pressures on the subjunctive [1] paradigm.

| | sbj.[1] expected | regularized >> \|pré/ə́kə-\| | prt.[3] \|prekə́-\| | sbj.[5] \|táka-\| | sbj.[5] \|ká/ə́tka-\| |
|---|---|---|---|---|---|
| 1sg. | **prékəw* | *prékəw* | *prekə́wa* | *tákaw* | *kátkaw* |
| 2sg | **préktə* | *prékt* | *prekə́sta** | *tákat* | *kátkat* |
| 3sg. | **prékən* | *prékən* | (*préksa*) | *tákan* | *kátkan* |
| 1pl. | **pərkə́m* | >> *párkəm* | *prekə́m** | *tákam* | *kátkam* |
| 2pl. | **párkcer* | *párkcer* | *prekə́s* | *tákacer* | *kátkacer* |
| 3pl. | **párkən* | *párkən* | *prekə́r* | *tákan* | *kátkan* |

Peyrot found an initial spark for the creation of initially-accented subjunctives in the paradigms of the type of subjunctive [5] with *a*-grade throughout the subjunctive paradigm and an accompanying identical preterite [1] stem (Table 5.18) above. From there, the drive for a differentiation between the subjunctive [5] stem and the preterite [1] stem in this type of verbs appears to have expanded and affected almost all verbs with identical subjunctives and preterite stems, both with and without final *-a*. The sound development that lies at the basis of this scenario is the progressive accent shift that is also in accordance with syncope and epenthesis as discussed in 5.2.

In Jasanoff's explanation of the accent, the stable first-syllable accentuation of full-grade forms of subjunctive [1] like act.3sg-3sg. *prékäṃ-ne* 'will ask', *tékäṃ-ne* 'will touch', *yópäṃ*-ne 'will enter' are regular following the "*yāmu*-rule". Initially accented zero-grade forms such as act.3sg-pl. *párkäṃ-me* 'will ask', on the other hand, would be analogical to the initial accent of the full-grade forms (Jasanoff 2015: 90–91).

The non-ablauting full-vowel roots *awn-* 'hit; begin', *nes-* 'be', *yam-* 'do', *yok-* 'drink', *ay-* 'give' and *eṅk-* 'seize' have the accent on the second syllable /ə/ instead of on the root vowel, contrary to what would be expected according to Jasanoff's *yāmu*-rule. Jasanoff assumes that *awn-* and *yok-* derive from athematic presents [class 1] and got their accentuation analogically from there (Jasanoff 2015: 91 fn. 9), although it remains unclear to me how it should have arisen even in that case within Jasanoff's framework. The accent of *yam-*, *ay-* and *eṅk-* would be analogous to their presents as well. Jasanoff assumes that the non-initial accent in subjunctive [1] forms of *yam-* 'do' is due to an analogical shift in the accentuation of the present [9] *yamə́sk-* from *yáməsk-* (expected according to Jasanoff's "*yāmu*-rule"), based on a generalization of non-initial accentuation in non-causative presents of this class [9] (Jasanoff 2015: 91 fn. 9). The motivations for this analogy are not enirely clear to me, though perhaps implied is a secondary differentiation from the causative presents of class [9].[62] All in all, Jasanoff's "*yāmu*-rule" remains inadequate also in the subjunctive class [1].

According to Winter, the initial accent found in the ablauting subjunctives of classes [1] and [5] is due to the earlier presence of a syllable before the stem: a lost reduplication syllable. The derivation of this category from the Indo-European perfect based on the ablaut TB *e* : *ə*, which could reflect the ablaut pattern of the PIE perfect **o* : **∅*. The perfect was reduplicated in Proto-Indo-European: **Ce-CoC-* : **Ce-CC-*. An initial reduplication syllable **Cə-* would have regularly disappeared by syncopation and simplification of the resulting geminate, i.e., **Cə-Cé-* > **CCé-* > *Cé-*, thus leaving a trace in Tocharian B only as fixed first-syllable accent (Winter 1994b: 305–308).

Malzahn follows Winter's connection with the perfect, but derives the subjunctives with initial accents from the perfect by addition of the present endings, rather than as continuing the perfect themselves in some way (Malzahn 2010: 309–310). While the preterite reduplication was later given up by analogy to non-reduplicated preterites, according to Malzahn, it was retained in the subjunctives, eventually leading to the first-

[62] There are eight other non-causative presents [9] with a full vowel in the root and accent on the second syllable /ə/: |yawkə́$^{§§ə}$/$_{ske}$-| to *yawkk-* 'use', |sompə́$^{§§ə}$/$_{ske}$-| to *samp-* 'take away', |spalkə́$^{§§ə}$/$_{ske}$-| to *spalk-* 'make an effort', |aksə́$^{§§ə}$/$_{ske}$-| to *aks-* 'announce', |arskə́$^{§§ə}$/$_{ske}$-| to *arsk-* 'finish', |yaskə́$^{§§ə}$/$_{ske}$-| to *yask-* 'beg', |aklə́$^{§§ə}$/$_{ske}$-| to *akl-* 'learn', |twasə́$^{§§ə}$/$_{ske}$-| to *twas-*'shine', all of which would also have their accent analogically changed, according to Jasanoff (2015: 93 fn. 15). |ásə$^{§§ə}$/$_{ske}$-| to *as-* 'fetch' appears to show initial accent per Jasanoff's "*yāmu*-rule", but Malzahn (2010: 533) reports that the only attested present form, *āṣṣäṃ*, is found in a metrical text, which means that it may rather reflect |asə́-ṣṣə-n|. At least the following seven presents [9] with root-final vowel *-a* would have regular second-syllable accent according to Jasanoff's "*yāmu*-rule": *aksa-* 'wake up', *nana-* 'recognize', *məlla-* 'oppress', *mləwta-* 'pluck (?)', *yənm*(*a*)- 'get', *pək‹n›a-* 'intend', *yək‹n›a-* 'be careless'. The accent of *ə*-internal |rəynə́$^{§§ə}$/$_{ske}$-| to *rəyn-* 'give up' and |wəsə́$^{§§ə}$/$_{ske}$-| to *wəs-* 'dwell' would also be regular.

syllable accent found there. Since some subjunctives [1] and [5] do not show initial accent, Malzahn assumes that there was reduplication "at least some of these different kinds of paradigms, or maybe in just one of the morphological categories involved" (Malzahn 2010: 310). The subjunctives that do not have initial accent would have lost their reduplication syllable early on due to their root-internal *ə*-vowel. This follows the reduplication pattern in preterite participles, where *ə*-roots show no reduplication while full-vowel roots do (Malzahn 2010: 310–312). Since the derivation of the subjunctive from the perfect is not compelling, nor is any trace of reduplication found besides the first-syllable accent, Peyrot has pointed out that this explanation is circular (Peyrot 2013: 513).

Peyrot's explanation aligns most closely with the data, since in those cases where the subjunctive does not have initial accent, the preterite is differentiated by some other means: either it is morphologically active next to a middle subjunctive, or it has a different stem formation, as discussed by Peyrot (2013: 513–514) and listed above. These apparent exceptions become understandable, if the type of analogy assumed by Peyrot was indeed the driving force behind the creation of the subjunctive types with fixed first-syllable accent. The accent would thus have become just one of the many ways in which stems are distinguished from one another in Tocharian B, as is also its synchronic function (Peyrot 2013: 511–512).

### 5.3.5 The accent in the causative present-subjunctive [9] and preterite [2]

The Tocharian B presents of class [9] are formed with the suffix *-sk-* and come in two types: causative and non-causative. The causative type has a stable first-syllable accent, even though all forms are longer than two syllables; the non-causative has the usual second-syllable accentuation. The causative present [9] furthermore functions as the subjunctive as well: it is a present-subjunctive. It is combined with a preterite of class [2], which also has a stable first-syllable accent. The causative preterite [2] is furthermore characterized by a stable *a*-grade and initial palatalization. These paradigms are illustrated in Table 5.27. In general, on the *sk*-present [9], see Krause & Thomas (1960: 209–213) and Malzahn (2010: 473–495). On the preterite [2], see Krause & Thomas (1960: 244–247) and Malzahn (2010: 170–189).

Table 5.27: Examples paradigms of a non-causative present [9] with accompanying preterite, contrasted with the causative present-subjunctive [9] and preterite [2].

| | *yam-* 'do' (non-causative) | | *wətk-* 'command' (causative) | |
|---|---|---|---|---|
| | prs.[9] | prt.[1] | prs-sbj.[9] | prt.[2] (3sg. + suffix) |
| 1sg. | *yamaskau* | *yamaṣṣawa** | *watkäskau* | *yātkawa* |
| | /yaməskaw/ | /yaməṣṣawa/ | /wətkəskaw/ | /yátkawa/ |
| 2sg | *yamast* | *yamaṣṣasta* | *watkäst* | *yātkasta* |
| | /yaməst/ | /yaməṣṣasta/ | /wətkəst/ | /yátkasta/ |
| 3sg. | *yamaṣṣän* | *yamaṣṣa* | *watkäṣṣäṃ* | *yātka, yātka-ne** |
| | /yaməṣṣən/ | /yaməṣṣa/ | /wətkəṣṣən/ | /yátka, yátkane/ |
| 1pl. | *yamaskem** | *yamaṣṣam** | *watkäskem** | *yātkam* |
| 2pl. | *yamaścer** | *yamaṣṣas** | *watkäścer-* | *yātkas* |
| 3pl. | *yamaskeṃ* | *yamaṣṣare* | *watkäskeṃ** | *yātkare* |

The preterite [2] of Tocharian B corresponds to a reduplicated preterite [2] in Tocharian A, e.g., *wotka-* from pre-TA **ẃaẃätkā-* corresponding to TB *yátka* in Table 5.27 above. The exact correspondence between the reduplicated forms of the Tocharian A preterite [2] and the unreduplicated forms of the Tocharian B preterite [2] is complicated, but due to their equivalence in the morphological system of the two languages, it is preferable to derive them from a common source, which was originally reduplicated as in Tocharian A (Malzahn 2010: 187–188). The first-syllable accent in Tocharian B can thus be attributed to a loss of the reduplication syllable in this type (Peyrot 2013: 490–491). Various accounts of the precise development of the Tocharian B preterite [2], with regards to its palatalization and how it lost its reduplication, can be found in Kim (2003; 2009: 38–41) and Malzahn (2010: 188–189).

The causative present-subjunctive [9] is not reduplicated in either Tocharian A or B, but it does have the same first-syllable accent as the corresponding preterite [2]. According to Malzahn, the word-initial accent in the causative present-subjunctives cannot be original, and "must have been transferred from another, more archaic type of present stem formation, [...]" (Malzahn 2010: 315–316). In particular, the source of the initial accent should be a type formed by adding present endings to reduplicated preterites. Peyrot (2013: 490–491) reconstructs a Proto-Tocharian stem shape **Cə-CəC-* for the causative present-subjunctives. The first syllable would regularly be lost by syncope in both Tocharian A and B, and because it was not a preterite formation, no analogical **Ca-*reduplication was (re-)introduced in Tocharian A.

Another notable feature of the present-subjunctives [9] is that they often have a *-ə-* before the *sk*-suffix, even if the root normally would be *a*-final; e.g., the causative present-subjunctive *wátkəsk-* derives the *a*-final non-causative *wətka-* 'be separated, decided'.

Since there is synchronically no way to turn the root-final *-a-* of the preterite stem (e.g., *yátka-*) into the *-ə-* found in the present-subjunctive (*wə́tkəsk-*), both have to be analysed as suffixes in this type of verbs (Peyrot 2013: 148–149). However, there are also several attestations of present-subjunctive [9] forms that do show a root final *-a-*, which must be considered an archaism (Malzahn 2010: 458–459; Malzahn 2011; see also Peyrot 2013: 488–489).

To explain the disappearance of expected root-final *-a* in the causative present-subjunctives [9] with the *sk*-suffix Peyrot departs from the attested Tocharian B initial accentuation. Based on the findings of Malzahn (2011), he derives the overwhelming presence of *-əsk-* forms derived from roots with original final *-a-* by way of a sound law that weakened *-a* whenever it followed the accent and stood before a full vowel in the ending. Thus, a form like **ánasken* of **ánask-* 'make breathe' would have regularly developed to **ánəsken*. This accounts for the fact that the suffix form **-aske-* is less common than **-aṣṣə-* in the data presented by Malzahn (Peyrot 2013: 488–489).

Jasanoff has an entirely different account for the first-syllable accent of both the causative present-subjunctive [9] and the preterite [2]. Instead, his "*yāmu*-rule" in combination with an analogical crossing of two different verb types should explain the Tocharian B situation, while the Tocharian A reduplicated preterite [2] would be a wholly different formation. Jasanoff derives the causative preterite [2] of Tocharian B from original lengthened grade forms, e.g., act.3sg. *śārsa* 'let know' from an earlier **kērsH-* (Jasanoff 2015: 94). Jasanoff's account for the initial accent in these forms is the same as that for the subjunctive [5]: a cross with an original *a*-less type. In this way, the regular accentuation of a root **prénkə-* 'reject', according to Jasanoff's "*yāmu*-rule", would have affected expected **kʸersá-* to undergo an analogical change to **kʸérsa-*, before the former was itself modified to **prénka-* with final *-a* on the basis of the latter.

| | initial | "*yāmu*-rule" | analogy 1 | analogy 2 | *a*-umlaut |
|---|---|---|---|---|---|
| *a*-final | *śersa-* | *śersá-* | *śersá-* | >> *śérsa-* | *śársa-* |
| *ə*-final | *prénkə-* | *prénkə* | >> *prénka-* | *prénka-* | *pránka-* |

Like in Jasanoff's account of the subjunctive [5], the necessary developments are not supported by the Tocharian data and rely on an unverifiable timing of the extension of final *-a-* and accent shift. The two steps of analogy assumed only seem to be motivated by the framework of Jasanoff's "*yāmu*-rule", and are not further explained, but since Jasanoff's "*yāmu*-rule itself is suspect (cf. the discussions in the previous subsections), this interpretation of the preterite [2] cannot be upheld. Furthermore, as mentioned above, the separation of the Tocharian A and B preterite [2] formations as necessitated by Jasanoff's scenario is undesirable based on the functional identity of the Tocharian A and B preterite [2] within the verbal system; this is argued in more detail by both

Malzahn (2010: 187–188) and Peyrot (2013: 490). An interpretation of the Tocharian B preterite [2] as a dereduplicated counterpart to the Tocharian A preterite [2] is thus preferable as a starting point, and this can also explain its first-syllable accent.

In Jasanoff's view, the initial accentuation of the causatives with the suffix *-əsk-* can also be explained with his "*yāmu*-rule" for those verbs that originally had no root-final *-a-*. When the suffix was generalized as *-əsk-*, even forms with original final *-a-* would have regularly acquired initial accent as well. The accentuation of zero-grade roots would be analogical (Jasanoff 2015: 93). The original situation in the present-subjunctive [9] is difficult to reconstruct with certainty, the existence of relic forms with *-ask-* (Malzahn 2011) show that the prevalence of *-əsk-* in the causative is a relatively more recent phenomenon, something which does not accord with Jasanoff's "*yāmu*-rule".

The causative present-subjunctive that accompanies the preterite [2] does not show any reduplication in either language, so that the presence of earlier present-subjunctive forms like **wəwátkəsk-*, whence TB caus. *wátkəsk-* 'command', cannot be confirmed and some caution is advised. However, a derivation of the Tocharian B initial accent in causatives from lost reduplication syllable remains the most likely account for the initial accent of both preterite [2] and present-subjunctive [9] causatives, since earlier reduplication is confirmed at least for one of these two classes, and they are very closely connected within the Tocharian verbal system. The accentuation of the causatives can thus be interpreted within the framework of a second-syllable accentuation deriving diachronically from a rightward accent shift without any further analogical accentual alterations.

### 5.3.6 The phonologization and morphologization of the accent

In the categories treated in the last few subsections, it is assumed that the Tocharian accent has developed from a phonologically regular starting point to become increasingly irregular and morphophonologically specified. The dominant pattern is that of first-syllable accent in disyllabic words and second-syllable accent in longer words, although several categories deviate from this norm in more or less systematic ways, often in combination with an analogical regularization in the accentuation of a paradigm. The question of why any analogical regularizations happened in the first place has not yet been adequately addressed. In this subsection, I will provide a short overview of likely causes.

I assume that the fixed first-syllable accent found in the middle presents [2] and [8], the great majority of sigmatic forms of the *s*-preterite [3], and a large number of subjunctives [1] and [5] resulted from analogical morphophonological rules that either disallowed accented middle endings (5.3.2), took the unsuffixed third person singular as the basis for the paradigm (5.3.3), or promoted the differentiation of subjunctive and preterite stem forms (5.3.4). None of these changes make sense from the simple starting

point of an original, regular phonological accentuation rule: first-syllable accent in disyllabic words, second-syllable accent in polysyllabic words. In principle, it should have been perfectly possible to maintain such a rule as long as the placement of the accent was phonologically regular. However, there were at least two categories that made a continued purely phonological status of the accent impossible: the causatives and, perhaps less significantly, Iranian loanwords.

In the causatives, it is assumed that the initial accent observed in Tocharian B was the result of a regular deletion of the original first syllable due to syncope (5.3.5). This is particularly likely to have affected the causative preterite [class 2]. Once this deletion had taken place, the accent of the causatives could not be explained by the general phonological rule anymore, and needed to be specified otherwise, i.e., morphophonologically. After this development the original phonologically regular accent assignment had become inadequate for this one verbal category, so that morphologization in other categories would have become possible, or even necessary, as well. Phonologically irregular initial accent was introduced in a minor way by loanwords from Iranian as well, as discussed further below in 5.4.6.

I think that this change would have opened the doors for much more far-reaching accentual developments and analogical morphophonological regularizations in a number of categories. The detachment of the accent from the phonological realm could allow for a differentiation between the stems of the subjunctive [5] and the preterite [1] and inhibit accentual movements in sigmatic forms of the *s*-preterite [3] on the basis of the initially-accented disyllabic third person singular forms like *preksa*.

## 5.4 The accent in other categories

Aside from the verbal categories with initial accent treated in the previous subsection, there are various other topics and categories in Tocharian B that should be briefly addressed due to their aberrant accentual behaviour or their involvement in ideas on the diachronic development of the accent. These are vowel contractions (5.4.1), nasal prefixes (5.4.2), kinship terms (5.4.3), secondary case forms (5.4.4), compounds (5.4.5), and, importantly, loanwords (5.4.6).

### 5.4.1 Vowel contractions and the accent in the preterite participle

The position of the accent was important for the contraction of original sequences such as **əwe* and **awe* to simple vowels in Tocharian B. In both cases, we observe a split development into two different reflexes, which can be predicted by the position of the respective sequences in the word. These types of contraction could thus provide additional information on the chronology and the diachronic development of the Tocharian accent.

**Developments of pre-TB **əwe***

The sequence **-əwe-* shows a split development in Tocharian B, yielding both TB *-o-* and *-uwe-*. Both outcomes of **-əwe-* are commonly found in two classes of preterite participles, one type represented by *kekamu* with the m.obl.sg. *kekamoṣ*, and the other by *ltu* with the m.obl.sg. *ltuweṣ*; see Table 5.28. The developments of the ending of the different classes of preterite participles has been analysed by Þórhallsdóttir (1988) as the result of various contractions across the glide **-w-*, which was part of the participle suffix.

Table 5.28: Developments of **əwe* in Tocharian B.

| development | examples | meaning |
|---|---|---|
| **əwe* > TB *o* | **pəweńə* > TB *poñ* | imp. of *weñ-* 'say' |
| | **kekəməweṣə* > TB *kekamoṣ* | prt.ptc.m.obl.sg. of *kəm-* 'come' |
| **əwe* > TB *uwe* | **təwe* > TB *tuwe* | pers.pron.2sg. 'you (sg.)' |
| | **lətəweṣə* > TB *ltuweṣ* | prt.ptc.m.obl.sg. of *lət-* 'go out' |

The examples in Table 5.28 show that this sequence developed into **o* when **əwe* made up the original first two syllables of a multisyllabic word or the third and fourth syllable. In disyllabic words and in the second to third syllable, **əwe* was retained and became TB *uwe*. Thus, while **kekáməweṣə* became *kekamoṣ* with contraction, **lətáweṣə* became *ltuweṣ*. The latter development is identical to that of the 2sg. pronoun *tuwe* from **táwe*. The important example *poñ* from **pə-wéńə*, adduced in this context by Peyrot (2010: 72), shows the same development as *kekamoṣ* from *kekáməweṣə* and indicates that **-əwe-* and **-əwé-* developed in the same way. This can be easily understood if we assume a secondary stress every other syllable after the main stress, thus **kekáməwèṣə*. It is clear that the primary accent must have already been on second syllable wherever possible at the time that the contraction from **-əwe-* to **-o-* took place.

Þórhallsdóttir (1988) first argued that the *o* in preterite participles of the type *kekamoṣ* was the result of a contraction from *ˊCəweC* to *ˊCoC*, next to a lack of contraction in *CáweC* to account for the type exhibited by *ltuweṣ* (Þórhallsdóttir 1988: 190). Peyrot (2010) interprets the development from **əwe* to TB *o* differently, assuming syncope with a subsequent shift from **CweC* to **CoC*. Thus, **kekáməweṣə* would first have become **kekámweṣ*, and from there *kekámoṣ* (= TB *kekamoṣ*). Peyrot (2010: 72) considers the imperative TB *poñ* 'say!' *poñ*, via **pweń* from **pə-wéńńə*, to be in support of this development. The change from **CweC* to **CoC* needs to be specified to a closed syllable in order to account for the lack of the same development in examples like *tweye* 'dust' and *twere* 'door'.

There are a few counterexamples, however. For TB *śtwer* 'four (4)', Peyrot assumes that the *-we-* can have been introduced analogically (2010: 72 fn. 6). Since *śtwer* derives

from PIE **kʷetwores*, there should have been a final *-ə*, i.e., **śətwerə*. If **CweC* became *CoC* in closed syllables before the late apocope of *-ə*, TB *śtwer* would naturally have remained unaffected. However, in that case, *poñ* from **pəweńə* should not have undergone this change either. Additionally, analogical restoration would be needed in *malkwer* /málkwer/ 'milk' and *ārwer* /árwer/ 'ready' (ibid.). Peyrot assumes that these words derive from earlier **məlkəwer* and **arəwer* respectively, but then the initial accent is unexplained: such forms should have yielded classical TB ***mälkuwer*, ***aruwer*. Thus, *malkwer* and *ārwer*, in terms of their accentual structure, must derive directly from disyllabic PT **məlkwer* and **arwer*, contradicting a sound change from **CweC* to *CoC*.

In favour of a development from **CweC* to *CoC* is the verb 'take away', with in Tocharian B prs.[9] stem /sompə́-/ next to sbj.[5] /sámpa-/ and prt.[1] /sampá-/, and in Tocharian A prs.[6] *sumnā-* and sbj.[5] *sumā-*. The *o* in the Tocharian B present can be understood if it derives from **swempə-*, and the gemination in the preterite participle *sassāmpaṣ* can be understood with the Tocharian B change from *-sw-* to *-ss-*. On the reconstruction of this verb in Proto-Tocharian and the relationship between the forms in Tocharian A and B, see Peyrot (2013: 829 incl. fn. 918).

The development of the adjectives in TB *-ssu* (e.g., TB *śaulassu* 'life-possessing') also deserves to be mentioned in this context, with their m.obl.sg. in *-ssont*. The geminate *-ss-* implies an earlier cluster **-sw-*, so that the **w* appears to be needed twice, both for the geminate and for the vowel *-o-* in *-ssont*. The developments can be imagined in two ways, according to our conception of the development from **əwe* to **o* as a contraction or as a change from **CweC* to *CoC*. Both scenarios are illustrated in Table 5.29.

Table 5.29: Two potential developments of the adjective suffix *-ssu*.

| nom. : obl. | → | → | → | TB |
|---|---|---|---|---|
| *-su* : *-səwentə* | *-su* : *-sontə* | *-su* : *-s*[*w*]*ontə* | *-su* : *-ssontə* | *-s*[*s*]*u* : *-ssont* |
| *-su* : *-səwentə* | *-su* : *-swentə* | *-su* : *-ssentə* | *-s*[*s*]*u* : *ss*[*w*]*entə* | *-ssu* : *-ssont* |

The relative dating of the assimilation of **-sw-* to *-ss-* and the development from **əwe* to *o* is different in these two scenarios. Both require an equal number of analogical changes, so that neither is more parsimonious than the other. On the basis of the development in *malkwer*, *ārwer*, and *śtwer*, the contraction set up by Þórhallsdóttir (1988) seems to be more appropriate. The change from **CweC* to **CoC* in the present TB *sompə́-* 'take away' from **swempə-* would then seem to be a separate development. Alternatively, the sequence **-wer* could potentially be excluded from the theoretical development from **CweC* to **CoC*, although I cannot yet see a further motivation for this.

However, the details should be interpreted, the development of the sequence **-əwe-* to TB *o* can be linked to the place of the accent on the **e*, thus **-əwé-*. When the accent

was on the first vowel of the sequence, **-áwe-*, the result is uncontracted TB *-uwe-* instead. Since original **-əwe-* in oblique preterite participles of the type *kekamoṣ* show the development associated with the accentuation **-əwé-*, rather than **-áwe*, while the type *ltuweṣ* shows the development associated with the accentuation **-áwe-*, it is clear that the relevant changes took place when the normal Tocharian B accentuation had already become established.

**Developments of pre-TB **awe* and **a.e***

The next type of vowel contraction is found in the two remaining classes of preterite participles, represented by *kälpau* with the oblique *kälpoṣ* and *papaikau* with the oblique *papaikaṣ*; see Table 5.30. In this case, the place of the accent at the time that the sequence **-awe-* contracted is not so easy to determine, even though it is clear that two different developments took place.

Table 5.30: Developments of **awe* and **a.e* in Tocharian B.

| development | examples | meaning |
|---|---|---|
| **awe* > TB *o* | **kəlpaweṣə* > TB *kälpoṣ* | prt.ptc. of *kəlpa-* 'obtain' |
| | **kʷəryawer* > TB *karyor* | 'trade, commerce' |
| **a.e* > TB *o* | **ana.elme* > TB *onolme* | 'living being' |
| **awe* > TB *a* | **śawemane* > TB *śāmane* | 'living' |
| | **papaykaweṣə* > TB *papaikaṣ* | prt.ptc. of *pəyk-* 'paint, write' |

The difference between these two types of preterite participles formed to *a*-final roots has previously been regarded as the result of one being originally **å*-final instead, from PIE **eh₂* (Þórhallsdóttir 1988: 190ff; Pinault 2008: 598–599). This would theoretically account for the difference in the type *kälpoṣ* from **kəlpåweṣə* when considered alongside *papaikaṣ* from **papaykaweṣə* without reference to the accent. In the finite forms, the **å*-final type would have merged with the *a*-final type, leaving the preterite participle as the only remaining trace (Þórhallsdóttir 1988: 190ff; Pinault 2008: 598–599). However, Saito (2006) connected the contractions in the various classes of preterite participles with the place of the accent instead, and this hypothesis is elaborated by Peyrot (2010), with further reference to words beside the preterite participle formations. This interpretation provides a more appropriate explanation than the assumption of a different second-syllable vowel that is not otherwise observable.

The distribution between the two reflexes TB *o* and *a* can thus be understood with natural differences in accentuation based on the length of the forms, but it is difficult to date the developments relative to the establishment of normal Tocharian B accentuation or accent shift. This is because there are, as far as I am aware, no examples whose accent

can only be interpreted in one way, i.e., of the type *pəweńə* to *poñ* 'say!' in the case of **-əwe-* above; cf. Table 5.31 with both alternative interpretations.

Table 5.31: Two possible scenarios for the interaction of the accent and contractions of **awe*.

| development if after accent shift | development if before accent shift |
|---|---|
| **áwe* > *o* **kəlpáweṣə* > TB *kälpoṣ* | **awè* > *o* **kálpaweṣə* > TB *kälpoṣ* |
| **awé* > *a* **śawémane* > TB *śāmane* | **áwe* > *a* **śáwemane* > TB *śāmane* |
| **awè* > *a* **papáykaweṣə* > TB *papaikaṣ* | **àwe* > *a* **pápaykàweṣə* > TB *papaikaṣ* |

One reason to posit this contraction before accent shift is the accentuation of TB *karyor* 'trade, commerce', which has a first-syllable accent /kə́ryor/. This is most easily accounted for if it derives from **kʷə́ryor* already at the time of accent shift, so that the contraction should have taken place in **kʷə́ryawer* rather than in **kʷəryáwer*. The latter should have regularly become *käryor* /kəryór/, which is actually attested as a synchronic derivative from the same verbal root. If the contraction took place after accent shift, the initial accent of TB *karyor* must be secondary. Assuming such an analogical change is not altogether too expensive, since it only amounts to a single case.[63]

Peyrot has further argued that the *-w-* did not have any influence in these developments, because similar contractions of *-a.e-* are found without such an intervening glide. For instance, *ścmoññа* 'basis' can be derived from **ścəma-eńńa* and *onolme* 'being' from **ana-elme*. The fact that development of the sequence **-a.e-* in these words is the same as that of *-awe-* in preterite participles like **kəlpáweṣə* becoming *kälpoṣ* suggests that the **-w-* was of no importance for this particular sound law (Peyrot 2010: 72).

**Similar contractions in imperatives**

Some other instances of contraction without an intervening glide are found with the imperative prefix **pə-*, which is never accented itself (Winter 1977; Pinault 2008: 605–606). The accent in the imperative stays on the initial syllable (e.g., *purwar-me*, not ***purwār-me*), which is also in line with the idea that a lost prefixed syllable caused the word-initial accent of certain other verbal categories like the causative preterite [2].

[63] On the basis of twice attested *śamāne* (in THT 88 a5 and in PK DA M 507.37 & .36 a39) besides more common *śāmane*, Winter (1988: 218) considers the contraction to have taken place before the accent was fixed on the second syllable, which would accord with *karyor* 'trade, commerce'. However, it is difficult to exclude that the accent of *śamāne* was analogical, as its placement is according to the general second-syllable accentuation rule. On the other hand, *śāmane* may also be analogical, since the accent regularly falls before the suffix *-mane* in most verbs (e.g., *akemane* /akémane/ from *ak-* 'lead').

Table 5.32: Contractions in the imperative of vowel-initial verbs.

| development | examples | meaning |
|---|---|---|
| **ə.a* > TB *o* | **pə-akse* > TB *pokse* | imp. of *aks-* 'announce' |
| **ə.e* > TB *e* | **pə-ersar* > TB *persar* | imp.2sg. of *er-* 'cause' |
| | **pə-enksat* > TB *peṅksat* | imp.2pl. of *enk-* 'take' |

The main element of interest is the imperative *pokse* from *aks-* 'announce', where the initial *a* is changed to an *o*. Both verbs with initial *e* keep this *e* in the imperative appear to indicate that **ə.é* and **əwé* underwent separate developments, unlike the sequences **awe* and **a.e* discussed above. However, it is of course very much possible that the imperatives *persar* and *peṅksat* had their vowel restored by analogy, so that they are not particularly reliable and no firm conclusions can be drawn.

**Contraction in the first syllable of preterite participles**

Further simplifications occurred in sequences like **eyə*, **ewə́*, and **ewə*, where the second vowel was a schwa. These all became simple diphthongs, as illustrated in Table 5.33.

Table 5.33: Developments of the sequences **eyə*, **ewə́*, and **ewə* in Tocharian B.

| development | examples | meaning |
|---|---|---|
| **eyə* > TB *ai* | **yeyətəwə* > TB *yaitu* | prt.ptc.m.nom.sg. of *yät-* 'decorate' |
| | **yeyəpəweṣə* > TB *yaipoṣ* | prt.ptc.m.obl.sg. of *yəp-* 'enter' |
| **ewə́* > TB *ai* | **wewə́ru* > TB *yairu* | prt.ptc.m.nom.sg. of *wər-* 'practise |
| **ewə* > TB *au* | **wewəsəwə* > TB *ausu* | prt.ptc.m.nom.sg. of *wəs-* 'wear' |

The development from **yeyə́pəweṣə* to *yaipoṣ* is sometimes formulated as the result of a special syncope rule of accented /ə́/ following a glide (e.g., Ringe 1989: 37). I think it could rather be seen as yet another type of contraction, whereby **-eyə́-* developed via **e.í* to **éi* and eventual TB *ái*. Thus, two contractions would have occurred in **yeyə́pəweṣə* to yield TB *yaipoṣ*: **-eyə́-* to **-éi-* and **-əwe-* to **-o-*. The other examples are the preterite participles *yaitu*, *yainmu*, *yairu*, *yailu*, *ausu*, *aultsu**, and *auṣu* (the last three with a development along the lines of **wewə-* > **wowə-* > **ow-* > **aw-*; cf. Ringe 1989). Unless the stem was analogically restored from the masculine nominative **pernewə*, the development of *pernauntsa* from **pernewəntˢa* implies that this development was independent of the accent, since **pernéwəntˢa* with **éwə* behaves like **wewə́səwə* with **ewə́*.

The preterite participles of this type thus have initial accent, but are no longer reduplicated in the usual sense as the result of a development from disyllabic sequences to monosyllabic diphthongs. There is an additional group of preterite participles that also have initial accent, but do not show any trace of reduplication. The flagship example in Jasanoff's account of the Tocharian accent is of course TB *yāmu*, the preterite participle of *yam-* 'do', from which the name of Jasanoff's "*yāmu*-rule" derives. With a reconstruction of *yāmu* as deriving directly from **yaməwə*, an inhibited accent shift from **a* to **ə* would theoretically account for this accentuation. Especially when considered next to the participle type of *ltu* from **lətáwə*, with shift from **látəwə*, such an analysis seems appropriate on the surface. However, we have already treated a number of issues with Jasanoff's "*yāmu*-rule" in the previous subsections, and the accent of *yāmu* can certainly be understood in different ways as well.

First, it is possible that *yāmu* was reduplicated at an earlier stage, but lost its reduplication through sound development.[64] According to the Tocharian B rules of formation of preterite participles, preterites that have ablaut or a stable full grade form a reduplicated preterite. The verb *yam-* 'do' has a stable *a*-grade in all forms, so that the preterite participle should be reduplicated **yeyaməwə*. According to Peyrot, there is no other way to explain the lack of reduplication in *yāmu*, and also *yāku* from *yok-* 'drink', except by assuming it resulted in some way from a sound change that interacted with the initial *y-* (Peyrot 2013: 186). Winter additionally adduces *yāṣṣu* 'begged food', which has become detached from the verbal paradigm of *yask-* 'beg' and shows the same lack of reduplication as *yāmu* and *yāku* (Winter 1994b: 300–301). Other preterite participles of verbs with initial *y-* that do show the expected reduplication, like *yayātau* from *yata-* 'be able' and *yayāṣṣoṣ* from *yask-* 'beg' to which *yāṣṣu* 'begged food' is related, would then have to be secondary, in accordance with the usual pattern (Winter 1994b: 301).

**The accentuation of vowel-initial preterite participles**

There is another group of participles for which Jasanoff's "*yāmu*-rule" would work well, namely the vowel-initial participles *āklu, -oṣ* from *akl-* 'learn', *ākṣu, -oṣ* from *aks-* 'announce', *aiku, -oṣ* from *ayk-* 'know', *aipu, -oṣ* from *ayp-* 'covered', *ārskoṣ* from *arsk-* 'finish', *ālu* from *al-* 'keep away', *aukṣu* from *awks-* 'grow', *āsu* from *as-* 'dry', *eru* from *er-*, *-oṣ* 'cause', and *eṅku, -oṣ* from *enk-* 'take' (see Peyrot 2013: 720–726).[65] These all have

[64] Jasanoff dismisses this possibility because of the verb *yəya-* 'go along' remained disyllabic, preserving the first syllable *yə-* (Jasanoff 2015: 88 fn. 6). However, this verb was also accented on the first syllable, i.e., sbj.act.3sg. **yáyan* > TB *iyaṃ*, while a theoretical **yəyaməwə* or (more appropriate) **yayaməwə* would have been accented on the second syllable. The comparison is thus uninformative.

[65] The only participle derived from a vowel-initial verb that does not appear to have initial accent, as far as I am aware, is *aunu* from *awn-* 'hit; begin', with m.pl. *aunūweṣ*, based on fragmentary

initial accent, which could theoretically be understood as a lack of shift from a full vowel to a following schwa, e.g., **áypəwə* > *aipu*, **énkəwə* > *eṅku*, etc. An alternative interpretation is that these vowel-initial roots were still reduplicated until after accent shift, so that, e.g., **e-enkəwə* regularly kept the accent on the root (so Peyrot 2010: 81). Alternatively, the accent assignment in these cases may have been analogical, based on an association between the full grade and root-accent in the preterite participle, illustrated in Table 5.34.

Table 5.34: The accentuation of preterite participles in Tocharian B.

| prt.ptc. | phonology | preterite stem | interpretation |
|---|---|---|---|
| *ltu* | /lətə́w/ | \|lə$^{cə}$/$_{te}$-\| | zero grade in prt. + accent on suffix |
| *kälpau* | /kəlpáw/ | \|kəlpá-\| | zero grade in prt. + accent on suffix |
| *peparku* | /pepə́rkəw/ | \|prek$^{∅}$/$_{sa}$-\| | full grade in prt. + accent on root |
| *papaikau* | /papáykaw/ | \|payká-\| | full grade in prt. + accent on root |
| *eṅku* | /énkəw/ | \|énksa-\| | full grade in prt. → accent on root |
| *aipu* | /áypəw/ | \|ayp$^{∅}$/$_{sa}$-\| | full grade in prt. → accent on root |
| *yáku* | /yákəw/ | \|ya$^{śə}$/$_{ke}$-\| | full grade in prt. → accent on root |
| *yāmu* | /yáməw/ | \|yaməśṣa-\|$^{(a)}$ | full grade in prt. → accent on root |

(a) This preterite in Tocharian B is secondary, derived from the present *yaməsk-*, next to a more original the preterite [3] in Tocharian A. The preterite stem at the relevant point in time might thus have been \|yam$^{∅}$/$_{sa}$-\| instead.

Whether *yāmu*, *yāku*, *āklu*, *eṅku*, etc., were still reduplicated or not at the time of accent movement is thus difficult to argue, since they also adhere to the synchronic pattern whereby the accent of the preterite participle is predictable on the basis of the vowel grade in the preterite: if the preterite active has gradation or a stable full grade, the root is accented in the preterite participle. This corresponds closely to the presence or absence of reduplication (Peyrot 2013: 186), of which the accentual pattern is naturally the result. I see no reason why this clear distribution could not have caused an accentual regularization based more on root ablaut than on syllable count, especially since root ablaut appears to have become vitally important in the accentuation of the subjunctives as well (cf. 5.3.4).

*aunū*[*w*]*e*/// in THT1394bb.a3 (Peyrot 2013: 724 fn. 37). This would then belong to the same participle type as *ltu*, *-uweṣ*. If vowel-initial verbs have first-syllable accent due to earlier reduplication, the accent of *aunūweṣ* would remain unexplained, and if the first-syllable accent is due to association with the full grade, it would be the only preterite participle that was not affected. In both scenarios, the aberrant behaviour of *aunu* seems strange, so that the accentual behaviour of *aunū*[*w*]*e*///, if correctly interpreted as *aunūweṣ*, remains unexplained.

**Summary and conclusion**

Summarising the different types of contraction up to this point, it can be said that **əwé* (along with *əwè*) became TB *o*, while **áwe* remained essentially unaffected. These developments took place when the accent had already moved to its eventual Tocharian B position on the second syllable. The contractions of **ewə*, **ewə́* and **eyə* also took place after the general Tocharian B accentual pattern had already become established.

The contractions of **awe* to **a* and **o* are difficult to relate to the position of the accent with certainty. Only the first-syllable accent of TB *karyor* /kə́ryor/ 'trade, commerce' implies that the contraction from **awe* to TB *o* may have happened before accent shift. Whether the various contractions were shared with Tocharian A, and thus dates back to (pre-)Proto-Tocharian, is also difficult to reliably establish, as will be discussed further in 5.5.2 below.

### 5.4.2 Nasal prefixes and the accent

Apart from the imperative prefix **pə-*, there is also the nasal prefix in its many guises *e*(*n*)-, *a*(*n*)-, *o*(*n*)-, which mainly functions as a negative prefix or an intensifying prefix and is also used to form adverbs. Its vowel depends on the vowel in the following syllable: it is basically *e*-, which becomes umlauted to *a*- or *o*- by respectively an *a* or *o* in the following syllable. Although the nasal was always retained before a vowel, it often disappeared before a lexical base starting with a consonant. Before *p*- and *k*-, both allomorphs with and without a nasal are found. Hilmarsson (1991) claimed that the position of the accent in the lexical base played an important role in the preservation or loss of the *n*, so that this phenomenon needs to be briefly addressed.

**The nasal of the prefix**

Hilmarsson (1991) set up the following rule to explain the presence or absence of a nasal in the prefix depending on the accentuation of the base. If the accent fell on the first syllable of the root, the nasal was lost (e.g., TB *ekálymi* < **en-* + *kálʸmi*), while it was retained in those cases where the accent of the lexical base was one syllable removed from the nasal prefix (e.g., TB *emparkre* < **en-* + *pərkáre*).

However, an example like *emparkre* would first require an extra accent shift to the new second syllable (assumed by Hilmarsson 1991: 171), so that the position of the accent in the prefixed form cannot actually be observed directly in the way that Hilmarsson's theory would require. The type of accent shift implied by forms like *emparkre* is not seen elsewhere, and it would entail a retraction from a non-final syllable. As far as can be observed, the nasal formations themselves generally point to the accent being on the first syllable of the base, directly following the prefix, so that no difference in accentuation in these forms can be observed to correlate with the presence or absence of the nasal; cf. Table 5.35.

Table 5.35: Examples of words with the nasal prefixes with and without a preserved nasal. The final column indicates whether the presence of the nasal accords with Hilmarsson's rule.

| Examples | Meaning | Base | Hilmarsson |
|---|---|---|---|
| *apākärtse* | 'manifest, apparent' | *pakartse* /pakə́rtˢe/ '± obvious' | NO |
| *apākri* | 'clearly, evidently'? | *pākre* /pákre/ 'visible, exposed' | YES |
| *aplāc* | 'in conversation' | *plāc* /plácə/ obl.sg. of *plāce* 'talk' | YES |
| *eprete* | 'resolute, steadfast' | *prete* /préte/ '?' | YES |
| *emparkre* | 'wide(ly)' | *pärkare* /pərkə́re/ | NO |
| *empele* | 'terrible, horrible' | *pele* /péle/ 'law' | NO |

Several examples here are as Hilmarsson's rule would predict, but there are also contradictions. There is no indication that, for example, *pele* 'law' and *prete* '?' were accented differently, in such a way as to produce *empele* 'terrible, horrible, awful' with a nasal and *eprete* 'resolute, steadfast' without a nasal. Furthermore, it is not obvious that *emparkre* /empə́rkre/ 'wide(ly)' should be derived from synchronic *pärkare* /pərkə́re/ with second-syllable accent, rather than from **empə́rkre* directly, and the nasal in this word seems to be in opposition to the lack of a nasal in *apākärtse* next to *pakartse* /pakə́rtse/, also with second-syllable accent.

Malzahn (2010: 269-273) also briefly addressed Hilmarsson's conclusions about the presence or absence of the nasal for the privative formations, and noted that that the attested accentuations of other subjunctives do not accord with Hilmarsson's distribution. The assumption that the privatives that do not adhere to the pattern are based on earlier versions of the subjunctive stems with different accentuations is circular. Table 5.36 and Table 5.37 show privatives with and without a nasal in the prefix together with the accompanying subjunctive stems and their accentuation.

Table 5.36: Privatives with a nasal in the prefix, predicted by Hilmarsson to be derived from stems with second-syllable accent.

| privative | base | sbj. stem | diagnostic forms |
|---|---|---|---|
| *aṅklautkatte* | *klawtka-* 'turn' | \|kláwtka-\| [5] | sbj.ger. *klautkalle* |
| *amplākätte* | *plak-* 'agree' | \|plakə-\| | — |
| *eṅkälpatte* (arch.) | *kəlpa-* 'obtain' | \|kəlpa-\|* [5][(a)] | — |

| | | | |
|---|---|---|---|
| *eṅkärstātte* (arch.) | *krəst-* 'cut off' | \|krə́sta-\| [5] | opt.mid.1sg. *karstoymar*<br>opt.mid.3sg. *karstoytär*<br>inf. *karstatsi* |
| *eṅklyauṣäcce* | *klyaws-* 'hear' | \|kĺawṣə́-\| [2] | sbj.vn. *klyauṣalñe* |
| *empakwätte* | *pəkʷ-* 'rely on' | \|pəkʷə́-\| [1] | sbj.vn. *pkwalñe* |
| *empalkaitte* | *pəlka-* 'see' | \|pə́lka-\| [5] | sbj.act.1pl. *palkam*<br>sbj.vn. *palkalñe*<br>inf. *palkatsi* |

(a) The synchronic TB subjunctive stem of *kəlpa-* is the subjunctive [6] *kəlla-*, which derives from an erstwhile nasal present [6], apparently replacing the stem that was used for the basis of *eṅkälpatte* (Hilmarsson 1991: 69–71).

Of the examples of privatives with a preserved nasal in the prefix, *eṅklyauṣäcce* and *empakwätte* are derived from stems without first-syllable accent, so that they would be regular according to Hilmarsson's rule. On the other hand, *aṅklautkatte*, *eṅkärstātte* (arch.) and *empalkaitte* are derived from stems with first-syllable accent, so that they should have lost the nasal from the prefix according to Hilmarsson's rule. The accentuation of the underlying stems of *amplākätte* and *eṅkälpatte* cannot strictly be determined.

Table 5.37: Privatives without a nasal in the prefix, predicted by Hilmarsson to be derived from stems with first-syllable accent.

| Privative | Base | sbj. stem | Diagnostic forms |
|---|---|---|---|
| *akākatte* | *kaka-* 'call' | \|káka-\| [5][(a)] | sbj.mid.3sg. *kākatär*<br>opt.mid.1sg. *kākoymar*<br>sbj.vn. *kākalñe*<br>inf. *kākatsi* |
| *akautacce* | *kawta-* 'split' | \|káwta-\| [5] | sbj.mid.3sg. *kautatär*<br>sbj.vn. *kautalñe*<br>inf. *kautatsi* |
| *akraupatte* | *krawpa-* 'gather' | \|kráwpa-\| [5] | sbj.mid.1sg. *kraupamar*<br>sbj.mid.3sg. *kraupatär*<br>sbj.mid.3pl. *kraupantär*<br>sbj.ger. *kraupalle*<br>sbj.vn. *kraupalñe*<br>inf. *kraupatsi* |

| Privative | Base | sbj. stem | Diagnostic forms |
|---|---|---|---|
| *apāṣṣätte* | *pask-* 'guard' | \|pa$^{ṣṣ\acute{ə}}$/$_{skə\acute{-}}$-\| [2] | sbj.ger. *paṣṣalle*<br>sbj.vn. *paṣṣalñe* |
| *ākkārpacce* (arch.) | *karpa-* 'descend' | \|kárpa-\| [5] | sbj.act.2sg. *kārpat*<br>sbj.ger. *kārpalle*<br>sbj.vn.gen.sg. *kārpalñentse*<br>inf. *kārpassi*/*kārpatsi* |
| *ekamätte* | *kəm-* 'come' | \|kəm-\| [1]* | — |
| *ekätkātte* (arch.) | *kətka-* 'strew' | \|kə́tka-\| [5] | sbj.act.2sg. *kātkat*<br>sbj.ger. *katkalyi*<br>sbj.vn. *katkalñe*<br>inf. *katkatsi* |
| *ekältte* (arch.) | *kəl-* 'bear' | \|kəl-\| [1] | — |
| *ekwalatte* | *k$^{w}$əla-* 'fail' | \|k$^{w}$əlá-\| [5] | sbj.mid.3sg.-pl. *k$_{u}$lātär-me*<br>opt.mid.3sg. *k$_{u}$loytär*<br>sbj.vn. *k$_{u}$lālñe*<br>inf. *k$_{u}$lātsi* |

(a) There is one attestation of a supposed sbj.mid.1sg. as *kakāmar* in PK AS 17C, which would seem to suggest second-syllable accent. However, this is probably a verbal noun derived from *kakāmau*, the preterite participle of the root *kama-*, suppletive from the verb *pər-* 'bring'; cf. the common absolutive *kakāmarmeṃ*.

Of the privatives without a nasal in the prefix, *akākatte*, *akautacce*, *akraupatte*, *ākārpacce* (arch.), and *ekätkātte* (arch.) are derived from stems with first-syllable accent, in accordance with Hilmarsson's rule. However, *apāṣṣatte* and *ekwalatte* are derived from stems with second-syllable accent. The accentuation of the stems from which *ekamätte* and *ekältte* (arch.) are derived is uncertain.

In terms of the accentuation of the privative itself, it is clear that they are all accented on the second syllable, as far as their accent can be determined; cf. the privatives in Table 5.36 and Table 5.37. This means that it is difficult to assume a deletion of the nasal in the privative formations themselves as the result of contrasting positions of the accent.

This also holds for privatives starting with other consonants or with vowels, as shown in Table 5.38.

Table 5.38: Further examples of privatives, consistently showing the accent on the first syllable of the stem.

| Privative | Base | sbj. stem | Diagnostic forms |
|---|---|---|---|
| *atākatte* | *taka-* 'be' | \|táka-\| [5] | sbj.act.2sg. *tākat*<br>sbj.act.1pl. *tākam*<br>sbj.act.2pl. *tākacer*<br>opt.act.2sg. *tākoyt*<br>opt.act.2pl. *tākoycer* |
| *atraikatte* | *trəyka-* 'go astray' | \|trə́yka-\| [5] | sbj.vn. *trikalñe* |
| *anākätte* | *nak-* 'blame' | \|nákə-\| [1] | sbj.vn. *nākälyñe* |
| *anāyätte* | *ay-* 'give' | \|ayə́-\| [1] | sbj.act.1sg.-3sg. *ayu-ne* |
| *anautacce* | *nawta-* 'disappear' | \|náwta-\| [5] | sbj.act.3pl.-3sg. *nautan-ne*<br>sbj.ger. *nautalle*<br>sbj.vn. *nautalñe* |
| *amāntatte* | *mənta-* 'stir' | \|mánta-\| [5] | sbj.mid.3sg. *māntatär*<br>sbj.ger. *māntalñe*<br>inf. *māntatsi* |
| *amaukacci* | *mawka-* 'desist' | \|máwka-\| [5] | sbj.vn. *maukalñe* |
| *ayāmätte* | *yam-* 'do' | \|yamə́-\| | sbj.ger. *yamalle*<br>sbj.vn. *yamalñe* |
| *alālätte* | *lal-* 'exert oneself' | \|lalə-\| | — |
| *alaitacci* | *ləyta-* 'fall down' | \|láyta-\| [5] | sbj.vn. *laitalñe* |
| *awāskatte* | *waska-* 'stir' | \|wáska-\|* [5] | — |
| *awlāwatte* | *wlawa-* 'control' | \|wláwa-\| [5] | sbj.mid.3sg. *wlāwatär*<br>sbj.vn. *wlāwalñe* |
| *aspāwatte* | *spawa-* 'subside' | \|spáwa-\|* [5] | — |
| *etaṅkätte* | *tənk-* 'hinder' | \|tə́nkə-\| [1] | sbj.act.3sg. *teṅkäñ-c* |
| *etreṅkätte* | *trenk-* 'cling' | \|trenkə-\| [1] | — |
| *eśuwacca* | *śəwa-* 'eat' | \|śəwá-\| | sbj.act.3pl.-3sg. *śwā-ñ*<br>sbj.ger. *śwālle*<br>inf. *śwātsi* |

It is best to assume, with Adams (2013: 88) and *contra* Hilmarsson (1991), that the nasal-less allomorph originated in instances where certain consonant clusters were simplified, from where it could spread by analogy. Roots with an initial resonant like *l-* (e.g., *alālätte*) or *y-* (e.g., *ayāmätte*) or *w-* (e.g., *awāskatte*) are not found with a nasal, so that such forms could provide a point of origin for a nasal-less variant.

**The nasal prefix, contraction and accentuation**

The accentuation of the privatives remains the same in two cases where the prefix contracts with the first root syllable with the change **eyə* to *ai* as discussed in the previous subsection for preterite participles like *yainmu*. As with the participles, the accent stays on the resulting diphthong and does not move to the new second syllable, even if that is the accented syllable in other derivations from the subjunctive stem.

Table 5.39: Privatives showing the development from **eyə* to TB *ai*.

| Privative | Base | sbj. stem | Diagnostic forms |
|---|---|---|---|
| *ainmacce* | *yənm-* 'achieve' | \|yənmá-\| [6] | sbj.ger. *yänmälle*<br>inf. *yänmätsi* |
| *airpäcce* | *yərp-* 'observe' | — | none |

Other formations with the nasal prefix show this type contraction as well, including the development from **ewə* to *au*, as seen in the examples in Table 5.40.

Table 5.40: Forms with a nasal prefix showing the development of **eyə* to TB *ai* and **ewə* to TB *au*.

| Example | Meaning | Base |
|---|---|---|
| *aikne* | 'duty' | *yakne* /yə́kne/ 'way' |
| *aiksnar* | '(all) together' | *yəksnar**, from *yəks-* 'grasp, enfold' |
| *aurtse* | 'broad, wide' | *wartstse* /wə́rtˢtˢe/ 'broad, wide' |
| *auspa* | 'truly, certainly' | *wəspa** cf. TA *wäṣpā* 'id.' |

We already saw that this development from **eyə* and **ewə* to *ai* and *au* postdates the movement of the accent to the second syllable, and *aiksnar* '(all) together' furthermore illustrates that it did not interfere with epenthesis. Otherwise, the cluster **-ksn-* should have been broken up (cf. 5.2.2) resulting in ***aiksanar* /ayksə́nar/. Other forms with a cluster *CCR* do not show epenthesis either: we find *emparkre* rather than ***empärkare* and *eṅkatkre* rather than ***eṅkätkare*. This can be understood by assuming that the accent remained on first syllable of the root, which was the second syllable due to the prefix and thus not susceptible to accent shift.

Any potential phonetic epenthesis that may have occurred earlier was lost, because the schwa did not become accented with accent shift from the first to the second syllable. The reconstructions are thus **eyə́ksnarə* (cf. Hilmarsson 1991: 161–166), **empárkre*, and **enkátkre*, which may have become **eyáksənarə*, **empárkərə*, **enkátkərə* due to epenthesis but did not undergo accent shift and thus lost their schwa again in accordance with the deletion of unstressed schwa in open syllables. The development of

*emparkre* and *eṅkatkre* can thus be understood next to *pärkare* and *kätkare*, which did undergo accent shift after epenthesis from **párkre* and **kátkre* via **párkəre* and **kətkáre*, consequently retaining the newly stressed epenthetic vowel.

**The nasal prefix and abnormal accentuation**

There are two words with a nasal prefix that, rather unusually, have third-syllable accent: *eneṅkāññe* 'internal' and *eṣerñāna* 'sisterly ones'. Their accent is difficult to explain within the framework of the normal accentuation rules and is probably analogical. For *eneṅkāññe* this can be illustrated quite clearly. The adverb *eneṃ* 'inside, within', already containing the nasal prefix, was combined with the emphatic suffix *-ka* to form *eneṅka* /enénka/ (Adams 2013: 88). The accentual movement to the right may have become semi-regular in the process of word formation. It was naturally also found in other derivations that use the same adjective forming suffix *-ñña*, such as the semantically parallel derivation *pärnāññe* /pərnáńńe/ 'external' from *parna* /párna/ 'outside' (Adams 2013: 381–382). Not all derivations of this type inspire accentual changes, however, cf. *aknātsañña* 'ignorance' from *aknātse* 'foolish, stupid' (Adams 2013: 3). This points to a degree of flexibility with regards to the interplay of the accent and derivation in Tocharian B.

**Brief conclusion**

As a category, the words with nasal prefixes do not clash with the rules of accentuation discussed so far and can be easily understood within a framework of accent movement from the first to the second syllable. Examples like *emparkre* and *eṅkatkre* enrich our understanding of the accentuation of epenthetic vowels, since they show that epenthetic vowels did not attract the accent, and only became accented secondarily as the result of regular accent shift. When they did not get accented, they were lost again.

### 5.4.3 The accent in kinship terms

A few of the Tocharian kinship terms deserve special attention, since they do not adhere to the expected accentuation patterns, or show unexpected developments more generally. Firstly, the plural *pātärñ* to *pācer* 'father' should regularly have been accentuated on the second syllable, i.e., ***patarñ*. However, its initial accent can easily be analogical to the oblique singular *pātär* from **patər*. This form does not directly continue the Proto-Indo-European accusative **ph₂ter-m*, since that would have yielded ***pācär* instead, but it may be understood as a reflex of a remodelled accusative in early pre-PT **ph₂tr̥-m*. In this way, the "aberrant" accent of the oblique *pātär*, which prompted Malzahn (2010: 6–7) and partly also Jasanoff (2015: 87–88) to posit a limitation on accent shift from a full vowel to a schwa, can be understood without recourse to a change that

is so ill-equipped to explain the initial accent of epenthetic vowels and of several verbal categories, as discussed at various points above.

The parallel kinship term *tkācer* 'daughter', pl. *tkātärñ* /tkátərñ/ has its accent on the original second syllable from earlier **təkátərnʸə*, as expected. The plural *pātärñ* may thus also have been influenced by this form. Both *pācer* and *tkācer* have an alternative plural with an accent shift, namely *pacera* /pacéra/ and *tkacera* /tkacéra/ respectively. These plurals of kinship terms are all recent creations, however, and thus not reliable for the history of the accent (see Peyrot 2008: 112–114). They are not shared with Tocharian A, which has *pācri* and *ckācri* instead. The plural of TB *mācer* is once attested as *mācera* /mácera/, with an initial accent that cannot be understood except by accepting that it is a transparent creation of the nominative singular *mācer* in combination with the plural suffix *-a* without application of the accent rule. The alternative plural *mātärñ* is like *pātärñ* and *tkātärñ*.

The Tocharian word for 'sister' is unexpectedly short: TB *ṣer*, TA *ṣar* reconstructible as PT **sʸer(ə)*. This derives from the PIE noun **suésōr* with a complete loss of the first syllable. In the original disyllabic nominative **suésōr* this development is strange, as this would be expected to yield PT **sʸwʸəsor*, whence TB ***ṣasor* /ṣə́sor/, TA ***(ṣᵤ)sar*. Adams departs from **suesēr* with **-ēr* from the other kinship terms like **mh₂tēr* 'mother', **ph₂tēr* 'father', **dʰugh₂tēr* 'daughter' and **bʰreh₂tēr* 'brother' (Adams 2013: 724), though this should probably still regularly yield TB *ṣaṣer* /ṣə́ṣer/.

Ringe assumes an early syncope between two sibilants, irrespective of the accentuation, and takes the oblique stem **suesor-* as the basis, rather than **suesōr* or Adams's analogical **suesēr* (Ringe 1987: 263–265). In Ringe's formulation, this would be parallel to Winter's rule of /ə́/-deletion between two sibilants, which was rejected in the discussion in (5.3.3) as inadequate. However, if we indeed depart from forms like the old accusative **suesorm* (cf. Skt. *svásāram*) or pl. **suesores* (cf. Skt. *svásāraḥ*), the result is a trisyllabic form **ṣwə́serə*, which, with regular second-syllable accent may have produced **ṣérə* as the result of syncope and regressive assimilation in the initial cluster. A paradigm of the type nom.sg. **ṣəser/ṣəṣer/ṣəsor* : obl.sg. **ṣerə* : pl. **ṣerə* would have been quite irregular, so that an analogical change from the isolated nominative **ṣəser* (if remade to **suesor*) or **ṣəṣer* (if remade to **suesēr*) or **ṣəsor* (if from **suesōr*) to **ṣer* can be understood.

### 5.4.4 The accent in the secondary cases

Nouns normally show accent movement from the first to the second syllable when case endings are added, e.g., *lakle* /ləkle/ 'sorrow' in the genitive singular *läklentse* /ləklentˢe/ and the plural *läklenta* /ləklénta/. However, no accent shift is observed when the secondary (local) case suffixes are added; see Table 5.41. For example, the "classic" locative of TB *lakle* 'suffering' is *laklene* /ləklene/ rather than expected ***läklene* /ləkléne/,

which is only found in archaic texts where ⟨ä⟩ can signify both /ə/ and /ə́/. Similarly, the comitative is *laklempa* /lə́klempa/ and the perlative *laklesa* /lə́klesa/, with stable first-syllable accentuation (e.g., Kölver 1965: 5). Only the ablative shows a preference for accent movement, though not entirely regularly (ibid.); cf. the paradigm of *āntse* /ántˢe/ 'shoulder' in Table 5.41.

Table 5.41: Two example paradigms showing the accentual behaviour in Tocharian B nominal inflection.

| | *lakle* /lə́kle/ 'sorrow, pain' | | *āntse* /ántˢe/ 'shoulder' | |
|---|---|---|---|---|
| | singular | plural | singular | plural |
| nom. | *lakle* | *läklenta* | *āntse* | *āntsi* |
| gen. | *läklentse* | *läklentaṃts* | *antsentse** | *antseṃts* |
| obl. | *lakle* | *läklenta* | *āntse* | *āntseṃ* |
| all. | *lakleśc** | *läklentaśc** | *āntseśc** | *āntseñśc** |
| loc. | *laklene* | *läklentane** | *āntsene* | *āntseṃne** |
| abl. | *läklemeṃ** | *läklentameṃ* | *āntsemeṃ* | *āntseṃmeṃ** |
| per. | *laklesa* | *läklentasa** | *āntsesa* | *āntsentsa* |
| com. | *laklempa* | *läklentampa** | *āntsempa** | *āntseṃmpa** |

One interpretation of the lack of accentual movement is that the secondary case suffixes were still phonologically relatively independent (Carling 2000: 372; cf. 6.2.2.4). If it is indeed true that the secondary case endings were not yet properly fused with the noun at the point in time that the Tocharian standard accentuation on the second of three syllables was established, then the creation of the secondary cases should have taken place relatively late in the development of Tocharian.

However, I do not believe that this behaviour of the secondary cases is actually informative for either their age or the age of the accent itself. The secondary cases are based on the oblique in a synchronically transparent way, a disyllabic oblique regularly have first-syllable accent. The accentuation of *laklene* can thus simply have been remade from perhaps expected ***läklene* on the basis of the oblique *lakle*. Similarly, while the plural secondary cases of *lakle* all show the same non-initial accent as the oblique plural *läklenta*, those of *āntse* are accented as the plural oblique *āntseṃ*.

The analogical nature of the secondary cases is further supported by the secondary case forms of Proto-Tocharian schwa-final nouns like TB *ost* 'house' from PT **wostə*: the locative is *ostne* /óstne/ (or *osne* with a reduction in the cluster), rather than expected ***ostane* /ostáne/ as would be derived from PT **wostə* + **ne* by sound law. In comparison with verbal forms like prs.act.2sg-3sg. *aista-ne* /aystáne/ and prs.act.2sg-1sg. *nesta-ñ* /nestáñ/ (from **nestə-ñə*), there is no reason to suppose a syncope in a sequence like **-stənV*, so that *ostne* cannot regularly derive from an earlier **ostəne*. The apocope of the

final schwa itself is certainly too late to put before the proper creation of the secondary cases, so that a recreation of the locative based on the synchronic oblique *ost* seems probable.

The only secondary case where accent shift is regularly observed is the ablative in *-meṃ*, as in *läklemeṃ* /ləklémen/ and *ostameṃ* /ostámen/, which thus seems to hold a special status. The reason for this is unclear, as the etymology of the ablative ending *-meṃ* itself is obscure (cf. 6.2.2.2). The fact that it is a particular Tocharian B formation means that it is more likely to be a relatively recent creation than most of the other cases, which may have caused its aberrant behaviour. On the whole, I think that the unexpected accentual behaviour of the secondary cases does not provide reliable data that can be used for the reconstruction of either the development of the Tocharian B accent or the history and development of the secondary cases themselves as a category.

### 5.4.5 The accentuation of compounds

The accent in compounds is different from the normal accent. This is clear from examples like *astre-pälsko* /astrépəlsko/ 'having a pure spirit', a combination of *astare* /astə́re/ 'pure' and *palsko* /pə́lsko/ 'mind, spirit; thought'. The accent in the compound *astre-pälsko* is in the "wrong" place from the perspective of both individual words. The general tendency is for compounds to be accented on the final syllable of the first element (Marggraf 1970: 64ff.). In words that are usually disyllabic, this is equal to the second syllable that is accented in multisyllabic words as per the usual accentuation rules, e.g., *totkā-yärm* /totkáyərm/ 'of small measure', from *totka* /tótka/ 'few, little, small' and *yarm* /yə́rm/ 'measure, size'.

Compounds thus give the impression of an unretracted final accent on the first member in accordance with Marggraf's synchronic rules, but final accentuation can have been generalized from instances with a disyllabic first-membe, like *totkā-yärm*, to examples with a trisyllabic first member, like *astre-pälsko* and also *śtwarā-känte*, where the etymological third syllable is accented, i.e., as if from **kʸətwerá-* (see also Marggraf 1970: 76). The type here illustrated with *astre-pälsko* cannot be understood as regular. If derived from **ástre-pəlsko*, the developments observed in the base **ástre* to TB *astare* /astə́re/ should have yielded ***astare-pälsko* /astərepəlsko/ with both regular epenthesis and accent shift. The problem remains if one assumes a derivation from forms with an original accentuation as **astré* or as **astə́re*.

In some cases, a compound seems to preserve an otherwise lost final schwa, as in, e.g., *osta-ṣmeñca* /ostə́ṣmeñca/ from *ost* 'house', originally PT **wostə*. In others, a similar etymological schwa is lost without a trace, e.g., *ñem-kälywe* not ***ñema-kälywe*. The loss of schwa in the latter is ascribed by Marggraf to the structure of the onset of the second member, in this case, *CV-* (Marggraf 1970: 68). However, accented /ə́/ is not normally lost in classical Tocharian B even in open syllables, so that an original **ńemə́kəĺwe* should

have become ****ñema-kälywe***. There thus needs to be some other explanation for *ñem-kälywe*. Compounds like *ñem-kälywe* can be synchronically understood as a combination of *ñem* and *kalywe* /kə́ĺwe/.

I think that analogical pressure from the uncompounded words probably influenced the shape of the first member of these compounds in combination with a synchronic tendency to put the accent on the last syllable of the first member. Such a rule could easily be derived from compounds with a disyllabic first member. Since analogical accentuation of the first member of compounds is likely, the accent of these words cannot be used to determine the general historical development of the Tocharian accent. On Tocharian nominal compounds in general, see Bernhard (1958).

### 5.4.6 The accentuation of loanwords

A number of loanwords from other languages in Tocharian B provide additional indications for the prehistoric position and development of the Tocharian accent. One class of words, considered of BMAC-related origin, typically have second-syllable accentuation, e.g., *kercapo* /kercə́po/ 'donkey', *witsako* /wəytˢə́ko/ 'root' and *ṣecake* /ṣecə́ke/ 'lion' (Bernard 2023: 215–218, 222–227, 228–230). These words thus correspond to the general accentual pattern of Tocharian B.

On the other hand, Old Iranian and Khotanese loanwords tend to have initial accent, e.g., *ainake* /áynake/ 'common (vulgar); bad, evil', *ākteke* /ákteke/ 'wonderful, astonishing' from Old Steppe Iranian (Bernard 2023: 171–173), *twāṅkaro* /twánkaro/ 'ginger' from Khotanese (Dragoni 2023: 125–127), despite also being trisyllabic. Unlike the BMAC-related borrowings, these trisyllabic Iranian examples deviate from the general accentual pattern of Tocharian B. This implies that they were borrowed when Tocharian B accent assignment to the second syllable was no longer operative. In the framework of a progressive accent shift, this means that the borrowing took place after the accent shift.

Trisyllabic Iranian loanwords do contrast with disyllabic loanwords, which are actually integrated in the general accentual pattern observed in Tocharian B nouns. For example *tāno* /táno/ 'seed, grain', *parso* /pə́rso/ 'letter', from Proto-Khotanese-Tumshuqese (Dragoni 2023: 122, 138) or *pāke*/páke/ 'part, portion' and *śāte* /śáte/ 'rich' from Old Steppe Iranian (Bernard 2023: 41–43, 74–75) show accentual movement in nom.pl. *tanāñ* /tanáń/, obl.pl. *pärsonta* /pərsónta/, nom.pl. *śateñ* /śatéń/, gen.sg. *pakentse* /pakéntˢe/; and similarly, *cākkär* from Sanskrit *cakra-* shows accent movement in the gen.sg. *cakkarntse* /cakkə́rntˢe/ (Winter 1990: 382). According to Winter, this suggests that accent retraction was a recent phenomenon (Winter 1990: 382), from which it could theoretically be inferred that the accent shift took place very late, after the adoption of these loanwords. However, a more likely interpretation is that the prevalence of disyllabic words undergoing accent shifts in their inflection served as a

basis for its application to disyllabic loanwords, so that this phenomenon ultimately does not tell us anything about the history of the accent.

Especially the initially accented trisyllabic Old Iranian loanwords are informative, because they deviate from the standard accentual pattern in Tocharian B. This suggests that they may have entered the language after the shift to second-syllable accent had taken place, which, in turn, would imply that the period in which second-syllable accent was first established lies farther in the past than the earliest contact with Iranian, i.e., before Proto-Tocharian. The Tocharian A counterpart to TB *ainake* does pose a problem here: TA *enāk* has not undergone regular vowel weakening to ****enak*, so that a post-Proto-Tocharian origin for this word in Tocharian A appears to be required; see Bernard (2023: 171–173) for a discussion.

## 5.5 Tocharian A accentuation

The accent in Tocharian A is more difficult to determine, especially synchronically, since it is not marked in writing. However, there are certain sound changes that occurred in the prehistory of Tocharian A and give strong clues that, at some point in pre-TA, the accent fell on the first syllable. It is unclear whether this should be regarded as a secondary initial accentuation, or if it is a sign that Tocharian A never underwent a shift to the type of accent pattern as we see in Tocharian B. The main argument that has been used to equate early pre-TA accentuation with that of Tocharian B is the supposed more limited application of *a*-umlaut in Tocharian A. I discussed this in 2.5.5, and concluded that it cannot be interpreted as has been claimed.

### 5.5.1 Apocope, vowel weakening and syncope in Tocharian A

While the accent of Tocharian A cannot be deduced from the writing, there are three sound developments that give us indirect information on the accentuation in pre-Tocharian A. They are apocope, vowel weakening, and syncope. Apocope and vowel weakening imply a first-syllable accent, while syncope additionally shows that the schwa, written as *ä* in Tocharian A, had become incapable of bearing the accent at some point.

**Apocope**

Apocope affected all Proto-Tocharian word-final monophthongs in Tocharian A, which were completely lost (e.g., Winter 1994a: 401). Thus, TB *aknātsa* 'foolish, ignorant' corresponds to TA *āknats*, TB *keme* 'tooth' corresponds to TA *kam*, and TB *klyomo* 'noble' corresponds to TA *klyom*. Only final monophthongs that resulted from contraction of two syllables remained, specifically the prt.act.1sg. suffix of *ā*-final verbs TA *-ā* from pre-TA **-āwā* and the perlative suffix TA *-ā*, potentially from (generalized) pre-TA *-a.ā*

(Winter 1994a: 404). Since a word-final second syllable in disyllabic words was lost, the accent must logically have been on the first syllable.

This change also took place if the only internal vowel was a pre-TA **ä* (PT **ə*), as shown in examples like TA *śärs* from **śärsā* 'knew 3sg.', *wäs* from **ẃäsā* 'gold', *kus* from **kʷäsa* 'who', etc. If Tocharian A accent at the time of apocope was not on the *ä* in the first syllable but on the heavy vowel in the second syllable, a development from **śärsā́* to *śärs*, **ẃäsā́* to *wäs*, or **kʷäsá* to *kus* is needed, with the deletion of the accented vowel, which would be unnatural. Thus, the accent in these words was **śā́rsā*, **ẃā́sā*, **kʷā́sa* at the time of apocope (cf. Pronk 2009: 103 fn. 54). This is in line with the accentuation seen in TB *śarsa* /śə́rsa/ and TB *yasa* /yə́sa/, but not TB *kᵤse* /kʷsé/, whose syncope must consequently be a post-Proto-Tocharian innovation.

**Vowel weakening**

The sound change that confirms a general first-syllable accent in pre-TA even in words of more than two syllables, is known as vowel weakening (Kim 2007b; Peyrot 2013: 58), vowel balance (Krause & Thomas 1960: 45–47; Malzahn 8–10) or "syncope" (Winter 1994a). Tocharian A vowel weakening worked as follows: any pre-TA **ā* in the second syllable was reduced to **a* when preceded by a heavy vowel pre-TA **ā*, **a*, **e* or **o*. If the pre-TA **a* resulting from vowel weakening was also followed by a heavy vowel (after apocope, e.g., Kim 2007b), it was further reduced to **ä* and could be deleted in open syllables.

For example, PT **aknatˢa* 'foolish, ignorant' became TA *āknats* rather than ***āknāts* without weakening, but it did not develop further to ***ākänts* with full weakening due to the original final **-a-*. Full weakening did occur in the nom.pl. *āknts**ā**ñ*, where the stem-final **-a-* was not apocopated due to the addition of the plural suffix *-ñ*. Since the vowel in the first syllable is never affected by these changes, while the vowel in the second syllable is reduced, it is commonly agreed upon that it was the first vowel that was accented when vowel weakening took place in pre-Tocharian A, at least when it was a heavy vowel (e.g., Malzahn 2010: 8–10).

**Syncope**

Finally, any pre-TA **ä* that stood in an open syllable was lost due to syncope. This syncope applied to pre-TA **ä* corresponding to both accented and unaccented *ə* in Tocharian B, in the first, the second or in any other syllable. Thus, the Tocharian A counterpart to TB *yasar* /yə́sar/ 'blood' is TA *ysār* 'id.', even though the original accentuation in pre-TA should also have been **yā́sār*, as per the apocope observed in **śā́rsā*, **ẃā́sā*, etc. (cf. above). This development implies that TA *ä* lost its accent to any surrounding full vowels; i.e., **yā́sār* > **yäsār* > *ysār*. The accent should logically have

remained on TA *ä* if it was the only vowel in the word, such as in *wäs* from **ẃā́sā* after apocope.

**Conclusions**

These vowel developments strongly suggest initial accentuation in the development of pre-Tocharian A after the Proto-Tocharian period (e.g., Malzahn 2010: 8–10). Since Tocharian B accentuation can be analysed as a result of shift from the first to the second syllable, especially on account of the accented epenthetic vowels and the lost second-syllable schwa from PIE **e* (5.2.5), it is theoretically possible that the pre-Tocharian A first-syllable accent indicated by the sound changes treated here also reflects Proto-Tocharian accentuation, and that the shift in Tocharian B only happened in pre-Tocharian B specifically. However, it cannot be excluded either that the accent of pre-Tocharian A represented a secondary development of first-syllable accentuation after a Proto-Tocharian accentuation in accordance with Tocharian B. Two further indications for the relationship between Tocharian A and B are treated in the next two subsections.

### 5.5.2 Vowel contractions in Tocharian A

Since contraction, or lack thereof, interacted with the place of the accent in Tocharian B, it is important for the relative chronology to take a look at similar contractions in Tocharian A. Unfortunately, analogy has caused a lot of regularizations (Ringe 1996: 155–157), cf. Table 5.42.

Table 5.42: A comparison of forms with contraction in Tocharian B and analogical remodelling in Tocharian A.

| reconstruction | TB with contraction | TA with analogy |
|---|---|---|
| **tetemәweṣә* | *tetemoṣ* 'be born' prt.ptc.m.obl.sg. | *tatmuṣ*, cf. nom.sg. *tatmu* |
| **pernewentә* | *pernent* 'glorious' m.obl.sg. | *parnont*, cf. nom.sg. *parno* |
| **kәlpaweṣә* | *kälpoṣ* 'obtain' prt.ptc.m.obl.sg. | *kälpoṣ*, cf. nom.sg. *kälpo* |
| **kekәmәwer* | *kekamor* 'arrival' | *kakmur-äṣ*, cf. prt.ptc. *kakmu* |

Tocharian A has thus removed all traces of contraction in the categories where it took place the most. The development of the imperative PT **pә-weñә* of *weñ-* 'say' to TA *peṃ* is also ambiguous. The vowel *e* can be the result of the palatalizing influence from the palatal nasal and derived from either pre-TA **poñ* or **pañ*, or the *e* could conceivably have been introduced analogically from other forms of the verb *weñ-*.

Two examples of the development of **awe* outside of the preterite participles are reflected in Tocharian A, and they happen to show two different reflexes, parallel to the split development of this sequence in Tocharian B; see Table 5.43.

Table 5.43: Developments of the sequence **awe* in Tocharian A.

| Development | Example | Meaning | cf. |
|---|---|---|---|
| **awe* > TA *a* | **kʷəryawer* > TA *kuryar* | 'trade, commerce' | TB *karyor* |
| **awe* > TA *ā* | **śawemane* > TA *śāmāṃ* | 'living' | TB *śāmane* |

The difference in the Tocharian A reflexes of **awe* in these words corresponds to the two different reflexes of **awe* in Tocharian B, which is most easily accounted for if the developments had taken place in Proto-Tocharian already. However, as discussed in 5.4.1, the simplifications of the sequence **awe* cannot strictly be related to the position of the accent. TB *karyor* /kə́ryor/ 'trade, commerce' with its first-syllable accent suggests that the contraction took place before accent shift to the second syllable, but analogy cannot be wholly excluded. TB *śāmane* /śámane/ is attested with another accentuation as TB *śamāne* /śamáne/ in two[66] cases, opposed to six[67] examples of TB *śāmane*. If *śamāne* reflects the more original form, it would, like TB *karyor* suggest contraction of **awe* before accent movement, but analogical accentuation of either *śamāne* or *śāmane* cannot be excluded. The vowel contractions that are shared between Tocharian A and B thus do not provide entirely reliable information about the position of the accent in Proto-Tocharian.

### 5.5.3 Indirect indications of pre-Tocharian A accent

It appears that none of the changes that interacted with the accent in Tocharian B can be securely established for Tocharian A. We are thus left to very indirect arguments, which are primarily based on Tocharian B accentual evidence. The accent of loanwords from Old Steppe Iranian may provide some indication (cf. 5.4.6). In Tocharian B, they have initial accent when trisyllabic, and thus should have been adopted after the accent shift had been completed. Otherwise, the initial accent should have shifted along with that of native words to land on the second syllable. Since a number of Old Steppe Iranian loanwords regularly correspond between Tocharian A and B in the same way as native Tocharian words, it is assumed that they were adopted in (pre-)Proto-Tocharian, rather than in the separate prehistories of pre-Tocharian A and pre-Tocharian B (Bernard 2023: 249–253). Since Old Steppe Iranian borrowings retain Old Steppe Iranian initial accent in Tocharian B, they probably entered the language after accent shift. These borrowings consequently imply that the accent shift took place in Proto-Tocharian already.

[66] In THT 88 a5, PK DA M 507.37 and .36 a39–40.

[67] In THT 85 b4, THT 235 a3, THT 541 a3, THT 1111 b4, THT 1193 and 2247 a5, PK AS 16.7 b2.

## 5.6 Conclusions on the accent

In this chapter on the Tocharian accent, I have reviewed the evidence regarding several phonological phenomena in Tocharian involving the accent, such as syncope and epenthesis, as well as the development the accentuation of several verbal categories such as subjunctive classes [1] and [5], and contractions found in the preterite participle and in other categories. I have attempted to integrate the various lines of evidence to come to an account of the development of the accent and a relative chronology of sound changes leading up to the attested situation in primarily Tocharian B, but also Tocharian A. An important assumption in my argumentation is that a non-existent syllable cannot have been accented, and thus epenthetic syllables will need to receive the accent by some mechanism. The default mechanisms for this should be either regular accent shift by sound law or analogical accent shift on the basis of other members of the same paradigm. Since words with epenthesis always have a stable accent on the epenthetic vowel, analogy can be ruled out in that category.

A prehistoric accent shift from the first to the second syllable is capable of bringing the accent to all epenthetic vowels that indeed show up as accented in Tocharian B. This implies an earlier first-syllable accent, which accords well with the early syncope of **ə* in open non-first syllables. Such a syncope would not have occurred had this **ə* been accented, since accented **ə* in open syllables remains in Tocharian B. Original first-syllable accent in combination with an accent shift to the second syllable if this was the non-final syllable can work with the various Tocharian B categories that have atypical accentuation, as detailed in (5.3) based on Peyrot (2013: 507–514) and Marggraf (1970: 21). An older stage of pre-Proto-Tocharian can thus be reconstructed with first-syllable accent.

I argue that it is at present impossible to determine with certainty whether the type of accent system of Tocharian B should also be reconstructed for Proto-Tocharian as well, contrary to previous assumptions (first by Cowgill 1967: 176–177). None of the attested sound changes are informative, and those that should be informative are unclear in Tocharian A and likely altered by analogy. Tocharian A only evidences an earlier word-initial accent, which may be secondary, but could in principle be more original and reflect the state of the accent system before the pre-Tocharian B accent shift. Most arguments for a reconstruction of Proto-Tocharian accentuation in the same way as found in Tocharian B are circumstantial. For instance, while the analogical changes connected to the accent that occurred in Tocharian B should require some time, and thus intuitively suggest a relatively early date for the establishment of the Tocharian B accent, that does not amount to hard evidence. The only relatively clear linguistic argument seems to me the initial accentuation of trisyllabic Old Steppe Iranian loanwords, which suggest that the Tocharian B accent shift had occurred when these

words were adopted. Since Tocharian A and B have to some extent the same loanwords that can be derived from Proto-Tocharian, Old Steppe Iranian contact, and thus the Tocharian B accentuation, can also be tentatively reconstructed for (pre-)Proto-Tocharian.

However, in a broader perspective, this can be regarded as secondary. The first-syllable accent that should be assumed for an earlier stage of pre-Proto-Tocharian on the basis of syncope and secondarily accented epenthetic vowels is the most interesting stage for the purpose of comparison with Uralic. Seeing as Proto-Uralic and Proto-Samoyed are reconstructed with stress on the first syllable (e.g., Janhunen 1982: 27; Sammallahti 1988: 480), this thus constitutes another phonological parallel between the two language groups. It has to be admitted that in itself this feature is not particularly telling, as stress-initial languages are common. However, changes in stress patterns are found in contact situations (see, e.g., Pronk 2018 on changes in the prosody of Slavic and Baltic languages due to contact), so that if this feature of pre-Proto-Tocharian is ascribed to contact, early Uralic seems to be an obvious source on account of the other features shared with that language family in particular. The initial accent reconstructed for Old Steppe Iranian by Bernard (2023: 171–173) cannot be the source of a pre-Proto-Tocharian initial accent, since the Tocharian accent had moved to the second syllable at the time of contact with Old Steppe Iranian (cf. 5.4.6). The accentual shift to earlier pre-Proto-Tocharian initial accent can thus be potentially attributed to a different contact language that predated Tocharian interactions with Old Steppe Iranian, plausibly early Uralic/Samoyed.

# 6 Agglutinative case systems[68]

Tocharian has developed an agglutinating case system that has been at the centre of theories on substrate influence for a long time (e.g., Krause 1951), because of its innovative, non-Indo-European character. Influence from the Samoyed case system may provide an adequate explanation for certain important aspects of the Tocharian case system, and thus support the prehistoric contact hypothesis. Both the Tocharian and Samoyed case systems must be discussed in detail from a historical perspective to make a comparison possible, and to decide if early Samoyed provides an adequate candidate as the source of this substrate influence in Tocharian.

## 6.1 Introduction

Tocharian case inflection is quite elaborate, but there are only four cases that can be considered the descendants of the cases reconstructed for Proto-Indo-European: the nominative, the accusative (called oblique in Tocharian), the genitive, and the vocative. The dative disappeared as a separate case, but it provided the genitive form of some nouns (e.g., Pinault 2008: 510–511, cf. 6.2.2.3). The Tocharian genitive combines both genitive and dative functions (e.g., Krause & Thomas 1960: 82). The non-grammatical cases such as the locative and the ablative were lost at some point during the prehistory of Tocharian. After this reduction of the case system, Tocharian innovated new cases in the period leading up to Proto-Tocharian and continuing in the individual histories of Tocharian A and B (e.g., Carling 2000: 381–383).

However, the newly formed Proto-Tocharian cases were not merely functional replacements of those that were lost in pre-Proto-Tocharian. In particular, the perlative and comitative are entirely new and do not functionally correspond to cases that can be reconstructed for Proto-Indo-European. With the addition of these two cases, the Tocharian system conforms well to Anderson's characterization of the case systems found in languages of Central Siberia, where the presence of a perlative (=prolative) case and a separation of instrumental and comitative functions are common, at least in the present day (Anderson 2006: 278–292).[69] The case systems of Proto-Indo-European, Tocharian and various other Indo-European languages are illustrated in Table 6.1.

[68] This chapter is an adaptation of my (unpublished) MA thesis from 2019, titled "Uralic influence on the Tocharian agglutinative case system?".

[69] Anderson (ibid.) also mentions an opposition between allative and dative cases as a third characteristic of Siberian case systems, but Pakendorf has pointed out that this seems to be mostly restricted to Tungusic, and is thus rather a specific feature of that language family (Pakendorf 2010: 715, 718).

Table 6.1: An overview of the nominal cases in various Indo-European languages; a "+" indicates that the language at the top of the column has that particular case and that it is inherited from PIE in at least some instances; "≠" indicates that the languge has such a case, but that it is not a continuation of the corresponding PIE case; "–" indicates that the case is absent.

| | PIE | Hittite | Sanskrit | Greek | Latin | Lithuanian (Old) | Tocharian A | Tocharian B |
|---|---|---|---|---|---|---|---|---|
| nominative | + | + | + | + | + | + | + | + |
| vocative | + | + | + | + | + | + | – | + |
| accusative | + | + | + | + | + | + | + | + |
| genitive | + | + | + | + | + | + | + | + |
| dative | + | | + | + | + | + | – | – |
| locative | + | + | + | – | – | + | ≠ | ≠ |
| ablative | + | + | + | – | + | – | ≠ | ≠ |
| instrumental | + | + | + | – | – | + | ≠ | – |
| allative | +(a) | + | – | – | – | (≠) | ≠ | ≠ |
| adessive | – | – | – | – | – | (≠) | – | – |
| illative | – | – | – | – | – | (≠) | – | – |
| perlative | – | – | – | – | – | – | ≠ | ≠ |
| comitative | – | – | – | – | – | – | ≠ | ≠ |
| causal | – | – | – | – | – | – | – | ≠ |

(a) If this is to be reconstructed as a proper case based on the Hittite allative; only in adpositions and adverbs in the other Indo-European languages.

The endings of the Tocharian secondary cases differ from the older Indo-European norm in that they are agglutinating elements that are added to the oblique case form. For example, the same Tocharian A allative case ending *-ac* is added to both the singular *käṣṣiṃ* 'teacher' → all.sg. *käṣṣin-ac* 'to the teacher' and to the plural *käṣṣis* → all.pl. *käṣṣis-ac* 'to the teachers'. The oblique forms on which these allatives are based, by contrast, show fusional inflection, with the *-ṃ* (/-n/) in obl.sg. *käṣṣi-ṃ* signifying both oblique case and singular number and the *-s* in obl.pl. *käṣṣi-s* signifying both oblique case and plural number. Furthermore, not all obliques are formed using the same suffixal elements, so that it is not an agglutinative formation. The same goes for the other primary cases, while the secondary cases all have a single, unchanging ending each.

Another characteristic of Tocharian nominal inflection is a phenomenon known as *Gruppenflexion* or "group inflection". This refers to the use of only a single secondary case ending at the end of a phrase of multiple nouns. All elements before the final noun can be rendered in the oblique; cf. example (1 below, where the single use of the instrumental ending *-yo* applies to *kuklas* 'cart:OBL.PL', *yukas* 'horse:OBL.PL' and *oṅkälmās*

'elephant:OBL.PL' (after Krause 1951: 185–186; similar descriptions in, e.g., Krause & Thomas 1960; Pinault 2008: 466–467).

1. Tocharian A (A 253 b2)

| *kuklas* | *yukass* | *oṅkälmās-yo* |
|---|---|---|
| cart:OBL.PL | horse:OBL.PL | elephant:OBL.PL-INS |

'... with carts, horses, and elephants.'

Schmidt lists the three features of Tocharian nominal inflection mentioned above as "the main features caused by substrate influence", namely (1) agglutinating inflection, (2) the enlargement of the local case system to include a (new) locative, ablative, allative and perlative, and (3) Gruppenflexion (Schmidt 1990: 195). To explain these innovative features of the Tocharian case system, some scholars have invoked influence from non-Indo-European languages (see already Krause 1951, who considered influence from Tibetan, Altaic, Dravidian, Caucasian and Finno-Ugric; see also Thomas 1994).

However, not all scholars agree that these constitute significant innovations particular to Tocharian, or that it is warranted to consider them to be the result of language contact. Kim, for instance, is very much opposed to the idea that the secondary cases of Tocharian would be the result of substrate influence. He points to cases such as Vedic *pat-táḥ* 'from the foot' and Greek οὐρανό-θεν 'from heaven', which "demonstrate that the expansion of adverbial morphemes to other nominal bases and reinterpretation as case-like markers with an adverbial (usually locational) sense was far from uncommon in older Indo-European languages, and hardly requires the assumption of contact with Uralic, Turkic, or any other non-Indo-European languages of an agglutinating type" (Kim 2013: 124–126).

It is certainly important to keep in mind that agglutinating elements or tendencies are not alien to early Indo-European languages, and it seems prudent to caution against interpreting any and all agglutinating features as the result of substrate influence. However, if we consider the specifics of the Tocharian agglutinating case system as a whole, we may be justified in investigating potential contact effects, cf. Peyrot (2019a: 94). There are also other Indo-European languages that, like Tocharian, formed more extensive and innovative case systems with agglutinating elements, and this type of innovation can often be plausibly ascribed to language contact. In Ossetic, an Iranian language with nine cases, it has been argued that Georgian influence played an important role in the formation of the case system (Belyaev 2010: 311). Similarly, presence of agglutinative cases in some New Indo-Aryan languages like Sinhala and Hindustani can be attributed to the influence of agglutinating Dravidian languages spoken in the same area (Krause 1951: 189; Kulikov 2012: 300), and the Old Lithianian extended case system with the added allative, adessive, illative and inessive cases has been plausibly

attributed to influence from Finnic languages with their extensive local case systems (Kulikov 2012: 298–299). The same type of development has taken place in Anatolian Greek, which has made its case system more agglutinating in contact with Turkic (Thomason & Kaufmann 1988: 219–220). There thus seems to be a correlation between Indo-European languages with more extensive agglutinative case systems and their contact with non-Indo-European languages with agglutinative nominal morphology.

Since the rise of agglutinative case systems in Ossetic, New Indo-Aryan languages, Old Lithuanian and Anatolian Greek can be understood by considering the agglutinating languages spoken in the vicinity, it seems reasonable to investigate whether the Tocharian agglutinating cases can be explained in terms of language contact as well. Uralic was considered as a potential source in the literature early on already (e.g., Krause 1951; Bednarczuk 2015; Peyrot 2019a), but Turkic has also been suggested (Kulikov 2012; Akao 2020). This chapter will mainly concern itself with the question whether Uralic is an appropriate source, but the arguments in favour of alternative Turkic influence will also be considered in the final discussion as well.

## 6.2 The Tocharian case system

The Tocharian case system is divided into two types of cases: primary cases and secondary cases. The primary cases—nominative, oblique (accusative), genitive and (TB only) vocative—are the grammatical cases, and they exhibit the fusional marking of number and case typical of Indo-European languages. Most of the secondary cases have local semantics, and they are all formed by agglutination of their respective invariant case endings to the oblique singular or plural (see, e.g., Krause & Thomas 1960: 78ff.; Van Windekens 1979: 165–168; Carling 2000: 1–2). This system is illustrated for both Tocharian A and B in Table 6.2, with the vocative left aside.

Table 6.2: Example nominal paradigms in Tocharian A and B, using the word TA *yuk*, TB *yakwe* 'horse' (partly constructed where attested forms are lacking but predictable).

| | Tocharian A | | Tocharian B | |
|---|---|---|---|---|
| | singular | plural | singular | plural |
| nominative | *yuk* | *yukañ* | *yakwe* | *yakwi* |
| oblique | *yuk* | *yukas* | *yakwe* | *yakweṃ* |
| genitive | *yukes* | *yukaśśi* | *yäkwentse* | *yäkweṃts* |
| allative | *yuk-ac* | *yukas-ac* | *yakwe-ś(c)* | *yakweṃ-ś(c)* |
| locative | *yuk-aṃ* | *yukas-aṃ* | *yakwe-ne* | *yakweṃ-ne* |
| ablative | *yuk-äṣ* | *yukas-äṣ* | *yakwe-meṃ* | *yakweṃ-meṃ* |
| perlative | *yuk-ā* | *yukas-ā* | *yakwe-sa* | *yakweṃ-tsa*[(a)] |

| | Tocharian A | | Tocharian B | |
|---|---|---|---|---|
| | singular | plural | singular | plural |
| comitative | *yuk-aśśäl* | *yukas-aśśäl* | *yakwe-mpa* | *yakweṃ-mpa* |
| instrumental | *yuk-yo* | *yukas-yo* | = perlative | |

(a) The *-t-* is epenthetic between the final *-ṃ* /-n/ of the oblique and the *-s-* of the ending.

The reconstructions of the Tocharian cases to Proto-Tocharian are especially difficult in those instances where Tocharian A and B show different suffixes, such as in the ablative and the comitative. Furthermore, the precise functions of the individual cases are a cause for confusion, with the locative expressing both location and movement. The original, (pre-)Proto-Tocharian situation must be recovered for a comparison with the Samoyed case system (treated in 6.3), and this will be the objective of the following subsections.

### 6.2.1 The functions of the Tocharian secondary cases

In this subsection, the functions of the Tocharian secondary cases will be briefly discussed and illustrated. Tocharian A and B largely agree, but there are some discrepancies that may be of interest for the reconstruction of the Proto-Tocharian case functions and thus affect the comparison with Samoyed. The Tocharian B causal case *-ñ* is rather marginal and derived from a genitive form (Krause & Thomas 1960: 90; Van Windekens 1979: 258–259), so it will not be considered here.

#### 6.2.1.1 Allative

The allative TA *-ac*, TB *-śc* describes movement towards a goal describes movement towards a goal, especially when the goal is a living being, or when the goal is not necessarily reached or entered (Kölver 1965: 78; Carling 2000: 53–54). . It is often used to translate a Sanskrit dative in bilingual texts (e.g., Kölver 1965: 77). Normally, the Tocharian genitive is used for this function instead (cf. Krause & Thomas 1960: 82), and the dative use of the allative is consequently not proper to native Tocharian.

2. Tocharian B (THT 500.1 a3; verse; trsl. CEToM)

| *klyomñai* | *ytāri* | *oktatsai* | *yamim* | *pākri* |
|---|---|---|---|---|
| noble:F.OBL.SG | path:OBL.SG | eightfold:F.OBL.SG | do:OPT.ACT.1SG | manifest |

| ***rī-ś*** | *ynūcai* | *nervāṇäṣṣai* |
|---|---|---|
| town:OBL.SG-ALL | go:ACT.PTC.F.OBL.SG | Nirvana:ADJ.F.OBL.SG |

'May I make manifest the noble eightfold path that goes **to the city** of Nirvāṇa!'

3. Tocharian A (A 9 b1; prose; trsl. after CEToM)

| *po(ṃ)ś* | ***lānt-ac*** | *kakmuṣ* | *lānt* | *pälkoräṣ* |
|---|---|---|---|---|
| all:M.NOM.PL | king:OBL.SG-ALL | come:PRT.PTC.M.NOM.PL | king:OBL.SG | see:ABS |

'(Thereupon) all came **to the king**, and having seen the king, ...'

Tocharian A and B share the same use of the allative, so that the same function can be reconstructed for Proto-Tocharian (Carling 2000: 384).

#### 6.2.1.2 Locative

The locative TA *-aṃ*, TB *-ne* is used primarily to denote a location inside or movement into some object (Carling 2000: 51–52). The latter use is unusual when compared to locatives in other Indo-European languages. Kölver remarks that the common distinction in Indo-European languages between location and direction is remarkably absent in the Tocharian locative, which can express both (Kölver 1965: 97).

According to Carling, a noun marked in the locative is typically a bounded space like a 'house', which sets the locative apart from the perlative (see 6.2.1.4), which is used primarily with unbounded spaces like a 'field'.[70] Melchert suggests that the locative is rather used for spatial objects without significant surface extension, as this better explains the use of the locative with abstract nouns (which have no literal spatial delimitation) and words meaning 'place' (Carling 2000: 186, 264; Melchert 2002: 107). The object of emotions like 'love, desire, hate, compassion', etc., is also mostly rendered in the locative case in both Tocharian languages (Kölver 1965: 123), but since Sanskrit uses the locative in the same way in such situations (cf. Macdonell 1927: 198), this could also be due to secondary Sanskrit influence.

4. Tocharian B (IOL Toch 246 b3; prose; trsl. CEToM)

| *se* | *ṣamāne* | *yaka* | *yaṣi-sa* | *lānte* |
|---|---|---|---|---|
| DEM.M.NOM.SG | monk:NOM.SG | still | night:OBL.SG-PER | king:GEN.SG |

| ***kercyen-ne*** | *yaṃ* |
|---|---|
| palace:OBL.SG-LOC | go:SBJ.ACT.3SG |

'If a monk, still during the night, goes **into the palace** of the king ...'

[70] Tocharian A and B mostly agree on which combinations of verbs and nouns are construed with the locative or the perlative. An exception is the word for 'mountain', TA *ṣul* ~ TB *ṣale*: position on a mountain is expressed with the locative in Tocharian A, while Tocharian B uses the perlative (Caring 2000: 258–259).

5. Tocharian A (A 5 a6; prose; trsl. based on CEToM)

| *pkäntäk* | ***waṣt-aṃ*** | *lake* | *raksā-ṃ* |
|---|---|---|---|
| separate(ly) | house:OBL.SG-LOC | bed:OBL.SG | spread:PRT.ACT.3SG-3SG |

'He spread out a bed for him in a separate place **in the house**.'

Since the uses of the locative in Tocharian A and B are very similar, the Proto-Tocharian function would likely have been the same: the locative primarily denotes (internal) location and arrival inside a location (Carling 2000: 384–387).

### 6.2.1.3 Ablative

The ablative case TA *-äṣ*, TB *-meṃ* is primarily used to indicate the starting point of some movement or an action, without differentiating between a point of origin 'inside' or 'by' the object in the ablative case (Carling 2000: 23). In Tocharian A, the ablative is also used to mark the standard of comparison (Kölver 1965: 145–146). Here the two languages are in disagreement, as Tocharian B normally uses the perlative for this purpose instead (cf. 6.2.1.4).

6. Tocharian B (THT 93 a5; prose; trsl. CEToM)

| *te* | *keklyauṣormeṃ* | *candramukhe* | *walo* | *ṣecakecce* |
|---|---|---|---|---|
| DEM.N.OBL.SG | hear:ABS | PN | king:NOM.SG | lion:ADJ.M.OBL.SG |

| ***asāṃ-meṃ*** | *ñor* | *klāya* |
|---|---|---|
| throne:OBL.SG-ABL | down | fall:PRT.ACT.3SG |

'When he had heard this, king Candramukha fell down from [his] lion-throne.'

7. Tocharian A (A 30 a1; prose/verse?; trsl. CEToM)

| *vipul* | ***ṣul-äṣ*** | *lyutār* | *tpär* | *top* |
|---|---|---|---|---|
| PN | mountain:OBL.SG-ABL | more | high:M.NOM.SG | heap:NOM.SG |

| *na*(*ṣ*) |
|---|
| be:PRS.ACT.3SG |

'There is a heap higher **than mount** Vipula

The ablative endings of Tocharian A and B cannot be easily reconciled. For this reason, Carling does not reconstruct an ablative for Proto-Tocharian (Carling 2000: 379). She does remark that it is unexpected for the ablative to be absent from the local case system and wonders if the genitive might have been used for this purpose instead, even though the genitive has no discernible ablative use in either Tocharian A or B (Carling 2000.: 280). The fact that Tocharian A and B have different suffixes for the ablative does not necessarily mean that both are post-Proto-Tocharian innovations: it is possible that one

of the two replaced the original ending while the other kept it, or even that both replaced an earlier ablative ending. It will appear in subsection 6.2.2.2 that the ablative ending in Tocharian A may be derived from a Proto-Indo-European element, which would imply that it goes back to Proto-Tocharian as well.

The full range of functions of the Proto-Tocharian ablative can still not be established with certainty. Aside from the canonical ablative semantics, it could be that the Tocharian A use of this case to mark the standard of comparison is old. The use of the perlative as the standard of comparison in Tocharian B (see 6.2.1.4) could conceivably have resulted from the replacement of the original ablative ending. Perhaps the use of the ablative in some exceptional cases in Tocharian B was due to Sanskrit influence, since Sanskrit also uses the ablative to mark the standard of comparison (Macdonell 1927: 191–192), like Tocharian A. All in all, an ablative with the standard ablative meaning 'from' can be reconstructed for a Proto-Tocharian; potentially this case was used to mark the standard of comparison as in Tocharian A.

#### 6.2.1.4 Perlative

The meaning of the perlative case TA *-ā*, TB *-sa* is complex, and in some functions, it is rather close to the locative. The canonical use of the perlative can be translated with 'along, by, across, through' (Carling 2000: 45–46, 49–50; Krause & Thomas 1960: 84–86). Similar to the locative, the perlative can be used to indicate either movement or a static location. As mentioned in 6.2.1.2, Carling characterizes the difference between the two cases as a matter of unbounded space (perlative) vs. bounded space (locative) (Carling 2000: 258–259), whereas Melchert considers the notion of significant surface extension to be the primary factor (Melchert 2002: 107).

Another function of the perlative is to mark the agent in a passive sentence (Kölver 1965: 52; Krause & Thomas 1960: 85), although the genitive is used for this more frequently (Kölver 1965: 19). The perlative can also express a cause or reason in both Tocharian A and B (ibid.: 53–58; Krause & Thomas 1960: 84). In Tocharian B, the perlative is also used as an instrumental, and to mark the standard of comparison (Kölver 1965: 43–53, 64–66; Krause & Thomas 1960: 86).

8. Tocharian A (A 1 b2; prose; trsl. CEToM)

| *päñ* | *kursärwā* | *ār(ṣ)lās-yo* | *rarkusāṃ* | ***tkan-ā*** |
|---|---|---|---|---|
| five | mile:OBL.PL | snake:OBL.PL-INS | cover:F.OBL.SG | ground:OBL.SG-PER |

*kälk*
go:PRT.ACT.3SG

'For five miles he went **over the ground** covered with snakes.'

9. Tocharian B (THT 107 b5; prose; trsl. after CEToM)

| *ckentse* | ***manarkai-sa*** | *nyagrot stām* | *ñor* | ***atiyai-sa*** |
|---|---|---|---|---|
| river:GEN.SG | bank:OBL.SG-PER | Nyagrodha-tree | under | grass:OBL.SG-PER |

*lyama*
sit:PRT.ACT.3SG

'**On the bank** of the river he sat down under the Nyagrodha-tree **on the grass**.'

10. Tocharian B (THT 88 a1; prose; trsl. after Huard 2022: 382)

| *tumeṃ* | *durmukhe* | *brāhmaṇe* | *uttare*(*-ṃ*)[71] | *śamaśke-ṃ* |
|---|---|---|---|---|
| thereupon | PN | brahmin | PN(:OBL) | boy:OBL |

| *kärwāṣṣai* | ***witsakai-sa*** | *räskare* | *tsopaṃ-ne* |
|---|---|---|---|
| reed:ADJ.F.OBL.SG | root:OBL.SG-PER | sharply | hit:PRS.ACT.3SG-3SG |

'Thereupon the brahmin Durmukha hits the boy Uttara harshly **with a reed root**.'

11. Tocharian B (THT 107 a9; prose; trsl. after CEToM)

| *nānda* | *nandābala* | *weñāre* | *se* | ***ci-sa*** |
|---|---|---|---|---|
| PN | PN | say:PRT.ACT.3PL | REL.NOM | 2SG.OBL.SG-PER |

| *śpālmeṃ* | *tākaṃ* | *cwi* | *aiskem* |
|---|---|---|---|
| excellent | be:SBJ.ACT.3SG | 3SG.M.GEN.SG | give:PRS.ACT.1PL |

'Nandā and Nandabalā said: he who is more excellent **than you**, to him we [will] give [it].'

The functions of the perlative in Proto-Tocharian must have included the core locational meanings as attested in both Tocharian A and B (Carling 2000: 384–384), as these are shared between both Tocharian languages. What additional functions it would have had is difficult to establish due to the disagreement between Tocharian A and B. The expression of a cause or reason might have been in use early on, and given the innovative nature of the Tocharian A instrumental, it is possible that earlier instrumental use of the perlative was displaced by the new instrumental case in Tocharian A (on which see 6.2.1.6). The use of the perlative as the standard of comparison in Tocharian B could have been promoted by the loss or replacement of the original ablative case form (cf. 6.2.1.3), so that it may not have been a part of the function of the perlative in Proto-Tocharian. Aside from the locational meaning, the use of the perlative to mark a cause or reason can be reconstructed, and perhaps its instrumental meaning was derived from this at an early stage already, as it is found in Tocharian B.

[71] The final *-ṃ* is absent in the manuscript, but grammatically expected.

#### 6.2.1.5 Comitative

The comitative TA *-aśśäl*, TB *-mpa* marks accompaniment. Although the forms of the suffixes are different in Tocharian A and B, the comitative functions in both languages correspond closely (Kölver 1965: 69–74). If either suffix goes back to Proto-Tocharian (see 6.2.2.2), its function is likely to have been the same as attested.

12. Tocharian B (THT 78 b3–4; prose; trsl. CEToM)

| *kā* | *twe* | *brāhmaṇeṃ-mpa* | *t=* | *epiṅte* |
|---|---|---|---|---|
| why | 2SG.NOM.SG | brahmin:OBL.SG-COM | DEM.N.NOM.SG | between |

| *sak* | *wī(na)* | *w(ä)rpāt(ai)* |
|---|---|---|
| happiness:OBL.SG | pleasure:OBL.SG | enjoy:PRT.MID.2SG |

'Why have you enjoyed happiness and pleasure **with the brahmin** in the meantime.'

13. Tocharian A (A 66 a3; verse; trsl. CEToM)

| *śmeñc* | *śtwar* | *wäkn-ā* | ***ratk-aśśäl*** |
|---|---|---|---|
| come:SBJ.ACT.3PL | four | way:OBL.SG-PER | army:OBL.SG-COM |

'They will come **with** [their] fourfold **army** …'

#### 6.2.1.6 Instrumental

Only Tocharian A has a separate instrumental case, ending in *-yo*. The instrumental is used to mark an instrument or a cause, the latter function being shared with the perlative (Kölver 1965: 11–27). The instrumental is commonly agreed to be a specific Tocharian A derivation from the conjunction *yo* 'and' (see, e.g., Koller 2018: 180–181 with references), and thus cannot be reconstructed as a part of the Proto-Tocharian case system.

14. Tocharian A (A 91 a3; prose; trsl. after CEToM)

| *wsā-yokās* | ***poken-yo*** | *añcäl* | *pañwä(ṣ)* |
|---|---|---|---|
| gold-coloured:F.OBL.PL | arm:OBL.DU-INS | bow:OBL.SG | stretch:PRS.ACT.3SG |

'**With** gold-coloured **arms** he draws the bow.'

### 6.2.2 The forms of the Tocharian cases

At first blush, the formal correspondences between the secondary case suffixes in Tocharian A and B seem to be limited to the locative, which is very transparently the same: TA *-aṃ*, TB *-ne* from PT **-ne*. However, with a reinterpretation of the original final **-s* of the oblique plural ending **-ns* in Tocharian B, both perlatives TA *-ā* and TB *-sa* can be derived from PT **-a*. With the same development, the allative TB *-śc* can be connected

with TA *-ac*, with a generalized vowel *-a-* in the latter, and reconstructed as PT **-cə* (Gippert 1987: 25–28). These three cases will be discussed together in 6.2.2.1.

The ablative TA *-äṣ*, TB *-meṃ* and the comitative TA *-aśśäl*, TB *-mpa* cannot be reconciled, and their reconstruction for Proto-Tocharian is thus more problematic. These will be discussed in 6.2.2.2. It has furthermore been proposed that early Tocharian preserved a separate dative-locative case that was later lost, see 6.2.2.3. Finally, the status of the secondary cases will be given further specification in 6.2.2.4, before a final overview in 6.2.2.5.

### 6.2.2.1 Where Tocharian A and B agree

The locative suffix is the only secondary case suffix with a generally agreed-upon etymology. It can be connected with the element found in Lith. *nuõ* 'from', OLith. illative *-n*(*a*), OCS *na* 'on, at'; (on)to' from PBSl. **-nō* from **$h_2$noh$_1$* (Kim 2013: 132 with references). However, it is impossible to directly equate the reconstruction of this Balto-Slavic element to the Tocharian suffix PT **-ne*, since PT **e* normally reflects PIE **o* rather than **oH*. As a Proto-Indo-European reconstruction, also an accusative case form **$h_2$no-m* has been adduced (Kim 2013: 133), and a form **$h_2$no* without an additional ending (Van Windekens 1979: 257).

As a cognate of the Proto-Tocharian allative **-cə*, Greek δε as in οἶκόν=δε 'home(wards)' has been proposed (Van Windekens 1979: 254), and this works well semantically. However, PIE **de* should have given PT **śə* instead of **cə*, as in TA *śäk*, TB *śak* 'ten (10)' from PIE **dek̑m̥*. Gippert accounts for this by proposing a progressive devoicing rule after the accusative plural ending **-ns*, i.e., **-ns-de* > **-ns-te* > PT **-nst$^{y}$ə*. From such a position, **-cə* could have been re-analysed and spread as the new allative ending (Gippert 1987: 31). Similar re-analyses and extensions did take place within the Tocharian A and B case endings after the Proto-Tocharian period, so that Gippert's suggestion is not unthinkable.

Kim instead proposes that the allative **-cə* is to be derived from PIE **=d$^{h}$e* as found in Lat. *un-de* 'whence' and (with an additional *-n*) Greek ablatival forms in -θεν (Kim 2013: 134). Vedic *kúha* 'where', OCS *kъde* from **k$^{(w)}$u-d$^{h}$e* (Mayrhofer 1992: 383; Derksen 2008: 259-260) reflect a static meaning from this element, which may represent an intermediary between the ablative meaning found in Latin and Greek and the allative meaning of Tocharian. Perhaps Ru. *kudá* 'where, whither' contains the same element **d$^{h}$e*, although its exact reconstruction in the second-syllable vowel is not clear (Derksen 2008: 242-243).

The perlative PT **-a* has been connected to Lat. *ad* 'to, at', OEng. *æt* 'at, to', from PIE **$h_2$ed* (Pedersen 1941: 92; Van Windekens 1979: 251–252). An interpretation as **-a* from a generalized instrumental singular **-oh$_1$* of the thematic stems has also been proposed (Klingenschmitt 1994: 342–344), but PIE **-oh$_1$* is not normally known to become *-a* in

word-final position (cf. subsections 2.3.1 and 2.4.5.1). Kim (2012: 135) considers the allative, reconstructed by him as *-*eh*$_2$, as an option, but this reconstruction is not agreed upon: the Hittite allative in -*a* (unaccented) or -*ā* (accented) is argued by Kloekhorst to go back to *-*o* instead, based on the correspondence Hitt. *parā* /prá/ 'forward' to Gr. πρό, Skt. *prá-*, Lat. *pro*, Goth. *fra-*, all from PIE **pro* (Kloekhorst 2008: 161). Furthermore, the sequence PIE *-*eh*$_2$ normally became PT *-*o* rather than *-*a* (cf. subsection 2.3.1). The most straightforward reconstruction for the perlative is thus the adposition **h*$_2$*ed*, as this is the only one that does not involve extra phonological difficulties.

#### 6.2.2.2 Where Tocharian A and B differ

The endings of the ablative (TA -*äṣ*, TB *meṃ*) and the comitative (TA -*aśśäl*, TB -*mpa*) do not match, which makes it difficult to reconstruct these cases for Proto-Tocharian. Several suggestions have been made for the origins of these case suffixes, but an adequate explanation for the majority of them is still lacking.

**Ablative**

The Tocharian A ablative in -*äṣ* has been explained by Jasanoff as a reflex of PIE *-*ti*, or the corresponding adverb **h*$_1$*eti* 'from this' if this was suffixed early enough, similar to the ablative ending as found in Hittite *antuḫšaz* (-*z* < *-*ti*) 'from a man' and in Armenian *i getoy* (*-*o-ti*) 'from a river', *i banē* (*-*e-ti*) 'from speech' (Jasanoff 1987: 109–112). The phonological development necessary is based on two verbal endings, and on the Tocharian A ablative itself (see subsection 2.3.4.2 for more details).

The Tocharian B allative in -*meṃ* is very much obscure. Pinault proposed a connection with the adverb *mante* /mə́nte/ 'upward' (Pinault 2008: 471–472). The vowel of -*meṃ* does not match this adverb, so that a related word with a different ablaut grade would have to be assumed (Kim 2013: 132). According to Pinault (2008: 471–472) and Kim (2013: 132–133), this was added to the original ablative ending as found in Tocharian A[72] to explain the optional accent movement in found with TB -*meṃ*.

The exact origins of the Tocharian B ablative -*meṃ* thus remain uncertain, but with a possible derivation from PIE *(*h*$_1$*e*)-*ti*, the Tocharian A ablative -*äṣ* can be cautiously reconstructed for Proto-Tocharian.

[72] Kim's scenario, whereby pre-PT *-*ti* yield both TA -*äṣ* and TB -∅ (Kim 2013: 132-133), is chronologically quite difficult, since the variants *-*ti* and *-*t* should have remained side by side as ablative markers in Proto-Tocharian, even as *-*ti* had become *-*ṣ* and *-*t* had become *-∅. That Tocharian B should then have generalized the unmarked zero-ending at the expense of *-*ṣ*, before replacing it with -*meṃ* seems to me extremely unlikely.

**Comitative**

The Tocharian A comitative ending *-aśśäl* has a clear connection with the preposition and compositional element TB *śale~śle*, TA *śla* '(along) with; likewise'. The entire group is etymologically rather obscure, however, and connections with other Indo-European words are either semantically dubious or formally unresolved. The Celtic words OG *céile*, W *cilydd* 'companion, other' are semantically attractive for a comitative, but formally represent either **keiḷo* or **kiḷo* respectively, which would have yielded TB ***śille* or ***śalle* /śə́lle/. A pre-form **kilo* would perhaps regularly yield TB *śale* /śə́le/, but such a formation would be unparalleled (Adams 2013: 680; cf. Matasović 2009: 199–200).

Kim has proposed to derive PT **kʸəle* from a combination of **kʸə* from PIE **kʷe* 'and' and **yəle* 'to go', a hypothetical old gerundive from PIE **h₁i-lo-* (Kim 2013: 134–136). The Proto-Indo-European clitic **kʷe* is not otherwise preserved in Tocharian, and the regular form of the gerundive of 'to go' is TB *yalle* ~ TA *yäl* from PT **yəlle*. Thus, while a sequence like $N_1$ $N_2$-**kʷe h₁ilo-* ±'$N_1$ and $N_2$ following' > $N_1$ $N_2$-*śə-yəle* might regularly become TA *-aśśäl* with assimilation of pre-TA syncopated *-śy-* to explain the geminate *-śś-*, the individual elements needed for this etymology lack proper support.

The development of the preposition from this original suffix or clitic element also remains difficult. Kim points, as a parallel, to Ossetic *æd* 'with' and *ænæ* 'without', which are the only prepositions in a language otherwise dominated by propositions, similar to TB *śale* 'with' and *snai* 'without' (Kim 2013: 136). This does not, however, amount to an explanation as to how or why *śale* would have become a preposition if it originated as a suffix (cf. the criticism of Koller 2018: 172). Additional evidence would be needed to support the notion that there is a general pressure for elements meaning 'with' or 'without' to precede the noun phrase they modify, regardless of their origins. As such, Kim's etymology for the Tocharian A comitative *-aśśäl* and the preposition TB *śale*, *śle* ~ TA *śla-* cannot easily be accepted.

Koller has proposed to explain the geminate *-śś-* in *-aśśäl* as the result of a combination of *-śäl*, from the adposition, with the ablative suffix *-ac*: **-ac-śäl* → *-aśśäl* (Koller 2018: 172–173). This would account for the parallel existence of a simpler comitative suffix *-śäl*, which is especially found in combination with obliques in *-n* (Koller 2018: 168–169). The preposition PT **kʸəle* would have developed into a postposition in Tocharian A, and optionally selected the allative case of the noun in governed (Koller 2018: 173–182). While this might provide a pathway for the development of the comitative suffix within Tocharian A, it does not bring us closer to its ultimate origins.

The Tocharian B comitative *-mpa* is quite as mysterious as the TB ablative *-meṃ*. Van Windekens once proposed a Uralic origin, from the suffix PU **-mpA* with original contrastive semantics (see Ylikoski 2018), but he later abandoned this etymology (Van Windekens 1979: 253), and with good reason, as the semantic gap is quite great. Kim (2014:

133–134) has suggested a derivation from the word for 'both' (TB *antapi*, TA *āmpi*) from PT **ant(ə)pi*. The problem is that the final **-a* of a hypothetical pre-form **ant(ə)pa* is not found in either Tocharian language. Kim's solution is to reconstruct **ant(ə)pay* with a reduction of **-ay* to **-a* in unstressed syllables; e.g., TB *oksaine* /oksáyne/ 'two oxen', *pokaine* /pokáyne/ 'two arms' vs. *ckāckane* /ckáckane/ 'to shanks' (ibid.). Neither the existence of a form **ant(ə)pay* nor the development from **-ay* to **-a* as described by Kim is certain, however. Especially in final position, the obliques TB *oksai* 'ox', *pokai* 'arm' and *ckāckai* 'shank' rather suggest that there was at least no regular simplification of an unaccented **-ay* in that position.

Adams suggests a formation based on PIE **me* 'in the middle, around, with' as in Gr. μετά 'among, with', Goth. *miþ* 'with' and some other particle (Adams 2013: 514). Akao dismisses the possibility of a nasal element altogether, with an interpretation of TB ⟨mp⟩ as [β] rather than [mp] (cf. Pinault 2009: 236–240) and the lack of a full vowel in the independent particle TB *mpa* ±'like, as if' rather than ***mapa* disqualifying any interpretation of this word containing a separate nasal element (Akao 2020). However, a similar situation is found in *spe* 'near' from PIE **supo-* 'under', which is not found as expected *sape** (Adams 2013: 788), so that the existence of independent *mpa* does not disprove the possibility of an earlier form **mVpa*. In any event, a convincing etymology has still not been found.

It is thus not at all clear-cut whether we should reconstruct a Proto-Tocharian comitative on the basis of TA *-aśśäl* or TB *-mpa*, seeing as no compelling etymology is available for either. It could be that both are independent innovations, and that Proto-Tocharian did not have a comitative. At the same time, we may consider that PT ±**-śəle* based on TA *-aśśäl* would have becomeTB ***-śle*, which is more salient than the expected Tocharian A reflex of a PT **-mpa* based on TB *-mpa*: this would have become TA ***-mp* or perhaps just ***-m* (see Peyrot 2022b: 164–165 on the correspondence of TB final *-mp* to TA *-m*). Just a final ***-m* latter is not particularly salient, and may have been more vulnerable to a secondary replacement with an innovative suffix like TA *-aśśäl* (cf. Peyrot 2019a: 93). For this reason, **-mpa* as found in Tocharian B could be regarded as the older suffix, although this is by no means assured.

#### 6.2.2.3 A Proto-Tocharian dative-locative?

Peyrot (2012a) has suggested that there existed also an old dative-locative in Proto-Tocharian. This is based on a correspondence between Tocharian B *ā*-stem oblique singular forms in *-ai* and Tocharian A *ā*-stem genitive singular forms in *-e*. The reconstruction of TB *-ai* and TA *-e* is PT **-ay*, which could go back to **-eh₂-i* or **-eh₂-ei*. The semantics of this suffix could not have been either genitive or oblique already, since Tocharian A and B disagree on this point. The Tocharian B oblique function could not be derived from a genitive, which normally remains clearly differentiated, but since the

genitive functions as a dative, the Tocharian A genitive function could be derived from an old dative without any issues.

An element *-*ay* is also used to form adverbs such as TA *spānte* 'confidently', TB *spantai* 'trustingly'. In Tocharian B, such adverbs with a locative meaning generally have a locative prefix like *en-* (e.g., *enestai* 'in secret'), so that the locative meaning in Tocharian B need not be derived from the ending. The same cannot be said for Tocharian A, however, which thus implies that an original locative function should be attributed to the ending after all. With this argumentation, Peyrot concludes that there may well have been an additional case in Proto-Tocharian, combining dative and locative functions, at least for *ā*-stems (Peyrot 2012a: 204–207).

#### 6.2.2.4 The status of the secondary cases in Proto-Tocharian

Up to this point, the identity of the secondary cases as proper case forms has been taken for granted. However, the relationship between the secondary case endings and the oblique used as the base are different in Tocharian A and B in some interesting ways, which could have implications for their status in Proto-Tocharian.

In Tocharian A, it is clear from the regular apocope of final vowels in, e.g., loc. TA *-ṃ* vs. TB *-ne*, that the secondary cases and the noun they are attached to form a single accentual unit. Otherwise, these monosyllabic elements should have kept their vowel, had they been independent (Carling 1999: 99). Furthermore, an epenthetic *-y-* was added between those secondary cases starting with a vowel and an oblique stem ending in *-i*, *-e* or *-o*; this is not a normal sandhi development between two words, and thus points to the status of the secondary cases forms of nouns as single units (Kölver 1965: 4–5). The fact that the secondary case endings could receive an analogical initial vowel that originally belonged to the end of the noun also supports this.

In Tocharian B, on the other hand, a lack of unity is indicated by several facts. The accent is not attracted by the addition of secondary case suffixes except the ablative, contrary to the usual pattern: TB *lakle* /ləkle/ has a gen.sg. *läklentse* /ləkléntse/, as expected, but a per.sg. *laklesa* /ləklesa/, com.sg. *laklempa* /ləklempa/, loc.sg. *laklene* /ləklene/ (see also 5.4.4 and the references therein). In a few cases, a secondary case ending also occurs separated from the oblique stem it belongs to, as in *ṣkas meñantse ne* 'on the sixth of the month' and *ṣkas meñantse meṃ* 'from the sixth of the month', where the locative *-ne* and the ablative *-meṃ* are separate from *ṣkas* 'sixth' by the intervening genitive *meñantse* 'of the month' (Carling 1999: 99–100). This may reflect a continuation of the original independence of the case ending (cf. also Carling 2012: 60). On the other hand, the development of **ns* to **nts* in the perlative plural is a word-internal development, indicating that there was no word boundary between the oblique and the secondary case suffix; e.g., per.pl. *ñäkteṃtsa* /ńəkténtsa/ 'among gods' from the obl.pl. *ñäkteṃ* /ńəktén/ + per. *sa*, (Carling 1999: 100).

Tocharian B thus presents us with a most contradictory picture. However, some aspects of the unusual behaviour of the secondary case endings can also be understood as later innovations. The apparent lack of accentual unity may be the result of a secondary development: the local case forms are clearly based on the oblique case, which usually does not show any accent movement, as in nom.sg. *lakle*, obl.sg. *lakle*, both /lə́kle/. If the secondary case endings were simply (re-)added to the synchronic oblique, no accent movement would be expected; and after all, the accent rules of Tocharian B were not synchronically automatic anymore (cf. 5.4.4). Those oblique singulars that do have a different accent from the nominative retain this in the secondary case forms; e.g., *saswe* /sə́swe/ 'lord' with the obl.sg. *säsweṃ* /səswén/, com.sg. *säsweṃmpa* /səswénmpa/, and abl.sg. *säsweṃmeṃ* /səswénmen/. The peculiar examples like *ṣkas meñantse ne* 'on the sixth of the month' might be understood if *ṣkas meñantse* 'the sixth of the month' and other, similar expressions of dates were conceived of as more of a single unit. In such a regular type of expression, this could have been a later development. I think that the general behaviour of the Tocharian secondary endings suggests that they were normally phonologically merged with the base noun and fully integrated with the nominal paradigm in Proto-Tocharian already.

#### 6.2.2.5 Overview

In Table 6.3 below, the secondary case endings of Tocharian A and B are presented with their Proto-Tocharian counterparts. In the case of the ablative, the suffix of Tocharian A (*-äṣ*) is given precedence and is reconstructed for Proto-Tocharian, on account of its plausible Proto-Indo-European origins. The comitative cannot be reconstructed with any semblance of certainty. Where an etymology is available, the Proto-Indo-European adposition or adverbial element is given as well. These elements were not really part of the Proto-Indo-European case paradigm, and only became case markings in the prehistory of Tocharian, after much of the original case system had been lost (cf. Carling 1999: 98).

Table 6.3: The reconstructed Proto-Tocharian secondary case endings together with their most likely Proto-Indo-European origins and the Tocharian A and B forms.

| | TA | TB | PT | PIE |
|---|---|---|---|---|
| allative | *-ac* | *-śc* | **-cə* | **=$d^he$* |
| locative | *-aṃ* | *-ne* | **-ne* | **$h_2$nom* |
| ablative | *-äṣ* | *-meṃ* | **-ṣə* | **-ti* (or *$h_1$eti*?) |
| perlative | *-ā* | *-sa* | **-a* | **$h_2$ed* |
| comitative | *-aśśäl* | *-mpa* | ?**-mpa* | ?? |

## 6.3 The Samoyed case system

We have seen the Tocharian secondary case system in the previous section, but to make a comparison it needs to be established what the corresponding system in Samoyed looked like. Some Uralic languages are famous for their extensive local case system, such as Hungarian with eighteen cases or Finnish with fifteen. Many of these cases denote a location 'at' and movement 'to' or 'from', as well as a distinction between 'in', 'on' or 'near'. However, such extensive inventories are not inherited from Proto-Uralic. Instead, various Uralic languages increased their local case systems from an original set of three: a primary directional or "lative" *-*ŋ*, a locative *-*nA* and an ablative *-*tA* (see, e.g., Ylikoski 2011: 235).

Most languages from the Samoyed branch of Uralic do not have case systems that are as elaborate as those of Hungarian and Finnish. In addition to the grammatical cases nominative *-*∅*, accusative *-*m* and genitive *-*n* inherited from Proto-Uralic, the lative, locative and ablative also each have their Samoyed counterpart. The original suffixes are only preserved on their own in the inflection of relational nouns used as postpositions and in adverbs. The case suffixes of regular nouns, by contrast, are extended with coaffixes PS *-*ntə* and *-*kə* (Janhunen 1998: 469). The Samoyed languages also have a prolative *-*mənA*, which cannot be reconstructed for Proto-Uralic. Further information on the functions of these various cases is provided in the subsections of 6.3.1.

For a comparison with the Tocharian secondary case system, it is necessary to gain a clear picture of the Samoyed situation, and to determine the age of the cases and their various uses. The present section will provide an overview of the different cases in the Samoyed languages and their functions. It appears that the Uralic languages in general, including Samoyed, do not have Gruppenflexion as found in Tocharian; only adjective agreement in Nganasan shows some vague parallels, since it only applies to the core cases, but adjective agreement of this sort is understood to be a late innovation in Samoyed (e.g., Rießler 2016: 218). The formal reconstruction of the Proto-Samoyed and pre-Proto-Samoyed case systems will be discussed in 6.3.2.

### 6.3.1 The Samoyed cases and their functions

In this section, the functions of the local cases will be considered, as well as various comitative expressions in Samoyed and some more restricted cases or case-like elements found in only some Samoyed languages.

The Samoyed local case suffixes are usually accompanied by an additional element, a coaffix *-*kə-* or *-*ntə-*. These coaffixes are not used on their own, only in combination with the case suffixes, hence their name; they will be further addressed below, in 6.3.2. Only the prolative does not have a coaffix.

#### 6.3.1.1 Lative

The lative in Samoyed typically marks a direction of movement or a goal (example 15). This is also the case that is generally used to indicate the recipient or the beneficiary (indirect object) of a ditransitive verb (example 16). Accordingly, it is often called "lative/dative" or "dative" in grammatical descriptions; I will only use the term "lative" to avoid confusion.

15. Tundra Nenets (Nikolaeva 2014: 62; glosses adapted)
    *lam°pa-n°h* *xərasyin°-m* *xəmta-q*
    lamp-LAT kerosene-ACC pour-SBC.IMP.2SG
    'Pour some kerosene into the lamp.'

16. Forest Enets (Khanina & Shluinsky 2023: 836; glosses adapted)
    *tʃukutʃi* *buniki-xiz* *miʔ-e-ɲitʃ*
    all dog-LAT.PL give-OBC.PL-1DU.PST
    'We gave them all to dogs.'

It appears that all Samoyed languages agree on the function of the lative as a directional case with both illative and allative functions, translatable as '(in)to'. Since the use of this case to mark the indirect object is also widely attested, this was probably a function of this case in Proto-Samoyed as well.

More information about the lative in the Samoyed languages and additional examples can be found in, e.g.: Wagner-Nagy (2019: 196-197) for Nganasan; Khanina & Shluinsky (2023: 801 and *passim*) for both Tundra and Forest Enets; Siegl (2013: 159-160) for Forest Enets; Nikolaeva (2014: 62) for Tundra Nenets; Sammallahti (1974: 36) for Forest Nenets; Kuznecova, Helimski & Gruškina (1980: 179–180) for Taz (Northern) Selkup; Künnap (1971: 72–77, 1999: 16) for Kamas.

#### 6.3.1.2 Locative

The locative case is primarily used to mark a location inside or on something. It also marks an instrument in the northern Samoyed languages, as these do not have a dedicated instrumental case. For human referents, the locative can be used to express a comitative in at least Nenets and Enets (Nikolaeva 2014: 64; Siegl 2013: 161).

17. Kamas (Künnap 1999: 17; glosses adapted)
    ***orō-γân*** *ćüpi* *būzüj* *iʔbə*
    hole-LOC wet calf lie:PRS.3SG
    'A wet calf is lying in the hole.'

18. Tundra Nenets (Nikolaeva 2014: 64; glosses adapted)
    *Wera-xəna* *to-°-∅*
    PN-LOC come-AOR-SBC.3SG
    'He came with Wera.'

The Proto-Samoyed locative probably had the same function as the attested locatives, expressing location. It is also possible that an instrumental meaning was already available, as a secondary development from the more primary locational semantics. It is easy to imagine a re-analysis of instances like 'catch in a net' → 'catch with a net', 'travel in a boat' → 'travel with a boat' and 'hold in hands' → 'hold with hands' as a way in which instrumental use of the locative might have spread. A separate instrumental cannot be reconstructed for Proto-Samoyed.

More information about the locative in the Samoyed languages and additional examples can be found in, e.g.: Wagner-Nagy (2019: 197–199) for Nganasan; Khanina & Shluinsky (2023: 801 and *passim*) for both Tundra and Forest Enets; Siegl (2013: 161) for Forest Enets; Nikolaeva (2014: 63–64) for Tundra Nenets; Sammallahti (1974: 38) for Forest Nenets; Kuznecova, Helimski & Gruškina (1980: 180–181) for Taz (Northern) Selkup; Künnap (1999: 16–17) for Kamas.

#### 6.3.1.3 Ablative

The ablative case is used to denote movement away from a place, marking the source of the movement. The standard of comparison is also marked in the ablative in all Samoyed languages for which comparative constructions are known; the formation and syntax of comparatives in Mator is unknown (Helimski 1997: 145).

19. Nganasan (Wagner-Nagy 2019: 199; glosses adapted)
    *ďesi̯-mə* *büü-ʔnar-u* ***gərədə-gətə***
    father-POSS.NOM.SG.1SG go-FREQ-AOR.3SG town-ABL
    'My father often leaves (goes **out of**) **town**.'

20. Selkup (Kazakevič 2022: 792; glosses adapted)

| | | | |
|---|---|---|---|
| *buran* | ***ɔːtæː-qɨnɨ*** | *tʃʲatkɨ-sɨmɨʎ* | *ɛːŋa* |
| PN | reindeer-ELA[73] | rapid-ADJ | be:3SG |

'A Buran (snowmobile) is faster **than a reindeer**.'

The Proto-Samoyed ablative must have had the same canonical ablative functions as attested in all Samoyed languages, including a use as marking the standard of comparison.

More information about the ablative in the Samoyed languages and additional examples can be found in, e.g.: Wagner-Nagy (2019: 199–200) for Nganasan; Khanina & Shluinsky (2023: 801 and *passim*) for both Tundra and Forest Enets; Siegl (2013: 162–163) for Forest Enets; Nikolaeva (2014: 64–65) for Tundra Nenets; Sammallahti (1974: 38–39) for Forest Nenets; Kuznecova et al. (1980: 181–182) for Taz (Northern) Selkup; Künnap (1999: 16–18) for Kamas.

#### 6.3.1.4 Prolative

The prolative marks movement across or along an object. This function of the prolative can be found in Nganasan, Enets, Nenets and Selkup, and can thus be straightforwardly assumed for Proto-Samoyed as well. In Kamas the prolative is only found in a few adverbs (Mikola 1988: 238; Klumpp 2022: 824). The prolative stands apart from the other local cases in that is has no coaffix.

21. Tundra Nenets (Nikolaeva 2014: 65; glosses adapted)

| | | |
|---|---|---|
| ***yəxa-w°na*** | *buksir* | *myiŋa-∅* |
| river-PRO | tugboat | go:AOR-SBC.3SG |

'A tugboat is moving along the **river**.'

More information about the locative in the Samoyed languages and additional examples can be found in, e.g.: Wagner-Nagy (2019: 200–201) for Nganasan; Khanina & Shluinsky (2023: 801 and *passim*) for both Tundra and Forest Enets; Siegl (2013: 165) for Forest Enets; Nikolaeva (2014: 65–66) for Tundra Nenets; Sammallahti (1974: 40) for Forest Nenets; Kuznecova, Helimski & Gruškina (1980: 183) for Taz (Northern) Selkup; Künnap (1999: 16–17) for Kamas.

[73] The Proto-Samoyed ablative became the Selkup elative, more specifically used for internal movement, and restricted to inanimates.

#### 6.3.1.5 Comitative constructions

There are various strategies to express a comitative in Samoyed. In Nganasan, the postposition *na* 'to, near' can be used as a comitative: *ńuə-nə-na* 'with my child' (Wagner-Nagy 2019: 188–189), or a sociative derivational suffix *-səbtə* may be employed (Wagner-Nagy 2019: 329). Nenets and Enets use the locative for a comitative meaning (Sammallahti 1974: 37–38; Nikolaeva 2014: 64; Burkova 2022: 684; Khanina & Shluinsky 2023: 801). A comitative or proprietive adjective can be derived in Enets with EnT *-sae*, EnF *-saj*, e.g., EnT *niɔsae* 'with a child, with children' (Khanina & Shluinsky 2023: 801). The Nenets counterpart to this is *-sawey*°, e.g., *yísawey*° 'intelligent' from *yí* 'mind', *nyesawey*° 'married' from *nye* 'woman' (Nikolaeva 2014: 32–33); the suffix can be straightforwardly reconstructed as **-sAmAjə̂*. Kamas and Selkup can use the instrumental as a comitative (Künnap 1971: 132; Kuznecova, Helimski & Gruškina 1980: 175; 2002: 88; Kazakevič 2022: 791; Klumpp & Budzisch 2023: 908–909).

22. Taz Selkup (Kuznecova et al. 1980: 175; glosses added)

| | | |
|---|---|---|
| *man* | ***ima-ny-sä*** | *ila-k* |
| 1SG | wife-GEN.POSS.1SG-INS | live-SBC.1SG |

'I live with my **wife**.'

No dedicated comitative case can be reconstructed for Proto-Samoyed.

#### 6.3.1.6 Additional cases

There are a few other suffixes with a more limited distribution in Samoyed that can be considered cases. There is the predestinative, which is found in the northern Samoyed languages and marks something for the benefit of someone. It is usually combined with the possessive suffixes, and it has been variously interpreted as nominal tense marking (e.g., Nikolaeva 2015) or as a separate declension type (e.g., Siegl 2013). Despite its limited distribution in only the northern Samoyed languages, the predestinative suffix can be reconstructed for Proto-Samoyed as **-tə-* and connected with the Proto-Uralic translative **-ksi*, which expresses changes of state: the phonological correspondence is entirely regular, and there is an overlap in the use of both suffixes (see Janhunen 1989; Salminen 2014; Ylikoski 2017, Aikio 2022a: 15). It is unclear what the status of the predestinative was in Proto-Samoyed and early stages of pre-Proto-Samoyed. It may have remained as a translative for a long period of time.

23. Tundra Nenets (Nikolaeva 2014: 76; glosses adapted)

| *tyuku°* | *wenyako* | *mənyaq* | ***wenyako-d°-naq*** | *xəya-∅* |
|---|---|---|---|---|
| this | dog | 1PL | dog-DST-POSS.GEN.SG.1PL | go:AOR-SBC.3SG |

'This dog became **our dog**.'

The southern Samoyed languages Selkup and Kamas have an instrumental case in Sk. *-sä/-se* and Km. *-ʒʲəʔ/-ʒʲiʔ* respectively (Klumpp & Budzisch 2023: 905; Klumpp 2022: 827). Selkup furthermore has a number of additional cases such as a caritive in *-qalɨk*, an essive-translative *-ŋgo* and a translative in *-wle* (e.g., Klumpp & Budzisch 2023: 905). An additional set of specifically animate local cases in Selkup is based on the Proto-Samoyed postposition **na-* 'at, near' (SW 99 under **nä-*); this same postposition is used in the northern Samoyed languages to periphrastically express the local cases of the dual (e.g., Wagner-Nagy 2019: 191 on Nganasan; Nikolaeva 2014: 57 on Tundra Nenets).

### 6.3.2 The Proto-Samoyed and Proto-Uralic case systems

In this subsection, I will give an overview of the origins of the Samoyed cases, summarising the literature on the topic. I put forward a slightly updated reconstruction of the plural local cases relative to Janhunen (1998) and Däbritz (2017), which allows for a more uniform interpretation of their formation in Nganasan, Enets and Nenets, and subsequently, a more coherent reconstruction of the Proto-Samoyed plural local cases as a morphological system. In this way, the relationship between the formation of the plural local cases and the plural genitive is highlighted, although it does not affect the details of the eventual comparison with Tocharian.

The lative, locative, ablative and prolative, found in all Samoyed languages, are reconstructed for Proto-Samoyed (Janhunen 1998: 469). The coaffixes **-ntə-* and **-kə-* are present in the lative, locative and ablative, while the prolative does not have a coaffix. Crucially, the Samoyed languages do not agree on which co-affix is used in the locative: Nganasan uses **-ntə-* there (Ng. loc. *-NTənU*), while the rest of the Samoyed languages use **-kə-* (e.g., NeT loc. *-xəna*, Mt. loc. *-kənA*). Janhunen interprets this as a sign that the local case system was not yet fully formed in Proto-Samoyed, before Nganasan split off from the others (Janhunen 1998: 469). In the plural, only the northern Samoyed languages have special case suffixes. Selkup and Kamas both add the singular case endings to the nominative plural in a fully agglutinating manner (cf. Klumpp 2022: 825; Klumpp & Budzisch 2023: 905). The Proto-Samoyed plural case endings can be reconstructed based on the northern Samoyed languages as in Table 6.4.

Table 6.4: Proto-Samoyed local cases, singular and plural, based on Janhunen (1998: 470), though with a slightly different reconstruction of the Nenets (and Enets) plural case suffixes, without *-*j*-.

| PS | singular | plural |
|---|---|---|
| nom. | *-∅ | *-*t* |
| acc. | *-*m* | *-*j* |
| gen. | *-*n* | *-*j-t* |
| lat. | *-*ntə-ŋ* | *-*ntə-j-t* (Ng) / *-*kə-j-t* (En-Ne) |
| loc. | *-*ntə-nA* (Ng) / *-*kə-nA* (rest) | *-*ntə-j-t-nA* (Ng) / *-*kə-j-t-nA* (En-Ne) |
| abl. | *-*kə-tə* | *-*kə-j-t-tə* |
| pro. | *-*mənA* | *-*j-t-mənA* |

The basis of the plural local cases in Proto-Samoyed was reconstructed as two sets by Janhunen (1998: 470), based on the plural nominative to account for Nenets, and the plural genitive to account for Nganasan. In my view, the plurals of the local cases as found in the northern Samoyed languages are all formed with the addition of the plural genitive PS *-*j-t*, most clearly in the case of the prolative. This is inserted after the coaffix in the case of the lative, locative and ablative, while it comes before the prolative suffix in its entirety. No *-*j*- is visible in Nenets, probably due to a sound change from *-*əjtC*- to pre-PNe. *-*itC*- to PNe. *-*əqC*- (cf. Salminen 2012: 344 on development from *-*i* to PNe. *-*ə* in non-initial syllables; both Enets and Nganasan must have obtained their vowel -*i*- from this *-*j*-). Meanwhile, no *-*t*- is visible in the locative plural of Nganasan and Enets[74], but this would probably have regularly disappeared in a sequence like *-*tn*-, meaning that it can be reconstructed. Cf. the cognate set for the verb 'lie down (of animals)' in the northern Samoyed languages (not in SW), where only Nenets shows a cluster -*qn*- (= -*ʔn*-):

> PS **wåtᵊnə̂*- 'lie down (of animals) : sich legen (von Tieren) : лечь, улежать (о животном)' || Ng. *bənəďa* (NgSl.) | EnT *bano*- (EnSl.) | EnF *banu*- (EnWl.) | NeT *waqnə*(*sy*°) (T65) | NeF *waqnăsy*° (NeSl.) || see Kaheinen (2023: 270)

Apparently, the lative plural did not contain the lative suffix *-(*ə*)*ŋ*, seeing as it is absent in both Nganasan and Nenets. Only Enets seems to have added it, as it has a lat.pl. EnT -*kizo*, EnF -*kiz* from PEn. *-*kitə*, where the *-*t* that became a word-final glottal stop in Ng. -*NTəʔ* and NeT -*xəq*, NeF -*xăq* is intervocalic instead (Däbritz 2017: 74). The singular lative suffix PS *-*ntəŋ* does not have an *-*ŋ* in EnT -*do*, EnF -*d*, so final -*ŋ* was apparently lost in Enets.

[74] Cf. Mikola (2004: 103) and Däbritz (2017: 74), who reconstruct only *-*j*-, i.e., *-*ntə-j-nA* for Nganasan (so, too, Janhunen 1998: 469) and *-*kə-j-nA* for Enets.

Table 6.5: Case endings in Nganasan, Enets and Nenets. In the notation of Nenets forms, the glottal stop is written as *q*. Tundra and Forest Enets from Khanina & Shluinsky (2023: 799–802), Nganasan from Wagner-Nagy (2019: 193), Tundra Nenets from Nikolaeva (2014: 59) and Forest Nenets adapted from Sammallahti (1974: 41).

| | Nganasan | Enets | Nenets |
|---|---|---|---|
| nom. | *-ʔ* | **-ʔ* > T *-ʔ*, F *-ʔ* | **-ʔ* > T *-q*, F *-q* |
| acc. | *-j* | **-ʔ* > T *-ʔ*, F *-ʔ* | **-j* > T *-∅*[(a)], F *-∅*[(a)] |
| gen. | *-ʔ*[(a)] | **-ʔ* > T *-ʔ*, F *-ʔ* | **-jʔ* > T *-q*[(a)], F *-q*[(a)] |
| lat. | *-NTiʔ* | **-kitəŋ* > T *-xizo*, F *-xiz* | **-kəʔ* > T *-xəq*, F *-xăq* |
| loc. | *-NTi$^{c}$nU* | **-ki(ʔ)na* > T *-xine*, F *-xin* | **-kəʔna* > T *-xəqna*, F *-xăqna* |
| abl. | *-Ki$^{c}$tə* | **-kiʔto* > T *-xito*, F *-xit* | **-kəʔtə* > T *-xət°*, F *-xăqt°* |
| pro. | *-ʔmənU*[(a)] | **-i(ʔ)mna* >> T *-ine*[(b)], F *-in*[(b)] | **-jʔməna* > T *-qməna*[(a)], F *-qmăna*[(a)] |

(a) The earlier presence of the plural suffix **-j* in these cases can be seen in the stem-final vowel alternations; e.g., Ng. *kümaa* 'knife' with acc.pl. *kümaa-j*, gen.pl. *kümau-ʔ*, pro.pl. *kümau-ʔmənu* (Wagner-Nagy 2019: 193).

(b) Olesya Khanina has pointed out to me that the Enets forms in *-ine* and *-in* are secondary, with older sources attesting to regular *-ʔone* and the intermediate *-ʔine*.

For the coaffix **-ntə-*, Mikola (2004: 99–100) reconstructs a sequence of two lative endings **-n* and **-tə*. Leaving the origins of **-n* aside, Mikola supposed that the second lative, **-tə*, might be connected to similar elements with an **s* in the western Uralic *s*-cases found in Saami, Finnic, Mordvin and Mari; after all, PU **s* is regularly reflected as PS **t*.

Like Mikola, Ylikoski suggests that there may be a connection between the *s*-cases and the Proto-Samoyed coaffix **-ntə-*. However, he derives the **s* in the *s*-cases not from a lative suffix, but rather from an old set of theoretical postpositions derived from a locational noun that became a coaffixal element. This provides a more explicit functional derivation of the different elements that make up these case suffixes. The lative PS **-ntəŋ* would then be direct cognate of SaaS *-se*, Fi. *-hVn*, MdE *-s* and MariE *-š*, from PU **-[n]sVŋ*, while the locative PS **-ntənA* reflected in Nganasan would be shared with SaaS *-sne*, Fi. *-ssA*, MdE *-se*/*-so* and MariE *-što* from PU **-[n]sVnA* (Ylikoski 2016: 53–59).

Gusev has instead argued that the coaffixes PS **-ntə-* and **-kə-* go back to derivational suffixes that formed (locational) denominal nouns: Ng. *ŋuəntəə* 'tsar' derived from *ŋuə* 'cloud, air, deity, spirit' with **-ntə-jə*, EnT *ŋaaxi* 'tsar' from *ŋaa* 'god, weather' with **-kə-jə*, NeT *numki°* 'star' from *num* 'weather, sky, heaven, god' with **-kə-jə* (Valentin Gusev: teaching material Helsinki Samoyed Summer School 2023). The position of the plural markers after the coaffix (i.e., **-kə-jt-* rather than **-jt-kə-*) lends credence to this theory, since derivational suffixes would more naturally precede the plural endings than would

an agglutinated relational noun. This is also seen in the fact that the plural morphology in the prolative precedes the entire suffix (cf. below on the origin of the prolative).

The discrepancy between Nganasan and the other Samoyed languages is still interesting. It is difficult to imagine a motivation for either the coaffix *-*kə*- or the coaffix *-*ntə*- to spread from its original locus in either part of the Samoyed area, since either direction of spread assumes that a co-affix that was originally present in only one case (the lative for *-*ntə*- and the ablative for *-*kə*-) could oust its (apparently) semantically equivalent but necessarily more well-represented counterpart. It could be that the Samoyed case system as we see it, with its fully grammaticalized coaffixes, was not yet formed in Proto-Samoyed, but that the use of (perhaps) denominal nouns in combination with the local case suffixes, the probable basis for the later coaffixes, was already commonplace. The choice of which type of coaffix to grammaticalize could then have been a later phenomenon. The agreement of the other Samoyed languages vis-à-vis Nganasan provides one of the more salient arguments for an early phylogenetic split of Nganasan from the rest of Samoyed (Janhunen 1998: 459–469).

In any event, the functional cases lative, locative and ablative can definitely be reconstructed for Proto-Samoyed and pre-Proto-Samoyed without problem, as all three derive from Proto-Uralic, from the lative *-*ŋ*, the locative *-*nA*, and the ablative *-*tA* (e.g., Janhunen 1998: 469; Aikio 2022a: 14). The plurals were formed based on the genitive plural suffix PS *-*j-t*, which was added before the local case suffix.

It is more difficult to determine the status of the prolative. Janhunen suggested that it is in origin a combination of the accusative *-*m* and the locative *-*nA* (Janhunen 1998: 469), but Jalava has recently proposed that it can be derived from the locative or the Proto-Uralic noun **muka* 'back', or perhaps **müŋä* 'back'. Only the former has an independent reflex in Samoyed, as PS **mə̑kå* 'back', which became, e.g., Ng. *məku*, NeT *məxa*, Mt. *baga*, etc. The original shape of the postposition **muka-na* lit. 'on the back' would have become truncated in the process of becoming a case ending to PS *-*mənA* (Jalava 2022). At what point in pre-Proto-Samoyed the suffixation of PU **mukana* (or **müŋänä*) as a prolative case suffix took place cannot be determined. An unrelated prolative in *-*ko* can be reconstructed for the shared ancestor of Saami, Finnic and Mordvin (Aikio & Ylikoski 2016: 61), but it does not seem to be reconstructible for Proto-Uralic with the current state of our knowledge; the Permic prolative continues the Proto-Uralic ablative *-*tA* (Riese 1998: 268–269).

Table 6.6: Proto-Samoyed case endings and their Proto-Uralic origins.

| | PS | PU |
|---|---|---|
| nominative | *-∅ | *-∅ |
| accusative | *-*m* | *-*m* |
| genitive | *-*n* | *-*n* |
| lative | *-*ntə-ŋ* / *-*ŋ* (postp.) | *-*ŋ* |
| locative | *-*ntə-nA* (Ng) / *-*kə-nA* (rest) / *-*nA* (postp.) | *-*nA* |
| ablative | *-*kə-tə* / *-*tə* (postp.) | *-*tA* |
| prolative | *-*mənA* | ?**mukana* / **müŋänä* |

As an additional, non-local case, the Proto-Uralic translative *-*ksi*, can be reconstructed for pre-Proto-Samoyed as well. This became the predestinative PS *-*tə*- (Janhunen 1989; cf. 6.3.1.6).

## 6.4 Comparison of the Tocharian and Samoyed case systems

In this subsection, the case systems of Tocharian and Samoyed will be compared in terms of their non-core case inventories, the functions of the individual cases, and the ways in which new cases were added to the system.

### 6.4.1 Case inventory

Considering the pre-Proto-Tocharian and (pre-)Proto-Samoyed case systems side by side, there are certainly points of agreement, but also some discrepancies. Among the local cases, the main difficulty is the age of the Samoyed prolative, which cannot be confidently determined. There furthermore seems to be no dedicated comitative at any time in the history of Samoyed, which means that the Tocharian comitative cannot receive an explanation from contact with this language group. The limited continuation of a dative-locative in Proto-Tocharian is not matched with Samoyed either, nor is the Tocharian Gruppenflexion.

Table 6.7: The Proto-Tocharian and Proto-Samoyed secondary case systems compared.

| PIE | PT | PS | PU |
|---|---|---|---|
| (allative) | *new* allative | lative | lative |
| dative | ?? *old* dative-locative | — | — |
| locative | *new* locative | locative | locative |
| ablative | *new* ablative | ablative | ablative |
| — | *new* perlative | prolative | — |
| — | ?? *new* comitative | — | — |

### 6.4.2 Functional comparison

If we turn to the functions of the individual cases as they can be reconstructed for Proto-Tocharian and Proto-Samoyed, and assume, for the sake of the argument, that the same functions can be attributed to their earlier stages, the correspondence is still imperfect. Due to the lack of a Samoyed comitative, only the local cases can be compared.

The allative in Tocharian is not used to indicate movement into an object, whereas this is one of the functions of the Samoyed lative. Tocharian furthermore does not use the allative to mark a recipient, unlike Samoyed, as in Tocharian this function is normally fulfilled by the genitive instead.

The Tocharian locative can express not only location, but also movement. This means that it functionally overlaps with both the Samoyed locative and lative. Expressions of static location and movement into and from are generally kept separate in Samoyed.

The Tocharian perlative also has more functions than can be reconstructed for the Samoyed prolative: unlike the Tocharian perlative, the Samoyed prolative is not used to mark a static location. Furthermore, no instrumental function is known for the Samoyed prolative, while this is one of the functions of this case in Tocharian B; in (northern) Samoyed, the locative is used as an instrumental instead. The additional functions of the Tocharian perlative can be considered as secondary derivation from the core perlative function (Krause & Thomas 1960: 84), which is shared with Samoyed.

Only the ablative case appears to have the same meaning in both Proto-Tocharian and Proto-Samoyed. If the Tocharian A use of the ablative for the standard of comparison is taken to be original, the match is quite exact.

These differences could be due to secondary innovations from a shared original meaning, in particular in the Tocharian locative and perlative. Specialized meanings, such as the Tocharian B instrumental perlative use and the (northern) Samoyed instrumental locative are clearly different, independent developments from the original core local meanings of these cases. Perhaps a certain semantic range was also dictated by the Proto-Indo-European origins of these secondary cases: for example, the adposition *$*h_2ed$*, which can be connected with the Tocharian perlative, is attested with both directional meaning (Lat. *ad* 'to, up to, into') and locational meaning (as in English *at*). It may therefore be expected that a case ending created on the basis of this adposition would also incorporate both meanings. Furthermore, from a canonical prolative meaning 'across', a more static sense of 'on' could be derived in the sense that when, for instance, a blanket is spread out over the grass, it is then also on the grass, and one has to be on the tundra to move across or over it.

Table 6.8: A comparison of the Proto-Tocharian and Proto-Samoyed case functions; the first function listed it he most original (local) function, with other (grammatical) functions listed second.

| Case / functions in: | Proto-Tocharian | Proto-Samoyed |
|---|---|---|
| (al)lative | direction of movement. | direction of movement; movement into; recipient. |
| locative | internal location; movement into. | location; instrument. |
| ablative | source; starting point; standard of comparison? | source; starting point; standard of comparison. |
| perlative | path or direction of movement; external location; cause; instrument? | path of movement. |
| (comitative) | (accompaniment.) | — |

### 6.4.3 Comparison of development

The precise details of how the case systems of Tocharian and Samoyed developed are not entirely clear or agreed on at the present state of research. However, it is clear that both recruited earlier postpositional elements to form new cases, a strategy that continued in Selkup with the creation of additional local cases for animate nouns based on the postposition PS **na-* 'near, to'. The same type of expansion of the case system is likely found in western Uralic languages (Aikio & Ylikoski 2016). A discrepancy is found between Tocharian and Samoyed in the noun case to which the postpositions were added: Tocharian secondary cases are based on the oblique/accusative, whereas in Samoyed, and Uralic more generally, the genitive is used before postpositions deriving from spatial nouns. The plural of the local cases also has the same plural marker as the genitive plural, namely PS **-jt*. This is inserted after the coaffixes in the lative, locative and ablative, but it immediately precedes the prolative suffix. The singular local cases are not based on the genitive singular, but rather on the stem.

The mismatch in the cases that are used as the basis for the formation of agglutinative local cases can receive a possible explanation: primary adpositions in Indo-European languages generally select the accusative, dative or locative, so that the use of one of these cases in Tocharian is as expected. The accusative is the only of these cases that has a clear, direct descendent, in the form of the Tocharian oblique. The strategies used to expand the local case systems in Tocharian and Samoyed are thus broadly the same, including the use of a non-nominative plural as the basis for the plural local cases.

It is also important to bear in mind that a prerequisite for the formation of new cases from postpositions is the presence of postpositions in the first place. Proto-Indo-

European is not generally reconstructed with postpositions so that the development of postpositions in pre-Proto-Tocharian constitutes an innovation that itself may have been due to substrate influence from a language like pre-Proto-Samoyed. A similar reverse development is seen in Saami and Finnish, Uralic languages that developed and are developing more prepositions due to contact with Indo-European languages with prepositions.

## 6.5 Conclusions on the agglutinative case systems

The core of the local cases is shared between Proto-Tocharian and Proto-Samoyed, with an allative/lative, a locative, an ablative and a perlative/prolative. The Tocharian locative and to some extent the perlative are semantically broader than the corresponding Samoyed cases, with additional illative use of the Tocharian locative, and a locational use of the perlative. If these are later, secondary innovations, the original local systems may have been more similar, but this cannot be determined with certainty. It could also be that the broader semantics of the Tocharian secondary cases result from the broader semantic range of the adpositions from which they were formed, in which case the system was from the outset an inexact match for Samoyed.

The Tocharian comitative case has no counterpart in Samoyed. Tocharian A and B do not agree on the suffix of this case either, so that it could be a late innovation. However, it could also be that either Tocharian A or B inherited the Proto-Tocharian comitative with an innovation in only one of the two (this is suggested by Peyrot 2019a: 93). A (pre-)Proto-Tocharian comitative case cannot be explained based on Samoyed influence.

The Tocharian Gruppenflexion, whereby only the final noun in a sequence of nouns needs to be inflected with a secondary case suffix, has no Samoyed correspondence either. Meanwhile, this feature is found in Turkic languages, including Old Uyghur, so that Turkic influence on Tocharian is more likely in this regard (cf. Akao 2020: 9–10). On the other hand, no perlative is found in Turkic, so that Turkic influence cannot explain the development of that particular characteristic local case (Peyrot 2019a: 92).

The main problems with the hypothesis that the Tocharian agglutinative case system is the result of contact with pre-Proto-Samoyed can be summarized in two points, namely that (1) the functional ranges of especially the locatives and the perlative/prolative do not fully overlap, and that (2) Samoyed does not have Gruppenflexion. If a comitative case should be reconstructed for Proto-Tocharian after all, that is another point of disagreement, but the comitative could also be a later addition to the Tocharian case system.

Not all aspects of the prehistory of the Tocharian and Samoyed case systems are yet fully understood. The etymology of the Tocharian comitatives TA *-aśśäl* and especially TB *-mpa*, as well as that of the Tocharian B ablative *-meṃ*, are still unclear. The age of the

Tocharian secondary cases is also difficult to determine, which means that they could theoretically be too young to be caused by Samoyed substrate influence. On the Samoyed side, the precise age and origins of the co-affixes *-*ntə*- and *-*kə*- remain uncertain, as various interpretations are available. It is furthermore unknown when the Samoyed prolative was first formed, how exactly it was formed, and whether it would have come into existence early enough to possibly inspire the formation of the perlative in pre-Proto-Tocharian.

The elements that gave rise to the Tocharian agglutinative case system stem from Proto-Indo-European, meaning that they would already have been present in pre-Proto-Tocharian before contact with any other language. It could be that various (or all?) prepositions in pre-Proto-Tocharian were changed into postpositions, possibly due to language contact, and that a select few that semantically corresponded reasonably well to the four pre-Proto-Samoyed local cases became fixed expressions that would give rise to the eventual Tocharian local cases.

Later developments, such as a semantic extension of the perlative and the creation of a comitative in both Tocharian A and B, may have further obscured the similarities between the pre-Proto-Samoyed and Tocharian local case systems. Some aspects of the case system, such as the use of Gruppenflexion, find much better parallels in Turkic, and could perhaps be an instance of Turkic influence on Tocharian at some perhaps late stage.

The case system thus provides us with a general parallel in type, connected also with the broader region of Siberia, and a slightly better match with pre-Proto-Samoyed than with Proto-Uralic. While, on its own, it does not constitute clear evidence for any specific prehistoric contact event, it may be considered to strengthen the evidence from the other parallels discussed in this dissertation.

# 7 Participial systems

The participles of Tocharian, while derived from Proto-Indo-European, do not align with the other Indo-European languages in terms of their orientation. Rather than being inherently oriented as active or passive participles, Tocharian participles instead show a tendency for contextual orientation. This is not easily established for all participles, however, so that a detailed discussion is required. An analysis of Tocharian participles as contextually oriented opens up a new avenue for comparison with Samoyed, since contextual orientation is also found in the majority of Samoyed participles. The Tocharian and Samoyed participial systems will be discussed in turn to more closely investigate the similarities between them, which appear to extend beyond the basic contextual orientation, and to determine if the innovative aspects of the Tocharian participial system can be understood as the result of early Samoyed influence.

## 7.1 Introduction

Three participles are usually reconstructed for Proto-Indo-European: an active participle with the suffix *-*nt*-, a middle participle with the suffix *-*mh₁no*-, and an active perfect participle with the suffix *-*uos*- (e.g., Beekes 2011: 279–280; Lundquist & Yates 2018: 2169–2170). The verbal adjective in *-*to*- creates participles in many branches as well. The labels active, middle and perfect are in line with the functions of these participles in Greek and Indo-Iranian, where they form a coherent system aligned with the corresponding active and middle voices and the (active) perfect tense. In many other Indo-European branches one or two of these participles remain only in relics, or were lost completely. Tocharian preserves all three formations, but their behaviour does not neatly correspond to that found in Greek and Indo-Iranian. Most importantly, the distinction between participles corresponding to the active or middle voice has been abandoned, as will be discussed in more detail below. In this chapter, the main focus will be on the contextual orientation that the Tocharian participles mostly have, and the possible connection of this feature with Uralic/Samoyed.

Other, originally non-participial formations have also gained more prominence in Tocharian and arguably became integrated into the participial system. First, there are two types of gerundives, based on the present and subjunctive stems, which could be regarded as necessitative and potential participles, and second, there is the Tocharian privative, which can be described as a negative participle. These are Tocharian innovations based on Proto-Indo-European verbal adjectives. Their syntactic behaviour is very much like that of the participles, however, so that they should be included in this study.

There is still much ongoing discussion about several aspects of the various Tocharian participles (e.g., Fellner 2014; Fellner & Grestenberger 2018; Friis 2021; Peyrot 2010, 2017; Pinault 2012; Saito 2006). In this chapter, I will approach the participles primarily from the point of view of their orientation, as this is one of the saliant discrepancies between Tocharian participles and their cognates elsewhere in Indo-European. In Tocharian, many participles are not only insensitive to the active or middle voice of the verb from which they derive, but they are in general more ambiguous than "classic" active or passive participles, in such a way that a single formation can be both active and passive depending on the context (Krause & Thomas 1960: 184–186; Seržant 2020: 226–230). This feature finds a close parallel in the Samoyed languages, where many participles are contextually oriented as well (Leisiö 2006; Shagal 2018).

In section (7.2) and its subsections, the various participle formations in Tocharian will be discussed in more detail, with special attention to their behaviour with regards to the agent and the patient. This is followed by an overview of Samoyed participles in section (7.3). As in other chapters, our focus will be on this easternmost branch of Uralic. Finally, the characteristics of the Tocharian and Samoyed participial systems will be compared in the final section of this chapter (7.4). In order to evaluate the similarities between Tocharian and Samoyed, and the significance of these similarities, the situation of both the larger Indo-European and Uralic language families, as well as the areal typological context need to be taken into account. First, however, the concept of orientation is briefly introduced in the next subsection, (7.1.1), as this will be an important concept throughout the chapter.

### 7.1.1 Participles and their orientation

Participles can be difficult to define and delineate from other, related formations. For example, in their role as adjectival verb forms, they are very reminiscent of verbal adjectives, and they have a tendency to function as converbs as well, or to undergo nominalization (Haspelmath 1994: 171–172; Shagal 2019: 14–44). This chapter will focus on those formations whose synchronic behaviour in Tocharian relatively clearly warrants the label of "participle", and formations that descend from participles in Proto-Indo-European, but may have taken on other functions as well (see especially Shagal 2019: 20–41 for a more detailed delineation of the category of participles).

Participles come in various types with different kinds of behaviour. Of prime importance here is that they do not all interact in the same way with the arguments of the verb from which they are derived. Active participles can modify a subject S or an agent A (1a. and 1b., respectively), while passive participles modify a patient P (1c.; Shagal 2019: 51). An active participle like Ru. *ubítyj* cannot modify an intended patient (1d.), and vice versa, a passive participle like *ubívšij* cannot modify the patient (1e.; see Shagal 2019: 61–74 on these types of participles). Despite the fact that both *ubítyj* and *ubívšij* are

derived from the same active transitive verb *ubít'* 'kill', their inherent orientation is different (Haspelmath 1994, Shagal 2019: 51–128).

1. Examples of active and passive participles from Russian.

a. *begúščij* *olén'*
run:PRS.ACT.PTC deer:NOM.SG
'The running deer (S)' ~ 'The deer (S) that runs'

b. *medvéd'* *ubívšij* *olénja*
bear:NOM.SG kill-PST.ACT.PTC deer:ACC.SG
'The bear (A) that killed the deer.'

c. *olén'* *ubítyj* *medvédem*
deer:NOM.SG kill-PST.PASS.PTC bear:INS.SG
'The deer (P) that was killed by the bear.'

d. ** *medvéd'* *ubítyj* *olénja*
bear:NOM.SG kill-PST.PASS.PTC deer:ACC.SG
Intended: 'The bear (A) that killed the deer.'

e. ** *olén'* *ubívšij* *medvédem*
deer:NOM.SG kill-PST.ACT.PTC bear:INS.SG
Intended: 'The deer (P) that was killed by the bear.'

There are more types of orientation than just active and passive. An agentive participle is essentially an active participle that can only be formed to transitive verbs (Shagal 2019: 75–77). Absolutive participles, in Shagal's terms, are also found in many Indo-European languages (Shagal 2019: 78–85). These differ from strictly active or passive participles in that they relativize either the subject S of an intransitive verb (with an active meaning) or the patient P of a transitive verb (in which case they are interpreted as passive), but not the agent A of a transitive verb. The English past participle is an example of this type, e.g., *the grown man* = the man (S) that has grown (active, from an intransitive verb) vs. *the thrown rock* = the rock (P) that has been thrown (passive, from a transitive verb). Haspelmath (1994: 159–161) calls this type of participle "resultative" rather than "absolutive", and the correspondence between this type of participle and perfectivity or resultativity is indeed cross-linguistically very common, Shagal also reports (Shagal 2019: 80). I will follow Shagal's terminology and use the term "absolutive" to describe this type of participle throughout the rest of this chapter.

The types of participles discussed so far are all inherently oriented, so that they have a predictable relationship to the participants of the verb action. The final type of orientation that is relevant in the following discussion on Tocharian and Samoyed is contextual orientation (Shagal 2019: 81–105). Participles with contextual orientation are more flexible than inherently oriented participles, as they can relativize both the agent

and patient of a transitive verb, depending on the syntactic context, as well as the subject of an intransitive verb (Haspelmath 1994: 154). Sometimes further oblique participants like locations and possessors can be relativized with contextually oriented participles as well (Shagal 2019: 91–96). Contextually oriented participles are not particularly common in Europe, but they occur widely elsewhere. Pakendorf (2012) discusses contextually oriented participles as the main type of participle in Siberian languages, and considers it to be a shared linguistic feature of the area (also Shagal 2018, and see further below).

To illustrate contextual orientation, Haspelmath (1994: 154) adduces some examples from the Nakh-Daghestanian language Lezgian, reproduced here in (2). The same imperfective participle *kx̂i-zwa-j* can be used to modify various participants of the event described, such as the agent (2a.), the patient (2b.), or a peripheral participant like the place or time (2c.–d.).

2. Lezgian contextually oriented participles used to modify the agent (a.), the patient (b.) and peripheral participants (c.–d.; Haspelmath 1994: 154).

a. *čar* *kx̂i-zwa-j* *ruš*
letter(ABS) write-IPFV-PTC girl
'The girl who is writing a letter.'

b. *ruš-a* *kx̂i-zwa-j* *čar*
girl-ERG write-IPFV-PTC letter
'The letter which the girl is writing.'

c. *ruš-a* *čar* *kx̂i-zwa-j* *stol*
girl-ERG letter(ABS) write-IPFV-PTC table
'The table on which the girl is writing a letter.'

d. *ruš-a* *čar* *kx̂i-zwa-j* *juǧ*
girl-ERG letter(ABS) write-IPFV-PTC day
'The day on which the girl is writing a letter.'

Shagal describes how sometimes absolutive orientation can be observed as a tendency in otherwise technically contextually oriented participles. For instance, the Ket contextually oriented participle, when formed to a transitive verb, is interpreted by default as passive. However, when a patient (direct object) is expressed, it is active instead. An active interpretation is also possible when the head noun is highly animate (e.g., a human), even without the expression of a direct object (Shagal 2019: 82–83). A similar situation can be observed in the Samoyed languages, such as in Selkup; cf. the following examples from Kuznecova, Helimski & Gruškina (1980: 254).

3. Active (a., c.) and passive (b., d.) use of participles in Selkup (Kuznecova, Helimski & Gruškina 1980: 254; glosses added)

a. *ɔ̄tä-p* *qət-pi̯l′* *qorqi̯*
deer-ACC kill-PST.PTC bear
'The bear that killed the deer.'

b. *qorqi̯-t* *qət-pi̯l′* *ɔ̄tä*
bear-GEN kill-PST.PTC deer
'The deer that was killed by the bear.'

c. *topi̯r-i̯p* *wətɛnti̯-ti̯l′* *ījat*
berry-ACC collect-PRS.PTC children
'The children that collect berries.'

d. *ījat-i̯n* *wətɛnti̯-ti̯l′* *topi̯r*
children-GEN collect-PRS.PTC berry
'The berries that are collected by the children.'

In these sentences the past participles in *-pi̯l′* (PS **-pA-jə*) and the present participles in *-ti̯l′* (PS **-ntA-jə*) modify the following nouns, but the participles themselves are ambiguous as to whether they express an active or passive relationship. This goes for all other types of participles found in Selkup as well (Kuznecova, Helimski & Gruškina 1980: 254). The argument structure is different between the two sets of sentences, which facilitates the different interpretations: the active reading of *qətpi̯l′* 'kill:PST.PTC' and *wətɛnti̯ti̯l′* 'collect:PRS.PTC' in (a) and (c) is necessitated by the presence of an object in the accusative case, namely *ɔ̄täp* 'deer:ACC' and *topi̯ri̯p* 'berry:ACC'. In sentences (b) and (d), the presence of agents in the genitive case, viz., *qorqi̯t* 'bear:GEN' and *ījati̯n* 'children:GEN', makes it abundantly clear that a passive interpretation is required instead. An active or passive interpretation of these contextually oriented participles is not inherent in the participles themselves, but, as the name suggests, derives from the context.

Contextually oriented participles are prominently present in both the Tocharian and the Samoyed participial systems, as will be illustrated below. With regards to the preterite participle, the Tocharian situation has been referred to as "the problem of the passive and active usage of the prt.ptc." (Peyrot 2007: 799). This "problem" concerns the contextual orientation of the participles (cf. Seržant 2020: 226–230), which differs in Tocharian from the Indo-European norm. In this chapter we will explore the extent to which contextual participles occur in Tocharian, taking into account various functions of the participle formations, including predicative, substantive and converbial use. However, the attributive function will be given the most weight, as that is a more canonical participial function.

## 7.2 Tocharian

Since all of the Tocharian participial formations have an Indo-European origin, it is possible to evaluate how the orientation of most of them has changed over time. In the following subsections, the Tocharian use of each individual formation and its relationship to cognate formations in other Indo-European languages will be treated systematically. We will start with the Tocharian reflexes of the Proto-Indo-European active voice participle (7.2.1), then the middle voice participle (7.2.2) followed by the perfect participle (7.2.3), and then the more innovative Tocharian participle formations, namely the gerundives (7.2.4) and the privative (7.2.5).

### 7.2.1 The *nt*-participle

A participle with the suffix PIE **-nt* forms active participles in almost every branch of Indo-European where it is attested; e.g., Gr. ἄγων, ἄγοντος 'leading' from ἄγω 'lead', Skt. *bhárant-* 'carrying' from *bhárati* 'carry', Lat. *sapiēns*, *sapientis* 'wise' from *sapiō* 'discern', Go. *nasjands* 'saviour' from *nasjan* 'save'. This participle is aligned with morphologically active voice of verbs in those languages that show a distinction between active and middle voice, and it is also actively oriented in most Indo-European languages, as it normally modifies the verb's subject or agent.

There are several phonologically similar Tocharian formations that continue the Proto-Indo-European *nt*-participle (on which in detail see Fellner 2014; 2017; Friis 2021). In Tocharian B, the most productive representative of this group is the participle suffix TB *-ñca*, which corresponds functionally to TA *-nt*. This participle has deviated to a certain extent from its status as a regular participle, as it does not show gender concord, and thus behaves more like a substantive in terms of its inflection. It can be nominalized as an agent noun, but attributive function is still attested (Fellner 2014).

4. a. Tocharian A (A 75 a2; verse; trsl. based on CEToM)

| *tsräṃ* | *päryo* | *āriñc* | ***wākṣantāṃ*** | *yäslunt* |
|---|---|---|---|---|
| sharp:M.OBL.SG | arrow:INS | heart:OBL.SG | split:ACT.PTC.OBL.SG | enemy:OBL.SG |
| ***koṣäntāṃ*** | *śarwnaṃ* | *pälkāc-äṃ* | *kāruṃ* | |
| kill:ACT.PTC.OBL | hunter:LOC.SG | see:IMP.MID.2PL-3SG | compassion:OBL | |

'See his compassion for the **murdering** hunter, [his] enemy, **who splits** [his] heart with a sharp arrow.'

b. Tocharian B (THT 31 a4–5; trsl. based on CEToM)

| *aiśaumyi* | *ce$_u$* | *pällāntär* | *krento* |
|---|---|---|---|
| wise:NOM.PL | DEM.M.OBL.SG | praise:PRS.MID.3PL | good: M.SG.OBL |
| *āstreṃ* | *śaul* | ***śayeñcai*** | *wnolme* |
| pure: M.SG.OBL | life:OBL.SG | live:ACT.PTC.OBL.SG | being:OBL |

'The sages praise him as being a being **who lives** a good, pure life, ...'

c. Tocharian B (PK AS 2B a2; trsl. CEToM

| *hiraṇḍaṣṣe* | *ṣalypempa* | *tetriwu* | *nemcekaṃñe* |
|---|---|---|---|
| ricinus.ADJ.SG.OBL | oil:COM.SG | mix:PRT.PTC.M.NOM.SG | infallible.NOM.SG |

| *po* | *tekanma* | ***näkṣeñca*** |
|---|---|---|
| all | disease:OBL.PL | destroy:ACT.PTC.NOM.SG |

'Mixed with ricinus oil [it is] an infallible **destroyer** of all diseases.'

It is often considered to be an agent noun instead of a participle, as in the translation in (4.c.) of *näkṣeñca* as 'destroyer' (see Pinault 2012; Peyrot 2017: 328–329), but as is clear from (4.a-b.), it still shows canonical adjectival functions as well, supporting a continued analysis of this formation as a participle (Fellner 2014). The orientation of the *nt*-participle is definitely active, and most examples are formed from transitive verbs (e.g., Fellner & Grestenberger 2018: 68). The few intransitive verbs that form *nt*-participles have a local exponent, as with intransitive *mäskeñca* 'being, staying' in example (5). Kuritsyna describes mainly agentive orientation of *nt*-participles, i.e., active participles formed to transitive verbs, with two potential examples of attributive *nt*-participles derived from TB *kəĺ-*, TA *käĺ-* 'stand' (Kuritsyna 2022). The Tocharian *nt*-participle thus overall has an active orientation.

5. Tocharian B (PK AS 17A b1; prose; trsl. CEToM)

| *t*(*ane*) | *ñak*(*e* | *pūrvavedīd*)*v*(*ī*)*pn*(*e*) | ***mäsk*(*e*)*ñca*** | *ñ*(*a*)*kt*(*e*) |
|---|---|---|---|---|
| here | now | PN:LOC | be:ACT.PTC | god |

| *pūrv*(*o*)*ttare* | *ñem* |
|---|---|
| PN | name |

'Here now, the god **who stayed** in Pūrvavedīdvīpa, Pūrvottara by name, ...'

In most Indo-European languages, the *nt*-participle is aligned with the morphological category of the active voice, and the orientation of this participle is typically active.[75] In

[75] Some examples like Gm. *die melkende Kuh* 'the cow that is milked, milch cow' are taken as a trace of ambiguous orientation in the *nt*-participle (Haspelmath 1994: 170), but it is unclear to me if this is not rather formed to an (earlier?) intransitive sense of a labile verb 'to milk' as found in the Icelandic cognate *mjólka*, which can mean both 'milk' and 'produce milk'. Fellner and

Tocharian, the connection with the finite active, while common (Fellner & Grestenberger 2018: 68), is not complete, and *nt*-participles can be formed to exclusively middle-voice verbs. There is one other branch of Indo-European where *nt*-participles are famously not actively oriented, namely in Anatolian. For example, when a Hittite *nt*-participle is formed from an intransitive verb, the interpretation is active, as expected, e.g., Hitt. *akkant-* 'having died', *pānt-* 'having gone'.[76] When formed to a transitive verb, the Hittite *nt*-participle is usually interpreted as passive, e.g., Hitt. *appant-* 'seized, taken', *piyant-* 'given' (Hoffner & Melchert 2008: 339–340). Active instances of seemingly transitive *nt*-participles are argued by both Brosman and by Lyutikova and Sideltsev to be restricted to transitives that can optionally be construed without an object, such as 'eat' (Brosman 2010: 23; Lyutikova & Sideltsev 2021).

The behaviour of the *nt*-participle in Hittite thus contrasts with the active nature of the *nt*-participle in the other Indo-European languages, including Tocharian (Fellner & Grestenberger 2018; Friis 2021). For Tocharian, we can consider the *nt*-participle to have retained its usage as an active participle as inherited from post-Anatolian Indo-European, although it did not maintain its connection with the active voice. This (lack of) development of the orientation of the *nt*-participle from Proto-Indo-European to Tocharian is summarized and visualized in the table below.

Table 7.1: Schematic change in the orientation of the *nt*-participle from Proto-Indo-European to Proto-Tocharian.

| Indo-European | | Tocharian |
|---|---|---|
| active (S/A) | = | active (S/A) |

### 7.2.2 The *m*-participle

The Proto-Indo-European participle in ***-mh₁no-*** is not as broadly attested as the *nt*-participle. It is associated with the middle voice in most of the languages where it is productive, that is, mainly in Greek and Indo-Iranian. As a middle voice participle it stands in morphological opposition to the active voice *nt*-participle. In terms of its orientation, the middle voice participle does not have "middle orientation", if participles

Grestenberger also discuss Vedic *járant* 'old', which cannot be regarded as the synchronic active participle to *járati* 'make old', but rather semantically belongs with the middle *járate* 'become old'; some *nt*-formations in Greek and Latin also belong with middle verbs rather than active verbs (Fellner & Grestenberger 2018: 64–65). However, synchronically cases like Vedic *járant* are lexicalized adjectives, and they do not directly concern the participial system (Lowe 2015: 286–287).

[76] Unlike the *nt*-participle in other Indo-European languages, these Hittite *nt*-participles refer to the past.

with such an orientation exist at all (see Haspelmath 1994: 152), but rather an active orientation. Any arising passive semantics encountered in middle participles are normally shared with the finite middle verb. For example, the middle voice of Greek φέρομαι can already have the passive meaning 'be carried off' beside the more canonically middle meaning 'carry (for oneself)', so that the participle φερόμενον in (6.a.) is actually active in its orientation. Example (6.b.) shows that the middle voice participle of the semantically active τίθεμαι 'put (for oneself, around oneself, on oneself)' is also actively oriented, rather than passively oriented.

6. a. Ancient Greek (Hdt., Histories 2.29.2; Grestenberger 2020: 226)

| τὸ | πλοῖον | οἴχεται | **φερόμενον** | ὑπὸ |
|---|---|---|---|---|
| DEF.N.NOM | boat:NOM | goes.off | carry:PRS.MID.PTC.N.NOM.SG | by |

| ἰσχύος | τοῦ | ῥόου |
|---|---|---|
| strength:GEN | DEF.M.GEN | current:GEN |

'... the boat gets lost, **carried off** by the strength of the current.'

b. Ancient Greek (Hom., Iliad 4.514–16; Grestenberger 2020: 225)

| τὸν | δ' | εὗρ' | ἀμφ' | ὤμοισι | τιθέμενον |
|---|---|---|---|---|---|
| 3SG.M.ACC | PTCL | found | around | shoulders | put:PRS.MID.PTC.M.ACC |

| ἔντεα | καλὰ |
|---|---|
| armour.ACC | beautiful.F.ACC |

'He found him **putting** his beautiful armour around his shoulders.'

Tocharian has also inherited the Proto-Indo-European formation in *-*mh₁no*- that is reflected in the Greek middle voice participles. This yielded Proto-Tocharian *-*mane* and eventually TB -*mane*, A -*māṃ* (Klingenschmitt 1975: 159–163), and is called the *m*-participle or the present participle. Just as the Tocharian active *nt*-participle is not directly connected with active voice verbs anymore (see subsection 7.2.1), the Tocharian *m*-participle is not intimately connected with the middle voice inflection either (Malzahn 2010: 480–481; Peyrot 2017: 328 fn. 3). For example, the intransitive active voice verb TB *nes*-, TA *nas*- 'be' is only attested with an *m*-participle in TB *nesamane*, TA *nasmāṃ*, while the middle voice verb TB *məska*-, TA *mäskā*- 'be, become' has an *nt*-participle in TB *mäskeñca*, TA *mäskant* discussed above. Semantically, the *m*-participle is perhaps more associated with the idea of a middle voice than is the *nt*-participle, since it can be passive as well as active, but this does not always correlate with the diathesis of the finite middle voice verb. A passive *m*-participle can exist next to a semantically active verb, as shown by examples like the passive *m*-participle TA *pāsmāṃ* '(being) protected' based on the active transitive middle voice verb *pās*- 'protect' and passive TA *sesmāṃ*

'being supported' based on active transitive middle voice *se-* 'support'; both are illustrated in (7).

7. a. Tocharian A, passive attributive *m*-ptc. (A 2 b3; trsl. based on CEToM)

| | | | |
|---|---|---|---|
| ***pāsmāṃ*** | *niṣpal* | *lo* | *näkṣäl* |
| guard:PRS.PTC | property | PTCL.PFV | lose:PRS.GER.M.NOM.SG |

'[Even] (well-)**guarded** property will (eventually) be fully lost.'

b. Tocharian A, passive converbial *m*-ptc. (A 73 a2; trsl. CEToM)

| | | | |
|---|---|---|---|
| *weyeṃ* | *nāṃtsunt* | *ālam-wcā* | ***sesmāṃ*** |
| amazed | be:PRT.PTC.F.NOM.PL | each.other:PER | support:PRS.PTC |

'Surprised and **supported** by each other...'

There is even a certain ambiguity, as can be illustrated with the *m*-participle TB *akemane*, TA *ākmāṃ* derived from TB *ak-*, TA *āk-* 'lead'. This formation is attested in both active and passive orientation. In Tocharian B, the middle voice finite *āśtär* has a passive meaning next to active voice *āśäṃ*, an aspect that is deemed indicative of a continued connection with the middle voice by Fellner and Grestenberger (2018: 76). However, with the parallel existence of an actively oriented *m*-participle from the same verb, there is still a partial mismatch between this participle and the finite verb. Both Tocharian A and B attest this active use of TA *ākmāṃ*, TB *akemane*, as well as the passive meaning associated with the middle voice, so that there is no clear one-to-one connection between the *m*-participle and the meaning of the finite middle voice (cf. the examples with TA *pāsmāṃ* and *sesmāṃ* above).

8. a. Tocharian A, passive converbial *m*-ptc. (A 81 a5)

| | | |
|---|---|---|
| (*ca*)*ndālāsā* | *kossi* | ***ākmāṃ*** |
| Caṇḍāla:PER.PL | kill:INF | lead:PRS.PTC |

'**Led** to be killed by the Caṇḍālas.'[77]

b. Tocharian B, passive attributive *m*-ptc. (THT 3 b5–6)

| | | | |
|---|---|---|---|
| *mäkte* | *kautsy* | ***akemane*** | *wnolme* |
| like | kill:INF | lead:PRS.PTC | being |

'Just like a being **led** to be killed.'

[77] Theoretically, this could be read as 'leading [X] to be killed by the Caṇḍālas' due to the fragmentary nature of the manuscript, but the rest of the context makes it clear that the subject of the sentence is not the one leading.

c. Tocharian A, active converbial *m*-ptc. (A 402 a1)

| *wartsi* | ***ākmāṃ*** |
|---|---|
| assembly | lead:PRS.PTC |

'**Leading** the assembly, ...'

d. Tocharian B, active converbial *m*-ptc. (THT 88 b5–6)

| *uttareṃ* | *mñcuṣkeṃ* | ***akemane*** |
|---|---|---|
| PN:OBL | prince:OBL | lead:PRS.PTC |

'**Leading** prince Uttara, ...'

If there were a strict correspondence between the finite middle voice and the *m*-participle in Tocharian, the *akemane* 'leading' in *uttareṃ mñcuṣkeṃ akemane* 'leading prince Uttara' (8d.) should have been passive just like the passive finite TB *āśtär*. What we see instead, is that *m*-participles based on transitive verbs can be either active or passive, regardless of whether the associated verb is active or middle voice.

All that being said, Friis has identified some traces of an earlier connection between the *m*-participle and the middle voice in a few verbs where the middle voice provides the intransitive counterpart to the active voice. For example, TB *päksemane oṅkorñai* (THT 107 a2) is a 'cooking porridge' in the intransitive sense, not a porridge that 'cooks (something)' or 'is being cooked'. This means that the *m*-participle *päksemane* is in line with the intransitive meaning of the middle voice *pək-* 'cook, ripen (itr.)', rather than with the meaning of the transitive active voice *pək-* 'cook (tr.)' (Friis 2021: 3–4 fn. 6). The *m*-participle thus derives diachronically from a true middle voice participle, and its eventual ambivalence to active or middle voice in Tocharian is a secondary innovation (ibid.; also Fellner & Grestenberger 2018).

There are various ways to interpret the ambiguous orientation of the *m*-participle, and it may be that the different uses of this formation, between attributive and converbial, muddy the water. In the examples discussed so far, it appears that an *m*-participle that is used as a converb is more often active than an attributively used *m*-participle. Kuritsyna has found that passive orientation prevails particularly in attributive use of *m*-participles formed from transitive verbs, pointing to a potentially absolutive orientation (Kuritsyna 2023: 139). Since attributive use is the more canonical function of participles, this situation merits special attention. In CEToM searches for "mane", "māne" and "māṃ" I was able to find eight examples of passive *m*-participles for Tocharian A, and two examples in Tocharian B.

Tocharian B examples of passive *m*-participles that I found are *akemane* (THT 3 b5-6) from *ak-* 'lead' and *tsäksemane* (IOL Toch 5 a4) from *tsək-* 'burn (tr.)'. The use of *akemane* is attributive, and so is that of *aikemane* (THT 559.2 a2) from *aik-* 'know', adduced by Kuritsyna (2023: 119). Accompanying the converbial *tsäksemane* there is an instrument in the perlative case, namely *pwārasa* 'by fire'.

9. a. Tocharian B, passive attributive *m*-ptc. (THT 3 b5–6)

| *mäkte* | *kautsy* | ***akemane*** | *wnolme* |
|---|---|---|---|
| like | kill:INF | lead:PRS.PTC | being |

'Just like a being (that is) **led** to be killed.'

b. Tocharian B, passive converbial *m*-ptc (IOL Toch 5 a4; trsl. CEToM)

| *onmiṣṣana* | *pwārasa* | ***tsäksemane*** | *marmanma* |
|---|---|---|---|
| remorse:ADJZ.OBL.PL | fire:PER.PL | burn:PRS.PTC | vessel:PL |

| *troṅktse* | *stām* | *ra* | *sälpiñ*[78] | *cittsa* |
|---|---|---|---|---|
| empty | tree | like | glowing:M.OBL.SG | spirit:PER |

*wolokmar*
stay:PRS.MID.1SG

'[**While** my] vessels **are burnt** by the fires of remorse, I dwell like an empty tree with a glowing spirit.'

Tocharian B active *m*-participles used in attributive function are mostly based on intransitive verbs of the type in *päksemane oṅkorñai* 'the cooking porridge' mentioned above. When formed to transitive verb, the resulting attributive *m*-participle is generally passive, in a pattern that conforms to Kuritsyna's interpretation of attributive *m*-participles as absolutive in orientation (Kuritsyna 2023: 139). However, I have found one active example of an attributive *m*-participle based on the transitive verb *lans-* 'build, work on, carry out'. The presence of an accompanying object, *yketseñe cake* 'a local river' makes this interpretation quite clear, see example (10).

10. Tocharian B, active attributive *m*-ptc. (Kizil WD-I-a4; trsl.[79] based on CEToM)

| *skaśce* | *ikäṃ-śtwerne* | *yketseñe* | *cake* | ***laṃsemane*** |
|---|---|---|---|---|
| 6:ORD.OBL.SG | 20-4:LOC | local | river:OBL.SG | build:PRS.PTC |

| *kapyāri* | *ṣārwai* | *śawāre* |
|---|---|---|
| worker:NOM.PL | meat(?):OBL | eat:PRT.3PL |

'On the 24th [day] of the 6th [month], the workers **who were repairing** the local river ate meat.'

Tocharian A *m*-participles found in passive function are at least *pāsmāṃ* (A 2 b3) from *pās-* 'protect', *kropmāṃ* (A 3 b1; cf. the parallel formation *kropnmāṃ* attested only as active in A 18 b6, A 19 a1, A 303 a2 and A 405 a1) from *krop-* 'gather (tr.)', *ypamāṃ* (A 17 b3) from *yäp-* 'make, do', *esmāṃ* (A 257 b1, YQ I.4 a6, YQ III.7 a3 and b3; also found active in A 301 a5, and possibly in A 232 a2) from *e-* 'give', *ākmāṃ* (A 81 a5; also found active in

[78] Final *-ñ* due to the following *c-*; otherwise *sälpiṃ*.

[79] Original translation in Chinese: '六月二十四日，正在修理当地河道的净人们吃了肉.'

A 402 a1) from *āk-* 'lead', *sesmāṃ* (A 73 a2) from *se-* 'support'. Some examples from this group are given here in (11).

11. (a) Tocharian A, passive attributive *m*-ptc. (A 3 b1; trsl. CEToM)

| *taṃne* | ***kropmāṃ*** | *niṣpalntu* | *ykoṃ* | *oṣeñi* |
|---|---|---|---|---|
| thus | gather:PRS.PTC | property:PL | by.day | by.night |

| *śamaṃtär* | *mäṃtne* |
|---|---|
| increase:PRS.MID.3PL | as |

'Possessions **that are gathered** in that way by day and night increase like ...'

(b) Tocharian A, passive attributive *m*-ptc. (A 17 b3; trsl. CEToM)

| *pñintu*(*yo*) | *ptāñkät* | ***yärkā ypamāṃ*** | *yatatär* |
|---|---|---|---|
| virtue:INS.PL | Buddha | venerate:PRS.PTC | be(come):PRS.MID.3SG |

'Through virtues one becomes the **venerated** Buddha.'

(c) Tocharian A, passive attributive *m*-ptc. (YQ III.3 a3; trsl. CEToM)

| *pissaṅkac* | ***esmāṃ*** | *tñi* | *kanak* |
|---|---|---|---|
| community:ALL.SG | give:PRS.PTC | 2SG.GEN | cotton.cloth |

'A cotton cloth **given** by you to the community.'

A number of Tocharian A passive *m*-participles are used attributively, e.g., in A 3 b1 *kropmāṃ niṣpalntu* 'gathered possessions' (11a.) and in A 2 b3 *pāsmāṃ niṣpal* 'guarded possessions' (see 7a. above). Some are furthermore accompanied by an agent in the perlative or genitive, e.g., in YQ III.7 a3 *esmāṃ tñi kanak* 'the cotton cloth given by you' (here in 11c.), and in A 73 a2 *ālam-wcā se*(*s*)*m*(*āṃ*) '(being) supported by each other' (7b. above).

In Tocharian A, as in Tocharian B, active *m*-participles used in attributive function are mostly based on intransitive verbs; e.g., A 3b4 *sälpmāṃ por* 'glowing fire', A 16b2 *smimāṃ akmalyo* 'with a smiling face'. One possible exception treated by Kuritsyna (2023: 117) is the *m*-participle *wlesmāṃ*, given here in example (12). As Kuritsyna discusses, interpretation of this *m*-participle as an attributive rather than an adverbial is not certain, however, as either the start of the sentence, *āsre päśre trse* (*kroś*)*śe* 'dry, hard, rough, cold', or the ending, *aṃśre nāṃtsuṣ* 'being unsteady', could in principle be interpreted as an apposition. If *aṃśre nāṃtsuṣ* 'being unsteady' is taken as the predicate, *tukri wlesmāṃ* 'working (the) clay' is converbial rather than attributive.

12. Tocharian A (A 23 b5; trsl. based on CEToM)

| *āsre* | *päśre* | *tsre* | (*kroś*)*śe* | *tukri* | ***wlesmāṃ*** | *ñi* | *tsaräṃ* |
|---|---|---|---|---|---|---|---|
| dry | raw | rough | cold | clay | build:PRS.PTC | 1SG.GEN | hand.DU |

| *aṃśre* | *nāṃtsuṣ* |
|---|---|
| unsteady | be:PRT.PTC.M.NOM.PL |

'Dry, hard, raw/rough [and] (col)d [are] my clay-**working** hands, they are unsteady.'
or: 'Dry, hard, raw/rough, (col)d, **working** the clay, my hands are unsteady'

Since there is also a clear example of an active transitive *m*-participle in Tocharian B (cf. example sentence 10 above), TA *tukri wlesmāṃ ñi tsaräṃ* may be a genuine example of the same phenomenon in Tocharian A as well, if it is to be interpreted as attributive. The number of examples remains low in any case, and the majority of attributive *m*-participles follow an absolutive pattern of orientation (Kuritsyna 2023: 139). However, considering the fact that the meaning of the Tocharian finite verb cannot supply the right interpretation for the orientation of any given *m*-participle, at least not synchronically, in combination with the more contextually oriented converbial behaviour of the *m*-participle (Kuritsyna 2023: 139) and the marginal active transitive attributive use in example (10), the general orientation of the *m*-participle formation can be characterized as contextual. The development from Proto-Indo-European is illustrated in the table below.

Table 7.2: Schematic change in the orientation of the *m*-participle from Proto-Indo-European to Proto-Tocharian.

| Indo-European | | Tocharian |
|---|---|---|
| active (S/A) | → | contextual (S/A/P) |

### 7.2.3 The preterite participle

The third Proto-Indo-European participle formation inherited in Tocharian is the perfect participle in PIE **-uos*, which became the Tocharian preterite participle (Krause & Thomas 1960: 261; Þórhallsdóttir 1988: 185; Saito 2006: 4–11). Unlike the *nt*-participle and the *m*-participle, the preterite participle retained its canonical participial function more faithfully. It is fully inflected for gender, case and number, like any adjective, and it can take objects in the oblique and agents in the genitive or perlative case.

There are four different classes of preterite participle formation in Tocharian, which differ on formal parameters such as presence or absence of reduplication and the quality of the stem-final vowel (Þórhallsdóttir 1988: 184–185), but these classes do not show any functional difference. When based on a transitive verb, a preterite participle can either be interpreted as active or passive (e.g., Krause & Thomas 1960: 185, 190-191, Saito 2006:

18, Malzahn 2010: 232). Naturally, when a participle is formed to an intransitive verb, only an active interpretation is available.

Although the question of whether any individual Tocharian preterite participle is active or passive cannot be answered based on the form alone, the surrounding context typically holds the answer: the standard interpretation seems to be passive, and typically only when an object—an oblique noun or pronoun—is expressed, an active interpretation is warranted instead. The preterite participle thus shows contextual orientation, as per Seržant (2020: 226–230). Seržant (ibid.) connects the use of either actively or passively oriented preterite participles with the topicality of the respective arguments: if the subject or agent of the clause is more topical, the participle is actively oriented, but if the object is more topical, the participle is constructed in such a way as to be passively oriented.

If we look at the construction of the individual clauses, actively used preterite participles are generally accompanied by an overt direct object, while passively used preterite participles are not. This pattern emerges from a survey of preterite participle of *yam-* 'do, make' in both Tocharian languages in the CEToM database. This particular verb was chosen because it is amply attested, and obligatorily transitive; a list is given below in (13–16). Seržant also lists *yāmu* as one of the most clearly unoriented verbs (Seržant 2020: 230). It appears that an active interpretation occurs when an object is explicitly stated within the same sentence. Among the Tocharian A examples there is one minor exception in (14a.), where a preterite participle takes the same object as a preceding absolutive without the use of a resumptive pronoun, so that the object of the participle and thus its active meaning needs to be derived from the immediately preceding context. Other sentences with active forms of TA *yāmu* do contain an oblique. The examples illustrating the use of active *yāmu* are given here for Tocharian B in (13) and for Tocharian A in (14).

13. Tocharian B active use of the prt.ptc. *yāmu*.[80]

    a. THT 24 a3 (Udānālaṅkāra; verse; trsl. CEToM)
    *śateñ wnolmi cai mā śaiṣṣe kca*(OBL) ***ynāñm yā(m)oṣ***
    'These rich beings **valued** nobody, …'

[80] In some cases, the precise use and translation of *yāmu* is too uncertain, mostly due to limited context resulting from the fragmentary nature of the manuscripts. Excluded from the list for this reason are forms of *yāmu* found in THT 41 (*no yneś yāmoṣ*), 123 a3–4 (*mäntrākka ñäkta kerekauna yāmo*/// passive acc. to Saito 2006: 496, trsl. 'O god! Just so a flood is made', but the agreement between *kerekauna* and *yāmo-*, probably *yāmoṣ*, is unclear, as *yāmoṣ* should be in the oblique singular to agree with singular *kerekauna*, while Saito's translation suggests a nominative function), 223.a a4 (///–*dhat yāmoṣ karuntsa*///), 251 b1–2, 304b3 (passive acc. to Saito 2006: 494), 321 b3 (///*āssa yāmu osne yänma*///), 321 b4 (///*āssa yāmu osne ṣmalle*///), 340b5 (///–*tsemi yamoṣ*

b. THT 41 a1–a2 (Udānālaṅkāra; verse; trsl. CEToM; cf. Saito 2006: 495)
(*śtarcce dhyāṃ*(OBL)) *no* ***yneś yāmoṣ***
'For they **have realized** (the fourth Dhyāna).'

c. THT 41 a6 (Udānālaṅkāra; verse; trsl. CEToM)
*piś bhūminta*(OBL) ***saim yāmu***
'... the one that **takes refuge** in the five Bhūmis ["earths"] ...'

d. THT 45 a2 (Udānālaṅkāra; verse; trsl. CEToM)
*tañ paine*(OBL) ***saim yāmu***
'**having** made your feet [my] **refuge**.'

e. THT 89 a2 (Araṇemijātaka; verse; trsl. CEToM)
*waiptār wloṣäṃ letseṃne po kälymintsa lwāsa ñi lestai*(OBL) ***yāmwa***
'in my dishevelled hair curls animals **have made** nest[s] everywhere'

f. THT 89 b3 (Araṇemijātaka; verse; trsl. CEToM)
*ñi sak lakle t*(*a*)*ñ w*(*i*)*ṣ*(*ai*) *tu* ***pāke***(OBL) ***yā***(***m***)***u***
'(it is) [both] happiness [and] suffering to me to **have partaken** in your domain'

g. THT 102 a4 (Subhāṣitagaveṣin-Jātaka; prose; trsl. after Saito 2006: 493)
///(*ka*)*rtse*(OBL) ***yāmu*** *westrä*
'He is called a **benefactor** (lit. "good-**doer**")'

---

*luwa*///; passive acc. to Saito 2006: 494), 348 a3 (///·*k*· *yāmu ste mā*///), 352 b4 (///(*pra*)*tiṃ ñiśämpa yāmoṣ epyac* –), 405 b5 (– *l yāmu yarṣtar*), 529a2 (*waipalau yāmu*), 541a3 (*aimameṃ yāmu*; active acc. to Saito 2006: 495), 542 b2 (*āklyi yāmoṣ*), 562 a4 (///*lmeṃts läktsauña yāmusa mäsketärne*), 1119 a1 (...*ñ yāmu*), 1374.p b2 (///– *saim yāmu* · –///), 1378.a a7 (///*ruṃ yneś yāmu*///), 1481.a a3 (///*ce nauṣ yāmu ṣai* –///), 1530 a1 (///·*sk*· *kau no yāmoṣ*), 1536.c b5 (///*yāmu kowcä weña*–), 1575.b a3 (///*warttone nesañ̈ñe yāmu*), 2476 a2 (///–––*yāmu*–///), 2599 b2 (///–*yāmu*–///), 4092 a2 (*yneś yāmoṣepi*), IOL Toch 27 a4 (...*nenne yāmoṣäṃ*), IOL Toch 43 a5 (*omp omp yāmuwa yāmuwa*; passive acc. to trsl. CEToM) IOL Toch 79 b2 (///–*reṃ yāmusa*), IOL Toch 102 a3 (///*yāmu kleś ka*///), IOL Toch 114 b3 ((*yo*)*lo yamorcci yāmo*(*ṣ*)///), IOL Toch 141 a3 (///(*a*)*pākärtse yāmu* : *cewsa śaiṣṣentse*///), IOL Toch 174 b5 (///*t*· *yāmoṣ ekañi*///), IOL Toch 201 b5 (– *wī* – *yāmu āryavaṃśänt*·///; probably *wīna yāmu* 'having taken pleasure'), IOL Toch 226 b2 (///*yāmu po cwī* –///), IOL Toch 235 a4 (///*lyñ*· *tane yāmu ste* · *m*· –///), IOL Toch 157 b3 (///(*y*)*āmoṣä*), IOL Toch 240 b4 (*cīne yāmu larauñe kalppanmasa mā*///), IOL Toch 255 b3 (...*na yāmu mäsketä*), IOL Toch 280 b6 (*lo yāmu ste*; active trsl. by CEToM), IOL Toch 367 b2 (///*yāmoṣäṃ*), IOL Toch 754 a2 (///·*e yāmu śaul*·///), IOL Toch 827 a2 (///·*n*· *saim yāmu* –///), IOL Toch 828 b2 (///(*sadga*)*t*(*i*)*m yāmoṣ*· ·///), PK AS 12K b1, PK AS 15H a3 (///*yāmu wräntsaimeṃ ā*///), PK NS 11 b4 (*yāmu*), PK NS 112 (///*māka pñi nauṣ yāmuwa saṃsārne kleśanma*; this TB fragment is somewhat puzzling due to the use of TA *pñi* rather than TB sg. *yarpo* or pl. *yärponta*; cf. THT 242 a4 *lkäntär-c yāmwa nauṣ yärponta*. The translation in CEToM is active, 'having previously done much merit, ... the passions in the saṃsāra ...', but it is unclear what *yāmuwa*, f.pl., agrees with; *kleśanma*?) PK NS 305 & 306 b2, SI 5873 a17, W 9 b4 (*sā* (*yā*)*musa ka*).

h. THT 107 b2 (Buddha legend; prose; trsl. based on CEToM)
*snai epiṅkte śuddhawāsäṣṣi ñakti rṣākäññe weṣ*(OBL) ***yāmoṣ*** *mā lauke stāna ñor lymāre*
'Without delay, Śuddhavāsa gods, having **donned** [lit. **made**] the guise of a sage, sat down not far away under [some] trees'

i. THT 133 a4 (Kāvya; verse; trsl. Adams 2013: 168; cf. Saito 2006: 493)
*krent yāmor*(OBL) *mā* ***yāmoṣäṃ*** *ce*(*n n*)*o śäccätsī pkate*
'He intended, however, to let [**those who had**] not **done** a good deed proceed'

j. THT 203 b5 & 204 a3–4 (Buddhastotra; verse; trsl. CEToM)
*tañ pernerñe*(OBL) ***saim yamoṣ*** *perneñc ka po klautkāre*
'All **who made** your glory their **refuge** have already turned glorious.'

k. THT 273 a5 (Buddhastotra; verse; trsl. CEToM):
*snai keś wā wes cī*(OBL) ***saim yāmoṣ*** *tallāñciśkaṃ*[81]
'For in countless numbers we miserable ones **have made** you our **refuge'**

l. THT 288 a1 (Kāvya; verse; trsl. CEToM)
(*ka*)*ruṃ*(OBL) ***nauṣ yāmu***
'**Having put** (lit. "done") compassion at the forefront ...'

m. THT 542 a3 (Doctrine; prose; trsl. based on Saito 2006: 493)
*avibhaktasya · mā putkoṣepi · pāke*(OBL) ***yāmoṣepi*** *wat*
'[Skt.] Of something that is not detailed · Of something that is not divided · or of something that is not **shared** (lit. "do part") in'[82]

n. THT 1102 a1 (Karmavācanā; prose; trsl. CEToM; cf. IOL Toch 827 a4)
*upādhy*(*ā*)*y*(*e*)*ṃ*(OBL) ***saim yā*(*mu*)** ///
'**Having made** the Upādhyāya (teacher) a **refuge** ...'

o. THT 1520 a2 (trsl. CEToM)
///(*kä*)*ly*(*m*)*īntsa irge*(OBL) ***yāmu***
'**having made/done** *irge* (?) in all directions, ...'

p. THT 92 a5
*añcalī ṣarne* ***yāmu***
'Having **put** (lit. "done") [his] hands in *añjali* position.'

Same expression with slight variations: 99 b1 *añcali ṣarne yāmu*, 100 b6 *añcalī ṣarne yāmu*, 109 a7 *añjal ṣarne yāmusa*, 261 a4-5 *añcāl ṣärne yāmusa*, IOL Toch 97 a3 *añcali* (*ṣa*)*rne yāmu*, 178 a4 *añcali ṣarne yāmu*, 261 a3 ///(a)ñcal ṣarne yāmu –, PK AS 13J b1 añcalyi ṣarne yāmu, PK NS 144 b3 ///(a)ñcalyi ṣarne yā(mu)///

[81] For *tallāñciśkañ*.
[82] See SWTF I: 174a.

q. IOL Toch 5 b4 (Confession; verse; trsl. CEToM):
*cine **yāmu** śrigupti*(OBL)
'Who **has taken** excellent protection in you, ...'

r. IOL Toch 31 a3 (Daśakarmapathāvadānamālā; verse; trsl. CEToM)
*///...sa cimpa āklyi*(OBL) ***yāmoṣ** wes*
'... we have **studied** (lit. "do learning") with you.'

s. IOL Toch 92 b2 (Karmavācanā; prose; cf. also IOL Toch 776)
(*sāṅ*)*k*(OBL) ***saim yāmu***
'**Having made** the community (into) a **refuge**.'

t. IOL Toch 152 b4 (Udānavarga; prose)
(*wī*)*na*(OBL) ***yāmu** mäsketrä*
'He **has found** pleasure.'

u. IOL Toch 152 a7 (Udānavarga; prose)
*wīna*(OBL) *y*(*āmo*)*ṣ*
'They, **having found** pleasure, ...'

v. IOL Toch 167 (Medicine/magic text; trsl. Peyrot 2013: 714)
*kāwalñesa rano traṅko*(OBL) ***(y)āmu** tākaṃ*
'Even if he has **committed** (lit. "done") an offence by craving, ...'

w. PK AS 1A a2 (Udānavarga; prose; trsl. based on CEToM)
*tusāksa ṣek ompalskoññene wīna*(OBL) *yāmoṣ kakraupaṣäṃ*
'Therefore, indeed always **finding** pleasure in meditation [and] being concentrated, ...'

x. PK AS 7C a4 (Karmavibhaṅga; verse; trsl. CEToM):
*tane ṣemi ksa onolmi yāmor*(OBL) ***yāmoṣ** nrayne cmelyesa kakraupau ṣpä tākan-me*
'[There are] some beings **who have done** a deed, and by being reborn in hell it will be further accumulated by them.'

y. PK AS 7F b3–b4 (Karmavibhaṅga; verse; trsl. CEToM)
***yāmu** yärponta*(OBL) *ṣe /// (kektsentsa) ra skw*(*a*)*ssu mā no palskosa*
'**Who has done** meritorious deeds ...[83] he [is] happy with the (body) indeed, but not with the mind.'

z. PK AS 7G b5 (Karmavibhaṅga; verse; trsl CEToM)
*tane ṣemi ksa wnolmi mäskenträ śconai-yparwe duśśīlñesa yolain yāmor*(OBL) ***yāmoṣä***
'There are some beings here **having done** bad deeds because of their ill conduct that originates in hatred.'

[83] Apart from restored *kektsentsa*, eight more syllables are missing.

aa. PK AS 16.2 b1 (Aggañña-Sutta; verse; trsl. CEToM)
*warto wṣeṃñai*(OBL) *saim yāmoṣ*
'**Having made** a forest shelter their **refuge**, ...'

bb. PK AS 17A a6–b1 (Supriyanāṭaka; prose; trsl. CEToM)
***kau(c) yāmu*** *ṣukt naumīyenta*(OBL) *kauṃ-pi*(*r*)*k*(*o*)*ṣṣai k*(*al*)*ym*(*ine*) *ma*(*sa*) *c*(*akravārt*)*t*(*i*) *w*(*a*)*lo ra*
'**Having lifted up** the seven jewels he (went) in eastern direction, like a *cakravartin* king.'

cc. PK DA M 507.37 and .36 a83–84 (Account; prose; trsl. Ching 2010: 215)
*saswa ce*(OBL) *wes*(*s*)*i* – (*pu*)*ttewante snai-paille* ***yāmu*** *ste*
'Lord, **what** Puttewante **has** lawlessly **done** to us, ...'

dd. PK NS 56 a5 (Doctrine; prose; trsl. CEToM)
/// *-ntane āklyi*(OBL) *yāmoṣ*
'**Having studied** in ...'

ee. SI B 119(2) b5 (Udānavarga; prose; trsl. CEToM)
(*pelai*)*kne*(OBL) ***yāmu*** *nauṣäñ̈ñe*
'**Having accomplished** the ancient Law'

14. Tocharian A active use of the preterite participle *yāmu*.[84]

a. A 9 a2 (Puṇyavantajātaka; prose; trsl. based on Sieg 1944: 12 and CEToM)
*wrasañ penu cam peke pälkoräṣ* ***yneś pälskaṃ yāmuṣ*** *träṅkiñc*
'And the people who saw the painting **took** [it] **for real**, they say ...'

b. A 13 b6 (Puṇyavantajātaka; verse; trsl. CEToM)
*pñintu*(OBL) ***yāmu*** *wrasom mā yäṣ tālorñeyaṃ mā yäṣ ñātseyaṃ*
'A person **that has practised** virtues does not come into misery, does not come into hardship'

[84] Not included here are instances of *yāmu* that are either unclear as to their diathesis, or that occur in a (fragmentary) context where I am unsure of a translation: A 45 a5 (*modin yāmunt*); 69 a4 (///*yāmu* ·); 90 a3 (*jñ·* – *ntu yāmu ṣñy akälyme*; active acc. to Saito 2006: 491); 121 a1 (*was sem yāmuṣ*); 123 a3 (///*ārwar yāmu ṣe*); 148 a6 (*mänt wäknā ne yāmu tā*; active acc. to Saito 2006: 490); 167 b5 (///*·yāmuṣ* : *ysāl·*///); 179.c a5 & a6 (///*yāmuṣ tmäk śweñc*///); 220 a3 (*kuc yāmu ā*///); 236 b7 (*wr··t·* – *m· yāmu* : *knānmāṃ sä* ///); 281 b2 (///*·yāmuṣ* :); 310 a6 (///*yāmunt ālyeksac··* –; passive acc. to Saito 2006: 490); 318/319 a7 (///*vidyādharñāñ prusaṃ yāmunt*///); 418 b2 (///*muntaṃ lok yām··* ///; active acc. to Saito 2006: 491); 460 a5 (///rūpā wañi yāmu –) & b2 (///*āṣtär yāmu* –; active acc. to Saito 2006: 491). THT 1412h a2 ///*tim yāmu* is probably active *pratim yāmu* 'having made the decision', and similarly THT 2449 ///*·lypuntu yāmu* : /// for *lyalypuntu yāmu* 'having done deeds'.

c. A 16 b4–b5 (Puṇyavantajātaka; verse; trsl. CEToM)
*puk ceṣ pñintu*(OBL) ***yāmuntāp*** *puṇyavāni parnontāp yokmaṃ klyaṃtär*
'But they are all standing at the gate of splendid Puṇyavān, **who has practised** virtues.'

d. A 22 a5 (Puṇyavantajātaka, Bṛhaddyutijātaka; trsl. CEToM; cf. also fragmentary A 280 b3)
*pariweṣ*(OBL) ***yāmu*** *koṃ-ñkätt oki*
'Like the sun, **having produced** a halo, ...'

e. A 78 a3 (Ṣaḍdanta-Jātaka; verse; trsl. CEToM)
*caṃ wraman*(OBL) *ne*(*ṣa*)*ṣi praṣṭaṃ āklye* ***yāmu***
'(He **who**) **has educated** himself in this matter in an earlier time.'

f. A 148 a1 (prose; see CEToM)
*ṣom caṃ lyalypu*(OBL) ***yāmu***
'**Having done** this one deed.'

g. A 240 a4 (prose; see CEToM)
(*a*)*surāsäṣṣ aci wrasañ añu*(OBL) ***yāmūṣ***
'From the Asuras to the beings, they **were** silent ...'

h. A 244 a3 (verse; trsl. based on Thomas 1957: 280)
*mañkal*(OBL) ***yāmu*** *tñi pälskaṣi ñemi*
'... Your bliss-**providing** pearl of a mind ...'

i. A 279 b7 = A 302 a1 (Maitreyasamiti-Nāṭaka; verse; trsl. = 302 a1)
*p*(*r*)*em wyār*(OBL) ***yām***(***u***) *tāṣ* ///
///*prem wyār*(OBL) ***yām***(***u***) *tāṣ* –
'If he has **erected** (lit. "made") a temple (?) and a cloister.'

j. A 291 b8 (Maitreyasamiti-Nāṭaka; verse; trsl. CEToM)
*yok*(OBL) ***yāmu*** *jambunāt wäss oki āṣtär lukśanu wsā-yok ya*(*ts*) ///
'[His] skin **has produced** a colour pure like Jambunāda gold, shining and gold-coloured.'

k. A 336 a6 (Dialogue; verse; trsl. CEToM)
(*śäk tampe*)*yntu*(OBL) *yneś* ***yāmu*** *kucn= ākṣ*(*i*)*ññā märkampal säm sutär tāk*
'**Having manifested** (lit. "made manifest") (the ten powers), what he proclaimed as the law, this became the Sūtra.'

l. A 336 b2 (Dialogue; prose; trsl. CEToM)
(*ā*)*klye*(OBL) ***yāmuñcäs*** *ñäktas napenäśśi ākṣiññā säm abhidharm-piṭak tāk*
'(**What**) he **proclaimed** to learned gods and men, this became the *Abhidharma-Piṭaka*.'

m. A 347 a4–b1 (verse; trsl. based on CEToM):
*śil pat nu pāpṣu **yāmu** pat spaktāṃ*(OBL) : *knānmuneyaṃ spaltäk **yāmu***
'... or **has observed** good behaviour, or has done service, or **has made** an effort for knowledge ...'

n. A 386 a2–3 (Abhidharmakośa; prose; trsl. CEToM)
*oktasāṃ ytār*(OBL) *yneś **yāmuntāp** kus ne säm ṣūrmaṣ tām praṣtaṃ tsārwo kācke kātäṅkāṣ säm prāmodya plānto träṅkträ*
'Which comfort and joy arise for that reason at that time **to the one who has manifested** the eightfold path, this is called *prāmodya* joy.'

o. A 394 a4 (verse; trsl. CEToM)
*waṣtaṣ läntässi pratim*(OBL) ***yamu***
'**Having made** the decision to leave home, ...'

p. THT 1152 b2 (verse? trsl. CEToM)
*koc kanweṃ*(OBL) ***yāmu***
'**Having lifted up** (lit. "done upward") [his] knees.'

q. YQ I.6 a8 (Maitreyasamiti-Nāṭaka; prose; trsl. CEToM)
*p$_u$kis puk essi pratim*(OBL) ***yāmu***
'I **made** the resolve to give all of them away to all.'

r. YQ II.2 b7 (Maitreyasamiti-Nāṭaka; prose; trsl. CEToM)
*///*(*pari*)*weṣ*(OBL) ***yāmu** puk yärśār*
'**Having made** a halo all around.'

s. YQ II.13 b4 (Maitreyasamiti-Nāṭaka; prose; trsl. CEToM)
*///nu ptāñkät käṣṣinaṃ wsokone*(OBL) *tsopatsäṃ **yneś yāmu***
'**Having made manifest** the great joy towards the Buddha-god the teacher.'

From the passive examples below, it can be seen that in the absence of an oblique, the participle is typically passive. In such instances an agent or instrument is often expressed using a genitive, a perlative (in TB) or an instrumental (in TA only), but this is not necessary.[85] Examples illustrating the passive use of *yāmu* are given here for Tocharian B in (15), and for Tocharian A in (16).

15. Tocharian B passive use of the prt.ptc. *yāmu*:[86]

a. THT 30 a3–4 (Udānālaṅkāra; verse; trsl. based on CEToM)

[85] According to Meunier (2017: 229-230) the use of the genitive is original to Tocharian, while the perlative and instrumental are rather due to Sanskrit influence.

[86] According to Saito 2006: 584, 36% of passive preterite participles are construed with an expressed agent in Tocharian B.

*snai keś cmela karsatsiśco ṣpä tne ytārye sā śpālmeṃ : śaiṣṣe kärsaucaisa*(PER.A)
***apākärtse yāmusa*** *klyomña*
'And to recognize the countless births, this path **was revealed** (lit. "made manifest") as the best one by the knower of the world, the noble (path)'

b. THT 46 a1 (Udānālaṅkāra; verse; trsl. CEToM)
(***rä***)***skre yāmu*** *erkatte r·w*[87]*=attsaik päst snai-*(*mai*)*yya*
'**Having been** extremely **mistreated** the powerless one completely despaired(?)'

c. THT 119 a4–5 (Rājavṛtti; verse; see CEToM)
(*k*$_u$*se wa*)*t no cmelane* ***yāmwā*** *yāmornta yolaina krenta wat p*(*o*)///
'Or which deeds **were done** in the births, bad or good, all ...'

d. THT 119 b1 (Rājavṛtti; verse; trsl. after Peyrot 2013: 713)
(*krenta yā*)*mornta mākaikne no śaumontse k*$_u$*se* ***yāmwa*** *tāk*(*aṃ*)
'Which (good) deeds **are done** for a person in many ways, ...'

e. THT 175a2 (Abhidharmakośa; prose; trsl. after Thomas 1957: 289)
*ce kārikne* ***vyākhyāṃ yāmu*** *päst tāka totä kca* ||
'... **has been explained** approximately so much in this Kārikā.'

f. THT 202 b1 (Abhidharmakośa; prose; see CEToM)
*na ca tatkaraṇaṃ mana · mā ṣpä ātmaṃttse*(GEN.A) ***yāmu*** *palsko*
'(Skt.) · and [the thing] **not done** by the self is the mind.' (?)

g. THT 212 a2–3 (Buddhastotra; verse; trsl. based on CEToM)
*karuṇäṣṣe bhṛṅkār eṅku ṣärsa ymeṣṣe maiyttärṣṣe no warsa ite po* ***yāmoṣä***
'Having seized the vessel of compassion with the hand of your mind, **filled** [lit. 'made full'] with the water of friendship, ...'

h. THT 242 a4 (Buddhastotra; verse; see CEToM)
*lkāntär-c*(PRON.SUFF.A) ***yāmwa*** *nauṣ yärponta*
'They see the good deeds **done** by you earlier.'

i. THT 257 b5 (Kāvya; verse; see CEToM)
/// (***y***)***āmu*** *ñäktents*(GEN.A) *mäsketrä* ///
'It is **made** by the gods.'

j. THT 299 a1 (Udānavarga; verse; see CEToM)
(*ke*)*te yolo* ***yāmu*** *yāmor kreṃtsa*(PER.A)
'By whom a deed **has been** wrongly **done**, [if he covers it up] with a good [one] ...'

k. THT 330 b2 (Doctrine; trsl. based on CEToM)
*su wasaṃnpāt* ***yāmu*** *mā mäsketär*

[87] The proper translation and restoration of this (presumable) verb, *r·w* is uncertain.

'He **has** not **been** ordained.'

l. THT 528 a4 (Doctrine; prose; see CEToM)
/// *ep*(*e*) *aly*(*e*)*ksa*(PER.A) ***yāmu***
'Or **done** by another'

m. PK AS 2C a6–b1 (Medicine/magic; prose; trsl. CEToM)
*toy ṣotruna ṣpä tuntse*(GEN.A) ***yāmuwa*** *po tekanmats rano ṣe-ñemmats*[88]
'And these symptoms **induced** by that [belong] also to all diseases which have the same name.'

n. THT 521 a3 (Karmavibhaṅga; trsl. CEToM)
(*mäksu no yā*)*mor* ***yāmu*** *ślek ṣpä kakraupau*
'But a deed **that is done** and accumulated'

o. THT 521 a4 (Karmavibhaṅga; trsl. CEToM)
*k*$_u$*se no su yāmor mā* ***yāmu*** *mā kākraupau*
'But which is the deed **which has** not **been done** and not accumulated?'

p. THT 1574 b1 (Letter; trsl. CEToM)
*iśśapakentse*(GEN.A) ***yāmoṣ*** *saṅkrāmne ñake päkwalñesa kartse – –*
'In the monastery **made** by Iśśapake, now with trust ... good ...'

q. PK AS 3B a2 (Medicine/Magic; prose; trsl. based on CEToM)
*istak* ***yāmoṣ*** *pili lakoy*
'As soon as boils **emerge** (lit. "are made"], one should examine [them].'

r. PK AS 3B b6 (Medicine/Magic; prose; trsl. CEToM)
*k*$_u$*se salturna* ***yāmu*** *tākoy*
'If somebody **has been made** *salturna* [= bewitched??].'

s. PK AS 6C a5 (Udānālaṅkāra; verse; trsl. CEToM)
*ṣim mā prākre aipu no iścemtsa oṃṣmeṃ mā* ***yāmu***
'But the roof [is] not firmly covered and not **done** with clay on top.'

t. PK AS 7C a6–b1 (Karmavibhaṅga; verse; trsl. CEToM)
*empele rano* ***yāmu*** *ṣey-ne*(PRON.SUFF.A) *yāmor su* : *onmin no postäṃ yamaṣate mrau*(*skāte*)
'Although such a horrible deed **had been done** by him, nevertheless he felt remorse afterwards [and] felt revulsion.'

u. PK AS 7C b2 (Karmavibhaṅga; verse; trsl. CEToM)
*empelona ra* ***yāmwa*** *tākaṃ yāmornta*
'Even [if] horrible deeds **have been done**, ...'

[88] sic. Perhaps for a genitive plural *ṣe-ñemnaṃts*.

v. PK AS 7C b3 (Karmavibhaṅga; verse; trsl. CEToM)
*se*(*ṃ*) *t*(*u-*)*yäknesa yāmor* ***yāmu*** *ket*(GEN.A) *tākaṃ*
'By whom a deed of this kind may **have been done**, ...'

w. PK AS 7C b4 (Karmavibhaṅga; verse; trsl. CEToM)
*intsu no yāmor* ***yāmu*** *mā no kakraupau*
'But which [is] the deed **done** but not accumulated?'

x. PK AS 7C b5 (Karmavibhaṅga; verse; trsl. CEToM)
*se yāmor* ***yāmu*** *mā no kakkraupau mäsketrä*
'This deed is **done** but not accumulated.'

y. PK AS 7C b5–6 (Karmavibhaṅga; verse; trsl. CEToM)
*k*(*ᵤs*)*e no sū yāmor kakraupau mā no* ***yāmu***
'But what [is] the deed accumulated but not **done**?'

z. PK AS 7E b6 (Karmavibhaṅga; verse; trsl. CEToM)
*– – täntär-me lau*(*ka*)*ñña*n*a* ***yāmuwa***
'... for them (the deeds) **having been done** in the past'

aa. PK AS 7E b6 (Karmavibhaṅga; verse; trsl. CEToM)
*palsko no yairu ailñene āyor nano-nano* ***āklyi yāmu*** *māka cmela-cmelane*
'But the mind has been trained in gift giving [and] again and again **instructed** many times birth after birth'

bb. PK AS 7F a3 (Karmavibhaṅga; verse; trsl. based on CEToM)
*mā āyorṣṣe kreṃt wāṣmoṃmpa* ***āklyi yāmu*** *cmelane*
'He **has** not **been educated** during the [previous] births with a good friend of a gift.'

cc. PK AS 7K a3 (Karmavibhaṅga; verse; trsl. CEToM; cf. THT 46 a1)
*srukalläññentse*(GEN.A) ***räs***(***k***)***r***(***e yā***)***moṣ*** *erkatte*
'By death they have **been** severely **maltreated**.'

dd. PK AS 17E a3 (Supriyanāṭaka; prose; trsl. CEToM)
*viśvakarmi ñäkteṃtse*(GEN.A) *ra nermit* (***y***)***ām***(***u***)
'It **has been fashioned** by the god Viśvakarman, as it were.'

ee. PK AS 17F a5 (Supriyanāṭaka; verse; trsl. CEToM)
*yaitkor āk*(*ṣ*)*o*(*r*) *seṃ* (*kᵤse ste maiy*)*y*(*a*) *kete*(GEN.A) *po* ***yāmu***
'(Which is) this order [and] instruction by which all the power has been **shown** (lit. "made")?'

ff. PK AS 18B a3–4 (Vinayavibhaṅga; prose; trsl. Adams 2013: 739)
*sa*(*ṅ*)*k*(*a*)*meṃ ṣamāni* ***samanit yāmoṣ*** *ṣeyeṃ*
'Monks from the community **were assembled** ...'

gg. PK NS 53 b4–b5 (trsl. CEToM)
*kaklāyaṣepi pīle **yāmu** ra ñem er(sn)a*
'Name and form [are] like the wound **inflicted** on the one who has fallen.'

hh. PK NS 399 a4 (Candraprabha-Avadāna; prose; trsl. CEToM)
*ṣ(u)ktañce kauṃ mäñcuṣkeṃtse lantuññe abhiṣek **yāmu***
'On the seventh day the royal consecration of the prince **is performed**.'

ii. IOL Toch 62 b3 (Bhikṣu-Prātimokṣasūtra; prose; trsl. after CEToM)
*k$_u$se ṣamāne mā kappi **yāmoṣ** wassi ya(masträ)*
'If a monk makes a garment **that has** not **been** properly **made**.'

16. Tocharian A passive use of the prt.ptc. *yāmu* (in sentences a.–f. with a genitive agent, in g.–i. with an instrumental, j.–l. with neither):[89]

a. A 7 b3–4 (Puṇyavantajātaka; prose; trsl. after Sieg 1944: 10 and CEToM)
*mā naṣ āñmaṃ wrasaśśi*(GEN.A) *āñmaṣi īme **yāmu** wäṣpā* (*pe*)*nu āñcäm mā naṣ*
'There is nothing in the self. The image of the self **has been made** by the beings. And truly, there is no self.'

b. A 14 b4 (Puṇyavantajātaka; prose; trsl. CEToM)
*puk krañcäs wrasaśśi*(GEN.A) *kāpñe **yāmunt** kra – –*
'(They.F) **are loved** (lit. "made beloved") by all good beings.'

c. A 49 a5 (Avadāna; prose; trsl. after CEToM)
*/// puṣyamitres lānt*(GEN.A) *kat **yāmunt** wyāräs saṅgrāmäs ///*
'... **having been** destroyed by king Puṣyamitra, (Kaniṣka had built) cloisters and monasteries'

d. A 49 b4 (Avadāna; prose; trsl. CEToM)
*märkampalṣināp śpeśuy lānt*(GEN.A) ***yāmunt** wyār*
'The cloister **founded** by the Law-abiding king Śpeśu.'

e. A 49 b5 (Avadāna; prose; trsl. CEToM)
*märkampalṣināp kanaṣkes lānt*(GEN.A) ***yāmunt** wyār*
'The cloister **founded** by the Law-abiding king Kaniṣka.'

f. A 59 a5 (Unmādayantī-Jātaka; prose; trsl. CEToM)
(*k$_u$leñci ñe*)*mi **ynāñm yāmus** śomiṃ ṣeṣ*
'She was a girl **honoured** as a pearl (of women).'

g. A 79 a5 (Ṣaḍdantajātaka; prose; trsl. CEToM)
*täm kaklyuṣuräṣ śaru bodhisatveṃ swäräṃ rakeyo*(INSTR) *wsok **yāmu***

[89] According to Saito 2006: 584, 39% of passive preterite participles are construed with an expressed agent in Tocharian A.

'When he had heard this, the hunter, **having been made** glad by the sweet speech of the Bodhisattva, ...'

h. A 79 b5 (Ṣaḍdantajātaka; verse; trsl. CEToM)
*kāswoneyo*(INSTR) ***yā***(***mu***) *nu päl ākntsāśśi sätkatär tri āpāytwaṃ*
'But a wound **which has** out of virtue **been made** to the ignorant ones extends over the three evil forms of existence.'

i. A 93 a5 (Nanda-Carita I; verse; trsl. based on Saito 2006: 490)
*sākkres parnoñcäśśi*(GEN.A) ***kāpñe yāmu*** *saṃkrām nyaghrodha*
'The monastery Nyagrodha, **loved** by the happy, glorious people.'

j. A 172 a6 (trsl. CEToM)
*-āp*(GEN.A) ***yāmunt*** *jetavaṃ sa*(*ṅkrāma*)
'The Jetavana monastery **founded** by ...'

k. A 231 a4 (verse; untrsl. in CEToM)
*neṣā* ***yāmunt*** *lyalypūntu*
'The deeds **that were done** earlier.'

l. A 270 b8 (Maitreyasamiti-Nāṭaka; prose; trsl. CEToM)
(*klyo*)*m tu neṣinäṃ ṣñi* ***yāmunt*** *pratim opyā*(*c päklār*)
'O noble one, remember the resolution **made** by yourself in the past!'

m. A 340 b1 (Koṭikarṇa-Avadāna; verse; trsl. CEToM)
*ṣñy entsyo*(INSTR) ***y***(***ā***)***munt*** *lyalypäntwā* (*was pretañ*)
'Because of our deeds **committed** due to our miserliness, (we [are] pretas).'

n. YQ III.6 b7 (Maitreyasamiti-Nāṭaka; prose; trsl. CEToM)
///(*pi*)*ssaṅk penu* ***yärkā yāmu*** *tāśśi*
'The community, too, will **be honoured** by you.'

The example sentences show forms of *yāmu* as both active and passive in various functions in the sentence, illustrating the contextual orientation of the preterite participle. It is contextually oriented regardless of whether it is used as an attributive, absolutive, or part of the predicate (cf. also Saito 2006: 489–497, who gives a few examples of each type).

Besides Tocharian, reflexes of the perfect participle in other Indo-European branches are found in Indo-Iranian, Greek and Balto-Slavic, with a few relic formations in Germanic and Celtic. It is not found in Anatolian, or in Armenian, Albanian and Italic (e.g., Rau 2017). This participle is reconstructed as actively oriented for Proto-Indo-European, although some labile verbs are also attested with a passive meaning in Ancient Greek (Kümmel 2020: 17, 24). The fully contextual orientation of this participle in Tocharian is not paralleled elsewhere in Indo-European, indicating that it is a

Tocharian innovation. The development in orientation from Proto-Indo-European is represented in the table below.

Table 7.3: Schematic change in the orientation of the perfect participle from Proto-Indo-European to the preterite participle in Proto-Tocharian.

| Indo-European | | Tocharian |
|---|---|---|
| active (S/A) | → | contextual (S/A/P) |

### 7.2.4 The gerundives: necessitative and potential participles

The two gerundives in Tocharian are fully inflected adjectives derived from the present and subjunctive stems with a suffix TB *-lle*, TA *-l* (**-l̥i̯o*). The present-based gerundive expresses necessity, while the subjunctive-based gerundive expresses possibility. The gerundives can be used attributively, and they also often occur as part of the predicate, with or without a form of the verb 'to be' in auxiliary function (Krause & Thomas 1960: 186–187). In a syntactical study of these formations, Thomas (1952: 22ff.) points out that their use is mostly passive, but that some occurrences are rather to be interpreted as active (Thomas 1952: esp. 22–26).

The gerundives behave very much like participles (Thomas 1952; Fellner 2013: 74), motivating their inclusion in this chapter. Meunier describes them as "indifferent to diathesis" (Meunier 2017: 222), by which she might refer to the fact that they are formed from both active and middle voice verbs, but Kuritsyna (2022) listed them together with the preterite participles as examples of contextually oriented formations in Tocharian. The sentences in (17) below give a brief overview of the uses of the gerundives.

The present gerundive is most often found with a passive interpretation when formed to a transitive verb, as in (17a); see Thomas (1952: 16–28). The subjunctive gerundive is found with good representations of both passive and active use, as illustrated in (17b) and (17e) respectively; see Thomas (1952: 28–47). The gerundives can be formed to intransitive verbs as well; in such cases they are by necessity active. This is shown in example (17c) with a present gerundive and in (17d) with a subjunctive gerundive (Thomas 1952).

17. a. Tocharian B (THT 600 b5; trsl. after Thomas 1952: 19)

| *kamarttañ̇ñe* | *wai* | *amāṃ* | *karuṇäntse* | *sañī* |
|---|---|---|---|---|
| rulership | and | pride | compassion:GEN.SG | enemy:NOM.PL |
| *bodhisatveṃts* | | ***wikṣalle*** | | |
| Bodhisattva:GEN.PL | | avoid:PRS.GER.M.NOM.SG | | |

'Rulership and pride, the enemies of compassion, [**are**] **to be avoided** by the Bodhisattvas.'

b. Tocharian B (THT 14 b1; trsl. after Thomas 1952: 32)

| (*lä*)*k*(*le*)*nta* | ***parskallona*** | *nraiṣṣana* | *lyakār*= |
|---|---|---|---|
| suffering:PL.OBL | fear:SBJ.GER.F.PL.OBL | hellish:F.PL.OBL | see:PRT.3PL |

*ākalṣlyi*
disciple:NOM.PL

'The disciples saw the **fearful** sufferings of hell.'

c. Tocharian A (A 66 a5; trsl. CEToM)

| *ṣäkk atsek* | *nāñi* | *brahmadatte* | *wäl* | *pats* | ***mäskal*** |
|---|---|---|---|---|---|
| certainly | 1SG.GEN | PN | king | husband | be:PRS.GER |

'Certainly, king Brahmadatta **shall be** my husband.'

d. Tocharian B (THT 88 a4–5; trsl. after CEToM)

| (*pā*)*lka* | *nai* | *mā-ṣekaṃñe* | *wäntarwats* | ***sparkālye*** *ā*(*ke*) |
|---|---|---|---|---|
| see:IMP.ACT.2SG | PTCL | inconstancy | thing:GEN.PL | with.perishable.end |

'Behold the inconstancy of things [**that**] **have** a **perishable** end.'

e. Tocharian B (THT 1574 a3; trsl. CEToM)

| *totka* | *śātresa* | *pruccamñe* | ***yamalle*** | *ṣait* |
|---|---|---|---|---|
| little | grain:PER | excellence | do:SBJ.GER.M.NOM.SG | be:IPF.ACT.2SG |

'With a little grain you **would be doing** something excellent.'

Thomas furthermore provides a few examples of active present gerundives, which are constructed with an oblique (Thomas 1952: 22–26). The clearest examples are all impersonal constructions, as in (18).

18. Tocharian B (IOL Toch 15 b4; trsl. based on Thomas 1952: 24)

| *tumeṃ* | *po* | *wäntarwa* | *paramānuntasa* | ***läkāṣäle*** |
|---|---|---|---|---|
| thereupon | all | thing:PL | atom:PL.PER | look:PRS.GER.M.NOM.SG |

'Thereupon, **one must look at** all things precisely.'
or: 'Thereupon, all things **must be looked at** precisely.'

The expected construction for a passive would have the gerundive agree with *wäntarwa* 'things' in the feminine plural form *lkāṣlona* or a spelling variant thereof.
Instances of potentially active present gerundives with an overt acting person are also generally difficult to interpret, and example (19) below is not without problems. For instance, *aiṣlyi* is to be read as *aiṣlye* (Thomas 1952: 25). There is another potentially active example of the present gerundive of this same verb 'to give' in (20), but its interpretation is not entirely certain either. A passive reading is perhaps also possible if the scope of the genitive *ket* is extended, or if an omitted genitive agent is assumed. The

translation could then be 'Therefore, the one whose wish [is] to be always happy, [and] by whom a gift should be given [while] always being joyful.' The present gerund *aiṣṣälle*, in the masculine singular, can agree with either *āyor*, an alternant singular noun, or with an unexpressed grammatically masculine entity implicated by *ket*.

19. Tocharian B (THT 492 a1–2; prose; trsl. after Peyrot 2013: 346)

| *sessatatte* | *rine* | *meskeṣṣeṃ* | *cāneṃ* |
|---|---|---|---|
| PN | town:LOC | connection:ADJZ.M.OBL.SG | coin:OBL.SG |
| ***aiṣlyi*** | *tākaṃ* | | |
| give:PRS.GER.M.OBL.SG | be:SBJ.ACT.3SG | | |

'Śeṣadatta **will have to give** the stringed coins in town.'

20. Tocharian B (PK AS 7E a5; trsl. CEToM)

| *tūsa* | *ket* | *āñme* | *skwassu* | *ṣek* | *nessi* |
|---|---|---|---|---|---|
| therefore | who:GEN | wish | happy | always | be:INF |
| ***aiṣṣäll***= | *āyor* | *katkemane* | *ṣek* | | |
| give:PRS.GER.M.NOM.SG | gift | be.happy:PRS.PTC | always | | |

'Therefore, the one whose wish [it is] to be always happy **should give** gifts [while] always being joyful.'

Instances of active use of the subjunctive gerundive are more numerous. These mostly occur in constructions with a copula, and they are characterized by Thomas as a periphrastic future or periphrastic subjunctive (Thomas 1952: 38–43). Several illustrative examples are of this type of construction in Tocharian B are to be found in the manuscript THT 107, where the two sisters Nandā and Nandabalā are looking for a worthy recipient of their special porridge.

21. a. Tocharian B (THT 107 b1–2; trsl. Peyrot 2013: 344)

| *cisa* | *kreṃnt* | ***kälālyana*** | *tākam* | *cwi* |
|---|---|---|---|---|
| 2sg.PER | good | bring:SBJ.GER.F.PL.NOM | be:SBJ.ACT.1PL | 3SG.GEN |
| *aiskem* | *ci* | *eṃṣke* | *tärkanam* | |
| give:PRS.ACT.1PL | 2SG.OBL | while | leave:PRS.ACT.1PL | |

'If we **can find** (lit. "bring up") somebody better than you, we will give [it] to him while we leave you [alone].'

b. Tocharian B (THT 107 b5–6; trsl. Peyrot 2013: 342)

| *comtsa* | *śpālmeṃ* | *dakṣināke̯m* | *mā* | *kälālyana* |
|---|---|---|---|---|
| 3SG.DIST.PER | excellent | worthy of gifts | NEG | bring:SBJ.GER.F.PL.NOM |

*nescer*
be:PRS.ACT.2PL

'You **can/will** not **find** (lit. "bring up") anyone better and worthier of gifts than him there.'

c. Tocharian B (THT 107 b10)

| *mā* | *tot* | *ñiś* | *pintwāt* | ***warpalle*** | *nesau* |
|---|---|---|---|---|---|
| NEG | so | 1SG.NOM | alms(OBL) | accept:SBJ.GER.M.NOM.SG | be:PRS.ACT.1SG |

'I **can**not **accept** any alms.'

This type of active use of the subjunctive gerund could be partially a result of the periphrastic constructions, but the following examples in (22) and (23) show that a predicatively used subjunctive gerundive can be both active and passive.

22. Tocharian B (THT 33 a5; trsl. based on CEToM)

| *makte* | *ksa* | *rintär* | *waipecc=* | *ekñenta* | *takarṣkñesa* |
|---|---|---|---|---|---|
| self | INDEF | give.up:SBJ.MID.3SG | possession | property | faith:PER |

| *mrauskalñesa* | *mā* | *ṣpä* | *sū* | *wteṃtse* | ***kraupalle*** |
|---|---|---|---|---|---|
| world.weariness:PER | NEG | CONJ | he | again | gather:SBJ.GER.M.NOM.SG |

'If someone gives up possession and property himself, out of faith and weariness of the world, he **will** not **gather** [it] a second time.'

23. Tocharian B (THT 205 a2; trsl. CEToM)

| *papalarsa* | *ci* | *ñakta* | *nervvaṃn* | *oko* | ***kälale*** |
|---|---|---|---|---|---|
| praising:PER | 2SG.OBL | god:VOC | Nirvāṇa | fruit | obtain:SBJ.GER.M.NOM.SG |

'By praising you, o god, the Nirvāṇa-fruit **can be obtained**.'

In general, attributive use of the gerundives is less common than predicative use, and it is to my knowledge not attested in active transitive function. Passive attributive use can be illustrated with example (24), where the agent is expressed in the genitive, as in the case of the other participles (Thomas 1952: 19).

24. Tocharian B (PK AS 17D b5–6; trsl. CEToM)

| *ñi* | ***yāmṣälye*** | *wäntare* | *pācer* | *walo* | *tu* | *yāmäṃ* |
|---|---|---|---|---|---|---|
| 1SG.GEN | do:PRS.GER.M.OBL.SG | thing(OBL) | father | king | it | do:SBJ.ACT.3SG |

'[My] father, the king, will do the thing **that would have to be done** [lit. has to be done] by me.'

In attributive function, the gerundives thus attest to at least an absolutive pattern of orientation. However, it turns out that these formations can additionally be used to relativize non-core participants in both Tocharian A and B. The first two example sentences below, (25a–b), show both a subjunctive gerundive, *kroplyāṃ* and *swāsallye* respectively, and a present gerundive, *eṣlyāṃ* and *yalye* respectively, with similar functions (cf. Thomas 1952: 26–27, 32–34).

25. a. Tocharian A (A11 a3–4; trsl. CEToM)

| *wartsi* | ***kroplyāṃ*** | *praṣṭā* | (*wa*)*rtsi* | *kot* | *ravaṇe* |
|---|---|---|---|---|---|
| assembly | gather:SBJ.GER.F.OBL.SG | time:PER | company | split:PRT.3SG | PN |
| *sne* | *knānmuneyā* | *nati* | ***eṣlyāṃ*** | *praṣṭā* | *nati* |
| without | knowledge:PER | power | give:PRS.GER.F.OBL.SG | time:PER | power |
| *kot* | *rākṣtsāśśi* | | | | |
| split:PRT.3SG | Rākṣasa:GEN.PL | | | | |

'At the time **that** the assembly **had to be gathered**, Rāvaṇa split the company out of ignorance. At the time **that** power **had to be given**, he split the power of the Rākṣasas.'

b. Tocharian B (IOL Toch 7 a1; trsl. CEToM)

| (*yente* | *ma*) | ***yalye*** | *īkene* | *ma* | ***swāsallye*** | *īkene* |
|---|---|---|---|---|---|---|
| wind | NEG | go:PRS.GER.M.OBL.SG | place:LOC | NEG | rain:SBJ.GER.M.OBL.SG | place: LOC |

'(Not) in a place **where** (the wind) **goes**; not in a place **where it rains**.'

c. Tocharian B (IOL Toch 7 a2; trsl. CEToM)

| *ma* | *wär* | ***tärkalye*** | *īkene* |
|---|---|---|---|
| NEG | water | release:SBJ.GER.M.OBL.SG | place:LOC |

'Not in a place **where** water **is released**.'

These constructions whereby non-core participants are relativized with a participle are found in other languages with contextually oriented participles as well, such as those cited by Haspelmath (1994: 154) for Lezgian and by Shagal (2019: 2) for Kalmyk. Both of these examples display contextually oriented participles used to modify not only subject,

agent or patient, but also a time or location at which an event is taking place.[90] These are repeated here in (26):

26. Lezgian (Haspelmath 1994: 154 (a.)) and Kalmyk (Shagal 2019: 2 (b.))

a. *ruš-a* *čar* *kx̂i-zwa-j* *juǧ*
girl-ERG letter:ABS write-IPFV-PTC day
'The day on which the girl is writing a letter.'

b. *küük-n-ä* *bičəg* *bič-ǯä-sən* *širä*
girl-EXT-GEN letter write-PROG-PST.PTC desk
'The desk at which the girl is writing a letter.'

In terms of the participants that can certainly be relativized using gerundives, there are thus examples of the subject being relativized, the direct object being relativized, and a non-core participant (location or time) being relativized. In predicative use, subjunctive gerundives are found as both active and passive, depending on the context. The orientation of the gerundives can be generally described as contextual, although present gerundives are practically always passive when formed to a transitive verb.

The orientation of the Proto-Indo-European precursor of the gerundives is controversial. The gerundive suffix for both the present and the subjunctive gerundive can be reconstructed as PT **-lle*, from PIE **-lḭo*. This is derived from a Proto-Indo-European adjective suffix **-lo* (Hackstein 2003: 54–55). It is uncertain whether such a verbal adjective would already have had a dedicated orientation, making any reconstruction difficult. The closest relative of the Tocharian gerundive suffix is the Armenian gerundive in *-li*, also derived from **-lḭo*, although the formation and/or specialization of this suffix may have been parallel in the two branches rather than inherited from Proto-Indo-European.

Similar to the Tocharian gerundives, the Armenian gerundive in *-li* can be either active or passive as well. Olsen gives some examples of the ambiguity of this type, as in *əndowneli* both 'acceptable' and 'accepting, receiving; receiver', and in *sireli* both 'lovable, beloved' and 'loving' (Olsen 1999: 336). Olsen also points out the similarities with the Tocharian formation (Olsen 1999: 398), but regards the development as secondary, writing that "if the type of inherited **lo*-stems, […], at some point indicated the simple notion of a verbal action, the conception of the original diathesis, which is apparently active (nomina agentis) in the proto-language, is easily neutralized […]" (Olsen 1999: 397).

[90] IOL Toch 141 a5 /// – *pentse yainmoṣ prekene · orto* /// is suggestive of the same possibility with preterite participles, translating to something like 'at the time when X was/were reached/obtained by …-*pe* (perhaps *kraupe* 'congregation'?), …' I am not aware of other examples of this type.

The same ambiguity towards passive or active orientation is found in the related Armenian infinitive, e.g., *gorcel* 'to work, working' or 'to be worked' as the context requires (Olsen 1999: 397). A third related formation is the past participle in Classical Armenian *-eal*. This was absolutive in its orientation ("passive-intransitive") when used adjectivally (Meyer 2017: 75–76).

In the other Indo-European languages, the **lo*-adjectives often have an active interpretation, such as in Latin *crēdulus* 'easy of belief, gullible' (Hackstein 2003: 55–57). Active use is also found in the Slavic *l*-participles, which became the basis for the active past tense; e.g., Russian active transitive *on ubíl* 'he killed' vs. active intransitive *on plyl* 'he swam'. Kuryłowicz has argued that the active orientation arose secondarily in these branches, with the original orientation being absolutive (Kuryłowicz 1970–72: 324). The active use found in Latin and Germanic **lo*-adjectives would have developed because verbal adjectives in **-to-* had taken over the passive function. Similarly, the exclusively active use of the Slavic perfect based on the same adjective suffix would be due to the introduction of a new mediopassive construction with the reflexive pronoun **sę*, which could have displaced the passive function of the **lo*-adjective (Kuryłowicz 1970–72: 326–327).

Evidence in favour of earlier contextual or absolutive orientation is restricted to Tocharian, where it may be secondary, and Armenian. In Armenian, the reverse argument can be made as for Germanic, Latin and Slavic, since the adjectives in **-lo-* took over from the adjectives in **-to-* (Meyer 2017: 76), so that the absolutive orientation there is just as suspect as the active orientation in the other branches.

If the gerundives formed to transitive verbs had in origin specifically active orientation, the contextual orientation of the Tocharian gerundives is a later innovation. This development would be in line with the new contextual orientation of preterite participles from earlier active perfect participles. A parallel development then has to be assumed for Armenian gerundives. It is noteworthy that the two languages where derivatives of the **lo*-adjectives show contextual orientation are spoken in areas where various neighbouring languages have contextually oriented participles (on which cf. below). The potential development from active or absolutive orientation to Tocharian contextual orientation is represented in a schematic way in Table 7.4 below.

Table 7.4: Schematic change in orientation from Proto-Indo-European to the Proto-Tocharian gerundives.

| Indo-European | | Tocharian |
|---|---|---|
| ? active (S/A) or absolutive (S/P) | → | contextual (S/A/P) |

### 7.2.5 The privative: a negative participle

The privative in Tocharian B functions as a negative participle, and it is formed by a combination of the negative prefix *en-* (with various allomorphs) and the suffix *-tte*. It is principally derived from the subjunctive stem of the verb. The origin of the negative prefix is PIE negative **n̥-*, and the suffix *-tte* may reflects PIE **-to-* (Hilmarsson 1991: 14–16 with further references). The gemination could be the result of analogy to that found in other common Tocharian adjectival suffixes like *-ṣṣe*, *-tstse*, *-lle*, *-ñña* (Hilmarsson 1991: 14–16). The derivation by Krause and Thomas (1960: 188) from **-tu̯o-* would account for the Tocharian B geminate by sound law, but is excluded due to the lack of a labial reflex in Tocharian A, where relic privatives end in simple *-t*, rather than ***-tu* (Hilmarsson 1991: 15-16).[91]

The privative is inflected as an adjective, agreeing with a head noun in gender, case and number. Unlike the preterite participle or the gerundives, the privative does not have a specific time reference, as it indicates that an action "is not to be, cannot or should not take place, or has not taken place" (Hilmarsson 1991: 16). This indifference to features like tense, aspect or mood is found in participial systems with a negative participle in several Uralic languages (Finnish, Komi-Zyrian, North Saami, Northern Khanty) and in the Dravidian language Telugu (Shagal 2019: 159). Similar to the participles discussed above, however, it also shows both active and passive meaning when formed to transitive verbs (cf. Hilmarsson 1991: 20–117 *passim* with 15 examples of passive privatives formed to transitive verbs and 5 active privatives from transitive verbs).

27. Examples of passive (a.–b), active (c.–d.) and intransitive (e.) use of various privatives:
    a. Tocharian B (THT 331 b1; trsl. Hilmarsson 1991: 21–22)
    *tusa olyapotse* ***akākatte*** *yaka ce prekṣäṃ*
    '[If he] **uninvited** asks him still more than this …'
    b. Tocharian B (PK NS 32 a2; trsl. CEToM)
    *pelaikneṣṣe laukañña=* ***eṣpirtacce*** *ṣparttaṣṣim cākkä*(*r*)
    'May I turn the wheel of the law **[that] has not been turned** for a long time.'
    c. Tocharian B (PK AS 12C b1–2; trsl. Pinault & Peyrot fthc.)
    (*samu*)*d*(*tä*)*rntse totteñña*(OBL) ***ekätkātte*** *se mā srukālñe yänmālle*
    '[Even] **without traversing** the limit of the ocean (of the [beings] of the five births) he can attain the [state] of not dying.'

[91] Fellner (2013: 85–89) proposes an alternative reconstruction for the suffix in **-teio-*, a modal derivative from a neutral abstract suffix **-ti-* found in Greek and Sanskrit.

d. Tocharian B (IOL Toch 9 b4; trsl. Broomhead 1962: 196)
*te **aikaccepi** kleśanma*(OBL)
'That [is the characteristic] **of one who is unable to destroy** kleśas.'

e. Tocharian B (THT 18 a3; trsl CEToM)
*ṣe **atākatte** waike tne weṣṣäṃ*
'One says an **unfounded** lie here.'

These examples show use of the privative as either passive (27a-b.) or active (27c-d.), depending on the context, so that it may be regarded as yet another contextually oriented participle formation in Tocharian. Example (27e.) contains the privative of an intransitive verb.

The Indo-European verbal adjective in **-to-*, from which the privative most likely derives, takes on absolutive orientation in most Indo-European languages, often as a resultative similar to the perfect. It was primarily patient-oriented, contrasting with the actively oriented perfect participle (e.g., Kümmel 2020: 18). It is possible that the original time reference of the Tocharian privative was likewise past, and its orientation absolutive, but this is in any case not the attested situation. Brosman (2010: 26) has argued for residual contextual orientation of *to*-derivatives in Ancient Greek ἄγνωτος, which can mean both active 'ignorant' and passive 'unknown'. Brosman connects the (re)orientation of the **to*-participle as absolutive with the newly active orientation of the *nt*-participle in the non-Anatolian Indo-European languages. Since the *nt*-participle in Tocharian had clearly become an active participle, it is possible that the original pre-Proto-Tocharian **to*-adjective was correspondingly already absolutive, as was its common function in the other non-Anatolian Indo-European languages.

If Tocharian did not share in the development of the *to*-adjective as an absolutively oriented resultative participle-like formation, the contextual orientation of the Tocharian reflex of the **to*-participle could be seen as an archaism. The basis for reconstructing absolutive orientation for this participle in Indo-European is much greater, however, and contextual orientation in Tocharian is in line with the development of other contextually oriented participles, where it is more clearly innovative. As long as Tocharian is not conclusively shown to have split off early based on other arguments, the orientation of the Tocharian formation itself is not a reliable basis for reconstructing the orientation in Proto-Indo-European. The development from absolutive to contextual orientation in Tocharian is schematically illustrated in Table 7.5 below.

Table 7.5: Schematic change in orientation from the Proto-Indo-European *to*-participle to the Proto-Tocharian privative.

| Indo-European | | Tocharian |
|---|---|---|
| absolutive (S/P) | → | contextual (S/A/P) |

### 7.2.6 Summary

Among the *nt*-participle, *m*-participle, preterite participle, the gerundives, and the privative, it turns out that only the *nt*-participle is consistently active in Tocharian. The others all more or less clearly show contextual orientation. The participial system of Tocharian can thus overall be characterized as mixed as regards participle orientation, with a preference for contextual orientation. Only the *nt*-participle has remained true to its Proto-Indo-European orientation, as illustrated in the table below.

Table 7.6: Overview of the development in orientation of the Tocharian participle suffixes.

| | PIE | | Toch. |
|---|---|---|---|
| **-nt* | active (S/A) | = | active (S/A) |
| **-mh₁no* | active (S/A) | → | contextual (S/A/P) |
| **-uōs* | active (S/A) | → | contextual (S/A/P) |
| **-to* | absolutive (S/P) | → | contextual (S/A/P) |
| **-lo* | ? act. (S/A) or abs. (S/P) | → | contextual (S/A/P) |

While many Indo-European languages do not have any such contextually oriented participles, Tocharian is not entirely unique. Some participles in Indic languages like Marathi are also contextually oriented (Shagal 2019: 99), and similar situations are found in Armenian as well with the participles in *-li*, derived from the same Proto-Indo-European source as the Tocharian gerundives (cf. Olsen 1999: 336, 337–338, and 7.2.4). What could be relevant in all these situations is the place where these languages are spoken. Armenian is spoken in the Caucasus, which is known to have many languages with contextually oriented participles in the Nakh-Daghestanian family (Shagal 2019: 1, 100; cf. also Authier 2012: 232 on contextual orientation in the Iranian language Apsheron Tat, spoken in the Caucasus); and similarly, the Dravidian languages, with which Indic languages have long been in contact, show a preference for contextually oriented participles as well (see Shagal 2019: 1, 98, 100).

Considering that changes in participle orientation are also thought to have occurred due to contact situations in the more western Uralic languages in contact with European Indo-European (Shagal 2018, see also below), these Indo-European languages with

contextually oriented participles that are spoken in areas where the surrounding languages have contextual orientation might be examples of the same kind of development. The prevalence of contextual orientation in Tocharian participles may thus likewise be due to foreign influence.

## 7.3 Samoyed

We have seen that the Tocharian participial system has innovated in two ways: most participles became contextually oriented, and new necessitative/potential and negative participles were created. To determine whether these changes can be understood from a Tocharian-Samoyed contact hypothesis, we must discuss the Samoyed participial system and its Uralic context. When looking at the Uralic language family in its entirety, the orientation of the participles is varied, with active participles, passive participles, absolutive participles and contextual participles attested in different Uralic languages (Shagal 2018). Within Samoyed, however, the orientation is generally contextual, and as in the other chapters, we will focus on this branch first and foremost. A number of different participles can be reconstructed for Proto-Samoyed, although correspondences in the morphological material between the separate branches show that distinction with other categories like verbal nouns seems to have been fluid (see Shagal 2018).

Discussions of the prehistory of the various participle formations in Samoyed languages can be found scattered in the literature (e.g., Janhunen 1998: 473–474, Mikola 2004: 131, Aikio 2022a: 19), and while a complete and up-to-date overview is lacking, this is not the place to dwell too much on the details. Some general points about the Samoyed participles and their mutual correspondences should be borne in mind, however.

1. Samoyed languages generally share a set of participle types consisting of a present/imperfective, a past/perfective, a necessitative and a negative (also called abessive or caritive) participle that can be reconstructed for Proto-Samoyed. The two extinct Samoyed languages, Mator and Kamas, do not have this full system and Enets lacks the necessitative participle. Nenets and Enets have an additional future participle, which may be an innovation.

2. The suffixes used to form the participles in the various Samoyed languages are not always the same; e.g., the Nganasan imperfective participle suffix *-ntUə* derives from a complex suffix PS **-ntA + -jə*, while the Tundra Nenets counterpart is underived *-n(y)a/-t(y)a* from PS **-ntA*. In some cases, the suffixes do not match at all, such as in the case of the negative participle NeT *-wədawey°* vs. Sk. *-kuńčiŋtiľ*. The former descends from Proto-Uralic **-mAktAmA* (cf. Fi. *-mAtOn*), while the origin of the latter is unknown, as far as I am aware.

3. Various participles have given rise to new finite verb tenses and moods in various Samoyed languages; for example, the Nganasan imperfective aorist suffix *-ntU* derives from the Proto-Samoyed participle **-ntA*, which was replaced in its original function by the aforementioned participle formation with the ending *-ntUə* from PS **-ntA-jə* (on this point, see Janhunen 1998: 471–472).

The wide variety of participle suffixes used in the Samoyed languages hinder the reconstruction of a single, coherent system. Table 7.7 shows participle suffixes found in the Samoyed languages, accompanied by reconstructions to make it clear which suffixes are actually related. The suffixes **-ntA*, **-sA*, **-mə*, **-pA* and **mətAmA* derive from Proto-Uralic participles suffixes **-ntA*, **-ćA*, **-mA*, **-pA* and **-mAktAmA* respectively (see Aikio 2022a: 19).

Table 7.7: Samoyed participle suffixes with (partly mechanical) reconstructions to Proto-Samoyed. Kamas only preserves reflexes of the imperfective participle *-na*/*ne* from **-ntA* and the perfective participle *-bi* from **-pA-jə*, and is omitted from the table to conserve space. Forest Nenets is in line with Tundra Nenets.

| | Ng | EnT | EnF | NeT | Sk. | Mt. |
|---|---|---|---|---|---|---|
| ipfv. | *-ntUə* | *-da*/*-de* | *-da* | *-n(y)a*/*-t(y)a* | *-(n)tʲiʲ* | *-ndA* |
| | **-ntA-jə* | **-ntA* | **-ntA* | **-ntA* | **-ntA-jə* | **-ntA* |
| pfv. | *-sUəďəə* | *-sij* | *-i* | *-wi°*, *-bey°* | *-(m)pʲiʲ* | *-m(e)* |
| | **-sA-jə-jəjə* | **-sA-jə-?* | **mə-jə* | **-mə-jə*, **-pA-jə* | **-pA-jə* | **-mə* |
| necess. | *-bsUtə* | – | – | *-bco* | *-(p)sɔ̄tʲiʲ* | *-pso* |
| | **-pso-tə* | | | **-pso* | **-pso-tə-jə* | **-pso* |
| neg. | *-mətUmA?A* | *-ɔđaj* | *-ođaj* | *-wədawey°* | *-kuńčʲiʲ* | – |
| | **-mətAmA-j-?* | **-mətAmA-jə* | | **-mətAmA-jə* | **???* | |

The synchronic behaviour of participles in the individual Samoyed languages is otherwise quite consistent. Participles formed from intransitive verbs are active, while those formed from transitive verbs can usually be either active or passive based on the context. Practically all Samoyed languages agree on this point, as contextually oriented participles are found in all living Samoyed languages, in Nganasan, Nenets, Enets and Selkup (Leisiö 2006: 214), and they existed in Kamas as well (e.g., Klumpp 2022: 832).

The following example sentences from Tundra Nenets have the same type of structure as the Selkup example sentences about the deer and the bear cited as example (3) in subsection (7.1.1). In the first example here, the accusative object of the participle, *ŋənom* 'boat', necessitates an active reading. A genitive agent of a passive participle is

expressed in the proper noun *Werah* in the second sentence, and with the plural genitive *pyilyuq* 'gadflies' in the third sentence.

28. Tundra Nenets (Nikolaeva 2014: 317 (a–b), 287 (c); glosses adapted)

a. *ŋəno-m* *syerta-wi°* *xíbya* *to-°-∅*
boat-ACC do-PFV.PTC who come-AOR-SBC.3SG
'The one who made the boat came.'

b. *Wera-h* *temta-wi°* *ŋəmke* *xana-d°*
PN-GEN buy-PFV.PTC what take-IMP.SG.OBC.2SG
'Take the thing that Wera has bought.'

c. *pyilyuq* *pæ°r-tya* *ti* *tæryi* *pər°dor-ŋa-∅*
gadfly:GEN.PL bother-IPFV.PTC reindeer just shake-AOR-SBC.3SG
'The reindeer bothered by the gadflies is shaking.'

Forest Enets participles are described by Siegl to be "unoriented" as well (Siegl 2013: 330), although the great majority of the example sentences with participles in his description display absolutive orientation; I could not actually identify straightforward examples of active transitive participles.

29. Forest Enets (Siegl 2013: 330, 224; glosses adapted)

a. *to-iʔ* *enčiʔ*
come-PFV.PTC person
'A person who has arrived.'

b. *paʤu-miʔ* *paʤur*
write-PFV.PTC letter
'A written letter.'

The agent can also be marked with a possessive suffix, especially in the northern Samoyed languages. This possessive suffix is attached to the participle in Nganasan and Enets, and to either the noun or the participle in Tundra Nenets.[92]

[92] In Selkup this strategy seems to be almost entirely absent. A search for the combination of participle and possessive in the INEL Selkup Corpus turned up only *tantį-sa-nį*, a necessitative participle of *tantį-* 'go out' with an oblique first-person singular suffix. Apart from that, an example might be seen in *apsōtįľ qēlįmį taľca* 'the fish that should have been eaten [+by me?], went bad'. That is, if the suffix *-mį* on *qēlį* 'fish' is indeed the first-person singular possessive suffix, and indicative of an otherwise invisible first-person agent with the necessitative participle *apsōtįľ* of *a-* 'eat'; such an analysis is not apparent from the translation provided by Čeremisina (1991: 46), 'рыба, которая должна была быть съедена, протухла', which leaves the suffix *-mį* on *qēlį-mį*

30. Nganasan (Leisiö 2006: 220; glosses adapted)

| | | | |
|---|---|---|---|
| *ťenʲ-ntʲə-mtə* | *ŋusʲ* | *tande* | *śüďa-mi-ndə* |
| know-PRS.PTC-POSS.ACC.SG.2SG | job:ACC | kin:GEN | end-VN-LAT |

*kamiatə-bahu-rə*
keep-RPRT-SG.OBC.2SG

'They say that you will keep up the job that you know until the end of mankind.'

31. Tundra Nenets (Nikolaeva 2014: 290; glosses adapted)

| | | |
|---|---|---|
| *səwa-w°na* | *syerta-wi°* | *ŋəno-r°* |
| good-PRO | do-PFV.PTC | boat-SG.POSS.2SG |

'The boat that you made well.'

32. Forest Enets (Khanina & Shluinsky 2023: 844; glosses adapted)

| | | | | |
|---|---|---|---|---|
| *teđaʔ* | *tɔɔ-j-ɲiʔ* | *sɛxɛri* | *me-ɔn* | *kanʲe-ɲi-eʔ* |
| now | come-PFV.PTC-POSS.OBL.SG.1DU | road | in-PRO | leave-SBJ-1PL |

'Now let's go along the road we came by.'

In Tundra Nenets and in Enets, both the participle and the head noun may also be marked in the same way.

33. Tundra Nenets from Nikolaeva 2014: 315, glosses adapted)

| | | |
|---|---|---|
| *narey°-h* | *syerta-we-naq* | *ŋəw°rə-naq* |
| spring-GEN | do-PFV.PTC-PL.POSS.1PL | food-PL.POSS.1PL |

'The food (pl.) which we prepared in spring.'

34. Forest Enets (Khanina & Shluinsky 2023: 845, glosses adapted)

| | | | | |
|---|---|---|---|---|
| *kada-oda-jʔ* | *ɔburu-jʔ* | *oka* | *ɲi* | *ŋa-ʔ* |
| take.away-FUT.PTC-SG.POSS.1SG | thing-SG.POSS.1SG | many | NEG.3SG | exist-CNEG |

'The things I will take with me are not numerous.'

Kamas participles like the aorist participle in *-na*/*-ne* originally showed the same contextual orientation as those in the other Samoyed languages, with intransitive

'fish-*mʲ*' unexplained, as far as I can tell. An independent pronominal genitive agent is found in for instance *man qët-pʲl' qël'ʲ-m* 'рыбу, которую я ловил : the fish that I caught' (Čeremisina 1991: 46).

examples like *tunona ine* 'galloping horse', *küne kuza* 'man who has died', passive examples like *ťapsəna uja* 'roasted meat' and active examples like *seləne pim* 'my sharpening stone' (Klumpp 2002: 103; 2016: 196; 2022: 832). The Kamas aorist participle in these examples is the cognate of the imperfective participle in Tundra Nenets (*-n(y)a/-t(y)a*), the present participle in Selkup (*-(n)tɨl'*) and the imperfective participle in Nganasan (*-ntUə*).

Klumpp describes how the passive function of this participle was lost in the Kamas of the 20th century. The standard interpretation had become an active one, and the passive meaning could be derived with the addition of the valency decreasing suffix *-Ō* instead: *amor-ōna ine* 'eaten horse'. Similarly, the past participle in *-bi* was found in the combinations *-Ōbi* (pass.ptc.) and *-Təbi* (act.ptc.), rather than on its own (Klumpp 2002: 82, 85). It is also the basis of the past tense in *-bi*, an example of the finitization process that participles have undergone in many different Samoyed languages (Klumpp 2022: 832)

For Mator, there are not enough participles attested to draw a firm conclusion about their orientation. The five[93] certain examples of Mator present participles are intransitive (Helimski 1997: 169–171). Of the past participle, four examples[94] are based on intransitive verbs, and four[95] are based on transitive verbs. All four past participles based on transitive verbs have a passive meaning, so that Helimski (1997: 170) also uses the term passive participle. However, since the number of examples is so low, the Mator use of the past participle may in actuality have been more in line with that of the other Samoyed languages. Its orientation is at least absolutive.

The situation in Nganasan is not easily generalized either, partly due to the existence of a dedicated passive participle *-məə*. This is the only participle in the language that is specifically passive. It is mainly used as a resultative, denoting a more permanent characteristic of the object (Leisiö 2006: 222). Usually, this construction lacks an expressed agent, although it can be supplied in the lative case, instead of in the genitive that is usually used to mark the agent of a participle (Leisiö 2006: 222–223).

---

[93] These are *iləndä* 'lebendig : living', *gaterínda* 'lucidus : bright', *kémende* 'тяжелый : heavy', *neéngende*/*néengende* 'слабый : weak' and *ńärməndä*, *närməndä* 'rot : red'.

[94] These are *heelìm-soiù* 'foenum : hay' lit. 'dried grass', *geálma* 'dried (itr.)', *kājma* 'tot : dead', *соимъ* <soim> 'родной : native'.

[95] These are *идымме* <idɨmme> 'бритый : shaven', *triptima* 'dried (tr.), *gàdaima* 'убит : killed', *картымапса* <kartɨmapsa> 'жареное мясо : roasted meat' (*-apsa* 'meat').

35. Nganasan (Wagner-Nagy 2019: 267; glosses adapted).

| *manakü?* | *ŋanuə* | *ďekə-məə* | *kuhu* |
|---|---|---|---|
| just.now | real | cut-PASS.PTC | skin |

'A freshly skinned pelt.'

The Nganasan passive participle in *-məə* (PS **-məjə*) is the direct cognate of the Tundra Nenets perfective participle in *-wi*° and the Forest Enets perfective participle in *-i*, which were illustrated exhibiting contextual orientation in example sentences (28a–b.) and (29a–b) respectively. The distinctly passive, resultative function in Nganasan may be a secondary specialization of the perfective participle, stemming from the relationship between perfective aspect and resultativity.[96]

An additional participle or participle-like formation in Nganasan is called the iterative participle by Wagner-Nagy, and it is formed with the suffix *-KUTə*. Helimski (1998: 510) uses the term *nomen amatoris* instead, indicating a more nominal analysis. Cognates of this formation is also found in Tundra Nenets *-xad*°, Enets *-kaď* inclinative adjectives like NeT *pyisy*°*ŋkad*° 'prone to laugh; merry person', EnF *piśigaď* 'funny; merry person', NeT *yar*°*xad*° 'crybaby', EnF *ďaruxaď* 'crybaby; whining, piteous', *yabyexad*° 'drunkard; constantly drunk', EnF *ďebixaď* 'drunkard'. Mator *-kAt*/*-gAt* is characterized by Helimski as an agent noun formation that can also form adjectives, e.g., *täl*(*ə*)*gät* 'thief', *bisingät* 'mocker', *bälimgät* 'annoyed, angry'. This Samoyed suffix has recently been connected by Aikio with a Saami adjective suffix PSaa. **-kkēs*, which also creates inclinative adjectives, such as SaaN *boagustahkes* 'prone to laugh', *suoládahkes* 'thievish' (Aikio 2022b). The Nganasan example in (36) shows that at least in that language, this formation can take an accusative object, displaying a verbal argument structure in the same way as a participle.

36. Nganasan (Wagner-Nagy 2019: 516–517; glosses adapted)

| *tə* | *ŋuʔəj-küə-j* | *śelu-j* | *təbtə* | *təʔ* |
|---|---|---|---|---|
| well | one-EMPH-ACC.PL | naked.horn.reindeer-ACC.PL | also | you.know |

| *kotə-gutə* | *i-kə-bambu-ŋ* |
|---|---|
| hunt-ITER.PTC | be-ITER-RPRT-SBC.2SG |

'They say that you like to hunt reindeer with naked horns.'

The remaining participles in Nganasan, when formed to transitive verbs, can be either active or passive, like in the other Samoyed languages. An agent in the genitive often signals passive use and an object in the accusative necessitates an active interpretation,

[96] This is further implied by Ng. *kondüʔməə* 'murderer, assassin; band leader; rascal', cited by Aikio (2006a: 15), derived with *-məə* from the verb *kondüs-* 'murder, assassinate'.

again similar to what we have seen in the Selkup, Nenets and Enets examples above. However, as Leisiö explains, many Nganasan nouns do not have separate nominative, genitive and accusative case forms, so that ambiguities may still arise (Leisiö 2006: 220).

The general picture of the Samoyed participles is that the participial forms of transitive verbs tend to have a passive meaning unless the object is expressed in the accusative. It is easier to find examples of the past participle being used passively, and of the active participle used actively, but this merely reflects a tendency (see, e.g., Helimski 1983: 45 on this phenomenon in Selkup). This is furthermore in line with observations made by Haspelmath on the nature asymmetric participial systems, where commonly active present participles are opposed to past passive participles, without corresponding active past and passive present participles (Haspelmath 1994: 155–157).

When a participle is used passively, the agent may be expressed in the genitive, and in a few cases also in the lative, which is the case used to express an instrumental in many Samoyed languages. In the northern Samoyed languages, if the agent can be represented by a pronoun, it may be expressed in the form of a possessive suffix on either the participle itself or on the head noun, or both, depending on the language and some internal variation (see Pakendorf 2012).

### 7.3.1 Uralic and areal context

Contextually oriented participles are quite widely attested in other Uralic languages as well, as they are known from Mansi, Khanty, Hungarian, Permic and Kildin Saami (Leisiö 2006: 214, with further references). Shagal (2018) has done a typological study of Uralic participles, and she shows that in the easternmost Uralic languages contextually oriented participles characterize the entire participial system, while some of the westernmost Uralic languages have predominantly inherently oriented participles (Shagal 2018: 75–76).

The contextually oriented participles of eastern Uralic languages align with the general trend in northern Asia, and may be regarded as one of the areal features of Siberia (Shagal 2018: 60; 71–72, 2023: 962–971; Pakendorf 2012). The Finnish type of participial system, with active and passive participles, aligns closely with Russian and Lithuanian (Shagal 2018: 69–71). The areal distribution of the types of participle orientation found in the Uralic languages suggests that this feature can be transmitted by contact, even if the etymological sources of the participles are inherited (Shagal 2018: 80).

Aside from Uralic, contextually oriented participles are common in other language families of Siberia as well. Pakendorf (2012) and Shagal (2018) list Samoyed, Ob-Ugric, Turkic, Mongolic, Tungusic and Yeniseian languages as making wide-spread use of this type of participle (cf. also Lehmann 1984: 52–55 on this type of participle in general and in Turkic). The feature clearly transcends the boundaries of language families in this part of the world.

The encoding of participial agents as possessors is found not only in the Samoyed languages, but more widely in other Uralic languages like Finnic, Mordvin, Saami and Mari as well. The position of the possessive suffix differs, however, as within Samoyed: in Finnish, for instance, the participle itself is marked, not the head noun that it modifies (Shagal 2018: 61).

General trends of the participial systems of Uralic languages are thus a predominance of contextual orientation in the east and more inherent orientation in the west, and a use of the genitive to mark the agent. The details of the implementation of these features correlate with the surrounding languages and language families. The Samoyed participial system furthermore has a four-way distinction between an imperfective or present participle, a perfective or past participle, a necessitative participle, and a negative participle.

## 7.4 Conclusions on the participial systems

In pre-Proto-Tocharian, a change occurred in the participial system. In the Proto-Indo-European ancestral to Tocharian (i.e., without Anatolian) there was a participial system that consisted of mostly actively oriented participles (*-*nt*-, *-*mh₁no*-, *-*uos*). The differentiation of the active participle in *-*nt*- and the middle voice participle in *-*mh₁no*- came down to the morphological specifications of the verb. There were also deverbal participle-like formations in *-*to*-, which were likely absolutive in their orientation. The deverbal adjectives in *-*lo*- may have been active, but it is difficult to be certain, and they could potentially have been more flexible as well.

In Tocharian, this system was transformed. The only participle that retained its original active orientation is the active participle in PIE *-*nt*-. This formation may have started to shift its function to that of an agent noun in Proto-Tocharian, and it does not agree in gender with the head noun in those cases where it still functions attributively. The actively oriented middle voice participle *-*mh₁no*- lost its connection to the middle voice, and became a present participle with contextual orientation. It furthermore became widely used as a converb as well.

The preterite participle is arguably the most versatile participle formation in Tocharian, and it is very clearly contextually oriented. This is an innovation compared to its origin as an active perfect participle in PIE *-*uōs*. The verbal adjective in *-*lo*, or rather its derivation *-*l̥₁o*, has expanded its function in Tocharian as well in the form of the gerundive suffix. The present and subjunctive gerundive are both more or less clearly contextually oriented. The final participle-like formation in Tocharian is the privative, which may be described as a negative participle. It has a contextual orientation, and derives from the Proto-Indo-European *-*to*-adjective with probable absolutive orientation.

The bulk of the participles of Tocharian have thus undergone a development in terms of their orientation, with a majority becoming contextually oriented. This constitutes a marked change, an almost complete shift in the character of the participial system. In Samoyed, as in early Uralic in general, the participial system has probably always been characterized by a preference for contextual oriented participles, so that pre-Proto-Samoyed is a prime candidate to have influenced the orientation of participles in in pre-Proto-Tocharian.

Table 7.8: Overview of the development of the orientation of participles in Tocharian in comparison to Proto-Samoyed.

| | PIE | | Toch. | PS |
|---|---|---|---|---|
| *-*nt*- | active | = | active | ? cf. act.iter.ptc. *-*kAtə* / agn. *-*jA* |
| *-*mh₁no*- | active | → | contextual | cf. contextual prs.ptc. *-*ntA* |
| *-*uōs* | active | → | contextual | cf. contextual pst.ptc. *-*pAjə*/*-*məjə* |
| *-*to*- | absolutive | → | contextual | cf. contextual neg.ptc. *-*mətAmA* |
| *-*lo*- | ? act./abs. | → | contextual | cf. contextual necess.ptc. *-*psO* |

Having described the situation, we must ask if the Tocharian innovations can be understood through the lens of substrate influence from Samoyed. In the original pre-Tocharian situation, it can be assumed that the active and middle voice participle carried out their respective functions as expected from the Indo-European point of view. At some point, they became separate from the active and middle voice of the finite verbs, and could be formed to both as an actively oriented and a general present participle respectively. Aside from the active participle in *-*nt*, all participles became contextually oriented, like in Samoyed.

The types of participles that exist in the Tocharian system also correspond quite well to the Samoyed system. In Samoyed, there is a general contrast between a present or imperfective participle, a past or perfective participle, a necessitative or future participle, and a negative participle. If the two Tocharian gerundives expressing necessity and possibility, both formed with the same suffix, are considered together, a similar set of participles is obtained in Tocharian. Only the Tochairan active participle has no obvious participial counterpart in Samoyed, although the inclinative adjective or iterative participle in PS *-*kAtə* may have fulfilled a similar role. Perhaps the agent noun in PS *-*jə* (e.g., Ng. *toľiə* 'thief' < PS **tålä-jə* lit. 'stealer', PU **sala-ja*) could also be considered here, as the same suffix is found in participles elsewhere in Uralic. Both Tocharian and Samoyed also agree in marking agents of passively used participles in the genitive case.

In conclusion, the Tocharian participial system shows a good deal of isomorphism with the Samoyed system, while deviating from the Proto-Indo-European situation.

Other languages of the wider central and north Asian area show a broadly similar participial profile, however, so that specific convergence to Samoyed cannot be conclusively established. Especially the known later contact with Turkic languages could also have led Tocharian on the path toward contextually oriented participles. That contact may have been too late to affect Tocharian A and B in the same way, however, so that an earlier point of contact with another language group than Turkic may be preferable. Yeniseian and Yukaghir also possess contextually oriented participles, but no other clear traces of substrate influence from these languages in pre-Proto-Tocharian have yet been identified. Further comparison the early participial systems of the other Siberian language families may more accurately determine the specificity of the comparison between the Tocharian and Samoyed participial systems discussed in this chapter.

Even if the specifically Samoyed or Uralic connection remains uncertain, the high degree of contextual orientation found in the participles connects Tocharian to the Siberian linguistic area, lending further credence to the hypothesis that the Tocharians took a route through southern Siberia to end up in the Tarim Basin. While it cannot be conclusively established whether it was contact with a Uralic speaking group like the early Samoyeds that caused the changes observed in the Tocharian participial system, this aspect of the grammar may be added to the list of possible traces of influence from such a source.

# 8 Object marking on the verb

Both Tocharian and Samoyed can mark the object of a verb with a set of special endings. This feature is certainly innovative in Tocharian and could thus result from language contact. In this chapter I investigate the verbal object marking in both Tocharian and Samoyed to determine whether the two phenomena are similar enough to be connected.

## 8.1 Introduction

As in other Indo-European languages, the Tocharian finite verb is inflected to agree with the subject of the sentence with personal endings. They come in different sets for the active and middle voice, and also depend on the verb tense and mood (e.g., Peyrot 2013: 16–17). These endings are illustrated for Tocharian B in the Table 8.1 below. At their core, they descend from Proto-Indo-European personal endings, although various phonological and analogical processes have reshaped them quite significantly (e.g., Malzahn 2010: 26–48; Peyrot 2013: 409–420 with literature). The table below provides a sample of the personal endings in Tocharian B.

Table 8.1: Tocharian B verbal endings (e.g., Krause & Thomas 1960: 254–256; Malzahn 2010: 28, 38–39; Peyrot 2013: 17).

| TB | prs/sbj.act. | prs.mid. | prt.act. | prt.mid. |
|---|---|---|---|---|
| 1sg. | *-w* | *-mar* | *-wa* | *-mai* |
| 2sg. | *-t* | *-tar* | *-sta* | *-tai* |
| 3sg. | *-n* | *-tär* | *-∅* | *-te* |
| 1pl. | *-m* | *-mtär* | *-m* | *-mte* |
| 2pl. | *-cer* | *-tär* | *-s* | *-t* |
| 3pl. | *-n* | *-ntär* | *-re*, *-r* | *-nte* |

A second argument of the verb can be marked with an additional set of suffixes, which are placed after the endings that agree with the subject. These suffixes derive from originally clitic personal pronouns that were added to the end of a finite verbal form. While the pronominal elements themselves descend from Proto-Indo-European, their suffixation specifically to the verb is not reconstructible for Proto-Indo-European, and thus constitutes a Tocharian innovation. There are separate suffixes for all three persons in the singular, while the plural has one single undifferentiated form for all persons. The shape of the pronoun suffixes in both Tocharian languages is given in Table 8.2 below.

Table 8.2: The pronoun suffixes of Tocharian A and B.

| | TB | TA |
|---|---|---|
| 1sg. | *-ñ* | *-ñi* |
| 2sg. | *-c* | *-ci* |
| 3sg. | *-ne* | *-n* |
| pl. | *-me* | *-m* |

In his investigation of possible Uralic influence on Tocharian, Peyrot (2019a: 97–99) made a comparison between the pronoun suffixes in Tocharian and the objective conjugation found in Uralic. The objective conjugation is a set of special endings present in a number of Uralic languages, with which transitive verbs can be inflected to refer to the object as well as the subject. The objective conjugation contrasts with the subjective conjugation, which only agrees with the subject. The origins for the objective conjugation are often sought in the addition of earlier clitic object pronouns at the end of the verb form as a suffix (e.g., Honti 1995; Abondolo 1998; Rédei 1998/99), similar to the Tocharian situation. Based on this view, Peyrot made a connection between the objective conjugation and the Tocharian pronoun suffixes (Peyrot 2019a: 97–99). Some of the general similarities between the verbal object marking of Tocharian and Uralic are also briefly remarked upon by Stefan Georg, although he adds that the specifics differ significantly between the two languages families (Georg 2023: 202).

It is clear that both languages have some way to mark objects in their verbal inflection, but do the similarities go beyond that? In order to answer this question, a more detailed look at the Uralic objective conjugation is necessary. This is because its origins are very much debated, including the manner of its formation, the point in time when it came into existence, and its original functional and syntactic scope. According to the theory mentioned above and adduced by Peyrot for a comparison with Tocharian pronoun suffixes, the objective conjugation descends from object pronouns that over time became attached to the verb, but the great similarities of the objective conjugation suffixes and the possessive declension of nouns instead suggest a close connection between the latter two categories (Janhunen 1982: 35). Moreover, a great difficulty with an origin theory of the objective conjugation as a straightforward addition of object clitics is that the objective conjugation endings of the first and second person are phonologically similar to the first and second person pronouns respectively, but agree with a third person object instead. This is illustrated in the table below with personal pronouns and verb endings from Nganasan. The objective endings like 1sg. *-mə* are phonologically related to the pronoun of the corresponding subject, but mark the presence of a third person object. Note that the *-r-* in the second person endings derives from a **-t-* (as does the palatalized *-t'-*).

Table 8.3: Nganasan personal pronouns and verbal endings (Wagner-Nagy 2019: 106, 229, 231).

| | pers. pron. | sbc. | sg.obc. | pl.obc. |
|---|---|---|---|---|
| 1sg. | *mənə* | *-m* | *-mə* | *-ńə* |
| 2sg. | *tənə* | *-ŋ* | *-rə* | *-ťə* |
| 3sg. | *sʲitʲ* | *-∅* | *-tu~tʲ* | *-ťü~ťi* |
| 1du. | *mii* | *-mi* | *-mi* | *-ńi* |
| 2du. | *tii* | *-ri* | *-ri* | *-ťi* |
| 3du. | *sʲti* | *-gəj* | *-ði* | *-ťi* |
| 1pl. | *mʲŋ* | *-muʔ~mʲʔ* | *-muʔ~mʲʔ* | *-ńüʔ~ńiʔ* |
| 2pl. | *tʲŋ* | *-ruʔ~rʲʔ* | *-ruʔ~rʲʔ* | *-ťüʔ~ťiʔ* |
| 3pl. | *sʲtʲŋ* | *-ʔ* | *-tuŋ~tʲŋ* | *-ťüŋ~ťiŋ* |

The age of the objective conjugation is another contentious issue that intersects with the discussion about its origins, addressed in more detail in section 8.3 below. I conclude that the origins of the objective conjugation lie in the distant past, in pre-Proto-Uralic times, which makes a precise reconstruction particularly difficult. Its general structure in the earliest stages of pre-Proto-Samoyed can still be approximated, however, and it does not quite match the functions of the Tocharian pronoun suffixes.

We must also address the fact that neither the Tocharian pronoun suffixes nor the Uralic objective conjugation can straightforwardly be described as being concerned only with "object marking", as the reality is more complex. We will need a more accurate understanding of the functions of both the Uralic objective conjugation (8.3.3) and the Tocharian pronoun suffixes (8.2.1) in order to arrive at a comparison between the two. The Tocharian pronoun suffixes have recently been investigated in detail by Teigo Onishi (2022), and the objective conjugation of various Uralic languages has been elucidated by the research of scholars such as Irina Nikolaeva (1999a, 1999b, 2001), Elena Skribnik (2001), and Susanna Virtanen (2014). It appears that a more general analysis of both as instances of topical object marking on the verb describes their use more accurately, especially in the case of Uralic, and that this may provide the basis for a functional comparison between the two.

## 8.2 The Tocharian pronoun suffixes

This section provides an overview of the Tocharian pronoun suffixes based on the recent literature and illustrates their use with a selection of examples sentences. The aspects of the pronoun suffixes that make them stand out as innovative are also addressed, as this provides the main reason for considering the possible effect of influence from other languages like Samoyed.

### 8.2.1 Functions of the pronoun suffixes

The pronoun suffixes in Tocharian are very versatile. As described by Onishi, "they may represent the direct object, the indirect object with various roles such as source, goal, addressee, recipient, beneficiary, experiencer, and location, the possessor, the subject of a non-finite verb, the argument of a predicate consisting of an adverb or postposition and a verb, and so on" (Onishi 2022: 42). These many functions are illustrated with ample examples by Onishi (2022: 42–76), and I will here cite a selection to provide a general overview with sentences (1-6).

1. Tocharian A (A 101 b5; trsl. Onishi 2022: 43; glosses adapted)
   *kuc* *yāmwā* *kᵢᵤyal* *pälkse-**ñi***
   what.OBL.SG do:PST.ACT.1SG why torture:PRS.ACT.3PL-1SG
   '... What did I do? Why do they torture **me**?'

2. Tocharian B (THT 107 a7–8; trsl. CEToM, cf. Onishi 2022: 5, 45; glosses adapted)
   *śilāre-**ne*** *oṅkarñai* *wñār-**ne*** *purwar*
   bring:PST.ACT.3PL-3SG porridge:OBL speak:PST.ACT.3PL-3SG accept:IMP.MID.2SG
   *wesanmeṃ* *pinwāt* *rṣāka*
   1PL:ABL alms sage:VOC
   'They brought **him** the rice porridge [and] said **to him**: "Accept the alms from us, o sage!"'

3. Tocharian A (A 92 a5; trsl. after Onishi 2022: 48; glosses adapted)
   /// *cam* *śkaṃ* *lo* *psumār-**ñi*** *kᵢᵤyal*
   DEM.M.OBL.SG CONJ away take.away:IMP.MID.2SG-1SG why
   *lykäly* *lykäly* *tuṣt-**ñi***
   fine fine burn:PRS.ACT.2SG-1SG
   '... and take that away **from me**! Why do you burn **me** extremely fine?'

4. Tocharian B (PK NS 34 b3; trsl. Onishi 2022: 59; glosses adapted)
   /// (*āklyi*) *krent* *yāmorne* *yāmtar* *kwri* *aum* *no*
   learning:OBL good:M.OBL deed:LOC do:SBJ.MID.2SG if then CONJ
   *ppo* *lau* *cmeln=* *alyek* *tākañ-**c*** *āklyi*
   all far birth:LOC other:OBL.SG be:SBJ.ACT.3SG-2SG learning:OBL
   *kreṃnt* *yāmor*
   good:M.OBL deed
   'If you exercise in good action, then in every distant [future] birth **you** will have (lit. "will be yours") the good action (as) a lesson.'

5. Tocharian A (A 215 a2; trsl. based on Krause & Thomas 1960: 163)
   *ārar-**ñi*** *puk* *niṣpalntu*
   cease:PRT.ACT.3PL-1SG all belonging:NOM.PL
   'All my possessions have come to an end.'

6. Tocharian B (THT 11 a4; verse; trsl. based on Krause & Thomas 1960: 163)
   *laute* *mā* *kätkoytär-**me***
   moment NEG pass:OPT.MID.3SG-PL
   '[The right] moment must not be passed (by) **by you** (pl.).'

The common functions of the pronoun suffixes align with those otherwise carried out by the Tocharian oblique/accusative case—mainly direct object—and the genitive-dative case—indirect object, possessor—(Krause & Thomas 1960: 162–163). Onishi further shows that, when the direct and indirect object of a verb are both pronominal, the pronoun suffixes refer to the indirect object, not to the direct object; the latter may then only be referred to with an independent pronoun, although it can also be omitted entirely (Onishi 2021: 155–159, 177). The pronoun suffixes are thus more strongly associated with the indirect object than with the direct object.

A pronoun suffix sometimes also co-occurs with an overt antecedent. For example, a pronoun suffix marking a direct object can appear within the same sentence where that object appears as a noun. Peyrot (2019a: 97–99) proposed to analyse this as a form of object agreement. Onishi has argued that there may be three different phenomena at play, and that all three can be analysed with reference to the concept of topicality (Onishi 2022: 182–253).

First, if the pronoun suffix co-occurs with an object in its normal position after the subject and before the verb, the pronoun suffix may indicate that this object functions as a secondary topic (Onishi 2022: 217–252). Second, the object may appear before the subject, in which case it functions as the primary topic (Onishi 2022: 197–207). Last, the object may appear after the finite verb that bears the pronoun suffix if it constitutes a discourse-old primary topic (Onishi 2022: 207–210). Since the use of a pronoun suffix in combination with an overt object is not particularly common, Onishi hypothesises that this strategy for making objects as secondary or primary topics never fully grammaticalized (Onishi 2022: 193). Sentences (7) and (8) provide examples of co-occurrence of a pronoun suffix and an object that it references or agrees with.

7. Tocharian A (A 13 a1–2; trsl. based on CEToM, Onishi 2022: 220; glosses adapted)

| *kaśśi* | *yokañi* | *pälkāt* | *cesäm* | *amoktses* | *kātse* |
|---|---|---|---|---|---|
| hungry | thirsty | see.PRT.MID.3SG | DEM.M.OBL.PL | artisan:OBL.PL | near |

| *kälymāṃ* | *cesm-äk* | *puk* | *śtwar* | *śälkās* | *pokeyo* |
|---|---|---|---|---|---|
| stand:PRS.PTC | DEM.M.OBL.PL-EMPH | all | four | together | paw:SG.INS |

| *wa*(*ltsurä*)*ṣ* | ***poñcäs*** | *kosā-**m*** | *tāpa-**m*** | *śkaṃ* | *lo* |
|---|---|---|---|---|---|
| crush:ABS | all:OBL.PL | kill:PRT.ACT.3SG-PL | eat:PRT.ACT.3SG-PL | and | PFV |

'Hungry and thirsty, it [i.e., a newly resurrected lion] saw these artisans standing nearby, and crushing those very four all together with [his] paw, he killed **them all** and ate **them** up.'

8. Tocharian B (THT 88 a1; trsl. after Huard 2022: 382; cf. Onishi 2022: 223; glosses adapted)

| *tumeṃ* | *durmukhe* | *brāhmaṇe* | ***uttare*(*-ṃ*)**[97] | ***śamaśke-ṃ*** |
|---|---|---|---|---|
| thereupon | PN | brahmin | PN:OBL.SG | boy:OBL.SG |

| *kärwāṣṣai* | *witsakaisa* | *räskare* | *tsopaṃ-**ne*** |
|---|---|---|---|
| reed:ADJZ.F.OBL.SG | root:SG.PER | sharply | hit:PRS.ACT.3SG-3SG |

'Thereupon the brahmin Durmukha hits **the boy Uttara** sharply with a reed root.'

To my knowledge, there is no indication that (secondarily) topical objects *require* the presence of a pronoun suffix if they remain themselves present in the sentence. When a nominal object, indirect object, or possessor is particularly topical it can be completely replaced by a pronoun suffix. One could thus say that highly topical zero anaphora tend to trigger object agreement with a finite verb in the form of a pronoun suffix, and that especially topical indirect objects and possessors do the same.

However, there are also some instances of zero anaphora. In the following sentence, the two sisters Nandā and Nandabalā are speaking to a disguised Indra about whom they should give a special porridge that they have made as alms to. The direct object of the verb *aiskem* 'we give' is not overtly present in the sentence, nor does it host a pronoun suffix indicating the indirect object, which is instead present as an independent pronoun *cwi* 'to him'.

[97] The final *-ṃ* is absent in the manuscript, but grammatically expected.

9. Tocharian B (THT 107 a9; trsl. CEToM; glosses adapted)

| *nānda* | *nandābala* | *weñāre* | *se* | *ci-sa* |
|---|---|---|---|---|
| PN | PN | say:PRT.ACT.3PL | REL.NOM.SG | 2SG.OBL.SG.PER |

| *śpālmeṃ* | *tākaṃ* | *cwi* | *aiskem* |
|---|---|---|---|
| excellent | be:SBJ.ACT.3SG | 3SG.M.GEN.SG | give:PRS.ACT.1PL |

'Nandā and Nandabalā said: he who is better than you, to him we give [it]'

After having been sent by Indra to the god Brahma, Nandā and Nandabalā go from Brahma to the Śuddhāvāsa-gods, also disguised as sages. The two sisters offer the porridge to them, but in case there be others that are more excellent, they ask:

10. Tocharian B (THT 107 b3–4; trsl. CEToM; glosses adapted)

| *wesi* | *pokses* | *posa* | *śpālmeṃ* | *rṣāke* | *intsu* |
|---|---|---|---|---|---|
| 1PL:GEN | proclaim:IMP.ACT.2PL | all:PER | excellent | sage | which |

| *ste* | *cwi* | *ka* | *ṣ* | *mant* | *klāskem* |
|---|---|---|---|---|---|
| be:ACT.3SG | DEM.M.GEN.SG | indeed | and | thus | bring:PRS.ACT.1PL |

'Tell us which sage is better than all, so that to him indeed we bring [it].'

In neither of these sentences where the porridge is the object do we see any pronoun suffix on the verb. It is unclear to me if, in these examples, the presence of the indirect object pronoun *cwi*, suppresses the use of a pronoun suffix on the verb, similar to how pronoun suffixes refer to the indirect object rather than to the direct object if both are available. A more complete account of Tocharian sentences without an overt object is still a desideratum (cf. Onishi 2022: 252), and I am unable to provide a full analysis here.

### 8.2.2 Origins and status of the pronoun suffixes

As seen in the example sentences in the previous subsection, the Tocharian pronoun suffixes are mainly used to anaphorically refer to an entity mentioned in the surrounding context. Most often, they appear in a sentence instead of the nominal object constituent that they refer to, but sometimes they also co-occur with noun phrases.

The shape of the first- and second-person singular suffixes is quite similar to that of the corresponding independent pronouns. The third person singular and general plural forms, on the other hand, do not have direct counterparts elsewhere in the Tocharian pronominal system (cf. Peyrot 2019a: 98). All pronoun suffixes can be derived from Proto-Indo-European sources, however, and with exception of the third person singular, all can be reconstructed as clitics at an early stage. The basic reconstructions are given in the table below; for more extensive discussions on the specifics of the reconstructions and developments in Tocharian, I refer to Onishi (2022: 38–41) with literature.

Table 8.4: Tocharian A and B pronoun suffixes and their possible origins.

| | TB | TA | PT | PIE |
|---|---|---|---|---|
| 1sg. | *-ñ* | *-ñi* | **-nʸə* | **-me* |
| 2sg. | *-c* | *-ci* | **-tʸə* | **-tu̯e* |
| 3sg. | *-ne* | *-n* | **-ne* | **Hono-* |
| pl. | *-me* | *-m* | **-me* | **-smos* |

The investigation recently carried out by Onishi has provided us with a more detailed analysis of the functions and behaviour of the pronoun suffixes in Tocharian from a synchronic perspective. The reasons why the Tocharian pronoun suffixes differ from similar pronominal clitics in other Indo-European languages are not yet fully understood, however, as a full diachronic account of their development is still lacking.

We may still attempt a comparison with the Uralic objective conjugation, even if only to see what kind of similarities or differences we can identify, and how we might approach a proper diachronic reconstruction. Peyrot (2019a: 97–99) has made a first pass at such a comparison already, noting some of the ways in which Tocharian pronoun suffixes are unusual among ancient Indo-European languages, and pointing out some parallels with the objective conjugation. One important feature is that pronoun suffixes form a single phonological word with the verb to which they are attached, as is apparent from the stress placement in Tocharian B (cf. chapter 5). The same is suggested by their participation in morphophonological alternations and assimilations in Tocharian A (Peyrot 2019a: 97, Onishi 2022: 32). In example (11), the final consonant of the verb *tāṣ* 'be:SBJ.ACT.3SG' is fused with the initial consonant of the 2sg. pronoun suffix *-ci* to yield *tāśśi*.

11. Tocharian A (A 81 a3; trsl. Onishi 2022: 32; glosses adapted)

| /// | (*mäski*)*tac* | *kakmuräṣ* | *träṅkäṣ* | *tärkor* |
|---|---|---|---|---|
| | prince:ALL | come:ABS | speak:PRS.ACT.3SG | permission |
| *tāśśi* | | *mäśkit* | *pläc* | *w*(*aṣtäṣ*) /// |
| be.SBJ.ACT.3SG-2SG | | prince | go.out:IMP.ACT.2SG | house:ABL |

'He (scil. the king) came to (the prince) and says: "**You** shall have permission, o prince. Become (a monk!) (lit. go out (from the house])!) ..."'

Onishi further specifies that the suffixes do or do not participate in assimilation depending on the type of word they are attached to. While a combination of a finite verb ending in *-ṣ* and the pronoun suffix *-ci* for the second person singular results in an assimilated cluster TA *-śś-*, as in example (11) above, the same combination resulting from a participle in *-ṣ* and the pronoun suffix *-ci* surfaces as *-ṣc-* (Onishi 2022: 32–33).

12. Tocharian A (A 125b2; trsl. Onishi 2022: 32; glosses adapted)

| *laltuṣ-ci* | *cesäm* | *pälkoräṣ* | *nu* | *tmä* /// |
|---|---|---|---|---|
| go.out.PRT.PTC.M.NOM.PL-2SG | DEM.PL.OBL | see.ABS | CONJ | ... |

'[Those who] went out from you ... But having seen them, ...'

The behaviour of the pronoun suffixes is mixed as regards to their status as verbal endings or clitics, and that any analysis will depend on one's theoretical framework (Onishi 2022: 34–35). Pronoun clitics in Tocharian A and B furthermore show differences in behaviour in some crucial ways, since the Tocharian A use of the pronoun suffixes on various non-finite verbs, like in example (12), is more canonically clitic-like than the Tocharian B ending-like restriction to finite verbs only (ibid.). This complicates a clear-cut analysis. Based on their high selectivity for (finite) verbs in both Tocharian languages, I refer to the Tocharian pronominal elements under discussion in this chapter as pronoun suffixes rather than pronoun clitics.

### 8.2.3 Similar pronominal elements in other Indo-European languages

The Tocharian pronoun suffixes descend from Proto-Indo-European pronominal clitics, as discussed above. In this subsection we will functionally compare the Tocharian pronoun suffixes with similar pronominal clitics in other ancient Indo-European languages, in order to get a better view of the position of Tocharian within the family in this regard. Onishi briefly summarizes the main differences (Onishi 2022: 6), and here I will illustrate each point with some examples.

I. Only one pronoun suffix may appear in a sentence in Tocharian, while multiple pronominal clitics may be found in a single sentence in other ancient Indo-European languages Latin and Hittite, as illustrated in the examples below.

13. Latin (Cic. ad Att., II, 19, 5; from Meillet & Vendryes 1948: 580)

| *Caesar* | ***mē*** | ***sibi*** | *vellet* | *esse* | *legātum* |
|---|---|---|---|---|---|
| PN | 1SG.ACC | REFL.DAT.SG | want:SBJ.IMPF.3SG | be.INF | legate:ACC |

'Caesar wants **me** to be **his** legate.'

14. Late Hittite (CTH 42.A KBo 5.3+ obv. ii 8; from Lyutikova & Sideltsev 2020: 34)

| *nu=**šmaš**=**at*** | *lē* | *āra* | *i̯yenzi* |
|---|---|---|---|
| CONN=2PL.DAT=3PL.ACC | PROHIB | right | do:PRS.3PL |

'They will not make **them** right **for you**.'

There are other Indo-European languages that allow only a single infixed or suffixed pronominal element per verb, such as Old Irish. Griffith has argued that their analysis as truly pronominal is only valid for earlier stages of the language, and that in synchronic Old Irish, they may rather be interpreted as agreement affixes instead (Griffith 2011). The questions of their classification bear some resemblance to the Tocharian situation, where the pronoun suffixes are variously considered to be clitics or agreement markers.

II. There is no contrast between the way a direct object and an indirect object is referred to using a pronoun suffix in Tocharian, whereas many other ancient Indo-European languages like Sanskrit, Greek, and Latin, do show such a difference. In Hittite, as in Tocharian, no difference is found in the first- and second-person enclitic pronouns. Only the third person anaphoric enclitic pronoun has separate case forms, including a nominative (Hoffner & Melchert 2008: 135–136). These differences in the clitic pronoun systems are illustrated in Table 8.5.

Table 8.5: Pronominal clitics in ancient Indo-European languages.

| | 1sg | | 2sg | | 3sg | |
|---|---|---|---|---|---|---|
| | acc. | dat. | acc. | dat. | acc. | dat. |
| TB | *-ñ* | *-ñ* | *-c* | *-c* | *-ne* | *-ne* |
| Hitt. | *=mu* | *=mu* | *=tta* | *=tta* | *=an* | *=šši/šše* |
| Skt. | *mā* | *me* | *tvā* | *te* | *sīm* | |
| GrH | *me* | *moi* | *se* | *soi/toi* | *he/min* | *hoi* |
| Lat. | *mē* | *mihi* | *tē* | *tibi* | *sē* | *sibi* |

The lack of any formal difference between accusative and dative clitics in Tocharian is reminiscent of Old Irish, where the normally accusative suffixed and infixed pronouns can combine with the substantive verb 'be' to express a dative function, e.g., the 2sg. -(*u*)*t* in *táthut* 'you (sg.) have' and *ro-t·bia* 'you (sg.) shall have'. This use is similar to, e.g., TB *nesañ-c* 'you (sg.) have' (THT 144a1). In Old Irish, this is rare with other verbs, but does sometimes occur, such as in *do-t·árfas* 'it has appeared to you (sg.)' (Thurneysen 1946: §409, §430).

Normally, the pronoun suffixes and infixes of Old Irish express only the object of a transitive verb, as in *ar-ro-t·neithius* 'I expected you (sg.)', or a first- or second-person subject of a passive verb; e.g., *carthair* 'he/she is loved' but *no-t-charthar* 'you (sg.) are loved' (Thurneysen 1946: §409–4010, §540). The latter use is contrary to that of Tocharian, where a pronoun suffix attached to a passively used middle can indicate the agent instead (cf. example 6 above, repeated here as 15)6.

15. Tocharian B (THT 11 a4; verse; trsl. based on Krause & Thomas 1960: 163)

| *laute* | *mā* | *kätkoytär-**me*** |
|---|---|---|
| moment | NEG | pass:OPT.MID.3SG-PL |

'[The right] moment must not be passed (by) **by you** (pl.).'

III. Clitics in ancient Indo-European languages usually appear in the second position in a clause, the so-called Wackernagel position (after Wackernagel 1892). In Tocharian B, pronoun suffixes instead only appear after a finite verb, while in Tocharian A they can attach to non-finite verbs as well. This is more in line with the behaviour of pronominal clitics in for instance the Romance languages, where they always attach to the verb (e.g., Viti 2016: 19–20). The Tocharian placement of the pronoun suffixes contrasts with the placement of pronominal clitics in Latin (13) and Hittite (14) already cited above, and in Sanskrit (16) and Ancient Greek (17).

16. Vedic Sanskrit (RV 7.63.3c from Hale 1987: 80)

| *eṣá=**me*** | *deváḥ* | *savitā́* | *cachanda* |
|---|---|---|---|
| DEM.M.NOM.SG=1SG.DAT | god | PN | appear:PF.ACT.3SG |

'He appeared **to me** as the god Savitar

17. Ancient Greek (Hdt. 1.35.2 from Goldstein 2016: 54)

| Κροῖσος=δέ=μιν | ἐκάθηρε |
|---|---|
| PN=PTCL=3SG.M.ACC | cleanse:AOR.ACT.3SG |

'Croesus purified **him**.'

A number of conjunctions in Tocharian do tend to appear in the second position, such as TA *nu*, TB *no* 'and, but' (e.g., Pinault 2008: 523–524). This is illustrated with an example from Tocharian B in (18), where *mā* 'not' stands at the beginning of a new clause, followed by the clitic conjunction *no*. The behaviour of such particles indicates that second position in Tocharian was still available for clitics, but not for pronominal ones, a fact that might be significant.

18. Tocharian B (THT 3 b6; trsl. Adams 2013: 370)

| *waimenetse* | *śaul* | *totk=* | *āttsaik* | *su* | ***ṣp*** | *laklempa* |
|---|---|---|---|---|---|---|
| difficult | life | short | certainly | DEM.M.NOM.SG | and | sorrow:COM |

| *rittowo* | *mā* | ***no*** | *wnolmy* | *aikenträ* |
|---|---|---|---|---|
| attach:PRT.PTC.M.NOM.SG | NEG | CONJ | creature.PL | know:PRS.MID.3PL |

'Life is difficult and short **and** (it [is]) bound up with suffering, **but** creatures do not notice.'

We have now seen that the Tocharian pronoun suffixes can fulfil a wide range of functions, not restricted to object-marking. Despite this, they show no formal differentiation to go along with their functional flexibility. They descend from Proto-Indo-European clitic pronouns, but their fixed position after the finite verb is an innovation. In order to make a comparison with Uralic, we will now turn to a discussion of the objective conjugation.

## 8.3 The Uralic objective conjugation

The specifics of both the formal and functional reconstruction of the objective conjugation are controversial. To properly understand the Uralic objective conjugation, a lot of information from various branches of Uralic needs to be reviewed and discussed. In this section, I will address some fundamental questions pertaining to the age and formation of the objective conjugation, necessary for a serious comparison with the Tocharian pronoun suffixes.

### 8.3.1 General overview of the objective conjugation

In most branches of Uralic there exists a set of verbal endings used to form a so-called objective conjugation (also known as the "definite" or "determinate" conjugation) beside another set that functions as the subjective conjugation ("indefinite", "indeterminate"). The endings of the objective conjugation indicate not only the person and number of the subject, but also at least the involvement of an object as well, while the subjective conjugation only refers to a subject. Such a split in conjugations is found on the eastern side of the Uralic language family in Samoyed, Mansi, Khanty, and Hungarian, and also in Mordvin on the western side. In the languages that have both conjugations, intransitive verbs are typically only used with the subjective conjugation, but transitive verbs can be inflected with either subjective or objective endings. The specifics of the objective conjugation differ both structurally and functionally from language to language, which complicates a reconstruction of the system to Proto-Uralic and renders it a contentious issue (e.g., Abondolo 1998: 29–30; Keresztes 1999: 102; Aikio 2022a: 18).

Even the most basic question, the degree to which the objective conjugation existed in Proto-Uralic already is disputed. Janhunen (1982) states that the objective forms may have existed in Proto-Uralic for all persons, perhaps as an elaboration from an earlier pre-PU 3sg. form only, but that the objective conjugation could also have been dialectally restricted. Keresztes (1999) regards the objective conjugations in Mordvin and the other Uralic language as largely independent innovations, but like Janhunen, he does allow for the possibility that the third person singular objective form already existed; cf. also Körtvély (2005: 44). It seems to be particularly the Mordvin objective conjugation that precludes the establishment of this category as inherited from Proto-Uralic, well as the

notion that it should have been lost in Saami, Finnic, Mari, and Permic (so, strongly, Havas 2004). Compare the following recent statement on the matter by Aikio in his handbook chapter on Proto-Uralic (with references to arguments from other scholars):

> "The Mordvin system of objective conjugation differs so radically from the others that it is highly probably an innovation (Keresztes 1999), but the objective conjugation in the easternmost branches [viz., Samoyed, Mansi, Khanty and Hungarian] might nevertheless represent a Proto-Uralic archaism that was lost in the more western branches (Janhunen 1982: 35). On the other hand, it is equally possible that the Saami, Finnic, Mari and Permic branches possess a more archaic verb system, and [that] the definite conjugation in the eastern branches is an areal innovation (Salminen 1996)." (Aikio 2022a: 18)

The existence of at least a third person singular objective conjugation suffix "**-sV*" in Proto-Uralic is often assumed, and this suffix is also the central character in discussions on the origin of the entire conjugation type. According to one theory, it is effectively a suffixed object pronoun, cognate to, e.g., SaaN *son*, Fi. *hän*, Hu. *ő*, etc. (e.g., Honti 1995, 1998/99; Rédei 1998/99; Abondolo 1998: 29–30). The other theory equates this same suffix with the third person singular possessive suffix PU **-sA*, found in, e.g., Fi. *-nsa*, Hu. *-(j)a*, PS **-tA* (Janhunen 1982; Zhivlov 2023: 155–156). The vowel actually corresponds to that of the possessive suffixes in those branches that retain the vowel quality more faithfully, so that the reconstruction often given as **-sV* can reasonably be written as **-sA* instead. At least on the surface, this strengthens that particular connection. And indeed, the objective conjugation suffixes are on the whole similar to the possessive suffixes used on nouns, although to which degree the two sets are really identical differs from language to language (cf. below).

To better understand the issues associated with the objective conjugation, we must take a close look at the objective conjugation paradigms in the individual languages, as well as their relation to the possessive declension paradigm on the one hand, and the subjective conjugation on the other. In the end, close scrutiny and comparison of the individual languages should inform us on the probable age of the suffixes, and also their likely origin. It is my opinion that the Mordvin objective conjugation is at its core not so alien as has been claimed, and that it in fact holds important information regarding the reconstruction of the objective conjugation endings in Proto-Uralic. It should be taken into account by any attempt to explain the origin of the objective conjugation endings from earlier syntactic constructions, an endeavour that will mostly fall outside the scope of the present investigation.

The two main issues regarding the Uralic objective conjugation are (1) whether it existed in Proto-Uralic already, and if so, whether it constituted a full paradigm or only the third person singular, and (2) whether it originates from suffixed object pronouns or

from possessive inflection. To practically work towards solving these issues, I think the following set of questions should be addressed:

1. Do the languages with an objective conjugation show non-trivial formal and structural correspondences?
2. Do the suffixes of the objective conjugation match either possessive suffixes or object pronouns in any consistent and non-trivial way?
3. And finally, is a relation to either the possessive declension or suffixed object pronouns functionally understandable?

I will address especially (1) and (2) on the basis of a closer look at the objective conjugation paradigms found in the individual branches (starting with Samoyed, then Mansi, Khanty, Hungarian, Mordvin), and also a brief summary of the verbal inflection found in the remaining branches (Saami, Finnic, Mari, Permic). This is the topic of subsection 8.3.2. From the discussion of the objective conjugation paradigms in especially Samoyed, Mansi and Mordvin, I conclude that there is indeed an additional corresponding formal feature beyond the third person singular suffix *-*sA*, namely the object plural *-*n*-, and that the two shared features together are non-trivial. Some of the functional considerations needed to address question (3) will be discussed separately in subsection 8.3.3, but its proper resolution lies beyond the scope of this chapter and is left to future research.

### 8.3.2 Formal reconstruction of the objective conjugation

There is no satisfactorily complete and detailed historical overview of the objective conjugation and related issues available at this moment. In this subsection I will go over the forms and structure of the objective conjugation paradigms as they occur in each of the Uralic branches that have this inflectional category. In some cases, I will also supply reconstructions of the objective conjugation paradigms to the lower-level protolanguages, in order to improve understanding. To make the similarities with possessive declension suffixes apparent, I will list those together with the objective conjugation. The subjective conjugation endings will be given as well, to make the formal contrasts and similarities between the two sets of endings clear. The tables are all composed from the information in the grammars of the various languages under discussion, recombined and rearranged for illustrative purposes. Overviews of conjugations in various Uralic languages can be found in, e.g., Keresztes (1999: 15–21), Körvtély (2005: 27–42), Trosterud (2006: 143ff.).

Table 8.6: Endings of the Proto-Uralic subjective conjugation, objective conjugation (singular object) and possessive declension (singular and plural possessum) (Aikio 2022a: 16, 18; Zhivlov 2023: 156-157).

| | sbc. | obc. | sg.poss. | pl.poss. |
|---|---|---|---|---|
| 1sg. | **-m, *-∅?* | **-mi* | **-mi* | **-ni* |
| 2sg. | **-n, *-t?* | **-ti* | **-ti* | **-nti* |
| 3sg. | **-∅* | **-sA* | **-sA* | **-nsA* |
| 1du. | **-mijn* | **-mijn* | **-mijn* | **-nijn* |
| 2du. | **-tijn* | **-tijn* | **-tijn* | **-ntijn* |
| 3du. | **-ki* | **-sAjn* | **-sAjn* | **-nsAjn* |
| 1pl. | **-mAk/t* | **-mAk/t* | **-mAk/t* | **-nAk/t* |
| 2pl. | **-tAk/t* | **-tAk/t* | **-tAk/t* | **-ntAk/t* |
| 3pl. | **-t* | **-sAk/t* | **-sAk/t* | **-nsAk/t* |

### 8.3.2.1 Samoyed

Within Samoyed, the objective conjugation paradigm is most extensive in the northern Samoyed languages like Tundra Nenets and Nganasan. In these languages, not only the number and person of the agent, but also the number of the object is marked with the verbal ending. I will start this discussion with the Tundra Nenets forms (Table 8.7) The possessive and objective endings are almost completely identical in this language, with apparently the exception of the plural. However, even there the difference is merely due to an interaction between the plural suffix **-j-* and the aorist suffix **-ŋA-*, which together yield Tundra Nenets *-yə-*. The verbal and possessive personal endings themselves would thus originally have been the same. The dual possessum and dual object are both marked with NeT *-xəyu-* (PS ±**-kəńəj-*), an extension of the dual suffix NeT *-xəh* (PS **-kəń*).

Table 8.7: Tundra Nenets endings of the subjective conjugation, possessive declension and objective conjugation (Nikolaeva 2014: 67, 78–80).

| | sbc. | sg.poss.=obc. | du.poss.=obc. | pl.poss. | pl.obc. |
|---|---|---|---|---|---|
| 1sg. | *-d°m* | *-w°* | *-xəyu-n°* | *-V-n°* | *-yə-n°* |
| 2sg. | *-n°* | *-r°* | *-xəyu-d°* | *-V-d°* | *-yə-d°* |
| 3sg. | *-∅* | *-da* | *-xəyu-da* | *-V-da* | *-y°-da* |
| 1du. | *-nyih* | *-myih* | *-xəyu-nyih* | *-V-nyih* | *-y°-nyih* |
| 2du. | *-dyih* | *-ryih* | *-xəyu-dyih* | *-V-dyih* | *-y°-dyih* |
| 3du. | *-x°h* | *-dyih* | *-xəyu-dyih* | *-V-dyih* | *-y°-dyih* |
| 1pl. | *-waq* | *-waq* | *-xəyu-naq* | *-V-naq* | *-y°-naq* |
| 2pl. | *-daq* | *-raq* | *-xəyu-daq* | *-V-daq* | *-y°-daq* |
| 3pl. | *-q* | *-doh* | *-xəyu-doh* | *-V-doh* | *-y°-doh* |

The first-person endings of the plural and dual objective conjugation show a nasal element *-n-* rather than *-m-*, which results from a fusion of the possessed plural suffix PU **-n* and the first person possessive suffix starting with **-m*; e.g., 1.g. pre-PU **-n-mi* > PU **-ni* > PS **-nə* > NeT *-n°* (Janhunen 1982: 31). The plural suffix *-n-* is not found in the second- or third person possessive suffixes, but it may originally have been present there as well, since its loss would have been the regular development of an **-n-* between two consonants. With the addition of the plural marker **-j*, the resulting cluster in, e.g., pre-PS 3sg. **-jntA* would have been regularly simplified to PS **-jtA*. The same type of changes caused the morphophonological alternation of the cluster-initial suffixes like present participle suffix **-ntA* with **-tA* after consonant-final stems (cf., e.g., Janhunen 1998: 468 on this phonological rule).

Just like Tundra Nenets, Nganasan shows the presence of the plural suffix **-j* in these paradigms, since the sequence **-jt-* yielded *-ťʼ-* and **-jn-* yielded *-ń-*; cf. the paradigm below (based on Wagner-Nagy 2019). The dual possessive and dual objective endings only differ in the syllabicity of the palatal of the dual suffix. The possessive dual suffix *-kəj-* is more original and derives directly from PS **-kəń*, while the dual element of the objective conjugation, *-kəi-* from **-kəń+əj*, is derived with the added, analogical introduction of plural **-j-*, like in Tundra Nenets *-xəyu-* (Wagner-Nagy 2019: 230).

Table 8.8: Nganasan endings of the subjective conjugation, the possessive declension and the objective conjugation (Wagner-Nagy 2019: 204–205, 229, 231).

| | sbc. | sg.poss./obc. | du.poss. | du.obc. | pl.poss./obc. |
|---|---|---|---|---|---|
| 1sg. | *-m* | *-mə* | *-kəj-ńə* | *-kəi-ńə* | *-ńə* |
| 2sg. | *-ŋ* | *-rə* | *-kəj-ťə* | *-kəi-ťə* | *-ťə* |
| 3sg. | *-∅* | *-tu~ʲ* | *-kəj-ťü~i* | *-kəi-ťü~i* | *-ťü~i* |
| 1du. | *-mi* | *-mi* | *-kəj-ńi* | *-kəi-ńi* | *-ńi* |
| 2du. | *-ri* | *-ri* | *-kəj-ťi* | *-kəi-ťi* | *-ťi* |
| 3du. | *-kəj* | *-ði* | *-kəj-ťi* | *-kəi-ťi* | *-ťi* |
| 1pl. | *-mu~ʲʔ* | *-mu~ʲʔ* | *-kəj-ńü~iʔ* | *-kəi-ńü~iʔ* | *-ńü~iʔ* |
| 2pl. | *-ru~ʲʔ* | *-ru~ʲʔ* | *-kəj-ťü~iʔ* | *-kəi-ťü~iʔ* | *-ťü~iʔ* |
| 3pl. | *-ʔ* | *-ðu~ʲŋ* | *-kəj-ťü~iŋ* | *-kəi-ťü~iŋ* | *-ťü~iŋ* |

The *-r-* in the second person forms of the subjective conjugation and the singular objective conjugation and possessive endings results from a change from PU **-t-* to PS **-r-* in intervocalic position after the second syllable. Comparison with the Tundra Nenets endings shows that the subjective conjugation second person endings of the dual and plural were affected by the NeT *-d-* (PU/PS **-t-*) that was preserved in the plural objective suffixes due to it being part of a cluster with **j* and/or **n* at the time of the change of intervocalic **-t-* to **-r-* (cf. above). The NeT *-d-*, Ng. *-t-* (PS **-t-*) in the third

person forms derive from PU **s* by regular sound law as well. The origin of the 3pl. suffix NeT *-doh*, Ng. *-ðu~ɟŋ* from PS **ton* is, to my knowledge, uncertain. It can be reconstructed mechanically as PU **son*, according to Körtvély (2005: 72), from a pronominal source where the final **-n* marks plurality, but the details are unclear to me.

The other northern Samoyed languages Forest Nenets and both Forest and Tundra Enets show essentially the same picture. The singular and plural objective conjugation endings of these languages are given in the next table.

Table 8.9: Forest Nenets (based on Burkova 2022: 688), and Forest and Tundra Enets objective conjugation (from Khanina & Shluinsky 2023: 820).

| | NeF sg. | NeF pl. | EnF sg. | EnF pl. | EnT sg. | EnT pl. |
|---|---|---|---|---|---|---|
| 1sg. | *-m°* | *-n°* | *-a, -u, -b* | *-n* | *-a, -ɔ, bo* | *-no* |
| 2sg. | *-ɬ°* | *-t°* | *-r* | *-z* | *-ro* | *-zo* |
| 3sg. | *-ta* | *-ta* | *-za* | *-za* | *-za* | *-za* |
| 1du. | *-j°* | *-j°* | *-jʔ, -biʔ* | *-ɲʲi* | *-jʔ, -biʔ* | *-ɲʲiʔ* |
| 2du. | *-ɬy°* | *-ty°* | *-riʔ* | *-ziʔ* | *-riʔ* | *-ziʔ* |
| 3du. | *-ty°* | *-ty°* | *-ziʔ* | *-ziʔ* | *-ziʔ* | *-ziʔ* |
| 1pl. | *-maq* | *-naq* | *-aʔ, -baʔ* | *-naʔ* | *-aʔ, -baʔ* | *-naʔ* |
| 2pl. | *-ɬaq* | *-taq* | *-raʔ* | *-zaʔ* | *-raʔ* | *-zaʔ* |
| 3pl. | *-tuŋ* | *-tuŋ* | *-zuʔ* | *-zuʔ* | *-zuʔ* | *-zuʔ* |

The Forest Nenets and Enets objective conjugation paradigms are basically the same as those of Nganasan and Tundra Nenets, and thus do not need to be discussed further. The northern Samoyed languages together provide a very clear picture of the objective conjugation, with three sets of endings for singular, dual and plural objects.

In the southern Samoyed languages Selkup and Kamas, the objective conjugation only marks the presence of an object, not its number. According to, e.g., Körtvély (2005: 70) and Wagner-Nagy and Szeverényi (2022), the lack of number agreement in the southern Samoyed languages means that the plural and dual object forms in northern Samoyed are a later innovation, after the end of the Proto-Samoyed period. I consider it more likely that part of the objective paradigm was lost in Selkup and Kamas (the Mator verbal endings are not well-attested), and that plural and dual object forms were already in existence in Proto-Samoyed, inherited and maintained in northern Samoyed (so Janhunen 2002: 81). A similar situation can be observed in Northern Khanty, which has lost the dual object forms; nevertheless, on the basis of the other Khanty varieties where dual object forms are found, their absence is taken as an innovation, a simplification (Honti 1976: 94; Csepregi 2023: 716).

Selkup and Kamas also show innovations in their dual and plural possessive declensions: Kamas does not mark possessed dual at all, and it uses the clearly innovative

suffix *-zaŋ-* in the plural; and both Selkup and Kamas lost the first-person plural possessive suffixes with an **-n-* (found in northern Samoyed and in Mator, and inherited from Proto-Uralic), and they generalized the second person *-l-* (PS **-r-*) to plural forms. With these changes, the system was reorganized to be more transparently agglutinating. In this instance the situation in the northern Samoyed languages is rightly considered to be more archaic (Däbritz 2017), and in the same way, the objective conjugation of the northern Samoyed languages is also more archaic than the one found in the southern Samoyed languages.

In Kamas, the objective conjugation forms of the first person singular merged with the subjective conjugation due the loss of final vowels, and the second person singular objective conjugation ending has become generalized to replace the corresponding subjective conjugation ending as well. As a result, only the third person has distinct objective conjugation suffixes (excepting the imperative, on which see farther below). Table 8.10 gives the Selkup and Kamas subjective and objective conjugation paradigms.

Table 8.10: Selkup and Kamas verbal endings of the subjective and objective conjugation (Selkup from Kazakevič 2022: 796; Kamas from Klumpp 2022: 829–830).

| | Selkup | | Kamas | |
|---|---|---|---|---|
| | sbc. | obc. | sbc. | obc. |
| 1sg. | *-ŋ* | *-m* | *-m* | *-m* |
| 2sg. | *-ntį* | *-l* | *-l* | *-l* |
| 3sg. | *-∅* | *-tį* | *-∅* | *-t* |
| 1du. | *-mɪː* | *-mɪː* | *-bəj* | *-bəj* |
| 2du. | *-lɪː* | *-lɪː* | *-ləj* | *-ləj* |
| 3du. | *-qɪ* | *-tɪː* | *-gəj* | *-dəj* |
| 1pl. | *-mįt* | *-mįt* | *-baʔ* | *-baʔ* |
| 2pl. | *-lįt* | *-lįt* | *-laʔ* | *-laʔ* |
| 3pl. | *-tįt* | *-tįt* | *-jəʔ* | *-dən* |

We can reconstruct a Proto-Samoyed objective conjugation paradigm based on Tundra Nenets and Nganasan, as presented in the Table 8.11 below. It is also possible to apply an additional step of internal reconstruction, already discussed above, namely the addition of plural **-n-* outside the first-person suffixes. This **-n-* is certainly needed in the first person, and the way in which it caused **-m-* to assimilate can even be reconstructed for Proto-Uralic (Janhunen 1982: 31). This same plural **-n-* will appear again outside the first-person forms in some of the objective conjugation paradigms of the other Uralic languages to be discussed further below, which provides comparative evidence for its reconstruction in non-first-person endings in pre-Proto-Samoyed. Table 8.11 contains Proto-Samoyed reconstructions for the objective conjugation, as well as reconstructions

showing the probable original structure of the endings. These should be taken as an illustration of the likely underlying structure of these endings, not as precise reconstructions at each stage.

Table 8.11: The development of the objective conjugation endings through time from (pre-)Proto-Uralic to Proto-Samoyed (cf. the Proto-Uralic paradigms given by Janhunen 1982: 32, 35; Aikio 2022: 16, 18; Zhivlov 2023: 156; the pre-Proto-Nenets paradigm given by Janhunen 1998: 471; and especially for the dual endings, Proto-Samoyed reconstruction by Urmanchieva 2023).

| | PS sg. | | pre-PS sg. | PS pl. | | pre-PS pl. | | PU pl. | | pre-PU pl. |
|---|---|---|---|---|---|---|---|---|---|---|
| 1sg. | *-mə* | < | *-mi* | *-j-nə* | = | [*-j*]*-nə* | < | *-ni* | < | *-n-mi* |
| 2sg. | *-rə* | < | *-ti* | *-j-tə* | < | [*-j*]*-n-tə* | < | *-n-ti* | = | *-n-ti* |
| 3sg. | *-tA* | < | *-sA* | *-j-tA* | < | [*-j*]*-n-sA* | < | *-n-sA* | = | *-n-sA* |
| 1du. | *-məjn* | < | *-məjn* | *-j-nəjn* | = | [*-j*]*-nəjn* | < | *-nijn* | < | *-n-mijn* |
| 2du. | *-rəjn* | < | *-təjn* | *-j-təjn* | < | [*-j*]*-n-təjn* | < | *-n-tijn* | = | *-n-tijn* |
| 3du. | *-tAjn* | < | *-sAjn* | *-j-tAjn* | < | [*-j*]*-n-sAjn* | < | *-n-sAjn* | = | *-n-sAjn* |
| 1pl. | *-mAt* | < | *-mAt* | *-j-nAt* | = | [*-j*]*-nAt* | < | *-nAk/t* | < | *-n-mAk/t* |
| 2pl. | *-rAt* | < | *-tAt* | *-j-tAt* | < | [*-j*]*-n-tAt* | < | *-n-tAk/t* | = | *-n-tAk/t* |
| 3pl. | *-ton* | < | *-son*? | *-j-ton* | < | [*-j*]*-n-*[*son*?] | < | *-n-sAk/t* | = | *-n-sAk/t* |

Just as the possessive paradigm in Samoyed is probably more archaic than the paradigms found in many other Uralic languages (Däbritz 2017), the (parallel) structure of the objective conjugation has also been very stable in this branch. This is partly the result of the generally phonologically preservative nature of the Samoyed branch (cf., e.g., Janhunen 2009: 67). The only changes seem to be the intrusion of extra markers for the dual (**-k+jn*) and plural (**-j-*) into the paradigm, and the third person plural ending PS **-ton*. At least the recharacterization of the dual is understandable if we consider the fact that a sequence like dual **-k-* and 2sg. **-ti* would have yielded PS **-tə* by sound law, without any remaining overt sign of the probably more original dual marker **-k-* (cf. subsection 3.5.2). This would have applied to all second- and third person forms, and perhaps even the first-person forms would have deleted the **-k-* in the expected cluster **-km-* (parallels for this cluster are lacking, so it is uncertain how it would have developed). The introduction of the nominal plural suffix **-j-* removed the plural suffix **-n-* proper to the possessive declension and the objective conjugation from the second- and third person forms, so that it is only retained in the first person.[98]

[98] Georg and Seefloth (2019) provide a different internal reconstruction of Samoyed than the one in Table 8.11, with the **-n-* restricted to the first person only, and compare the resulting paradigm with that of Eskimo-Aleut. The presence of **-n-* in the other persons is, however, supported by the rest of the Uralic languages that preserve plural possessive marking, as discussed in the following

#### 8.3.2.2 Mansi and Khanty

While the objective conjugation suffixes in Samoyed display some degree of fusion (mainly in the first- and second person forms), the paradigms in the Ob-Ugric languages Mansi and Khanty are almost entirely agglutinating. Since one fusional aspect, the first-person plural *-*n*- from *-*n-m*-, can be reconstructed for Proto-Uralic (e.g., Janhunen 1982), the strong agglutinating character of the objective conjugation in these languages is at least partially innovative, assuming that the objective conjugation is inherited.

Just as in (northern) Samoyed, the objective conjugation in Mansi and Khanty can agree with singular, dual and plural agents in person and number, and with singular, dual and plural objects in number only. The objective conjugation suffixes in both languages resemble the possessive suffixes, and they are generally considered to be of a common origin (Liimola 1973, Keresztes 1998: 411; possessive suffixes of Mansi are discussed in detail by Liimola 1963: 202–241).

The main differences between the possessive and objective suffixes in Mansi are found between singular possessum and singular object forms: the object forms are marked with an element *-l-*. This is also one of the more salient differences between the Mansi and Khanty objective conjugations. The origin of this suffix *-l-* is not agreed upon. Keresztes (1998: 403) marks it as "of uncertain origin", while Liimola (1973) proposed that the *-l-* is a deverbal nominalizing suffix originally, which was used to differentiate the first- and second person objective forms from those of the subjective conjugation. Rédei, meanwhile, derived the *-l-* from the momentative suffix PU *-*li* (Rédei 1989).

I am unsure which reconstruction is preferable, but the motivation for the introduction of the element *-l-* is clear. Without the *-l-* marking the singular, the objective conjugation endings of the first and second person would look identical to the subjective conjugation, which in turn looks actually the same as the possessive forms in Mansi. Liimola also describes a different strategy for differentiating subjective and objective conjugation forms, found in the Pelym dialect of Mansi. Aside from the forms with *-l-*, analogical addition of the final vowel from the 3sg.obc. *-te* to the other person forms is found as well: 1sg.obc. *-me* and 2sg.obc. *-ne* (Liimola 1973: 204).

Since the possessive forms are consistently different from the subjective conjugation in the other Uralic branches, at least in the singular and in third person forms, the identity of the first- and second person singular subjective and possessive endings in Mansi must be an innovation. Obviously, it resulted from the apocope of final vowels,

---

subsections. The *-*j*- is reconstructed by Georg and Seefloth only for the third person forms originally, but the fact of the matter is that it is present in all persons in Samoyed, and that it would have regularly removed the *-*n*- by sound law except in the first person. As a consequence, I do not think that the pre-Proto-Samoyed reconstruction proposed by Georg and Seefloth accords with the available Uralic comparative evidence.

which rendered, e.g., PU 1sg. *-*m* and *-*mi* identical (Liimola 1973). The subjective suffixes are still different from the possessive ones in the third person, namely sbc.3sg. -∅, 3du. -*ɣ*, 3pl. -*t* as opposed to sg.poss.3sg. -(*t*)*e*, 3du. -*aɣe*, 3pl. -*ane*; and the latter are identical to the corresponding objective conjugation suffixes. Besides the singular object marker -*l*-, which, as discussed, is innovative in Mansi, the paradigm of the possessive declension and the objective conjugation only show minor differences.

Table 8.12: North Mansi possessive and verbal endings (Bakró-Nagy, Sipjocz & Skribnik 2022: 547).

| | sg.poss. | sg.obc. | du.poss. | du.obc. | pl.poss. | pl.obc. | sbc. |
|---|---|---|---|---|---|---|---|
| 1sg. | -*m* | -*ləm* | -*aɣəm* | -*aɣəm* | -*anəm* | -*anəm* | -*m* |
| 2sg. | -*n* | -*lən* | -*aɣən* | -*aɣən* | -*anən* | -*an*(*ən*) | -*n* |
| 3sg. | -(*t*)*e* | -*te* | -*aɣe* | -*aɣe* | -*ane* | -*ane* | -∅ |
| 1du. | -*men* | -*lamen* | -*aɣmen* | -*aɣmen* | -*anmen* | -*anmen* | -*men* |
| 2du. | -*en* | -*lən* | -*aɣen* | -*aɣən* | -*anen* | -*an*(*ən*) | -*n* |
| 3du. | -(*t*)*en* | -*en* | -*aɣen* | -*aɣen* | -*anen* | -*anen* | -*ɣ* |
| 1pl. | -*w* | -*ləw* | -*aɣəw* | -*aɣəw* | -*anəw* | -*anəw* | -*w* |
| 2pl. | -*en* | -*lən* | -*aɣen* | -*aɣən* | -*anen* | -*an*(*ən*) | -*n* |
| 3pl. | -(*a*)*nəl* | -*anəl* | -*aɣanəl* | -*aɣanəl* | -(*a*)*nanəl* | -*anəl* | -*t* |

The endings of the 3sg. and 3du. possessor, -(*t*)*e* and -(*t*)*en*, descend from PU *-*sA* and *-*sAjn*, with the Mansi change of PU *-*s* to *t*. The endings without the -*t*- are post-consonantal allomorphs (e.g., Keresztes 1998: 397). The -*t*- is also lacking after the dual and plural object suffixes -*ɣ*- and -*n*- in -*aɣe* and -*ane*. Since the -*t*- of the suffix disappeared after consonants, these third person endings can be reconstructed as an agglutinative series in pre-PMs., i.e., 3sg. *-*te*, *-*ɣ-te*, *-*n-te*, and 3du. *-*ten*, *-*ɣ-ten*, *-*n-ten*.

The ending of the first-person plural *-*w* in Mansi is shared with Khanty, and constitutes one of the potential shared innovations of Ob-Ugric in the phylogenetic sense (e.g., Honti 1976: 88). The ending of the second person dual and plural are merged, a syncretism that is shared with Khanty. Only in Mansi does this situation extend to the subjective conjugation as well, while the Khanty subjective conjugation has preserved the expected 2pl. suffix *-*təɣ* from PU *-*tAk*, as opposed to the 2du. *-*tən*, which is also as expected from PU *-*tijn*.

The origin of the Mansi 3pl. suffix -*anəl* is rather unclear (e.g., Keresztes 1998: 403). Liimola (1963: 230–231) connects the -*l* with the -*l* (or -*ɬ*) found in the plural possessed forms in Khanty, but does not explore the ultimate origin of this suffix further (see below on Khanty for a suggestion). In favour of a Khanty origin may be the appearance of this verbal suffix in those Mansi dialects that have undergone strong Khanty influence (Liimola 1973: 200-201). Plural possessive forms in simply -*an* are found in some varieties

of Konda Mansi, and could be more archaic, perhaps supporting the suggestion that the *-l-* was taken over from Khanty (Liimola 1963: 231). However, the precise origin of this *-an* does not seem to be entirely clear either. Liimola suggests that it is simply the possessive plural suffix *-*n* on its own (Liimola 1973: 231–232).

The origin of the *a*-vowel at the beginning of the dual and plural objective and possessive suffixes is also mysterious, but Honti connects it with the *a*-vowel found in most of the same suffixes in Khanty; only the placing of the vowel is different, since in Khanty it appears after the dual and plural co-affixes, while in Mansi it stands before them (Honti 1976: 112).

Table 8.13: A generalized abstracted structure of Mansi possessive and verbal endings; cf. Honti's (1998: 342) table of the Proto-Mansi person markers.

| | sg.poss. | sg.obc. | du.poss. | du.obc. | pl.poss. | pl.obc. | sbc. |
|---|---|---|---|---|---|---|---|
| 1sg. | *-m* | *-lə-m* | *-a-ɣ-əm* | *-a-ɣ-əm* | *-a-n-əm* | *-a-n-əm* | *-m* |
| 2sg. | *-n* | *-lə-n* | *-a-ɣ-ən* | *-a-ɣ-ən* | *-a-n-ən* | *-a-n-ən* | *-n* |
| 3sg. | *-te* | *-te* | *-a-ɣ-te* | *-a-ɣ-te* | *-a-n-te* | *-a-n-te* | *-∅* |
| 1du. | *-men* | *-l-a-men* | *-a-ɣ-men* | *-a-ɣ-men* | *-a-n-men* | *-a-n-men* | *-men* |
| 2du. | *-en* | *-lə-n* | *-a-ɣ-en* | *-a-ɣ-ən* | *-a-n-en* | *-a-n-ən* | *-n* |
| 3du. | *-ten* | *-ten* | *-a-ɣ-ten* | *-a-ɣ-ten* | *-a-n-ten* | *-a-n-ten* | *-ɣ* |
| 1pl. | *-w* | *-lə-w* | *-a-ɣ-əw* | *-a-ɣ-əw* | *-a-n-əw* | *-a-n-əw* | *-w* |
| 2pl. | *-en* | *-lə-n* | *-a-ɣ-en* | *-a-ɣ-ən* | *-a-n-en* | *-a-n-ən* | *-n* |
| 3pl. | *-anəl* | *-anəl* | *-a-ɣ-anəl* | *-a-ɣ-anəl* | *-a-nəl* | *-a-nəl* | *-t* |

The Khanty possessive and objective endings are again very similar to one another, although some important differences can be found. In most of the first and second person forms, the possessive and objective suffixes are, like in Mansi, shared with the subjective conjugation, but the third person endings and the 2pl. of the subjective conjugation are clearly opposed to the corresponding suffixes in the possessive declension and objective conjugation. Table 8.14 and Table 8.15 below illustrate the endings of various Khanty varieties. An important feature of the Khanty possessive declension and objective conjugation is the *-ɬ-*/*-l-* in forms marking dual and plural possessions or objects.

Table 8.14: Possessive and verbal suffixes of Vasjugan Khanty (Filchenko 2007), Surgut Khanty (Csepregi 2023), Kazym Khanty (Sipos 2022) and Obdorsk Khanty (Nikolaeva 1999a). Person/number of the subject/possessor in the far-left column, declension /conjugation type at the top. Dual object/possessee endings below the row ">du."; they are combined with the dual suffix supplied in that row. Plural object/possessee endings are listed following ">pl." and combined with the plural suffix of that row instead of the dual suffix. The singular object/possessee endings are not marked with an additional suffix.

| | Vasjugan | | Surgut | | Kazym | | Obdorsk | |
|---|---|---|---|---|---|---|---|---|
| | poss. | obc. | poss. | obc. | poss. | obc. | poss. | obc. |
| 1sg. | *-əm* | *-ɨm* | *-əm* | *-em* | *-em* | *-em* | *-em* | *-em* |
| 2sg. | *-ən* | *-ɨn* | *-ən* | *-e* | *-en* | *-en* | *-en* | *-en* |
| 3sg. | *-əl* | *-tə* | *-əɬ* | *-təɣ* | *-ɬ* | *-əɬ-ɬe* | *-l* | *-lli* |
| 1du. | *-mɨn* | *-ɨmən* | *-imən* | *-temən* | *-emən* | *-mən* | *-emən* | *-emən* |
| 2du. | *-tən* | *-ətən* | *-in* | *-ttən* | *-n* | *-ɬən* | *-lən* | *-lən* |
| 3du. | *-tən* | *-ətən* | *-in* | *-ttən* | *-n* | *-ɬ-ɬən* | *-lən* | *-lən* |
| 1pl. | *-uɣ* | *-uɣ* | *-əw* | *-təw* | *-eβ* | *-ɬeβ* | *-ew* | *-ew* |
| 2pl. | *-tən* | *-ətən* | *-in* | *-ttən* | *-n* | *-ɬən* | *-lən* | *-lən* |
| 3pl. | *-təl* | *-il* | *-iɬ* | *-iɬ* | *-ɬ* | *-eɬ* | *-el* | *-el* |
| >du. | *-ɣ(ə)l-* | | *-ɣəɬ-* | | *-ŋəɬ-* | | *-ŋil-* | |
| 1sg. | *-əm* | *-am* | *-am* | *-am* | *-am* | – | *-am* | *-am* |
| 2sg. | *-ən* | *-an* | *-a* | *-a* | *-an* | – | *-an* | *-an* |
| 3sg. | *-∅* | *-∅* | *-∅* | *-∅* | *-aɬ* | – | *-al* | *-li* |
| 1du. | *-əmɨn* | *-amən* | *-amen* | *-amen* | *-əmən* | – | *-mən* | *-mən* |
| 2du. | *-ɨn* | *-ɨn* | *-ən* | *-ən* | *-ən* | – | *-lən* | *-lən* |
| 3du. | *-ɨn* | *-ɨn* | *-ən* | *-ən* | *-ən* | – | *-lən* | *-lən* |
| 1pl. | *-uɣ* | *-əɣ* | *-əw* | *-əw* | *-əβ* | – | *-uw* | *-uw* |
| 2pl. | *-ɨn* | *-ɨn* | *-ən* | *-ən* | *-ən* | – | *-lən* | *-lən* |
| 3pl. | *-əl* | *-al* | *-aɬ* | *-aɬ* | *-aɬ* | – | *-al* | *-al* |
| >pl. | *-l-* | | *-ɬ-* | | *-ɬ-* | | *-l-* | |
| 1sg. | *-əm* | *-am* | *-am* | *-am* | *-am* | *-ɬam* | *-am* | *-am* |
| 2sg. | *-ən* | *-an* | *-a* | *-a* | *-an* | *-ɬan* | *-an* | *-an* |
| 3sg. | *-lə* | *-lə* | *-∅* | *-∅* | *-aɬ* | *-ɬe* | *-al* | *-li* |
| 1du. | *-əmɨn* | *-amən* | *-amən* | *-amən* | *-əmən* | *-ɬəmən* | *-əmən* | *-əmən* |
| 2du. | *-ɨn* | *-ɨn* | *-ən* | *-ən* | *-ən* | *-ɬən* | *-ələn* | *-əllən* |
| 3du. | *-ɨn* | *-ɨn* | *-ən* | *-ən* | *-ən* | *-ɬən* | *-ələn* | *-əllən* |
| 1pl. | *-uɣ* | *-oɣ* | *-əw* | *-əw* | *-əβ* | *-ɬəβ* | *-uw* | *-uw* |
| 2pl. | *-ɨn* | *-ɨn* | *-ən* | *-ən* | *-ən* | *-ɬən* | *-ələn* | *-əllən* |
| 3pl. | *-əl* | *-al* | *-aɬ* | *-aɬ* | *-aɬ* | *-ɬaɬ* | *-al* | *-al* |

Table 8.15: Subjective conjugation endings of Khanty (Filchenko 2007; Csepregi 2023; Sipos 2022; Nikolaeva 1999a).

| | Vasjugan | Surgut | Kazym | Obdorsk |
|---|---|---|---|---|
| 1sg. | *-əm* | *-əm* | *-m* | *-m* |
| 2sg. | *-ən* | *-ən* | *-n* | *-n* |
| 3sg. | -∅ | -∅ | -∅ | -∅ |
| 1du. | *-mən* | *-mən* | *-mən* | *-mən* |
| 2du. | *-tən* | *-ttən* | *-tən* | *-tən* |
| 3du. | *-ɣən* | *-ɣən* | *-ŋən* | *-ŋən* |
| 1pl. | *-oɣ* | *-əw* | *-β* | *-w* |
| 2pl. | *-ətəɣ* | *-təɣ* | *-ti* | *-ti* |
| 3pl. | *-ət* | *-ət* | *-t* | *-t* |

As in Mansi, the second person singular of Khanty has **-n* throughout, but the dual and plural subjective endings do show the **t* familiar from the other Uralic languages in 2du. *-tən* and 2pl. *-təɣ*/*-ti* (cf. Table 8.14). The 3sg. possessive suffix and the 3sg. objective conjugation endings are consistently differentiated in the majority of Khanty varieties, since an added object pronoun **-ɬəɣ* is suffixed in the objective conjugation (Honti 1976: 95; Raun 1988b: 564).

The dual co-affix in Khanty is **-ɣəɬ-*, as opposed to Mansi *-ɣ-*. This is different from the regular Khanty dual suffix as it is found in the unpossessed dual of nouns and in the subjective dual: KhVj.Sur. *-ɣən*, KhKaz.O *-ŋən* from PKh. **-ɣən*. Since there is no corresponding **-n-* or **-ɬ-* in the Mansi dual suffix *-ɣ-*, we should consider how the Khanty suffix arose. The **-n* in the Khanty dual suffix **-ɣən* was presumably taken over from the personal suffixes 1du. **-mən*, 2du. **-tən*, obc.3du. **-ɬən*, whence it spread to all dual forms; cf. also the Samoyed dual **-kəń*, which probably originated as the result of a similar, parallel development due to analogical pressures from PS 1du. **-məń* and 2du. **-təń*.

The plural in Khanty has the co-affix **-ɬ-*, instead of the **-n-* found in Mansi. It is also distinct from the unpossessed plural in Khanty, which has the plural suffix **-t* (as in Mansi, from PU **-t*). Honti relates the dual/plural co-affixal element **-ɬ-* to the toponym suffix *-lA* of Finnic, which developed into a plural suffix in Karelian (Honti 1998: 341).

However, since we expect both the dual and plural co-affixes to end in **-n*, we may consider whether this could have been regularly lost or removed. I think that a probable scenario is as follows: the original 3sg. dual and plural forms in pre-Proto-Khanty were **-ɣənsA* and *-nsA* respectively, corresponding to pre-Proto-Mansi **-ɣsa* and **-nsA* with spread of the **n* to the dual in Khanty as seen in the dual ending **-ɣən*. The cluster **-ns-* in these suffixes may well have been simplified to just **-s-*, eventually yielding the **-ɬ-*

found in the Khanty dual and plural co-affixes *-*ɣəɬ*- and *-*ɬ*-.[99] From an original (and expected) locus in the third person forms, these elements may then have spread throughout the paradigm by analogy.

Possibly in support of this hypothesis is the fact that the third person singular forms of the dual and plural objective conjugation do not have additional suffixes in a number of Khanty varieties: e.g., KhVVj. obc.du.3sg. -*ɣəl*-∅, KhTrj. -*ɣəɬ*-∅, and KhV obc.pl.3sg. -*l*{-∅/*əl*} (facultative zero-ending), KhTrj. -*ɬ*-∅. Alternative explanations are also available, however, such as analogical removal of the ending *-*əɬ* to differentiate it from the third person plural endings in *-*äɬ*, or, as per Honti (1976: 94-95), due to haplology. If, however, as I propose, the reflexes of *-*sA* and *-*nsA* merged as PKh. *-*ɬ*, the lack of an additional element is as expected.

This may also fit with the full vowel -*ä*- found after the dual and plural co-affixes in Khanty in most persons. This vowel is connected by Honti to the co-affix itself rather than to the personal ending, with which it only got associated due to metanalysis (Honti 1976: 112–113). The second-person forms in the dual and plural could potentially derive from assimilated *-*ɬ-tən*, as there is no trace of the expected suffix-initial consonant. Based on the rest of the paradigm, we might have expected obc.du.2du **-*ɣəɬätən* and obc.pl.2du. **-*ɬätən*. If such forms once existed, they might have undergone syncope of the vowel between two voiceless elements and simplification of the resulting cluster; i.e., **-(*ɣə*)*ɬätən* > **-(*ɣə*)*ɬtən* > *-(*ɣə*)*ɬən* (cf. MdE 3.obc.2pl. -*nk* < -*ntak* with a similarly conditioned truncation).

The singular possessive third person dual endings in -*ən* are extracted from the plural possessive third person in -*ɬən* by analogical removal of the plural co-affix -*ɬ*- (Honti 1976: 96). Expected is PKh. *-*ɬən* from PU *-*sAjn*, and this is actually reflected in the corresponding objective conjugation endings in some varieties like KhO -*lən*. The plural objective conjugation third person plural -*ɬ-aɬ* (historically *-*ɬa-ɬ*) looks like extended *-*nsa*(*n*)*sa* in Proto-Uralic terms.

[99] Honti reconstructs the co-affix as Proto-Khanty *-*l*- rather than *-*ϑ*- (=*-*ɬ*-), but the co-affix consistently has the same reflex as the third person possessive suffix in *-*ϑ*-, so that a reconstruction as *-*ɬ*- seems possible.

Table 8.16: Proto-Khanty possessive and verbal endings, after Honti (1998: 343).

| | sg.poss. | sg.obc. | du.poss/obc. | pl.poss/obc. | sbc. |
|---|---|---|---|---|---|
| 1sg. | *-(i)m* | *-im* | *-ɣə-ɬä-m* | *-ɬä-m* | *-m* |
| 2sg. | *-(i)n* | *-in* | *-ɣə-ɬä-n* | *-ɬä-n* | *-n* |
| 3sg. | *-ɬ* | *-ɬ* | *-ɣə-ɬ* | *-ɬ* | *-∅* |
| 1du. | *-(i)mən* | *-imən* | *-ɣə-ɬä-mən* | *-ɬä-mən* | *-mən* |
| 2du. | *-n* | *-tən* | *-ɣə-ɬ-ən* | *-ɬ-ən* | *-tən* |
| 3du. | *-n* | *-ɬən* | *-ɣə-ɬən* | *-ɬən* | *-ɣən* |
| 1pl. | *-(i)w* | *-iw* | *-ɣə-ɬä-w* | *-ɬä-w* | *-w* |
| 2pl. | *-n* | *-tən* | *-ɣə-ɬ-ən* | *-ɬ-ən* | *-təɣ* |
| 3pl. | *-(i)ɬ* | *-iɬ* | *-ɣə-ɬä-ɬ* | *-ɬä-ɬ* | *-t* |

Honti (1998: 342) gives a further reconstruction for Proto-Ob-Ugric, but for our purposes all relevant points have now been treated. It is clear that Mansi and Khanty both have objective conjugations with endings that are shared to a large extent with their respective possessive declensions. Mansi shows the plural morpheme **-n-* throughout the paradigm, also present in the first-person forms in Samoyed. In Khanty the earlier presence of this **-n-* may have been obscured by both sound law and analogical processes.

#### 8.3.2.3 Hungarian

The Hungarian objective conjugation is also known as the definite conjugation, as it agrees with definite nouns in Modern Hungarian. In Old Hungarian, it did not always agree with non-topical definite objects, however (É. Kiss 2013). Unlike most other Uralic languages with an objective conjugation discussed so far, the objective conjugation in Hungarian only marks the presence of the (definite) object, not its number. The function as marking definite objects rather than topical objects in modern Hungarian is probably a corollary of the fact that Hungarian has innovated definiteness as a category in the form of definite articles, combined with the semantic closeness of topicality and definiteness (É. Kiss 2013). Just as with Selkup and Kamas, I think that it is possible to take the absence of dedicated objective forms for plural objects as an innovation in Hungarian

The origins of many of the Hungarian verbal endings are rather difficult and obscure, however. The 1sg. present tense subjective ending *-k*, as well as the 1sg. present tense suffixes *-sz* and *-l* and the precise shape of the 1pl. endings *-(j)uk* and *-unk* have long been a matter of debate. The Hungarian data is not very useful from a comparative point of view for the purposes of a formal reconstruction, due to the uncertainty about many of the developments, so I will here only give a brief overview.

Table 8.17: Endings of the Hungarian subjective (indefinite) conjugation, objective (definite) conjugation, and possessive declension; only back vocalic allomorphs are shown (after Kenesei & Szécsényi 2022: 643–644).

| | sbc. | obc. | sg.poss. | pl.poss. |
|---|---|---|---|---|
| 1sg. | *-k*/*-m* | *-m* | *-m* | *-im* |
| >2sg. | — | *-lak* | — | — |
| 2sg. | *-sz*/*-l* | *-d* | *-d* | *-id* |
| 3sg. | *-∅* | *-(j)a*/*-i* | *-(j)a*/*-i*[(a)] | *-(j)ai*/*-i*[(a)] |
| 1pl. | *-unk* | *-(j)uk* | *-unk* | *-ink* |
| 2pl. | *-tok* | *-(j)átok* | *-(j)atok* | *-(j)aitok* |
| 3pl. | *-nak* | *-(j)ák* | *-(j)uk* | *-ik* |

(a) The suffix *-i* is found in possessed form in both singular and plural function in Old Hungarian (Papp 1968: 137–138).

From our point of view, the most interesting ending in the Hungarian objective conjugation is 2sg. *-d*. This suffix cannot directly continue PU **-ti*, since that would have become assibilated to ***-z*. Rather, it must derive from a cluster with a nasal, **-Nt*. Within the objective conjugation itself, this would of course only be expected in the plural objective forms, with the plural suffix **-n-*, of which otherwise no trace remains in Hungarian. The 2sg. *-d* is shared with the possessive declension, however, where it can be derived either from the plural (Kulonen 1993: 71–72) or from the old genitive or accusative (cf. Holopainen 2023). The possibility that the possessive declension had some effect on the Hungarian definite conjugation is not excluded, although there do not seem to be other clear traces of cross-pollination between the two sets of endings.

In this context it is noteworthy, however, that the Hungarian possessive declension does not mark plural possession with the Uralic suffix **-n-* at all, but rather with *-i-*. This element has been connected with the plural suffix PU **-j-* found in Saami, Finnic and Samoyed (e.g., Korhonen 1981: 209; Abondolo 1998: 21; Aikio 2022a: 15), but its origin is disputed. One problem with this connection is that the plural suffix **-j-* is otherwise completely absent from Hungarian, as well as from its most probable closest relatives Mansi and Khanty. It is therefore unclear when and from where this suffix should have been recruited for the possessive declension, before being lost in all other functions. Taking this into consideration, perhaps an internal source for the Hungarian marker *-i-* should be preferred (so argues Kulonen 1993: 56–57).

As an alternative, it has been suggested that it reflects a differentiation of an earlier polysemic and polymorphic 3sg. possessive suffix. This might be supported by Old Hungarian evidence, where at least the variant suffix *-i* can refer to both singular and plural possessions (Kulonen 1993: 57; Papp 1968: 137–138). The logical next question would then be how such polysemy arose in the first place; if the entire plural possessive

declension collapsed in Hungarian, there simply might not have been a way to mark the number of the possession at some point in time, so that it should later be reinvented completely. However, we might consider whether the expected plural *-*nsA* would have preserved the *-*n*- at all. The cluster *-*ns*- is otherwise unknown, so that the fate of such a suffix in Hungarian is unclear. Perhaps *-*nsA* and *-*sA* initially merged, as seems to have happened in Khanty -*ɬ*- (see 8.3.2.2), leading to the polysemy of the suffix observed in Old Hungarian, and allowing for a similar spread throughout the possessive paradigm as I suggested for Khanty.

Many of the other endings in Hungarian are obscure. The first person ending -*k* of the subjective conjugation has been variously identified as cognate with the plural suffix -*k*, the deverbal noun forming suffix -*k*, and as an instantaneous suffix; Abaffy (2001: 322) summarizes the previous literature. Havas instead suggests that it is connected with the medial suffix -*ik* (Havas 2004: 123ff.). It has also been identified with the Proto-Uralic present tense suffix -*k*- seen in Mansi, and with the first-person singular ending -*g* of the Permic negative verb (Zhivlov 2023: 157). Zhivlov argues that this connection would make sense given the fact that Hungarian 1sg. -*k* is not used outside of the present, with -*m* fulfilling the same function in the past tense. Both second-person singular endings -*sz* and -*l* have long been considered to have their origin in frequentative suffixes (Abaffy 2001: 322 with references). Abaffy herself considers -*l* to be the reflex of a medial suffix instead (ibid.: 326–327).

Rédei has attempted to explain the first-person plural suffixes -*unk* and -(*j*)*uk* as derived from the Proto-Uralic participle suffix *-*pA*, which became early Hungarian *-*w*-. for example, an indefinite verb form like *várunk* 'we wait' would, according to Rédei be formed like **várə-wə-mək*, with -*wə*- being the participle suffix and -*mək*- the actual 1pl. ending. The -*u*- found in these suffixes would thus be a reflex of their participial origin. Abaffy argues against such a derivation, on the basis of Old Hungarian data more often showing a mid-vowel -*ok*/-*ëk*/-*ok* and -*onk*, and forms like *vagmuc* for modern *vagyunk* 'we are' showing syncope of the stem-final vowel. The participle suffix *-*w*- normally yields a long vowel *ú* or *ű*, which should have been immune to syncope, and could also not have been shortened regularly (Abaffy 2001: 329). Honti connects the rounded vowel in the Hungarian first-person plural suffixes with the first person plural ending *-*w* of Mansi and Khanty; this would account for the absence of a nasal in the definite suffix *-(*j*)*uk*. Havas would rather get rid of the nasal due to intervocalic lenition of *-*m*- in a sequence of the type *várə-jə-mək* to *várə-jə-wək*, eventually *várjuk*, where syncope would have operated differently from the indefinite type *várə-mək* to *várunk* (Havas 2004: 132–133).

A more detailed discussion of the various considerations and arguments goes beyond the purpose of our present investigation into the Uralic objective conjugation as a whole, and therefore I will leave the matter with the scholars referred to and those of the future.

In general, it is clear that the Hungarian endings of the objective conjugation do not provide us with much of the comparative material that we need for our formal reconstruction. Only the 2sg. suffix *-d* potentially contains trace of a theoretically more informative earlier situation, but its value is diminished significantly due to uncertainty regarding its exact formation.

#### 8.3.2.4 Mordvin

Mordvin is the only branch belonging to the western group of Uralic that has an objective conjugation; and it has the most elaborate system of object marking of all Uralic languages, with suffixes for first and second person objects as well as those for third person objects. The paradigms in the two Mordvin languages Erzya (E) and Moksha (M) contain the same number of functional endings, but they are different in their form to some extent. The entire paradigm appears to be heavily restructured when compared to the other Uralic languages, to the point that it has been considered to be a separate innovation almost in its entirety; so Kereszstes (1999), who does not reconstruct a complete Proto-Mordvin paradigm and ascribes the difference in the conjugations of the Erzya and Moksha to independent, parallel developments based on an original defective partial paradigm. I would rather follow Bergsland's analysis (1973), which focuses on the similarities between the Mordvin and Samoyed objective conjugations, as well as the ways in which the Mordvin endings may have been on the one hand extended and on the other hand recharacterized. Bergsland provides pre-Proto-Mordvin reconstructions of some individual endings, but no coherent reconstruction of Proto-Mordvin itself. Bergsland's arguments are furthermore presented in a rather compact way that is at points difficult to follow. I hope here to provide a clear picture of Proto-Mordvin and integration with the endings of the other Uralic languages.

**Past tense**

It is especially in the past tense that the Mordvin endings resemble those in the Samoyed languages (Bergsland 1973: 51), and also to an extent those of Mansi. I will therefore first discuss the past tense paradigm. One thing to look out for is once again the presence of an element *-n-* in certain plural forms, in addition to the more notorious suffix **-sA* of the third person singular.

To understand the morphological nature of the past tense endings, it is important to know that the main past tense morpheme is *-i-* in Erzya, and *-j-* or *-∅-* in Moksha (PU **-j-*); sometimes in Moksha this past morpheme is only reflected in the palatalization of the first consonant of the ending. In the third person singular and third person plural of the subjective conjugation, the past tense suffix *-ś-* is used instead. For ease of comparison, I also give the possessive paradigms again, although it should be borne in mind that they are marked with an additional suffixed demonstrative element in Moksha

(cf. below). The choice of a vowel *e* or *o* in Erzya depends on vowel harmony. In combination with the past tense morpheme *-i-* only front harmonic allomorphs exist, with palatalization of alveolar consonants where possible.

Table 8.18: Erzya (E) and Moksha (M) Mordvin endings of the possessive declension (with the oblique forms added in parentheses if different from the nominative), the past tense objective conjugation and the past tense subjective conjugation. Past tense morphemes are separated from the person endings with a hyphen wherever the two are not wholly fused. The subject/possessor person/number is given on the left, the declension/conjugation type and object/possessee person/number at the top. This information can be found in, e.g., Raun (1988a: 102–103, 106–107); Keresztes (1999: 18–19, 67–68); Hamari & Ajanki (2022: 403, 408–412).

| | | sg.poss. | pst.3sg.obc. | pl.poss. | pst.3pl.obc. | pst.sbc. |
|---|---|---|---|---|---|---|
| 1sg. | E | *-m* (*-n*) | *-i-ja* | *-n* | *-i-ń* | *-i-ń* |
| | M | *-źä* (*-n*) | *-j-ńä* | *-ńä* | *-j-ńä* | *-ń* |
| 2sg. | E | *-t* | *-i-k* | *-t* | *-i-tʹ* | *-i-tʹ* |
| | M | *-ćä* (*-t*) | *-j-tʹ* | *-tʹńä* | *-j-tʹ* | *-tʹ* |
| 3sg. | E | *-ze*/*o* (*-nze*/*o*) | *-i-ź́e* | *-nze*/*o* | *-i-ńź́e* | *-ś* |
| | M | *-c* (*-nza*) | *-źä* | *-nza* | *-źəń* | *-ś* |
| 1pl. | E | *-ne*/*ok* | *-i-ńek* | *-ne*/*ok* | *-i-ńek* | *-i-ńek* |
| | M | *-ńkä* (*-nk*) | *-ś-k* | *-ńkä* | *-ś-k* | *-mä* |
| 2pl. | E | *-nk* | *-i-nk* | *-nk* | *-i-nk* | *-i-ďe* |
| | M | *-ńťä* (*-nt*) | *-ś-tʹ* | *-ńťä* | *-ś-tʹ* | *-ďä* |
| 3pl. | E | *-st* | *-i-ź́* | *-s-t* | *-i-ź́* | *-ś-tʹ* |
| | M | *-sna* (*-st*) | *-ź́* | *-sna* | *-ź́* | *-ś-tʹ* |

The possessive suffixes in Erzya continue the Proto-Uralic suffixes rather straightforwardly, with the caveat that the plural possessive forms were generalized also to the singular when the possessor was plural; e.g., for sg.poss.1pl. *-ne*/*ok* one would expect ***-me*/*ok* instead, from PU **mAk*, but it was levelled with the pl.poss.1pl. *-ne*/*ok* < PU **-nAk*.

The Moksha possessive suffixes are remade with the addition of demonstrative pronouns (e.g., 2sg. *-t* + *śä* → *-ćä*; see Bereczki 1988: 325–326). Because of this, the nominative possessive suffixes of Erzya and Moksha are remarkably dissimilar to one another. However, in the oblique the possessive suffixes of Erzya and Moksha are more alike: 1sg. EM *-n*, 2sg. EM *-t*, 3sg. E *-nze*/*o* ~ M *-nza*, 1pl. E *-ne*/*ok* ~ M *-nk*, 2pl. E *-nk* ~ M *-nt*, 3pl. EM *-st* (Raun 1988a: 103). These suffixes are identical to the plural possessive suffixes due to the oblique (genitive) morpheme being the same as the possessive plural morpheme, namely PU **-n-*. The Proto-Mordvin possessive paradigm was thus basically

the same as the one found in Erzya, while the Moksha nominative forms are innovative with the addition of demonstrative pronouns.

From among the singular endings of the past objective conjugation, only the 3sg. E *-i-źe* ~ M *-źä* immediately seems to be truly archaic (Keresztes 1999: 107). This regularly continues the ending **-sA* as seen in the other branches that have an objective conjugation, namely Hu. *-(j)a*, Ms. *-te*, Kh. *-ł*, PS **-tA*. The forms for the plural subjects are again analogically levelled to be the same as those with plural objects (this is the general pattern in Mordvin, cf. above on the possessive endings).

The third person plural object endings of the past objective conjugation may also constitute a particularly archaic part of the paradigm, however: 3pl.obc.1sg. E *-i-ń* ~ M *-j-ńä*, 3pl.obc.2sg E *-i-ť* ~ M *-j-ť*, 3pl.obc.3sg. E *-i-ńźe*, 3pl.obc.1pl. E *-i-nek* and 3pl.obc.2pl. E *-i-nk* can all be regularly derived from past **-j-* + PU **-ni*, **-nti*, **-nsA*, **-nAk* and **-ntAk* respectively (Bergsland 1973: 50–51 on the first- and second-person forms). These suffixes contain the same plural suffix **-n-* as also seen most clearly in Mansi, and in the first person in Samoyed, although it is only preserved in the third person singular, the first-person plural and the second person plural. The use of this particular plural suffix indicates a strong connection with the possessive paradigm. While Keresztes (1999: 114) interprets this as a secondary transfer of the possessive endings to the objective paradigm where no endings existed yet, in my opinion, the fact that same structure is clearly paralleled outside of Mordvin in the other Uralic branches rather suggests that these plural objective endings are particularly old.

The plural objective forms in the preterite are partially syncretic with the subjective preterite paradigm, which causes problems for a straightforward reconstruction. This is particularly the case in the Erzya 1sg. *-iń*, 2sg. *-iť* and 1pl. *-ińek*. According to Keresztes (1999: 114), this means that the past plural objective paradigm simply did not exist originally, and that therefore the preterite subjective forms were extended to be used as objective endings as well. However, in that case it is unclear why only a few specific endings were taken over from the subjective paradigm in this way. The syncretism may rather be the result of a merger by sound law, since, if we assume a plural objective paradigm as indicated by Samoyed and Mansi for pre-Proto-Mordvin as well, the Erzya endings (except the clearly innovative third person plural) are all as expected: 1sg. **j-ni* would become Erzya *-iń* with loss of the final vowel, and 2sg. **-j-nti* would regularly lose the **n* in the word-final cluster **-jńť* to become Erzya *-iť*. At the same time, the subjective first-person singular is also expected to become *-i-ń* from PU **-j-m* with the change of final **-m* to *-n* and palatalization caused by **-j-*, and, likewise, the subjective second-person singular *-i-ť* from PU **j-t* would be regular as well (cf. Bergsland 1973: 49–50 and further below).

Table 8.19: Expected and attested subjective conjugation singular and plural objective conjugation singular subject endings.

| S ↓ | pst.sbc. (quasi-)PU | expected PMd. | Erzya | Moksha |
|---|---|---|---|---|
| 1sg. | **-j-m* | > **-j-ń* | = *-iń* | = *-ń* |
| 2sg. | **-j-t* | > **-j-t'* | = *-it'* | = *-t'* |
| 3sg. | **-ć* | > **-ś* | = *-ś* | = *-ś* |
| S ↓ | pst.3pl.obc. (quasi-)PU | expected PMd. | Erzya | |
| 1sg. | **-j-ni* | > **-j-ń* | = *-iń* | ~ *-jńä* |
| 2sg. | **-j-nti* | > **-j-t'* | = *-iť* | ~ *-jt'* |
| 3sg. | **-j-nsA* | > **-j-ńźA* | = *-ińźe* | >> *-źəń* |

Moreover, the subjective 1pl. **-ińek* in Erzya is innovative, and taken over from the objective conjugation. The corresponding subjective ending in Moksha, *-mä*, is isolated and more archaic, reflecting expected PU **-j-mAk* with apocope of final **-k*. The original Proto-Mordvin first-person plural endings would thus have been different in at least the preterite subjective **-mAk* (as reflected in Moksha) and the preterite objective plural **-nAk* (as reflected in Erzya). This is not taken into account by Keresztes (1999: 112–115), who interprets the differences between Erzya and Moksha as the result of later creations of new, independent forms.

The Erzya sbc.2pl. *-i-d'e* is also clearly different from the plural objective suffix *-i-nk*, and together with Moksha *-d'ä* it represents an archaic suffix descending from, in PU terms, **-j-tAk* (cf. the Erzya present sbc.2pl. *-tado* from PMd. **-ta-dAk*). The opposition between the subjective conjugation and objective conjugation endings here is as expected, and thus likely inherited.

The subjective third person preterite endings sg. *-ś* and pl. *-śt'* are both distinct from the corresponding objective endings as well, but the 3pl. *-ź* was secondarily taken from a perfect participle or gerund in *-ź* (Keresztes 1999: 95)[100], in line with the general tendencies of Uralic languages to innovate third-person verbal endings. The element *-ź-* is included in a number of other plural endings (also in the first- and second-person object forms discussed below), among them the innovative Moksha 1pl. *-śk* and 2pl. *-śt'* with regressive voice assimilation (Keresztes 1999: ibid.).[101] The archaic forms M *-jńək* (=

[100] The current function of *-ź* outside of the objective conjugation paradigm is to form gerunds expressing simultaneity, but it is considered to have been a perfect participle originally (Zaicz 1998: 205).

[101] The nature of the *-k* in 1pl. *-śk* and the *-t'* in 2pl. *-śt'* are not addressed by Keresztes (1999: 95). The exact mechanism for the formation is not immediately obvious, but *-k* and *-t* are found in the

E *-ńek*) and M *-jəńť* (= E *-ink*) are preserved in southern and northern/north-eastern dialects (Keresztes 1999: 114, 118, 222, 228). The presence of the originally participial suffix *-ź-* in the objective conjugation is certainly secondary, and it thus obscures the ending would have existed for the third person plural at an earlier stage in pre-Proto-Mordvin: the expected objective ending of the third person plural acting upon the third person singular or plural is probably ***-śť*, parallel to the corresponding possessive ending E *-st* with the additional palatalizing effect of the past tense morpheme, i.e., **-j-st*. This is obviously the same as the third person plural subjective ending, which may have been cause for a secondary replacement.

Table 8.20: Expected and attested subjective conjugation plural and plural objective conjugation plural subject endings.

| S ↓ | pst.sbc. (quasi-)PU | expected PMd. | Erzya | Moksha |
|---|---|---|---|---|
| 1pl. | **-j-mAk* | > **-j-mA*(*k*) | >> *-ińek* | = *-mä* |
| 2pl. | **-j-tAk* | > **-j-ďA*(*k*) | = *-iďe* | = *-ďä* |
| 3pl. | **-ś-t* | > **-ś-ť* | = *-śť* | = *-śť* |
| S ↓ | pst.3pl.obc. (quasi-)PU | expected PMd. | Erzya | Moksha |
| 1pl. | **-j-nAk* | > **-j-ńAk* | = *-ińek* | archaic = *-jńək* ; >> *śk* |
| 2pl. | **-j-ntAk* | > **-j-ń*(*t*)*k* | = *-ińk* | archaic = *-jəńť* ;>> *śť* |
| 3pl. | **-j-nsAt*/*k* | > **-j-st* | >> *-ź* | >> *-ź* |

The origin of the pst.3sg.obc.1sg. *-i-ja* is quite obscure. According to Keresztes (1999: 112), it is analogical to the present *-sa*. Bergsland's account appears to be contradictory: he derives E *-ija* from a particle *-jak* (*-gak*, *-kak*) added to a "dialectal" preterite in *-i* (Bergsland 1973: 48), but the Moksha ending *-jä* (dialectally restricted to the south, see Keresztes 1999: 207) is taken by Bergsland directly from PU **-j-mi*. Bergsland posits an opposition between Moksha sbc.1sg. *kunda-ń* 'I catch' vs. 3sg.obc.1sg. *kundajä* 'I catch it' vs. 3sg.obc.1sg. *kundajńä* 'I catch them' as the original situation, and a reflection of the Proto-Uralic distinction in the suffixes as **-j-m* vs. **-j-mi* vs. **-j-ni* (Bergsland 1973: 49). The opposition between the suffixes *-jä* an *-jńä* does appear to be present in any single dialect, however (see Keresztes 1999: 207, 209). The final **-ńä* of the Moksha first person objective preterite endings can also reflect, in part, the plural pronominal stem *ńe*, as seen in the possessive paradigm (cf. M pl.poss.1sg. *-ńä*, there from **-n-ńe*). The expected

possessive suffixes (oblique) 1pl. *-nk*, 2pl. *-nt*. Expected for these past tense objective endings would be ***-ńk* (or perhaps ***-ńť*) and ***-ńť*.

outcome of the past third person objective conjugation first person ending would probably have been **-*i*-*m* (cf. sg.poss.1sg. E -*m*), which thus seems to have been replaced.[102]

The -*k* in the second person singular was introduced from the imperative 2sg. -*k* (Bereczki 1988: 329), and the pst.3sg.obc.2sg. E -*i*-*k* is identical with the objective imperative. In fact, all imperative forms are more or less syncretic with the corresponding preterite endings, with the sole exception of the imp.1sg.2sg. -*mak* (cf. the table below). According to Bergsland, this -*mak* derives from a spread of the imperative suffix *-*k* in phrases with a negative imperative like **iľa moń kundak* 'don't catch me' → **iľamak kundak*, which by analogy yielded forms like *kundamak* instead of expected **kundak moń* (Bergsland 1973: 48). Perhaps a sequence **kundak moń* through univerbation could also give **kundamań* (with simplification of the cluster *-*km*-), whence analogical addition of the regular imperative suffix would have yielded *kundamak* as well. Whichever sequence of changes facilitated the creation of the suffix -*mak*, an addition of personal object pronouns to verb forms and subsequent analogical processes needs to be assumed for the first- and second-person objective suffix, on which see farther below.

Table 8.21: The imperative endings of Erzya (E) and Moksha (M) Mordvin (e.g., Keresztes 1999: 68; Hamari & Ajanki 2022: 411). Subject person/number on the left, conjugation type and object number/person at the top.

| S ↓ | | sbc. | 1sg.obc. | 3sg.obc. | 1pl.obc. | 3pl.obc. |
|---|---|---|---|---|---|---|
| 2sg. | E | *-k*/*t*/*ť* | *-mak* | *-ik* | *-miź* | *-iť* |
| | M | *-k*/*t*/*ť* | *-mak* | *-k* | *-maśť* | *-jť*/*iť* |
| 2pl. | E | *-do*/*ďe* | *-miź* | *-ink* | *-miź* | *-ink* |
| | M | *-da* | *-maśť* | *-śť* | *-maśť* | *-śť* |

[102] One could come up with a sequence of changes from *-*jmə* to E -*ja* if *-*jmə* first yielded *-*jm*, and if this cluster underwent epenthesis to *-*jəm*. It would probably have been the only instance of such a cluster in word-final position in the language, so that it is difficult to corroborate this hypothetical change. With analogical removal of the final nasal, as also posited by Keresztes (1999: 112) for the corresponding present ending *-*sa*, this *-*jəm* would yield E -*ja*. A parallel for analogical removal of the final nasal in a first-person ending may be found in M prs.2sg.obc.1sg. -*ťä* without a final nasal next to E -*tan* with a final nasal. The vowel -*a*- in Erzya if often of epenthetic origin, such as in the second person plural ending -*tadok*, morphologically from -*t* + *dok* (Ravila *apud* Bergsland 1973: 44–45; Bereczki 1988: 326). While such a derivation of -*ja* might be possible in principle, it remains much too speculative.

In Erzya the difference between objective imperative and preterite forms is only maintained in the first-person object forms, as they lack the suffix *-*i*-. In Moksha, the preterite suffix is absent in the corresponding forms in the preterite as well, having merged with the final consonant. The Moksha imp.3sg.obc.2sg. is simply -*k* as opposed to prt.3sg.obc.2sg. -*jť*. Unlike the subjective conjugation imperative, however, this -*k* does not partake in alternations based on the shape of the stem, so that an intervening vowel as seen in Erzya must be reconstructed for Moksha as well.

**Present tense**

Now that we have seen the past tense suffixes, we can move on with an overview of the present. The endings are given in Table 8.22 below. The endings of the optative are added here as well, since they are very similar to the present tense endings. Optative forms of the objective conjugation are only really found in older material and as such they are not written in the same phonological transcription as the past and present endings. The spelling is fairly easy to convert, however, as for instance 3sg. -*zynze* is to be analysed as -*zińźe*, analogous to the corresponding present -*sińźe*.

Table 8.22: Erzya (E) and Moksha (M) Mordvin endings of present and optative objective conjugation and the present subjective conjugation. Subject person/number and object person on the left, tense/mood, conjugation type and object number at the top.

| S ↓ | | prs.sg.obc. | opt.sg.obc. | prs.pl.obc. | opt.pl.obc. | prs.sbc. |
|---|---|---|---|---|---|---|
| 1sg>3. | E | -*sa* | -*za* | -*s-i-ń* | -*zyn* | -*an* |
| | M | -*sa* | -*zainä* | -*sa-j-ńä* | -*zainä* | -*an* |
| 2sg>3. | E | -*sa-k* | -*zak* | -*s-i-ť* | -*zt* | -*at* |
| | M | -*sa-k* | -*zait* | -*sa-j-ť* | -*zait* | -*at* |
| 3sg>3. | E | -*s-i* | -*zazo* | -*s-i-ńźe* | -*zynze* | -*i* |
| | M | -*s-i* | -(*za*)*zä* | -*si-ńä* | -*zasɨn* | -*aj*/-*j* |
| 1pl>3. | E | -*s-i-ńek* | -*zynek* | -*s-i-ńek* | -*zynek* | -*tano*/-*ťano* |
| | M | -*sa-ś-k* | -*zaśk* | -*sa-ś-k* | -*zaśk* | -*tama*/-*ťama* |
| 2pl>3. | E | -*s-i-nk* | -*zynk* | -*s-i-nk* | -*zynk* | -*tado*/-*ťado* |
| | M | -*sa-ś-ť* | -*zaśt* | -*sa-ś-ť* | -*zaśt* | -*tada*/-*ťada* |
| 3pl>3. | E | -*s-i-ź* | -*zyze* | -*s-i-ź* | -*zyze* | -*iť* |
| | M | -*sa-ź* | -*zaz* | -*sa-ź* | -*zaz* | -*ajť*/-*jť* |

The present objective paradigm is certainly completely remade, or newly made, based on an element *-*sa*-. It therefore has very little bearing on the original connection of the Mordvin objective conjugation with the other Uralic languages. According to Keresztes, the element *-*sa*- originates from the optative paradigm, and consists of a combination of the optative suffix -*za* and the 3sg. objective suffix -*ze/o*. Geminates are always

voiceless in Mordvin, so the result of a sequence *-*zVzV*- would be *-*ssV*- after syncope, whence regularly Mordvin -*sV*- (Keresztes 1999: 110–111).

The suffixal element -*sa*- would thus essentially include the same suffix twice, in Proto-Uralic terms *-*sA*-*sA*. On its own, *-*sA* came to be used as a general optative suffix (PMd. *-*zA* > E -*zo/e*, M -*za*; Bereczki 1988: 328), apparently only retaining its association with the objective conjugation in the past tense. From there the new present tense objective conjugation in *-*sA*-*sA* was derived from the optative with the addition of the ending from the preterite (Bergsland 1973: 48). The discrepancy in the vowels between -*sa*- and -*zo* is explained by Bergsland as the result of analogy with the vowel in the endings of the subjective present 1sg. -*an* and 2sg. -*at*. The preterite objective suffix would thus have been in existence and functional when the present objective endings were first formed (ibid.).[103]

The objective conjugation of the optative itself disappeared in both Erzya and Moksha, and only Erzya still has a complete subjective optative paradigm, while 1pl. and 2pl. suffixes are missing in modern Moksha (Keresztes 1999: 96; Hamari & Ajanki 2022: 412). The optative suffix is EM -*za*-, both in the objective and subjective forms. Paradigms of the definite optative were recorded before they disappeared (Keresztes 1999: 96, with further references). The general pattern is that the optative endings are the same as the present, but with a -*z*- instead of an -*s*-. The only exceptions to this are the third person singular object suffixes in both Erzya and Moksha, and in Moksha the first- and second-person singular.

**First- and second-person forms**

Now turning briefly to the first- and second-person objective forms, we can see that only those suffixes referring to singular objects are truly distinctive from one another. Most forms involving a plural (either as the agent or the object) are syncretic in some way or another, merging reference to distinctive plural subject persons in one form. In Table 8.23, endings that are equal to the ones to the left are replaced with an "=" sign to reduce clutter. Without parallels in the other Uralic languages that have an objective conjugation, these forms are clearly innovations specific to Mordvin and do not represent the Proto-Uralic state of affairs. Moreover, the personal object forms look like they are made by agglutination of various recognisably Mordvin personal pronouns, and much analogical modification (see Bergsland 1973).

[103] It is suspicious that the objective conjugation contains beside -*sa*- the elements -*ma*- and -*ta*- for the first and second person respectively, from secondary univerbations with pronominal *moń* and *toń*; in parallel, pronominal *soń* would probably have yielded -*sa*- as well if incorporated into the present objective conjugation. This would have created essentially a two-tier system with old objective endings in the past and new objective endings in the present, influencing each other over time to create the present-day Mordvin objective conjugation paradigm.

Table 8.23: Erzya (E) and Moksha (M) Mordvin objective conjugation endings with reference to first- and second-person objects. Subject person/number and object person on the left, tense and object number at the top.

| S ↓ | | pst.sg.obc. | pst.pl.obc. | prs.sg.obc. | prs.pl.obc. |
|---|---|---|---|---|---|
| 1sg>2. | E | *-i-ť-iń* | *-i-ď-iź* | *-ta-n* | *-ta-diź* |
| | M | *-j-ť-äń* | *-ď-äź* | *-ťä(-ń)* | *-ťä-ďäź* |
| 2sg>1. | E | *-i-m-ik* | *-i-m-iź* | *-sa-m-ak* | *-sa-m-iź* |
| | M | *-m-ajť* | *-m-aś-ť* | *-sa-m-ak* | *-sa-m-aś-ť* |
| 3sg>1. | E | *-i-m-im* | *-i-m-iź* | *-sa-m-am* | *-sa-m-iź* |
| | M | *-m-ań* | *-m-aź* | *-sa-m-ań* | *-sa-m-aź* |
| 3sg>2. | E | *-i-ńź-iť* | *-i-ď-iź* | *-ta-nza-t* | *-ta-d-iź* |
| | M | *-ńźä* | *-ď-äź* | *-ta-nza* | *-ťä-ď-äź* |
| 1pl>2. | E | *-i-ď-iź* | = | *-ta-d-iź* | = |
| | M | *-ď-äź* | = | *-ťä-ď-äź* | = |
| 2pl>1. | E | *-i-m-iź* | = | *-sa-m-iź* | = |
| | M | *-m-aś-ť* | = | *-sa-m-aź* | = |
| 3pl>1. | E | *-imiź* | = | *-sa-m-iź* | = |
| | M | *-maź* | = | *-sa-m-aś*[*ť*] | = |
| 3pl>2. | E | *-iďiź* | = | *-ta-d-iź* | = |
| | M | *-ďäź* | = | *-ťä-ď-äź* | = |

The origin of the past plural ending 2.obc.pl. E *-iďiź* ~ M *-ďäź* is relatively easily understandable: to the pst.sbc.2pl. ending PMd. **-ďa* the suffix **-ź* of the pst.obc.3pl. was added (Bergsland 1973: 46). In Erzya this was depalatalized to *-tadiź*. The same type of development can be assumed for the corresponding pst.obc.1pl. in Erzya *-imiź* (i.e., **-j-mə+j+ź*), while the Moksha form *-maśť* is extended with **-t*. The present suffixes were then formed with the addition of the present suffixes **-sa-* for the first-person object forms (as in the rest of the present objective conjugation paradigm) and **-ta-* for the second-person object forms.

The ending of the pst.2sg.obc.1sg. E *-iťiń* ~ M *-jťäń* looks similar to the second person genitive pronoun EM *toń*, which is also the expected form the accusative; according to Bergsland, it can be directly derived from this (Bergsland 1973: 47). In principle the suffix could have been extracted from constructions such as **eźiń toń ńeje* 'I did not see you', which became **eźi*(*ń*)*ťań ńeje*, and **ńejiń toń*, which became **ńeji*(*ń*)*ťań* 'I saw you', both with reductions resulting from the suffixation process (namely, **-nť-* to *-ť-* and **o* to *a*). The present suffix counterpart E *-tan* could be based on the opposition in palatalization on the other first-person suffixes between past *-ń* and present *-n* (Bergsland 1973: 47). Segmentation of this form as *-ta-n* then allowed for the analysis of *-ta-* as a marker of the

present + second person, and in that function, it is inserted into E *-tadʹiź* ~ M *-tʹädʹäź* (cf. above). The Moksha form *-ťä(ń)* seems to be based only on the removal of the preterite morpheme *-j-*, but without depalatalization. The final *-ń* was further removed everywhere except in some northern and central dialects (see Keresztes 1999: 234).

At least the Moksha past objective conjugation ending referring to a first-person singular object *-mań* looks like it could be of a similar origin, i.e., **eś moń ńeje* 'he did not see me' > **eźəmań ńeje*, and parallelly *ńejemań* 'he saw me'. Bergsland rather derives this suffix by a replacement of the final *-k* in the first-person objective second person singular subject ending *-mak* with *-ń* from the possessive paradigm (Bergsland 1973: 47); the same development would yield the final *-m* in the corresponding Erzya suffix pst. *-imim*, prs *-samam* (ibid.). Bereczki (1988: 329) takes this element as a doubling from original **-imi*, **-sama*, and derives the Moksha final **-ń* from the subjective conjugation. However, since I would rather assume suffixation of the oblique pronoun *moń*, as seemingly reflected in the Moksha form, the final *-m* in E *-imim*, *-samam* is, in my opinion, the result of an assimilation from more expected E ***-imiń*, ***-samań* to the preceding *-m-*.

The 2sg.obc.3sg. endings containing **-nzA* are difficult to account for: this is the expected 3pl.obc.3sg. suffix, and in that function this sequence is found in the Erzya past objective conjugation ending 3pl.obc.3sg. *-ińźe*. How it came to refer to a second-person singular object, in Moksha generally and in Erzya only with added *-tʹ/t*, is not entirely clear (Keresztes 1999: 118). In the present tense, the *-ta-* from the 2sg.obc.1sg. endings **-tan* and **-tań* is generalized throughout all second-person object forms. Bergsland therefore considers the present in M *-tanza* to be more original than **-tanzat*, with past *-nza* the result of analogical removing of the *-ta-* element marking the present (or its replacement with the past tense suffix *-j-*). The basis for *-tanza* itself could be the 2sg.obc.1sg. in **-tan* with the addition of the 3sg.obc.3sg. suffix *-za* (Bergsland 1973: 46). E *-tanzat* is formed with an additional *-t* marking the second person; this is also found in the corresponding past ending E *-ińźitʹ* and in some Moksha varieties in an ending *-ńźäť* (Keresztes 1999: 41).

**Proto-Mordvin reconstruction**

Based on a comparison of the Erzya and Moksha forms and consideration for possible or probable origins of the various elements contained in the suffixes, I think that a full Proto-Mordvin paradigm can be reconstructed. Bergsland (1973) gives several very useful structural pre-Proto-Mordvin reconstructions, but no overview. Keresztes does not provide such a reconstruction either, since he believes that "paradigm construction was started on the grounds of a few fundamental pillars, which was, in turn, slowly and gradually extended" (Keresztes 1999: 120).

I think that the Mordvin conjugations are so similar that they can be derived from a full Proto-Mordvin paradigm. This is presented in Table 8.24. I use **A* to indicate the harmonizing vowel E *e*/*o* on etymological grounds (= PU **A*). The vowel that turns up as *a* in both Mordvin languages is represented here as the lower case **a*. It occurs in epenthesis and weakening environments (e.g., in agglutinated object pronouns), so it may go back to a reduced vowel like **ə* or **ɐ*.

Table 8.24: Proto-Mordvin verb endings and their reflexes. Secondary, analogical changes are marked in square brackets. Subject person/number and object person on the left, tense and conjugation type, including object number, at the top.

| S ↓ | pst.sbc. | pst.sg.obc. | pst.pl.obc. | prs.sbc. | prs.sg.obc. | prs.pl.obc. |
|---|---|---|---|---|---|---|
| 1sg>3. | **-jń* | **-j(a?)* | **-jń* | **-an* | **-sa(n)* | **-sajń* |
| E | *-iń* | *-i[ja]* | *-iń* | *-an* | *-sa(n)* | *-siń* |
| M | *-ń* | *-j[ńä]* | *-j[ńä]* | *-an* | *-sa[]* | *-saj[ńä]* |
| 1sg>2. | — | **-jťajń* | **-jďAjź* | — | **-tan/ń* | **-tadAjź* |
| E | | *-iťiń* | *-iďiź* | | *-tan* | *-tadiź* |
| M | | *-jťäń* | *-ďäź* | | *-ťä(ń)* | *-ťäďäź* |
| 2sg>3. | **-jť* | **-jk* | **-jť* | **-at* | **-sak* | **-sajt* |
| E | | *-ik* | *-iť* | *-at* | *-sak* | *-siť* |
| M | | *-jť* | *-jť* | *-at* | *-sak* | *-sajť* |
| 2sg>1. | — | **-jmajk* | **-jmajź* | — | **-samak* | **-samajź* |
| E | | *-imik* | *-imiź* | | *-samak* | *-samiź* |
| M | | *-majť* | *-maś[ť]* | | *-samak* | *-samaś[ť]* |
| 3sg>3. | **-ś* | **-jźA* | **-jńźA* | **-(a)j* | **-saj* | **-sajńźA* |
| E | *-ś* | *-iźe* | *-ińźe* | *-i* | *-si* | *-sińźe* |
| M | *-ś* | *-źä* | *-[]ź[ń]* | *-(a)j* | *-si* | *-si[ńə]* |
| 3sg>1. | — | **-jma(j)ń* | **-jmajź* | — | **-samań* | **-samajź* |
| E | | *-imi[m]* | *-imiź* | | *-sama[m]* | *-samiź* |
| M | | *-mań* | *-maź* | | *-samań* | *-samaź* |
| 3sg>2. | — | **-jńźA* | **-jďajź* | — | **-tanzA* | **-tadajź* |
| E | | *-ińź[iť]* | *-iďiź* | | *-tanza[t]* | *-tadiź* |
| M | | *-ńźä* | *-ďäź* | | *-tanza* | *-ťäďäź* |
| 1pl>3. | **-jmA(k)* | **-jńAk* | **-jńAk* | **-TamAk* | **-sajńAk* | **-sajńAk* |
| E | *-i[ń]ek* | *-ińek* | *-ińek* | *-Ta[n]o* | *-sińek* | *-sińek* |
| M | *-mä* | *-[ś]k* [(a)] | *-[ś]k* [(a)] | *-Tama* | *-sa[ś]k* | *-sa[ś]k* |
| 1pl>2 | — | **-jďajź* | **-jďajź* | — | **-tadajź* | **-tadajź* |
| E | | *-iďiź* | *-iďiź* | | *-tadiź* | *-tadiź* |
| M | | *-ďäź* | *-ďäź* | | *-ťäďäź* | *-ťäďäź* |

| S ↓ | pst.sbc. | pst.sg.obc. | pst.pl.obc. | prs.sbc. | prs.sg.obc. | prs.pl.obc. |
|---|---|---|---|---|---|---|
| 2pl>3. | *-*jďA*(*k*) | *-*jnk* | *-*jnk* | *-*TadAk* | *-*sajnk* | *-*sajnk* |
| E | *-*iďe* | *-*ink* | *-*ink* | *-*Tado* | *-*sink* | *-*sink* |
| M | *-*ďä* | -[*ś*]*t*′ (a) | -[*ś*]*t*′ (a) | *-*Tada* | *-*sa*[*ś*]*t*′ | *-*sa*[*ś*]*t*′ |
| 2pl>1 | — | *-*jmajź* | *-*jmajź* | — | *-*samajź* | *-*samajź* |
| E | | *-*imiź* | *-*imiź* | | *-*samiź* | *-*samiź* |
| M | | *-*maś*[*t*′] | *-*maś*[*t*′] | | *-*samaź* | *-*samaź* |
| 3pl>3. | *-*śt*′ | *-*jź* | *-*jź* | *-*(a)jt* | *-*sajź* | *-*sajź* |
| E | *-*śt*′ | *-*iź* | *-*iź* | *-*it*′ | *-*siź* | *-*siź* |
| M | *-*śt*′ | *-*ź* | *-*ź* | *-*(a)jt*′ | *-*saź* | *-*saź* |
| 3pl>1. | — | *-*jmajź* | *-*jmajź* | — | *-*samajź* | *-*samajź* |
| E | | *-*imiź* | *-*imiź* | | *-*samiź* | *-*samiź* |
| M | | *-*maź* | *-*maź* | | *-*samaś*[*t*′] | *-*samaś*[*t*′] |
| 3pl>2. | — | *-*jďajź* | *-*jďajź* | — | *-*tadajź* | *-*tadajź* |
| E | | *-*iďiź* | *-*iďiź* | | *-*tadiź* | *-*tadiź* |
| M | | *-*ďäź* | *-*ďäź* | | *-*ťäďäź* | *-*ťäďäź* |

(a) Dialectal forms of these endings in Moksha are more similar to the Erzya endings: M 1pl. dial. *-jńək* ~ E *-ńek*, M 2pl. dial. *-jəńť* ~ E *-ink*. This is interpreted by Keresztes as a preservation of the more archaic situation (Keresztes 1999: 114).

Seeing as both Erzya and Moksha show syncretism of the plural subject/object forms, this can be reconstructed for Proto-Mordvin as well. In Moksha, the endings were remade to M 1pl. *-śk*, 2pl. *-śť*, so that the Erzya endings in *-ne*/*ok* and *-nk* (with syncope from PU *-*ntAk*) are to be regarded as more original. These are matched by dialectal M *-jńək* and *-jəńť* respectively. In the subjective conjugation, we must reconstruct pst.sbc.1pl. *-*j-mA*(*k*) on the basis of M 1pl. *-mä*; E *-ińek* is analogically taken over from the plural object ending (Bereczki 1988: 327). Similarly, subjective PMd. 2pl. *-*j-dA*(*k*) forms can be reconstructed based on M 2pl. *-ďä*, E 2pl. *-ďe*. Here we have the expected counterparts to, e.g., PSaa. 1pl. *-*mēk*, 2pl. *-tēk* and Samoyed 1pl. *-*mAt*, 2pl. *-*rAt*~*-*tAt* from PU 1pl. *-*mAk*, 2pl. *-*tAk* (Zhivlov 2023: 156–157). The loss of final *-*k* in certain endings does not seem to be entirely regular, but it is observed in other places as well, cf. dialectal variation in Erzya prs.1pl *-tano* ~ *-tamo* ~ *-tamk* and prs.2pl. *-tado* ~ *tadk* (Bereczki 1988: 326).

The Erzya pst.3sg.obc.1sg. *-ja* might be innovative from archaic *-j* (as per Keresztes 1999: 28), but its exact origins are not clear (cf. the conflicting accounts of Keresztes 1999: 112; Bergsland 1973: 48–49). If the ending was *-*ja* in Proto-Mordvin already, we have to assume levelling in Moksha. A suffix M *-jä* like the one found in Erzya does exist in some Moksha varieties (see Keresztes 1999: 28).

In the second-person singular and plural object endings (**-jk* and **-jt*) regularly merged in Moksha as *-jť* by sound law, so that the Erzya differentiation between singular object *-ik* and plural object *-iť* can be reconstructed. Those Moksha varieties that differentiate the two in the same way as Erzya does are characterized by Keresztes as mixed dialects (Keresztes 1999: 176, 178).

In the 1sg.obc.3sg., I would reconstruct pst. **-jma(j)ń* and prs. **-samań* with a final *-ń* as in Moksha instead of *-m* as in Erzya, because the original shape **-mań* allows for a connection with the expected genitive-accusative personal pronoun. The final *-m* in Erzya pst.1sg.obc.3sg. *-imim*, prs.1sg.obc.3g. *-samam* may be explained as the result of assimilation. Bereczki (1988: 329) would reconstruct endingless forms instead, while neither Bergsland (1973: 47) nor Keresztes (1999: 114ff.) reconstruct anything beyond the attested variation.

The *-j-* in the present plural object endings **-sajn* and **-sajt* may have been taken over from the preterite as part of the personal endings, but alternatively, it might reflect an extension of the subjective present ending **-j* — this is especially likely in the third person singular, by which the other forms were likely influenced. Whichever is the case, the element was already present in Proto-Mordvin, as it is abundantly reflected in both Erzya and Moksha.

The pst.2sg.obc.3sg. ending can be reconstructed as identical to that of the pst.pl.obc.3sg., namely as **-j-nzA*. Erzya *-ińźiť* looks like a re-characterization of this with an added plural (or second person) element **-T*. In Moksha, the ending is as would be expected for the pst.3pl.obc.3sg. form, which is instead made up from the ending M *-źä* from the singular object ending and the plural demonstrative pronoun **ńə* to yield M *-źəń*. The expected function of **-j-nzA* as an ending of the plural objective conjugation third person singular is found in Erzya, but since it is part of the second person object ending in Erzya *-ińźiť* and found as the second person object ending in Moksha *-ńźä*, which shows a clearly re-made pst.3pl.obc.3sg. ending *-źəń*, the ending PMd. **-j-nzA* can be reconstructed as polysemic for Proto-Mordvin. Both the extension in the Erzya suffix prs.2sg.obc.3sg *-ińźiť* and the replacement in Moksha with new prs.3pl.obc.3sg. *-źəń* are then clearly motivated.

According to Keresztes (1999: 108), the objective forms with third person plural subject in E *-iź* ~ M *-ź* show a differentiation, with secondary **-j-* only in Erzya to form pre-E **-jź*, but seeing as **-jń* develops to M *-ń* with loss of **-j-*, I see no reason not to reconstruct **-jź* as the ancestor of both the Erzya and the Moksha endings. **-j-* in Moksha was preserved in pst.3sg./pl.obc.2sg. *-jť*, merged from PMd. **-jk* and **-jt*, but elsewhere Erzya *-i-* from **-j-* is not reflected by a distinct segment in Moksha, so that we can assume that it was regularly lost before voiced consonants. This means that the Proto-Mordvin paradigm can be reconstructed as more uniform than suggested by Keresztes (1999: 107–121 *passim*).

**Connection with Proto-Uralic**

As discussed above, the past tense objective conjugation appears to contain the most archaic endings. Many of the suffixes there are as expected in either Erzya or Moksha, or in both, so that also the syncretism with part of the subjective paradigm can be the result of regular sound change. This is illustrated in the table below. The plural objective forms of the first and second person were generalized in favour of the plural object forms, so the expected singular object forms are lost.

Table 8.25: Proto-Uralic verbal suffixes and their expected Proto-Mordvin reflexes. Which suffixes are reflected in Erzya (E) and/or Moksha (M) is marked with the '=' sign. Cf. Table 8.19 and Table 8.20.

| S ↓ | PU sbc. | expected in PMd. | PU sg.obc. | expected in PMd. | PU pl.obc. | expected in PMd. |
|---|---|---|---|---|---|---|
| 1sg. | *-*m* | *-*n* (= EM) | *-*mi* | *-*m* (≠) | *-*ni* | *-*n* (= E, M?) |
| 2sg. | *-*t* | *-*t* (= EM) | *-*ti* | *-*t* (≠) | *-*nti* | *-*t* (= E, M?) |
| 3sg. | *-∅ | *-∅ [pst. -*ś*] | *-*sA* | *-*zA* (= EM) | *-*nsA* | *-*nzA* (= E) |
| 1pl. | *-*mAk* | *-*mAk* (= M) | *-*mAk* | *-*mAk* (≠) | *-*nAk* | *-*nAk* (= E, arch. M) |
| 2pl. | *-*tAk* | *-*dAk* (= EM) | *-*tAk* | *-*dAk* (≠) | *-*ntAk* | *-*nk* (= E, arch. M) |
| 3pl. | *-*t* | *-*t* [pst. -*ś*-*t*] | *-*sAk* | *-*st* [→ -*ź*] | *-*nsAk* | *-*st* [→ -*ź*] |

The third person singular objective conjugation plural subject endings are all identical with the corresponding third person plural objective conjugation plural subject endings, but they correspond to the latter category as expected when compared to the plural subject endings of the Samoyed and Mansi objective conjugations. The endings that are syncretic between the subjective and objective conjugations thus largely occur where expected by regular sound law. Only the first- and second person singular show a distinct lack of direct reflexes: the nature of pst.3sg.obc.1sg. E *-ija* is unclear. The corresponding M *-jńä* probably contains an additional pronominal element **ńe*. The **-k* of the pst.3sg.obc.2sg. in **-jk*, E *-ik* ~ M *-jt'* was taken from the second-person singular imperative.

Some of the finer details of the developments of the Mordvin objective conjugation may still need to be worked out further, but I hope that this subsection provides the necessary impetus for work in this direction after Keresztes's (1999) study. The Erzya and Moksha endings have too much in common for the system to not be a Proto-Mordvin creation. In my opinion, Bergsland's (1973) analysis of the pre-Proto-Mordvin structure of these suffixes provides a more promising avenue for further research into the relation of the Mordvin objective conjugation to the objective conjugation in the eastern Uralic languages. Following Bergsland's approach, I have argued that Mordvin can be added to the same list as Samoyed and Mansi for showing the plural suffix *-n-* to mark the plural number of the object in both the possessive declension and the objective conjugation,

indicating that the connection between the suffixes of these respective paradigms is very old.

#### 8.3.2.5 Saami, Finnic, Mari and Permic

The remaining Uralic languages, Saami, Finnic, Mari and Permic, do not have a separate objective paradigm. Table 8.26 shows the possessive and "subjective" verbal endings of Proto-Saami, Finnic and Mari.

Table 8.26: Possessive and verbal endings of Proto-Saami (possessive endings from Sammallahti 1998: 69; verbal endings from Koponen 2022: 109; 3du. *-kān* as per Aikio 2022a: 18–19), Finnish (Laakso 2022a: 245–247; 2022b: 260; third person endings imported from finitized participles marked in square brackets) and (East) Mari (after Saarinen 2022: 448).

| | Proto-Saami | | | Finnish | | Mari (E) | |
|---|---|---|---|---|---|---|---|
| | poss. | pl.poss. | vb. | poss. | vb. | poss. | vb. |
| 1sg. | *-*mę* | *-*nę* | *-*m* | -*ni* (sg. *-*mi*) | -*n* | -*m* | -*m* |
| 2sg. | *-*tę* | *-*ntę* | *-*k* | -*si* (pl. *-*nsi*) | -*t* | -*t* | -*t* |
| 3sg. | *-*sē* | *-*ssę* | *-∅ | -*nsA*ʔ~-*Vn* | -*V* (*-*vi*) | -*žo*/*ö* | -∅/*š* |
| 1du. | *-*męn* | *-*nęn* | *-*męn* | — | — | — | — |
| 2du. | *-*tęn* | *-*ntęn* | *-*tęn* | — | — | — | — |
| 3du. | *-*sēn* | *-*ssēn* | *-*nęn*, *-*kān* | — | — | — | — |
| 1pl. | *-*mēk* | *-*nēk* | *-*mēk* | -*mme*(ʔ) | -*mme*(ʔ) | -*na* | -*na* |
| 2pl. | *-*tēk* | *-*ntēk* | *-*tēk* | -*nne*(ʔ) | -*tte*(ʔ) | -*da* | -*da* |
| 3pl. | *-*sēk* | *-*ssēk* | *-*k*, prt. *-*n* | -*nsA*ʔ~-*Vn* | -*vAt* | -*št* | -*t* |

Saami clearly shows the presence of the suffix *-*n*- in the plural possessive paradigm (*-*ss*- derives from *-*ns*-). This element is also reflected in a number of the Finnish forms, namely in the 1sg. -*ni* (rather than *-*mi*, which is only reflected as a relict in dialects and old texts; Laakso 2022a: 245), in the 3sg/pl. -*nsAʔ* and in the 2pl. -*nneʔ* (from *-*ntek*; rather than **-*deʔ* from *-*tek*; Kulonen 1993: 72). The Mari 1pl. -*na* in the possessive paradigm also derives from the plural *-*nAk* rather than singular *-*mAk* (Bergsland 1973: 49; Saarinen 2022: 443). Whether the presence of the same person ending in the 1pl. of the verb could constitute a direct reflex of an earlier objective conjugation in pre-Proto-Mari is more difficult to tell, since an alternation between a generalized plural possessive ending *-*na* and the more expected (subjective) verbal ending *-*ma* may also have been levelled at some point, given the general identity of most other possessive and verbal endings in this language (Bereczki 1988: 342).

Bergsland (1973: 52) suggests that one aspect of Saami verbal inflection might reflect an earlier presence of the objective endings, namely the 1sg. and 2sg. forms of the *ē*-stems; cf., e.g., the North Saami 1sg. *guottán* and 2sg. *guottát* of *guoddit* 'carry, bear' (PSaa. stem

**kuontē-*). The forms *guottán* and *guottát* contain the vowel *-á-*, which is usually only found instead of **ē* (SaaN *i*) before a PSaa. **-ę-* in the following syllable. This can be illustrated with the possessive form of, e.g., *goddi* 'deer': 1sg. *goddán* 'my deer' and 2sg. *goddát* from PSaa. **kontāmę* and **kontātę*. These forms are opposed to the accusative singular *gotti* from PSaa. **kontēm* (SaaN *-tt-* is the weak grade of *-dd-* from PSaa. **-nt-*). If the verb forms of the type *guottán* and *guottát* were derived from PSaa. **kuontē+m* and **kuontē+k*, with the normal subjective endings, they should have become SaaN ***guotti* and ***guottit* respectively.

On the other hand, hypothetical earlier objective suffixes would have yielded PSaa. **kuontā-mę* and **kuontā-tę*, whence SaaN ***guoddán*, **guoddát*. These latter forms have the correct second-syllable vowel, but the wrong consonant grade. If any reflex of an objective conjugation ending is preserved in these forms, as tentatively suggested by Bergsland, it can be only by way of early analogical influence having changed, e.g., pre-PSaa. sbc.1sg. **kuontē-m* to (phonologically irregular) **kuontā-m* on the basis of pre-PSaa. *sg.obc.1sg. **kuontā-mę* (phonologically regular). The latter form would subsequently have been lost. The unexpected vowel in these forms is explained by Korhonen (1967: 190–192) as an adoption from the third person, e.g., SaaN *guoddá* from **kuontā*, whence it spread as a present marker in the singular (Korhonen 1967: 230–233). The final vowel of certain some other Saami verb endings is also difficult to explain, such as the vowel of 3du. **-pān* in the great majority of varieties, rather than expected ***-pēn* (Korhonen 1967: 284–287), so that this suggested connection with an earlier objective conjugation may not be the correct approach.

Several other potential connections have been made in the literature between endings and these branches and the objective conjugation, which I will here summarize. I leave a detailed (re-)evaluation of the nature of these endings to future research. The past plural ending **-n* of Saami has been connected with the objective conjugation and possessive declension on account of the plural co-affix **-n-* in those paradigms, but the precise connection is difficult to understand (Korhonen 1967: 338–339, 349 with references); dialectal Mansi 3pl. *-an* might be similar, but is also poorly understood.

In the Saami third person imperatives there is an element *-s-* that has been connected with the possessive and objective conjugation endings PSaa. 3sg. **-sē*, 3du. **-sēn*, 3pl. **-sēk* (Korhonen 1967: 234, 292–293; 340–342), as would be regular from PU **-sA*, **-sAjn*, **-sAk*. A similar suffix is found in the Finnic third person imperative 3sg. **-hen* from **-sen* and 3pl. **-het* from **-set* (Viitso 1998: 111–112), cf., e.g., Fi. imp.3sg. *menköön* < **men-kö-hen*, 3pl. *menkööt* < **men-kö-het* of *mennä* 'go'. The Mari third person desiderative and imperative suffixes 3sg. *-že/o*, 3pl. *-št* correspond to this as well (cf. Bereczki 1988: 347, who compares in the first place Mordvin optatives MdE 3sg. *kundazo* 'he should take', 3pl. *kundast* 'they should take').

The final branch to be briefly discussed is Permic; the possessive and verbal endings of Komi and Udmurt are given in Table 8.27 below. In both Komi and Udmurt, the system of singular possessum vs. plural possessum (with the plural suffix *-n* as in many other branches) was simplified, but it is reconstructed for Proto-Permic to explain the *-n-* in the plural possessor endings of Komi. Modern Komi and Udmurt both use the plural suffix Komi *-jas*, Udm. *-jos* to indicate the plurality of the possession, which Csúcs (2005: 200) reconstructs as part of an unstable system that was levelled in different ways in the two branches (cf. also somewhat differently, Rédei 1988: 385).

Table 8.27: Proto-Permic, Komi Zyrian and Udmurt possessive and verbal endings (Csúcs 2005: 199, cf. also Rédei 1988: 384).

| | Proto-Permic | | | Komi Zyrian | | Udmurt | |
|---|---|---|---|---|---|---|---|
| | sg.poss. | pl.poss. | vb. | sg.poss. | vb. | sg.poss. | vb. |
| 1sg. | **-e* | **-ne* | **-∅* | *-ę(j)* | *-a* | *-e/-į* | *-∅* |
| 2sg. | **-įt* | **-nį-t* | **-n* | *-įd* | *-an* | *-ed/-įd* | *-d* |
| 3sg. | **-įs* | **-nį-s* | **-∅/-s* | *-įs* | *-ę/-s* | *-ez/-įz* | *-∅/-z* |
| 1pl. | **-mį* | **-nį-mį* | **-mį* | *-nįm* | *-am* | *-mį* | *-mį* |
| 2pl. | **-tį* | **-nį-tį* | **-tį* | *-nįd* | *-a(nnį)d*[a] | *-dį* | *-dį* |
| 3pl. | **-sį* | **-nį-sį* | **-∅/-sį* | *-nįs* | *-nį/-snį* | *-zį* | *-o/-zį* |

(a) The variant ending 2pl. *-annįd* is made from the more original ending 2pl. *-ad* in combination with the poss.2pl. *-nįd*; older Komi sources also attest to unassimilated *-adnįd* (Rédei 1988: 390).

The lack of fusion with the first person *-m-* in the Komi possessive ending 1pl. *-nįm* suggests some restructuring, similar to what we see in Mansi (cf. Ms. 1sg. *-anəm*). Due to various reductions in clusters (cf., e.g., Zhivlov 2023: 134) and apocope of final vowels, the expected paradigm for plural possessum in (pre-)Proto-Permic would probably be **-n*, **-t*, **-s*, **-nį*, **-tį*, **-sį*. This would have been almost entirely undifferentiated from the singular possessum forms **-m*, **-t*, **-s*, **-mį*, **-tį*, **-sį*. The irregularity of having **n* in the first-person forms may initially, and understandably, have been mended by simply adding the corresponding singular possessum endings (thus 1pl. **-nį+mį* = **-nįmį*).

From that point, reanalysis of **-nį-* as a plural possessum morpheme would have been quite straightforward, allowing it to spread to the second- and third-person forms as well. Word-final **-m* was lost in Proto-Permic (e.g., Csúcs 2005: 199, 205), and apparently this loss applied to both original Proto-Uralic word-final **-m* (in the ending of the subjective conjugation and the accusative, PU **-m*) and to secondary word-final **-m* after apocope (as in the possessive suffix from PU **-mi*). The **-m(i)* of the first person is only preserved in combination with some case endings, such as the illative and accusative (Rédei 1988:

384–385). According to Rédei, Udmurt also preserves verbal 1sg. *-m* in questions, principally before the question particle *-a* (Rédei 1988: 389).

The Permic third person singular verbal ending **-s* (KoZ *-s*, Udm. *-z*) is furthermore interesting, as this could be a remnant of the objective conjugation PU **-sA* (Csúcs 2005: 260–261 with further references). In the past tense in Komi, the use of third person *-s* is apparently more prevalent with transitive verbs than with intransitive verbs, which fits with its interpretation as a relic from an earlier objective conjugation (ibid.). A similar difference between the Komi and Udmurt verbal endings is found in the third person plural as well, where Komi has *-nį* in the present and *-snį* in the past and future, and Udmurt has *-o* in the present and *-zį* in the past and future. The latter would of course be the regular reflex of the possessive ending PU 3pl. **-sAk*, which may also have functioned as the corresponding objective conjugation ending.

The Komi endings look like an extension of the possessive plural suffix *-nį-* on the one hand to a form with a zero ending (perhaps expected from PU sbc.3pl. **-t*, possibly regularly lost in word-final position in Permic, as the Proto-Uralic plural **-t* has left no trace), and on the other hand to a form ending in **-s*, as expected for the counterpart of Udmurt **-zį*. The Komi forms could also have been based directly on the singular forms, but the fact that Udmurt shows a similar distribution of two 3pl. endings indicates that there were two endings in Proto-Permic already.

The Permic languages thus preserve a few endings that look like the objective conjugation suffixes in exactly those places where they would not have been lost by sound law. In the first person singular, regular sound change would have deleted both the subjective and the objective ending. This is illustrated in the following table. The vowels marking the first singular and third person forms are innovative in their function as verbal endings, as they arose after the loss of word-final consonants; they are here left out of consideration for clarity.

Table 8.28: A reconstruction of Proto-Permic with separate, though largely syncretic, subjective and objective endings. The elements in square brackets were added or altered secondarily in this scenario.

| | Proto-Permic | | Komi | | Udmurt | |
|---|---|---|---|---|---|---|
| | sbc. | obc. | prs. | fut./pst. | prs. | fut./pst. |
| 1sg. | *-∅ | *-∅ | -∅ | -∅ | -∅ | -∅ |
| 2sg. | *-*n* | *-*t* | -*n* | -[*n*] | -[*d*] | -*d* |
| 3sg. | *-∅ | *-*s* | -∅ | -*s* | -∅ | -*z* |
| 1pl. | *-*mį* | *-*mį* | -*m* | -*m* | -*mį* | -*mį* |
| 2pl. | *-*tį* | *-*tį* | -*d*[*nįd*] | -*d*[*nįd*] | -*dį* | -*dį* |
| 3pl. | *-∅ | *-*sį* | -[*nį*] | -*s*[*nį*] | -∅ | -*zį* |

Association of the objective conjugation endings with the future and the past tenses also makes sense in the light of similar phenomena in Mordvin: especially the past tense most faithfully preserves the objective conjugation, while the expected present tense 3sg. ending has developed into a (non-objective) suffix for the optative and the present tense endings were completely remade. Optative or imperative function also seems to be the only potential reflex of *-*sA* in Saami, Finnic and Mari.

#### 8.3.2.6 Proto-Uralic

As we have seen above, views on a Proto-Uralic reconstruction of the objective conjugation have often been quite pessimistic. Körtvély and Kövesi, for example, write:

> "In those Uralic languages that have a determinative [=objective] conjugation, the morphological structure of the verb forms is approximately similar, but not to the extent that the verb endings could be traced back unequivocally to the same source." (Körtvély 2005: 42)

> "In den Sprachen mit zwei Konjugationen sind die Abweichungen – sowohl in der Struktur der objektiven Konjugation als auch im Ursprung und in der Fügung der einzelnen »Bauelemente» – so gross, dass man im Zusammenhang mit ihnen nicht an eine genetische Verbinding denken kann, sie können höchstens als gewisse auf der Gleichheit der Anschauungsweise beruhende parallele sprachliche Erscheinungen betrachtet werden." (Kövesi 1973: 98)

Thus, it is acknowledged that there are parallels in morphological structure in the objective conjugations found in the different Uralic languages, while on the other hand, the reconstruction of a single, coherent paradigm purely on the basis of the attested suffixes is impossible. However, the points of agreement between the attested suffixes may just be enough to reconstruct the conjugation's original character, from which its formal features can be abstracted.

From the structural comparison of the different Uralic languages that have an objective conjugation, it is clear why at least the third-person singular form is considered to go back to Proto-Uralic: every branch with an objective conjugation shows reflexes of the suffix *-sA*, and even some other Uralic languages without an objective conjugation have possible traces of it, chiefly Permic. However, there is also another common element to be found in the plural object suffixes, with a functional dimension of the objective conjugation in all branches except Hungarian. The plural object is marked with an *-*n*- in Samoyed (visible in first person forms only), Mansi (in the entire paradigm, partly recreated, but seemingly old in third person) and Mordvin (visible in 3sg., 1pl. and 2pl.). In Hungarian it may, formally speaking, have been preserved in the 2sg. ending *-d*,

although this does not have plural reference and other interpretations of this *-d* are also possible (see 8.3.3.3). This particular way of marking the plural is shared with the possessive declension, where **-n-* indicates the number of the possessum. The table below gives just those suffixes of singular and plural object forms in each branch with an objective conjugation that show (potential traces of) a difference between singular and plural objects.

Table 8.29: Mordvin, Hungarian, schematic Ob-Ugric and Proto-Samoyed objective conjugation endings. On the left in each cell are given the 3sg. object forms, to the right the 3pl. object forms. The subject person/number is in the far-left column. Secondary additions are marked in square brackets, while forms showing the contrast between singular -∅- and plural **-n-* are given in bold.

| S ↓ | Mordvin | Hungarian | Ob-Ugric | Samoyed |
|---|---|---|---|---|
| 1sg. | -[∅] : -*n* | -*m* : – | *-***m*** : *-***n***[*sA*-*m*] | *-***mə*** : -[*j*]***nə*** |
| 2sg. | -[*k*] : -*t* | -*d* ← : – | *-[*n*] : *-[*nsA*-*n*] | *-*rə* : -[*j*]*tə* |
| 3sg. | -***zA*** : -***nzA*** | -*ja* : – | *-***sA*** : *-***nsA*** | *-*tA* : -[*j*]*tA* |
| 1pl. | -[*n*]*Ak* : -***nAk*** | -[*j*]*uk* : – | — | *-***mA***[*t*] : -[*j*]***nA***[*t*] |
| 2pl. | -[*n*]*k* : -***nk*** | -[*já*]*tok* : – | *-*tA*[*n*] : *-*ntA*[*n*]? | *-*rA*[*t*] : -[*j*]*tA*[*t*] |
| 3pl. | — | -*ják* : – | *-*sA*[*n*]? : *-*nsA*[*n*]? | — |

To reiterate, the two common elements that appear in almost every branch are (1) the third-person singular suffix **-sA* (in all), and (2) the possessive plural suffix **-n-* (at least in Mordvin, Mansi and Samoyed). The former is usually given as the only actually reconstructible aspect of the objective conjugation for Proto-Uralic (e.g., Keresztes 1999, Körtvély 2005; Abondolo 1998, etc.), while I have not come across particular mention of the latter.

As briefly summarized above, there are two main theories on the origins of the objective conjugation, one interpreting it as the product of suffixed object pronouns, and the other connecting it with the possessive declension. The former interpretation is closest to the Tocharian situation (cf. 8.2.2 and Peyrot 2019a), but it has a number of flaws. It is often assumed that the ending **-sA* could be in origin an anaphoric object pronoun **sA*, suffixed to the unmarked 3sg. present; e.g., **kala-m kanta-∅* 's/he carries a fish-ACC' next to **kanta-∅-sa* 's/he carries it', where the object **kala-m* is replaced by an anaphoric pronoun in the second example (e.g., Abondolo 1998: 29–30, Rédei 1998/99; Keresztes 1999: 107, Honti 1998/99: 117). The assumed anaphoric object pronoun **-sA* appears to be only accusative in value, not in form, since it does not show any accusative suffix. If there was an accusative suffix at the end of this pronoun originally, it has been lost. Rédei (1998/99) argues that loss is impossible, because no final accusative **-m* or **-t* could have been lost by sound law in the majority of the relevant languages. According to Rédei, the absence of an accusative suffix can be understood because no further reference to the

object was necessary. The use of the pronoun suffix by itself supposedly already indicated the presence of a definite object.

According to Honti, on the other hand, there was actually an earlier accusative suffix, and it would be the ultimate cause of the difference between the objective and subjective conjugation suffixes (Honti 1998/99: 116–117); e.g., 1sg.sbc. *-m* vs. 1sg.obc. **-mi* as if from **-mV* and **-mVt* respectively.[104] The phonologically unexpected loss of this accusative marker would simply be because grammatical morphemes may follow different sound laws than "normal" words (Honti ibid.).

Whether or not the original anaphoric pronoun was explicitly marked with an accusative suffix, the creation of the third person singular objective ending singular is intuitively easy to understand in this scenario. Over time, the combination of an endingless third person singular verb form and object pronoun were univerbated, creating the first beginnings of the objective conjugation as we know it. The first and second person endings, as well as the marking of the dual and plural objects, would have been created afterwards. The details of this process are very vague, however. As also pointed out by Havas (2004), an immediately apparent issue with the origin of the objective conjugation suffixes as erstwhile object pronouns is the fact that the first and second person suffixes (**-mi*, **-ti*, etc.) look like the corresponding pronouns as well, but never refer to first or second person objects. Instead, they invariably index first or second person subjects/agents and third person objects.

In an attempt to account for this, Keresztes (1999: 105), discussing the origins of the Mordvin objective conjugation specifically, reconstructs some considerable semantic ambiguity: "[...] **kunda-sV* could mean 'he/she catches ~ he/she catches him/her/it; he/she/it is caught (passive meaning)', **kunda-mV* 'I catch ~ he/she catches me; I am caught (passive meaning)', **kunda-tV* 'you (Sg) catch ~ he/she catches you (Sg); you (Sg) are caught (passive meaning)'." However, it is not at all clear why the forms should have such a wide range of possible interpretations, and why only the active interpretation is ever preserved

There are also other difficulties with the idea that objective conjugation endings constitute suffixed object pronouns. The plural marker **-n-*, as well as dual markers containing the dual **-k-* (preserved in Mansi and Khanty, reinstated in Samoyed, but lost in Mordvin and Hungarian) appear before the objective endings that index the agent, which is not the place where we would expect them if the object suffix like **-sA* were, etymologically, a suffixed object. The plural of a suffixed object like that should be ***-sA-*

[104] The 3sg. objective conjugation suffix would originally also have been different from the corresponding possessive suffix by virtue of a different ending, namely acc. **-sVt* vs. gen. **-sVn*, according to Honti (1998/99: 117). However, this is rather an attempt at internal reconstruction, since neither *-t* nor *-n* are actually found in either of the endings.

*t* or **-*sA-j*, or the like, not *-*nsA*. Since a formal identity in the structure of the possessive declension and objective conjugation paradigms can (and in my opinion, should) be reconstructed for Proto-Uralic on the basis of Mordvin, Mansi and Samoyed, any functional derivation of these suffixes in the pre-Proto-Uralic period should depart from that formal identity and explain how it arose and why these two categories are so intimately connected to each another (see also 8.3.3.5 on this question).

Table 8.30: The Proto-Uralic subjective and objective conjugations. Most endings of the objective conjugation cannot be reconstructed directly based on the endings preserved in the daughter languages and are instead extrapolated based on the general structure.

| S ↓ | sbc. | 3sg.obc. | 3du.obc. | 3pl.obc |
|---|---|---|---|---|
| 1sg. | *-*m* | *-*mi* | *-*kV-mi* | *-*ni* (< *-*n-mi*) |
| 2sg. | *-*n*/*t* | *-*ti* | *-*kV-ti* | *-*n-ti* |
| 3sg. | *-∅ | *-*sA* | *-*kV-sA* | *-*n-sA* |
| 1du. | *-*mijn* | *-*mijn* | *-*kV-mijn* | *-*nijn* (< *-*n-mijn*) |
| 2du. | *-*tijn* | *-*tijn* | *-*kV-tijn* | *-*n-tijn* |
| 3du. | *-*k* | *-*sAjn* | *-*kV-sAjn* | *-*n-sAjn* |
| 1pl. | *-*mAk* | *-*mAk* | *-*kV-mAk* | *-*nAk* (< *-*n-mAk*) |
| 2pl. | *-*tAk* | *-*tAk* | *-*kV-tAk* | *-*n-tAk* |
| 3pl. | *-*t* | *-*sAk* | *-*kV-sAk* | *-*n-sAk* |

### 8.3.3 Functional reconstruction of the objective conjugation

We have now had ample illustrations of what the objective conjugation looks like in the various Uralic branches, and so we come to its function. The function of objective conjugation in Samoyed, Mansi and Khanty is centred around the same pragmatic feature: topicality. If an object is topical (either the primary or secondary topic of a sentence), the verb will agree with it using verbal endings from the objective conjugation. In Mordvin and Modern Hungarian the objective conjugation is used with definite objects instead (cf. below), but even in earlier Hungarian, the use of the objective conjugation did not align with definiteness per se (Marcantonio 1985). In this section I will summarize the functions of the objective conjugation in each branch where it occurs, based on the available literature on the topic.

#### 8.3.3.1 Samoyed

The functions of the objective conjugation are fairly constant throughout the Samoyed languages, although there are still a few gaps in our knowledge of some specifics. In this subsection, I will mainly focus on the situation in Tundra Nenets, as the use of the objective conjugation has been researched quite extensively by especially Irina

Nikolaeva, and the same general patterns as in Tundra Nenets seem to be found in the other Samoyed languages.

Dalrymple and Nikolaeva (2011) provide a very clear description and overview of the use of the Tundra Nenets objective conjugation in relation to the topicality of the object. I will here reproduce some of their examples to help illustrate the functions of the objective conjugation. Dalrymple and Nikolaeva explain that whenever there is no particular pre-established background to a statement, or if the background includes only information about the subject and not the object, the objective conjugation cannot be used. Thus, the sentence in (19) with the verb in the subjective conjugation can answer questions of the types "what happened?", "what did a/the man do?" or "what did a/the man kill?"

19. Tundra Nenets (Nikolaeva & Dalrymple 2011: 128; glosses adapted)
    *xasawa* *ti-m* *xada-°-∅* / **xada-°-da*
    man reindeer-ACC kill-AOR-SBC.3SG kill-AOR-SG.OBC.3SG
    'A/the man killed a/the reindeer.'

It cannot, however, answer the question "what did a/the man do to the/a reindeer?" Such a question already presupposes the existence of a reindeer, and in that case, the verb has to be in the objective conjugation, as in (20).

20. Tundra Nenets (Nikolaeva & Dalrymple 2011: 128; glosses adapted)
    *xasawa* *ti-m* *xada-°-da* / **xada-°-∅*
    man reindeer-ACC kill-AOR-SG.OBC.3SG kill-AOR-SBC.3SG
    'A/the man killed a/the reindeer.'

Both the subject and object may remain unexpressed in such a case a well, as they are both readily available from the context and referenced with the suffix on the verb; cf. (21).

21. Tundra Nenets (Nikolaeva & Dalrymple 2011: 128; glosses adapted)
    *xada-°-da*
    kill-AOR-SG.OBC.3SG
    'He killed it.'

The use of the objective conjugation in Tundra Nenets is not dependent on the status of the subject as a topic, only that of the object (Dalrymple & Nikolaeva 2011: 129). One significant exception to the pattern described so far pertains to first- and second-person pronouns: those are never accompanied by a verb in the objective conjugation, even if they constitute a topical object (Dalrymple & Nikolaeva 2011: 130). This restriction is

shared by other Samoyed languages (e.g., Wagner-Nagy 2019: 375–376 on Nganasan; Budzisch 2021: 146 on Selkup).

At least for Tundra Nenets, Nikolaeva points out that there is also an implication that the action described with a verb in the objective conjugation has also reached its completion, i.e., a connection with perfective aspect, as in (22) (Nikolaeva 2014: 207). I have not seen mention of a similar situation in the other Samoyed languages, but it does find a parallel in Mordvin (see 8.3.3.4).

22. Tundra Nenets (Nikolaeva 2014: 207; glosses adapted)

a. *nyísya-myi* *ŋəno-m* *syerta-°-da*?
father-SG.POSS.1SG boat-ACC make-AOR-SG.OBC.3SG
'Has my father finished making the boat?'

b. *nyísya-myi* *ŋəno-m* *syerta-°-∅*?
father-SG.POSS.1SG boat-ACC make-AOR-SBC.3SG
'Did my father make a boat?'

While the objective conjugation is closely connected with the direct object, there is no associations between the objective conjugation and the indirect object (Dalrymple & Nikolaeva 2011: 132). The indirect object of a ditransitive verb like 'give' is in the dative/lative (a local case), which cannot agree with the objective conjugation. Only the direct object in the sentence below determines whether the verb is in the subjective or objective conjugation

23. Tundra Nenets (Nikolaeva 2014: 272; glosses adapted)

*Petya* *Maša-n°h* *ti-m* *myiq-ŋa-∅* / *myiq-ŋa-da*
PN PN-LAT reindeer-ACC give-AOR-SBC.3SG give-AOR-SG.OBC.3SG
'Petya gave a/the reindeer to Masha.'

The Enets languages make use of the objective conjugation in a similar fashion. Its use is described by Tereščenko (1973) and Sorokina (2010) as connected to whether the "logical stress" falls on the object or not, a general description with which Khanina and Shluinsky broadly agree (Khanina & Shluinsky 2015; see now also Khanina & Shluinsky 2024).

Khanina and Shluinsky have specifically investigated the use of the objective conjugation along several parameters, principally its associations with word order and definiteness, and show that standard (SOV) word order and indefinite objects are correlated with the subjective conjugation, while non-standard word order and definite objects are correlated with the objective conjugation. These correlations are especially strong if they occur together, so that both definite and indefinite nouns in standard word order are less likely to be associated with the objective conjugation than they would be

in non-standard word order, or if other constituents intervene between the object and the verb (Khanina & Shluinsky 2015: 408). This is shown in (24) for Forest Enets and in (25) for Tundra Enets.

24. Forest Enets (Khanina & Shluinsky 2015: 404; glosses adapted)[105]
    a. *tɔrse* *ɔburu* *kɔmita-∅*
       such thing like-SBC.3SG
       'He likes such a thing.'
    b. *tɔrse* *ɔburu* *ŋulʲ* *kɔmita-đa*
       such thing much like-SG.OBC.3SG.
       'He likes such a thing very much.'

25. Tundra Enets (Khanina & Shluinsky 2015: 404–405; glosses adapted)[106]
    a. *tʃike* *anʲ* *pe* *mua-∅*
       this and wood do-SBC.3SG
       'She put the wood down again'
    b. *dʲaudʲa* *sɔrja* *sudʲeʔ* *mu-da-đa*
       flour bowl inside do-FUT-SG.OBC.3SG
       'She will put the flour in a bowl'

Wagner-Nagy describes for Nganasan that the use of the objective conjugation is not determined by features like definiteness or animacy of the object (Wagner-Nagy 2019: 230). Rather it can be used to refer anaphorically to an object that was previously mentioned (Wagner-Nagy 2019: 338–339). Pronominal objects do not co-occur with a verb in the objective conjugation (Wagner-Nagy 2019: 339).

26. Nganasan (Wagner-Nagy 2019: 339; glosses adapted)
    *śelujkʲa-ðį* *küüü-ʔa-∅* *koða-ʔa-tu*
    naked.horn.deer-SG.POSS.3SG die-AOR.PFV-SBC.3SG kill-AOR.PFV-SG.OBC.3SG
    'The reindeer with naked horns is dead, he killed it.'

In Selkup, the precise relationship between the objective conjugation and the status of the object as either (in)definite or (non)topical has not yet been entirely elucidated, and various scholars have described the phenomenon in different and sometimes

[105] The original translations in Russian are, for a., 'Такую вещь он любит', and for b., 'Такую вещь он очень любит'.

[106] The original translations in Russian are, for a., 'Она опять положила дрова', and for b., 'Она муку в миску положит'.

contradictory ways (see Budzisch 2021: 145–147 with references). The relation between subjective conjugation and an indefinite object does seem clear, and conversely, any object that is definite requires the objective conjugation to be used. This means that only indefinite objects can be combined with either conjugation (Budzisch 2021: 147–148; 151–152). However, there are also a number of attestations of intransitive verbs inflected in the objective conjugation, which have not yet received a satisfactory explanation (Budzisch 2021: 149). I will leave the intricacies of the Selkup objective conjugation aside for now.

#### 8.3.3.2 Mansi and Khanty

The use of the objective conjugation in the Ob-Ugric languages largely agrees with that of Samoyed, and has been described as centred around the topicality of the object (Nikolaeva 1999a, 1999b, 2001; Skribnik 2001; Virtanen 2014). The objective conjugation is not used when the focus is on the object, e.g., when it is contrastive or newly introduced, or in questions (e.g., Nikolaeva 1999a: 64ff. with a detailed description for Khanty; cf. also example (19) for Tundra Nenets above). If an object is the secondary topic of a sentence, the verb is generally used in the objective conjugation. If the object is the primary topic, however, another strategy can be used, namely passivization (Kulonen 1989, Skribnik 2001).

One additional difference between the use of the objective conjugation in Mansi and Khanty as opposed to that in Samoyed concerns its use in combination with personal pronouns. In Samoyed, object personal pronouns (especially of the first and second person) always require the verb to be in the subjective conjugation, even if the first and second person constitute a topical entity in the sentence (cf. above). In Ob-Ugric, on the other hand, object personal pronouns of the first- and second-person can agree with verbs in the objective conjugation (e.g., Nikolaeva 1999a: 65; Virtanen 2014: 406). There are no special endings in the objective conjugation to mark this agreement with non-third person objects, however.

In both Mansi and Khanty there are some further particular features of the objective conjugation. Eastern Mansi combines its use of the objective conjugation with an accusative in *-m*. This accusative is obligatory with verbs in the objective conjugation in this variety of Mansi (Virtanen 2014: 399). The variation in object marking is based on the topicality of the object (Virtanen 2014: 392–393). Some example sentences where the accusative *-m* and the objective conjugation are combined are given in (27) and (28).

27. Eastern Mansi (Virtanen 2014: 406; glosses adapted)

| *sos-mø* | *uus* | *köält-øs-tø* |
|---|---|---|
| moose-ACC | again | frighten-PST-SG.OBC.3SG |

'He frightened the moose again.'

28. Eastern Mansi (Virtanen 2014: 406; glosses adapted)
*ään-øm* *jål=ääl-ääl-øn*
1SG-SG.POSS.1SG down=kill-IMP-SG.OBC.2SG
'Kill me!'

The accusative ending *-m* is not used in combination with the subjective conjugation (Virtanen 2014: 407). This is shown by Virtanen with various examples, of which one is reproduced here as (29).

29. Eastern Mansi (Virtanen 2014: 407; glosses adapted)
*kom* *jowt-nyõõl* *wø-s-∅*
man bow-arrow take-PST-SBC.3SG
'The man took a bow and arrow'

In the last example sentence, the bow and arrow had not been previously mentioned, and *they* are thus non-topical. Virtanen describes the connection between the objective conjugation and topicality as follows: "the subjective conjugation is for focal constituents, the objective conjugation for topical ones. In the latter one, zero anaphora express high topicality" (Virtanen 2014: 403). Virtanen reports that in the data on Eastern Mansi, 35% of sentences with a verb in the objective conjugation do not contain an overt nominal object. In such cases, the object has usually been mentioned recently and does not need to be activated again, such as in example (30), where a hen, the object, has been caught and killed in the preceding narrative (Virtanen 2014: 403).

30. Eastern Mansi (Virtanen 2014: 403; glosses adapted)
*kom* *juw* *tåt-øs-tø,* *pon-øs-tø* *ølø*
man home bring-PST-SG.OBC.3SG put-PST-SG.OBC.3SG away
'Th[e] man brought it (i.e., the hen) home and put it aside.'

Second person objects can also be omitted, although their omission is less common than that of third person objects. The fact that the objective conjugation then agrees with a covert second person object instead of a third person object is not marked any different (Virtanen 2014: 404–405).

The use of the objective conjugation in Northern Mansi, which lacks the accusative in *-m* completely, can be understood in similar terms as that of Eastern Mansi in all other regards (Skribnik 2001). Skribnik further addresses a phenomenon whereby datives can be promoted to object function, in which case they agree with the verb in the objective conjugation. Example (31), reproduced from Skribnik (2001: 228), illustrates this.

31. Northern Mansi (Skribnik 2001: 228; glosses adapted)

a. *am* *tawe-n* *mōjt* *mōjt-eɣ-um*
1SG 3SG-DAT tale tell-PRS-SBC.1SG
'I tell him a tale.' (< 'What do you do for him?')

b. *am* *tawe* *mōjt-əl* *mōjt-i-lum*
1SG 3SG.ACC tale-INS tell-PRS-SG.OBC.1SG.
'I tell him a tale.' (< 'What do you tell him?')

As a part of this same phenomenon, instruments can be promoted in the same way (Skribnik 2001: 229). Virtanen describes these constructions, and their interplay with the passive, for Eastern Mansi as well. The main factor affecting the choice between the different possible constructions is the topicality of the various participants (Virtanen 2012).

In her description of Northern Khanty, Nikolaeva (1999a: 65) points out that there are sentence pairs where the only difference is the conjugation of the verb (subjective or objective), while the object itself remains formally the same. This shows that the choice of the verbal inflection is determined by something other than formal characteristics of the object. Nikolaeva explains that focused objects do not trigger agreement, while topical objects do, in the same way as in Mansi. Focus often coincides with the focus position directly in front of the verb, but it need not, indicating that agreement does not strictly depend on the position of the object in the sentence (Dalrymple & Nikolaeva 2011: 144). Khanty, like most Mansi varieties, does not preserve the accusative in *-m*, so that the object remains unmarked. This is illustrated with the example sentences in (32) below.

32. Northern Khanty (Nikolaeva 1999a: 64; glosses adapted)

a. *ma* *tam* *kalaŋ* *we:l-s-əm*
1SG DEM reindeer kill-PST-SBC.1SG
'I killed this reindeer.'

b. *ma* *tam* *kalaŋ* *we:l-s-e:m*
1SG DEM reindeer kill-PST-SG.OBC.1SG
'I killed this reindeer.'

c. *ma* *tam* *kalaŋ* *we:l-s-ələm*
1SG DEM reindeer kill- PST-PL.OBC.1SG
'I killed these reindeer.'

Focus objects are opposed to topical objects, which are accompanied by a verb in the objective conjugation. Nikolaeva reports that, in a corpus of Khanty folklore texts, 89%

of transitive clauses with a subjective verb have an object referent that is newly introduced (focus), while the object of 83% of the objective verbs is already established in the discourse (non-focus). Some objects that are technically newly introduced may also trigger objective agreement, but in this case, they are related to another established noun in the discourse: for instance, possessed nouns are related to their possessor by inference, as in example (33) below (Nikolaeva 1999a: 75).

33. Northern Khanty (Nikolaeva 1999a: 75; glosses adapted)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *loːw-əl* | *wu-s-li* | *pa* | *oːx-əl* | *joxəs* | | | |
| horse-SG.POSS.3SG | take-PST-SG.OBC.3SG | and | head-SG.POSS.3SG | backwards | | | |
| *kir-s-əlli* | *pa* | *paj-əl* | *oːlŋəl* | *kir-s-əlli* | | | |
| harness-PST-SG.OBC.3SG | and | tail-SG.POSS.3SG | forwards | harness-PST-SG.OBC.3SG | | | |

'He took the horse and harnessed its head backwards and its tail forwards.'

If a possessor of an object is topical, the verb has to be in the objective conjugation as well, despite the object noun itself not being topical (Nikolaeva 1999b: 346).

34. Northern Khanty (Nikolaeva 1999b: 346; glosses adapted)

| | | | | |
|---|---|---|---|---|
| *Juwan* | *motta* | *xot-əl* | *kăśalə-s-em* | / **kăśalə-s-əm* |
| John | before | house-SG.POSS.3SG | see-PST-SG.OBC.1SG | see-PST-SBC.1SG |

'I saw John's house before'

Mansi is also sensitive to such considerations, and entities that are associated to the topic by a possessive relationship are considered eligible for topic status as well upon their first introduction, so that they can immediately combine with a verb in the objective conjugation or the passive (Skribnik 2001: 231).

#### 8.3.3.3 Hungarian

In Modern Hungarian, the objective conjugation is used in combination with definite objects. As such, it is generally called the definite conjugation in this language. This first example is taken from Marcantonio (1985: 268), and illustrates the most basic dichotomy in the use of the subjective and objective conjugations in Hungarian. It is required in the first sentence on account of the possessive suffix and the accompanying definite pronoun.

35. Hungarian (Marcantonio 1985: 268; glosses adapted)

a. *az* *újság-om-at* *kér-i*
DEF newspaper-SG.POSS.1SG-ACC want-OBC.3SG
'He wants my newspaper'

b. *egy* *magyar* *újság-ot* *kér-∅*
INDEF Hungarian newspaper-ACC want-SBC.3SG
'He wants a Hungarian newspaper'

A definite pronominal object can also be omitted from a sentence, as its presence is marked on the verb ending in the objective conjugation already (36a). First- and second-person pronouns in do not trigger object agreement with the verb, with the exception of the special suffix *-lek*/*lak* used specifically in the case of a first-person subject acting upon a second person object, as in (36b). However, it is also possible to omit first- and second-person pronouns in combination with the subjective conjugation, as in examples (36c) and (36d) (Janda, Laakso & Metslang 2022: 896).

36. Hungarian (a., c., d. from Janda, Laakso & Metslang 2022: 896; glosses adapted; b. constructed after the same pattern)

a. *azért* *sír-∅* *mert* *nem* *szeret-i* *senki*
for.that cry-SBC.3SG because NEG love-OBC.3SG nobody
'He/she is crying because nobody loves him/her'

b. *azért* *sír-sz* *mert* *nem* *szeret-lek*
for.that cry-SBC.2SG because NEG love-2SG.OBC.1SG
'You are crying because I do not love you.'

c. *azért* *sír-ok* *mert* *nem* *szeret-∅* *senki*
for.that cry-SBC.1SG because NEG love-SBC.3SG nobody
'I am crying because nobody loves me'

d. *azért* *sír-sz* *mert* *nem* *szeret-∅* *senki*
for.that cry-SBC.2SG because NEG love-SBC.3SG nobody
'You are crying because nobody loves you'

The obligatory connection between the objective conjugation and the (overt) definiteness of the object found in Modern Hungarian is not reflected in the Old Hungarian material, as shown by Marcantonio (1985) and É. Kiss (2010/11) with examples like the following. In (37) the object is possessed, and therefore morphologically definite, but in spite of this, the verb is in the subjective conjugation.

37. Early Hungarian (Marcantonio 1985: 290; glosses adapted)

| *es* | *tarsibeli* | *kiral-ok* | *es* | *ziget-ek* | *aiandok-ai-t* | *aiandokoz-nak* |
|---|---|---|---|---|---|---|
| and | friendly | king-PL | and | islander-PL | gift-PL.POSS.3-ACC | give-SBC.3PL |

'... and friendly kings and islanders give their gifts.'

Conversely, the objective conjugation is also attested with objects that do not take a definite article, as in the following example (38), where the lance had already appeared in the preceding text as well.

38. Early Hungarian (Marcantonio 1985: 289; glosses adapted)

| *sebes* | *számszereǵ-et* | *raǵad-t-a* | *vala* |
|---|---|---|---|
| quick | lance-ACC | grasp-PST-OBC.3SG | be.3SG |

'He gripped the quick lance.'

Marcantonio's analysis of the Old Hungarian objective conjugation as originally agreeing with topical objects is generally supported by É. Kiss (2010/11). However, while Marcantonio separates the formation of the Hungarian objective conjugation from that of the other Uralic languages and posits a new background for the formation of the entire conjugation, É. Kiss further points to the agreements in the basic structure of the 3sg. suffix (cf. 8.3.2) and the syntax with regards to topicality as described for Khanty by Nikolaeva (cf. 8.3.3.2) to argue for a common formation of the beginnings of the objective conjugation with at least Ob-Ugric and Samoyed (É. Kiss 2010/11: 135–137).

#### 8.3.3.4 Mordvin

In Mordvin the objective conjugation can only be used in combination with definite objects (e.g., Alhoniemi 1996: 145), and it correlates with perfective aspect (Bernhardt 2020: 21). The definiteness of the object is also marked on the object noun itself, which will stand in the genitive (or genitive-accusative) case *-ń* if definite, and remain unmarked if indefinite. The genitive suffix is typically enlarged by either a further definite marker or a possessive ending when applied to regular nouns; especially proper nouns and most pronouns do not receive this suffix, and it can also be absent in certain folklore texts (Bernhardt 2020: 22–23, with the examples reproduced here in (39)).

39. Erzya Mordvin (Bernhardt 2020: 22–23; glosses adapted)

a.

| *ram-i-ń* | *alaša* |
|---|---|
| buy-PST-SBC.1SG | horse |

'I bought a horse.'

b. *ram-i-ja* *alaša-ńť*
buy-PST-OBC.3SG.1SG horse-GEN.DEF
'I bought the horse.'

c. *ram-i-ja* *alaša-nzo*
buy-PST-OBC.3SG.1SG horse-GEN.SG.POSS.3SG
'I bought his/her horse.'

d. *ńe-i-ja* *Maśa-ń*
see-PST-OBC.3SG.1SG PN-GEN
'I saw Masha.'

Not all objects that appear in the genitive and marked with the nominal definite marker require the verb to be in the objective conjugation. The subjective conjugation can be used to allow for an imperfective interpretation. Use of the objective conjugation makes only a perfective interpretation possible (Bernhardt 2020: 23–24, with the examples in (40)).

40. Erzya Mordvin (Bernhardt 2020: 23–24; glosses adapted)

a. *veď-eńť* *kand-i-ń*
water-GEN.DEF carry-PST-SBC.1SG
'I was carrying the water / I carried the water.'

b. *veď-eńť* *kand-i-ja*
water-GEN.DEF carry-PST-OBC.3SG.1SG
'I carried the water.'

To describe imperfective events with definite objects, the object may be placed in the inessive case. In such instances, it is combined with the subjective conjugation (Bernhardt 2020: 24).

The link between perfectivity and the objective conjugation is not absolute, however. For example, certain verbs describing cognition or perception, like E *ńejems*, M *ńäjəms* 'see' and EM *sodams* 'know' are often used in the objective conjugation, irrespective or their aspectual interpretation, and especially in Moksha, frequentative verbs may be inflected in the objective conjugation without losing their imperfective interpretation (Alhoniemi 1996: 147–151).

#### 8.3.3.5 Proto-Uralic

Samoyed, Mansi, Khanty and Old Hungarian share their use of the objective conjugation in combination with topical objects, as opposed to focus objects. Modern Hungarian and Mordvin, on the other hand, use their objective conjugation in combination with

definite objects. Topicality and definiteness are closely linked conceptually, so that it is easy to imagine the one use changing into the other. Since both Modern Hungarian and Mordvin are among the only Uralic languages to have encoded definiteness in their grammar, and seeing as Old Hungarian shows an association of the objective conjugation with topicality, it is preferable to reconstruct the topic-associated function still found in the Siberian Uralic languages as original.

A minor point of disagreement found between Samoyed and Hungarian on the one hand and Mansi and Khanty on the other is the compatibility of the objective conjugation with personal pronouns. Samoyed and Hungarian disallow the use of objective conjugation in combination with first and second person pronouns, while in Mansi and Khanty such pronouns do trigger objective agreement on the verb. Hungarian here agrees with Samoyed, and those two have probably developed independently from each other for a long time, certainly, much longer than Mansi and Khanty, who are known to have been in close contact for hundreds of years. It is therefore more likely that Samoyed and Hungarian together preserve the original situation.[107] This means that in Proto-Uralic, as in Samoyed and Hungarian, object pronouns of the first and second person would not have triggered the use of the objective conjugation, even though they are conceptually topical/definite as referents to speech act participants.

Dalrymple and Nikolaeva reconstruct the opposite direction of development, arguing that first- and second-person pronouns are more likely to be primary topics than secondary topics, so that they should, at some stage, have been deemed to be out of line with third person objects, leading to a restriction of the use of the objective conjugation (Dalrymple & Nikolaeva 2011: 200–201). However, the fact that first- and second-person objects can occur as secondary topics with the objective conjugation in Mansi and Khanty would seem to suggest that the appropriate situations do exist.

Dalrymple and Nikolaeva (2011: 200) also adduce the fact that Mansi and Khanty often prefer to create a passive construction when a first- or second-person object is the primary topic, promoting it to the subject-role, but such constructions are not used in the same way in the Samoyed languages. One situation in which object agreement with a first- or second-person might thus be expected based on topicality has thus been largely removed in the Ob-Ugric languages, while an awkward gap in topic agreement is maintained in Samoyed and Hungarian. I think, therefore, that the original function of

[107] Coppock and Wechsler argue for such a reconstruction in a very different way, based on an idea of feature loss in various stages. Mansi and Khanty would have lost the earlier third-person selecting feature of the objective conjugation that is still active in Samoyed and Hungarian (Coppock & Wechsler 2010). Their entire analysis is concerned purely with functional considerations, however, without any reference to form or arguments from phylogeny.

the objective conjugation in Proto-Uralic would have been to index topical third person objects only.

An original connection with topical objects further accords well with the conclusions in subsection 8.3.2.6 above, that the objective conjugation endings are closely allied with the endings of the possessive declension: the genitive (possessive) is also used to mark the agents of passively used participles, which are often finitized in Uralic (cf. 7.3 *passim*), and the use of the passive voice is cross-linguistically connected with the topicality of the object (e.g., Cooreman 1987, Givón 1994, Imamura 2022). The use of (accusative) objects in combination with verbs in the objective conjugation can be seen as the result of a reinterpretation of the construction from passive to active, which may have taken place in Proto-Uralic already. Further research is required to work out the details.

With regards to the function of the objective conjugation in pre-Proto-Samoyed, it is likely that the objective conjugation already marked topical objects and was opposed to the subjective conjugation, which was employed in those cases where the object constituted a non-topical or focus constituent; and of course, when the subject was the only participant to an intransitive verb action. Functionally, this is the objective conjugation that should be compared with Tocharian.

## 8.4 Comparison between Tocharian and Uralic

In this final section of chapter 8, we will compare the verbal object marking found in Tocharian and Uralic. After our detailed look at both language families, we will be able to better evaluate the comparison put forward by Peyrot (2019a), according to which the Tocharian pronoun suffixes may have been innovated under the influence of the objective conjugation found in early Samoyed.

There are various immediate structural discrepancies between the respective object-marking strategies. The Tocharian pronoun suffixes are clearly extra elements that are added after the normal verbal endings agreeing with the subject. In the Uralic objective conjugation, the suffixes are not "stacked" in the same way. Unlike most of Uralic, Tocharian is furthermore not limited to third-person reference, with its specific first- and second person suffixes in the singular. In Uralic, an objective conjugation with special first- and second-person object suffixes is only found in Mordvin, where it has been innovated as an expansion of the objective conjugation with the addition of personal object pronouns. Some of these differences are illustrated in Table 8.31 below with some verbal forms from Tocharian B and Tundra Nenets.

Table 8.31: TB *nek-* 'destroy' past tense forms, and NeT *xada-* 'kill' aorist forms with reference to the object. Singular only.

| | Tocharian B | | | Tundra Nenets | | |
|---|---|---|---|---|---|---|
| S | 1sg. O | 2sg. O | 3sg. O | 1sg. O | 2sg. O | 3sg. O |
| 1sg. | – | *nekwā-c** | *nekwā-ne** | – | – | *xadaə-w°* |
| 2sg. | *nekasta-ñ** | – | *nekasta-ne** | – | – | *xadaə-r°* |
| 3sg. | *neksa-ñ** | *neksa-c** | *neksa-ne** | – | – | *xada°-da* |

The Tocharian pronoun suffixes clearly descend from earlier clitic (direct and indirect) oblique pronouns. This is reflected in both their form and function, and can be understood with reference to the other Indo-European languages: clitic pronouns can be reconstructed for Proto-Indo-European, and they were thus inherited into the earliest Tocharian. It is not obvious that the Uralic objective conjugation should similarly be derived from clitic object pronouns, although this possibility does constitute one of the major theories on its origins in the deep past. Within any reconstructible paradigm, however, the first- and second-person objective conjugation endings always phonologically correspond more closely to the agent than to the patient, making a simple analysis of those suffixes as clitic object pronouns very difficult (cf. 8.3.2.6 for more details).

At the same time, it must be pointed out that clear object pronouns were added in the objective conjugation of some Uralic languages. Most strikingly in Mordvin, first and second (and perhaps third) person pronouns greatly expanded the objective conjugation, and in the Khanty third person singular was enlarged with a third person singular object as well. This shows that, while a strict analysis of every objective conjugation ending as an object pronoun is difficult, such structures could be innovated within the basic framework of the Uralic objective conjugation in some instances. As initial formations, the Tocharian pronouns suffixes and the Uralic objective conjugation do appear to be quite different.

Functionally there is not a full match between Tocharian and Uralic either. The Tocharian pronoun suffixes are much more versatile than the Uralic objective conjugation, since they agree not only with direct objects, but also with indirect objects, possessors of objects and indirect objects, agents of passive verbs, locations, sources, etc. The objective conjugation agrees with the object and the subject, but not with the other categories that the Tocharian pronoun suffixes can index. The Ob-Ugric languages do have some possibility for agreement with an indirect object, but only if the indirect object is promoted from its usual status as a dative argument to be marked as a direct object instead (see 8.3.3.2). Use of the objective conjugations in such constructions is obligatory, so that semantically, one could argue that a verb in the objective conjugation

agrees with an indirect object. However, this phenomenon is still structurally quite different from the Tocharian usage of the pronoun suffixes to agree with an indirect object. In Mansi, the original direct object is demoted to an instrumental instead, while all case assignment in Tocharian remains as in any ditransitive construction.

The use of the Tocharian pronoun suffixes to mark a non-subject participant as a secondary topic most closely aligns with the functions of the objective conjugation, and a great similarity between the Tocharian pronoun suffixes and the Uralic objective conjugation is that both are used as anaphoric reference to non-subjects on the verb. We could imagine that the Tocharian change to host the anaphoric pronouns on the verb rather than elsewhere in the sentence was inspired by Uralic, where non-subject anaphoric reference is hosted on the verb in the form of the objective conjugation. The additional functionality of the Tocharian pronoun suffixes may then be interpreted as a holdover from their original flexibility, transferred from their original more general clitic position elsewhere in the sentence to the verb specifically, as illustrated in Table 8.32.

Table 8.32: Schematic mechanism for the conversion of Proto-Indo-European pronominal clitics to Tocharian pronoun suffixes on a Uralic model.

| anaphoric reference... | U | T | IE | |
|---|---|---|---|---|
| ...to the direct object | + | + | + | pivot |
| ...to the indirect object | - | + | + | transferred |
| ...to a possessor | - | + | + | transferred |
| ...on the verb | + | + | - | changed |

To summarize, we can conclude that the main changes from Proto-Indo-European to Tocharian in the use of anaphoric pronouns has been that they are hosted on the verb, generally in the standard sentence-final position. Few traces are left of their original Proto-Indo-European behaviour as clitic elements, so that they can reasonably be considered suffixes to the verb. Their function remains mainly anaphoric, although some agreement is found as well, especially when the referent is (secondarily) topical.

Both the structural and functional nature of the Tocharian pronoun suffixes can be understood quite well when considering their origins as Proto-Indo-European pronominal clitics. These could be used anaphorically, and occurred in sentences together with finite verb forms agreeing with the subject. The position of the Tocharian pronoun suffixes as quite strictly post-verbal is an innovation, however, which is not accounted for directly by any aspect of their early sources.

The Uralic objective conjugation anaphorically marks objects with the verb in its standard sentence-final position, like in Tocharian. Agreement with overt object nouns seems to be more common than in Tocharian, but its occurrence is triggered by similar

considerations of topicality. Formally, the endings of the objective conjugation could hardly be interpreted as pronominal object clitics, but the position of the object pronouns in Tocharian as verbal suffixes could conceivably have been inspired by the position of object-agreement suffixes on the verb in Uralic. Most other aspects of the Tocharian pronoun suffixes are probably best understood as continuations from the state of affairs already present in Proto-Indo-European, as illustrated above in Table 8.32 above. On its own, it cannot be said that the Tocharian pronoun suffixes provide clear evidence of Uralic influence, but in combination with the other cumulative evidence, they constitute a plausible additional contact feature.

# 9 Potential loanwords

The most obvious sign of contact between languages often comes in the form of lexical borrowings. While most of the potential traces of contact between Tocharian and Uralic are structural, there are also a few possible loanwords, which will be discussed in this chapter.[108]

## 9.1 Introduction

Janhunen was the first to make lexical comparisons between Tocharian on the one hand and Samoyed on the other (Janhunen 1983). The limits of this line of research were immediately acknowledged by him, as the inherent difficulties are manifold. For instance, there may have been early Indo-European languages spoken in South Siberia that were closely related to the ancestor of Tocharian, or there may have been branches that disappeared without leaving any descendants. To make matters more difficult, the Samoyed languages themselves have undergone further changes since the relevant period in time, so that "there is no reason to assume that very many traces of Indo-European influence would be left in the modern Samoyedic idioms" (Janhunen 1983: 118). The time depth is of course significantly large.

Nevertheless, Janhunen discussed two potential loanwords from specifically pre-Proto-Tocharian into pre-Proto-Samoyed, namely pre-PS **säjt³wə* 'seven' and **wäsa* 'metal, iron' (Janhunen 1983: 5–7). A few more suggestions have been made over the years, with especially Napol'skikh (2001) and Kallio (2004) dedicating articles to a discussion of lexical comparisons between Tocharian and Samoyed specifically or Uralic generally. Napol'skikh makes use of the concept of "para-Tocharian", a hypothetical sister language of attested Tocharian, to explain certain unexpected sound correspondences present in his corpus. Kallio, meanwhile, criticizes the earlier loan etymologies, but does provide the new comparison in the form of PS **wẹn* 'dog' and pre-PT **kwënə* 'dog (obl.sg.)'. The pool of suggestions has not greatly increased in the last 40 years, and in his treatment of the hypothesis that early Tocharian was in contact with early Samoyed, Peyrot lists only four that he deems "relatively good", namely PS **säjt³wə* '7', PS **wẹn* 'dog', PS ±**mänäwjə* 'full moon', and PS **wäsa* 'metal, iron' (Peyrot 2019a: 100–101).

I will give an evaluation of the etymologies that have been adduced in the publications mentioned, and added a few additional suggestions on the basis of not only the reconstructed Proto-Samoyed lexicon (Janhunen 1977), but also the individual

[108] I presented this topic at the Finno-Ugric Congress of 2022 in Vienna. This chapter is an adaptation of the resulting article (Warries fthc.), which will appear in the proceedings. I thank Tapani Salminen and an anonymous reviewer for their comments on the article version.

northern Samoyed languages. I searched through several dictionaries to consider per word if a Proto-Samoyed reconstruction matched a Tocharian word. I used Tereščenko (1965) for Tundra Nenets, Sorokina & Bolina (2001) for Forest Enets, and Kosterkina, Momde & Ždanova (2004) for Nganasan; for Tocharian I mainly consulted Adams' (2013) dictionary of Tocharian B, which means that I had a bias in favour of that language, possibly to the detriment of lexical material only found in Tocharian A. Despite my best efforts, I may still have missed some possible connections, and a study of more extensive (future) publications on the lexicon of especially Enets and Nganasan may yet yield a few additional results.

The methodology I used to identify possible comparisons was quite strict. Since there is always the possibility of accidental resemblances in both form and meaning, a relatively high degree of similarity is to be desired at the current stage of research. Perhaps if it is established with greater certainty that the two languages were in contact, and a larger number of comparisons is already established, one could venture to see different patterns, but at the present moment I looked for a number of specific criteria. Of a root with the shape CVC(CV), the reconstructed consonants should match, or there should be a reasonable explanation as to why they do not fully match (e.g., if one side of the comparison is known to have reduced or disallowed a specific type of cluster that is found on the other side). The vowels should at least be similar, e.g., be similarly front or back, so far as we can understand their earlier values. Precise phonetic information unavoidably remains irretrievable, so that at best we can arrive at an approximation based on the pronunciation of the relevant phonemes in the attested languages and their more distant cousins. Full semantic identity is of course desirable, but a plausible shift can be argued for and may be permitted. Additionally, a verb in the donor language should ideally correspond to a verb in the recipient language, and a noun or adjective to the like as well.

## 9.2 Earlier comparisons

### 9.2.1 pre-PT **septə(m)* 'seven' ⇒ pre-PS **säjt³wə* 'id.'

The first word adduced as a loanword from Tocharian into Samoyed by Janhunen (1983) is the word for 'seven': PS **säjt³wə* in connection to PT **sʸəptə* from PIE **septṃ*. This is later repeated by Janhunen as the most notable representative of Samoyed-Tocharian contact (Janhunen 1998: 477). The etymology requires a number of complicated steps to work, however, such as metathesis of **-pt-* to **-tp-* and a weakening of **-p-* to **-w-*. The vowel correspondence of simple **e* with a sequence **äj* (written by Janhunen still as **ej*, as per the old reconstruction of Proto-Samoyed; cf. Helimski 2005, Aikio 2006a) is also rather strange.

It appears furthermore that PS **säjt³wə* 'seven' can be derived from the same PU **ćäjć(ć)imä* as Fi. *seitsemän* 'seven', according to Aikio (2020: 110–111). The original value of PS **t³* in **säjt³wə* cannot be determined from Samoyed, but it may indeed represent PS **s* from PU **ć*; the other possibilities are PS **c* (from PU **č*), PS **t* (PU **s* or *t*) and PS **k* (PU **k*). The **-m-* was apparently irregularly lenited to **-w-* in the Samoyed reflex (Aikio 2020: 110–111). Despite this irregular development, I would agree with Aikio that the Uralic etymology is to be preferred over a loan etymology that shows several unexplained modifications of the source word.

### 9.2.2 pre-PT **wesa* 'gold' and PS **wäsa* 'metal, iron'

Janhunen also discusses the similarity of the Tocharian words for 'gold' (TA *wäs*, TB *yasa* from PT **wʸəsa*) to Samoyed words for 'metal, iron' (Ng. *basa*, NeT *yesya*, PSk. **kuə̑sə*, Km. *baza*, Mt. *be/äsä*). The story is quite complicated, however, as Janhunen admits (Janhnen 1983: 120–121; cf. also Kallio 2004: 132–133). The Samoyed words are connected with metal names in other Uralic languages, such as SaaN *veaiki* 'copper', Fi. *vaski* 'id.', Hu. *vas* 'iron', pointing to a reconstruction as approximately PU **wäćkä*, although Hungarian rather indicates back vocalism; some other Uralic branches have similar words (Grünthal et al. 2022; supplemental material). The loss of **k* in the cluster PU **ćk* in Samoyed is regular (cf. 3.5.2), but the vowels are troublesome. Not only do the Saami, Finnic and Hungarian vowels not match nicely, also Ng. *basa* seems to imply a reconstruction PS **wäsa* with second-syllable **-a* of unexplained origin, as according to Aikio (2015: 42–43). Additionally, Aikio argues that the Selkup and Kamas forms obtained their first-syllable vocalism due to some effect caused by this same **-a* as seen in Nganasan.

However, according to Grünthal et al. the second-syllable *-a* in Nganasan may also be of analogical origin due to paradigmatic alternations between final *-i̯* (usually PS **-ä*) and *-a* (usually PS **-a*) in that language. Grünthal et al. also consider the vowel development of the initial sequence **wä-* in the Nganasan, Selkup and Kamas words to be regular without influence from a second-syllable following **-a* (Grünthal et al. 2022; supplemental material). This would make the Samoyed reconstruction PS **wäsä* as expected from PU **wäćkä*. The complicated situation involving this word has inspired its analysis as a Wanderwort that entered different Uralic languages independently at an early stage of differentiation, and/or was borrowed from branch to branch (Aikio 2015: 43).

For the purposes of a comparison to Tocharian, an important aspect is that a **k* in the Finno-Ugric languages is not easily explained as a reflex of a pre-PT **wesa*, so that a loan etymology of PS **wäsa* from Tocharian would seem to necessitate a thorough separation of the Samoyed word from its Finno-Ugric counterparts. This seems undesirable. An additional complication is that PT **wʸəsa* does not have a fully ironclad Indo-European etymology (on both points, cf. Thorsø & Wigman et al. 2023). A borrowing in the opposite

direction, from early Samoyed into Tocharian can also be considered, although it would be so far the only example of its kind. If this route is taken, it would really be early Samoyed, rather than another Uralic branch, that was the source of the borrowing, on account of the lost **k* (again, cf. Thorsø & Wigman et al. 2023). An important issue with this option, however, aside from the lack of comparable loans in the same direction, is that the loss of **k* in the cluster PU **ćk* cannot be established to have occurred particularly early in pre-Proto-Samoyed (see 3.9). This may have been a late change, potentially taking place after any possible contact with early Tocharian.

Even after all this time, this etymology is still rife with uncertainty. The Tocharian word may be innovative, or could be derived from Proto-Indo-European, while the Samoyed word cannot easily be derived from Tocharian and a borrowing from Samoyed into Tocharian seems unlikely. This means that, while a connection looks very attractive at first, due caution is advised.

### 9.2.3 pre-PT **mēnē-* 'moon' ⇒ PS ±**mänä/iwjə* 'full moon'

A Proto-Samoyed word for 'full moon' can be tentatively reconstructed based on NeT *myenyuy°* and EnT *menio*, both of the same meaning. No cognates are known from the other Samoyed languages, but it bears a strong semantic and phonological resemblance to the Indo-European and Tocharian word for moon, cf. TB *meñe*, TA *mañ* (Blažek *apud* Napol'skikh 2001). This derives via a pre-PT **mēnē-* from PIE **meh₁nē(n)-* 'moon', with a probably early change of PIE **eh₁* to pre-PT **ē*. The Samoyed formation looks like it requires two extra suffixes, however, seemingly first a **-w* and then the suffix **-jə*, which is used to form adjectives, but also secondary nouns. Especially the latter suffix is very common in the Samoyed languages, but the need for extra derivational material to arrive at the attested word certainly does not add to the plausibility of this loan etymology.

There is a parallel of another Uralic language borrowing the Indo-European word for 'moon', namely in Saami: SaaN *mánnu* < PSaa. **mānō*, from Old Norse (Sammallahti 1998: 253–254). It looks striking that the rounded vowel in the second syllable is matched in both cases, but in Saami it derives from an **-ō* in Germanic; such a source does not seem to be available for Samoyed, if it is based on pre-PT **mēnē* (the ancestor to TB *meñe*). The semantic and formal match of these Samoyed and Tocharian words is quite close, so that a borrowing etymology of the base for NeT *myenyuy°*, EnT *menio* is overall possible, but not entirely compelling.

### 9.2.4 pre-PT **monko-* 'lack, deficit' ⇒ PS **måŋkə̑* 'poverty, need'

PS **måŋkə̑* 'poverty, need' is another one of the words listed by Napol'skikh (2001), this time originally from Helimski. Representative Samoyed reflexes are Ng. *məŋkə*, NeT

*maŋk*°, all of the same basic meaning (Ru. 'нужда, бедность'). [109] The semantics are virtually identical to TA *maṅk* 'lack, deficit'. Napoľskikh somewhat confusingly compares the noun PS **måŋkə̑* 'poverty, need' to the verb root TAB *mənk-* 'to lack', with supposed adoption of "para-Tocharian" **ə* as Samoyed **å* after **m-*. This sound substitution rule with reference to "para-Tocharian" does not lend to the loan etymology's credibility, and it is probably the reason why this example is not listed by Peyrot as one of the "relatively good" examples of loan etymologies (Peyrot 2019a: 100–101).

However, Napoľskikh's workaround is unnecessary. The Tocharian A noun *maṅk*, from PT **menke*, can derive regularly from an *o*-stem nominalization early pre-PT **monko-*, which provides a quite adequate source for the Proto-Samoyed noun **måŋkə̑*; especially when considering that the common change from PU **o* to PS **å*, if it applied after this borrowing, would allow for an even more faithful adoption of the pre-Proto-Tocharian vowel. The phonological and semantic match between the Tocharian and Samoyed words is thus very high, and I am not aware of any cognates for PS **måŋkə̑* in other Uralic languages. Borrowing of this word from pre-Proto-Tocharian into pre-Proto-Samoyed is therefore probable.

### 9.2.5 Pre-PT **kwënə(m)* acc. of 'dog' ⇒ pre-PS **wẹn* 'dog'

The Proto-Samoyed word **wẹn* 'dog', with reflexes such as Ng. *baŋ*, EnF *buʔ*, NeT *weh*, SkTym der. *kanak*, has no Uralic etymology. Kallio (2004) has proposed that it could be a borrowing from the Tocharian word for dog, not from the nominative TAB *ku*, but from the oblique TB *kweṃ*, TA *koṃ*. The oblique forms go back to PT **kʷenə*, from pre-PT **kwënə*(*m*), PIE **ḱu̯onm̥*. The initial cluster **kw-* could not be borrowed as such in Samoyed, which lacks word-initial clusters altogether, so it may have been simplified to a simple **w-*. There are no known parallels for this, but the development is plausible. The Samoyed vowel **ẹ* may point to the de-rounded **ë* (±*[ʌ]) of pre-Proto-Tocharian, from earlier IE **o* (Kallio 2004: 133-134). The semantic match is of course exact.

Interestingly the vowel in PS **wẹn* does not correspond to the vocalism of the previous etymology PS **måŋkə̑*, where PIE **o* would rather be reflected as PS **å* (via a pre-PS **å* or **o*). Perhaps the difference could be due to an earlier adoption of **måŋkə̑* than of **wẹn*, with in between the pre-Proto-Tocharian sound shift of **o* to **ë*, but at this point it is difficult to say.

[109] There is also Enets *maggô* 'arm : poor' in Mikola (1995: 141), and Sorokina & Bolina (2001: 66) give derived *magutuda* 'бедный : poor' and *magutut'* 'бедно : poorly'.

## 9.3 New comparisons

### 9.3.1 pre-PT **lonko-* 'steep slope; cliff' ⇒ PS **lâŋkâ* 'steep bank; steep'

The underived Samoyed word is found in Nenets, NeT *ləŋk*°, NeF *łăŋk*° 'обрыв, яр; крутой : precipice, cliff, steep bank; steep', and derivations are found in EnT *logori*, EnF *logaŕ* 'гора : mountain' and Ng. *ləŋkəri* 'порог : rapids'. The Tundra Nenets acc.pl. is *ləŋko* (Salminen 1998). This means that it is of the type where nom.sg. -° changes to acc.pl. *-o*, i.e., PS *-*â* to *-*âj*, which points to the word being back vocalic PS **lâŋkâ*. Front vocalic ***ləŋkə* should have yielded an acc.pl. ***ləncye* instead in Tundra Nenets.

This fits with back vocalism in pre-PT **lonko-*, the source for TB *leṅke* 'slope', which is very similar in both meaning and form to the Nenets words. It is an *o*-stem nominalization from the verb TAB *lənk-* 'hang, dangle', from PIE **h₁leng$^{wh}$-* 'move lightly' (Peyrot 2013: 476), thus quasi-PIE **h₁long$^{wh}$o-*. The initial **l-* of PS **lâŋkâ* before a vowel other than PS **ę* or **į* points to this word being borrowed into Samoyed after Proto-Uralic initial **l-* had changed into **j-*, as in, e.g., PU **lupsa* 'dew > PS **jâptå* 'id.' (Janhunen 1981: 223). Since the semantic and formal match with Tocharian is so close, this could provide a suitable source of borrowing. Especially the consonant skeleton is the same, only the vowels do not match up exactly. If we project PS **lâŋkâ* back to pre-Proto-Samoyed, we get **luŋkå*. The vowels **u* and **o* are close neighbours, so that the discrepancy may be allowed, but it is notable that the vowel reflex is not identical to that seen in PS **måŋkâ* from a pre-PT **moŋko-* (cf. above).

Anikin and Helimski connect the Nganasan (*ləŋkəri*) and Tundra Enets words (*ləŋkəri*, *logori*) with an Evenki word *лэӈирӯ* 'колода : (chopping) block (of wood)', and consider it to be a borrowing from Tungusic into northern Samoyed (Anikin & Helimski 2007: 125).[110] Evk. *лэӈирӯ*, in turn, they derive from *лэӈэ-/лаӈа-* 'сломать (зуб) : break (a tooth)'. I think that the semantic connection for this loan etymology is more distant, and it separates the Nganasan and Tundra Enets words from the semantically more closely related Nenets word. The Evenki/Tungusic loan etymology can therefore be rejected, with a possible Tocharian etymology to replace it.

### 9.3.2 pre-PT **lenk-* 'dangle, sway, hang' ⇒ pre-PS **laŋkå-* 'hang, dangle'

Tocharian has a verb *lənk-* 'dangle, sway, hang', which is very similar to a word of the same meaning found in Samoyed. In Tundra Nenets, there are only derivatives attested: NeT *lyaŋkabta-* 'повесить, развесить; понести что-л. свешивающееся : hang (tr.); carry sth. that is hanging down', *lyaŋkada-* 'развесить что-л. : hang up sth.', *lyaŋkasa-* 'висеть (в большом количестве); бродить без дела : hang (in great number); wander around',

[110] I am grateful to Kaisla Kaheinen for pointing this out to me.

etc. The underived root would be *lyaŋka-*, which can be mechanically reconstructed as PS **laŋkå-*.[111] Outside of Tundra Nenets, a direct cognate to *lyaŋkabta-* is found in NeF *lyaŋkapta-* with the same meaning, and a borrowing from Tundra Nenets is found in EnF *ľagata-* 'id.'.

Due to its absence in the other Samoyed languages, the distribution of this word leaves something to be desired. Nevertheless, the correspondences in form and meaning between Tundra Nenets and Tocharian are quite close, so that we may suppose that the word was lost elsewhere in Samoyed (or remains unattested), similar to the word for 'full moon', and the underived word for 'steep slope'. Like with the previous word, the initial **l-* in Samoyed points to this word being borrowed at some point in the specific prehistory of the Samoyed branch, as it cannot be inherited from Proto-Uralic; pre-Proto-Tocharian seems to be a possible source.

### 9.3.3 pre-PT **der-* 'split, separate' ⇒ pre-PS **tar-* 'divide, separate'

Reflexes of PS *tar-* 'divide, separate' is attested in all branches of Samoyed, with pretty much the same meaning everywhere: Ng. *tar-*, NeT *tyar-*, EnF *tär-*, Sk. *tār-*, Km. *tarə-*, Mt. *tā̈r-*. In SW (154-155), the reconstruction is given as **tär- ~ *tärə̂-*, which in the new vowel system notation is **tar(ə̂)-*. The form with second-syllable **-ə̂* is for the Kamas reflex, and may well be secondary. The word also has a suggested Uralic etymology in UEW: **šurɜ-* 'schneiden, teilen' (UEW: 503–504). However, PU **u* does not become PS **a*. An alternative etymology can thus be proposed, namely a borrowing from the ancestor of PT **tˢər-* 'be separated; caus. separate' from PIE **der-* 'split, flay, tear'. Semantically, the Tocharian and Samoyed verbs are very similar. In Tocharian texts the verb is often used to refer to people being separated from one another, but given the original more physical meaning of 'split, tear' attested in other Indo-European languages, Samoyed 'divide, separate' may represent a very fitting intermediary stage also for the Tocharian semantic development. The Samoyed meaning is more distant from, for instance, Gr. δέρω 'flay', Skt. *dṛṇā́ti* (root *dar-*) 'rend, split, burst (tr.)', and also English *tear*. The early Indo-Iranian **dar-*, from which the Sanskrit derives, could otherwise theoretically constitute an alternative source for the Samoyed word PS **tar-*, if Samoyed is regarded to independently have undergone a similar semantic shift as seen in Tocharian **tˢər-*.

The development of PIE **d* to Tocharian *tˢ*, as in this word, is not well understood as a diachronic change, especially as regards intermediate stages and phonetic interpretation (see 2.3.4.3). It is thus uncertain how PIE **der-* would have been pronounced at the time that this word could be borrowed into early Samoyed, although

[111] The Tundra Nenets morphological dictionary (Salminen 1998: s.v.) gives an uncertain analysis of the root-final vowel as either *-ə* or *-a*, but Tapani Salminen (p.c.) has informed me that the correct interpretation of the root is *lyaŋka-* with an *-a*.

some kind of dental is likely. In theory, even starting from pre-PT ±**t$^{s}$er-*, one could also posit a borrowing in the shape of pre-PS **sar-*, whence PS **tar-* by the well-known Samoyed change of PU **s* (and **š*) to **t*, depending on the time depth of the relevant changes (not easily established, see 3.4.1 and 3.9).

Formally, PS **tar-* 'divide, separate' can be derived from the ancestor of PT **t$^{s}$ər-* 'separate', or from a Proto-Indo-Iranian source **dar-* 'split, tear', both of which descend from PIE **der-* 'split, tear'. Semantically, the Tocharian verb provides the better match with the Samoyed, so that this is the more likely source.

### 9.3.4 pre-PT **newo-* 'new' ⇒ Pre-PS **ńuwV* 'child, young boy'

The Samoyed word for child is reconstructed in SW as **ńuə̑* (p. 111). However, a vowel sequence like **uə̑* often derives from sequences **VwV*, **Vji* and **Vγi* (Aikio 2012), which allows for an internal reconstruction of **ńuə̑* as pre-PS **ńuwV*. This is very close to pre-PT **newo-* (perhaps already **[ɲewo]*, but the status of phonetic palatalization at the possible time of contact is unclear, see 4.4), whence TB *ñuwe*, TA *ñu*. The semantics are not identical: the Tocharian word really means 'new' rather than 'young', but the latter meaning is attested for cognates in other Indo-European languages, e.g., Gr. νέος 'young, youthful', Skt. *náva-* 'new, young', and the semantic connection between the concepts 'new' and 'young' is close in any case.

Terms for children seem to be borrowed relatively easily: compare for instance the North Saami words *bárdni* 'boy, son', and *gánda* 'boy' with Swedish *barn* 'child' and Old Norse *ganti* 'boy' respectively (Sammallahti 1998: 230, 243) and Dutch colloquial *kids* 'children' from English. On account of this, borrowing of a word for 'child' would not be out of the ordinary.

Nevertheless, the etymology remains tentative due to the lack of semantic exactness, as well as the theoretically possible alternative internal reconstructions for pre-Proto-Samoyed, namely **ńuji* or **ńuγi*. The Samoyed word might be derived from early Tocharian, but it might also be of some yet unknown different origin.

### 9.3.5 pre-PT ?**mëywya* 'power, strength' ⇒ pre-PS **mᵩə̑* 'sturdy, solid'

We can reconstruct a word PS **mᵩə̑* 'sturdy, solid, strong', with reflexes in all the northern Samoyed languages: Ng. der. *maagəə*, NeT der. caus. *mey°ra-* 'strengthen', der. *meyo*, NeF der. *miju* 'id.', EnF *mui*. It apparently mainly refers to 'strong' in the sense of 'sturdy, solid' ('крепкий, прочный, твёрдый' are the translations given, notably not 'сильный'). This is slightly different from the Tocharian word TB *maiyya* 'power, might,

strength', which rather seems to be linked to some kind of general strength as potency.[112] However, the different shades of 'strong' are semantically close, and may have shifted somewhat over time. The word could in principle be borrowed from pre-Proto-Tocharian with loss of *-w-* in an original cluster *-ywy-* reconstructed by Adams (2013: 508), if it had not yet been simplified on the Tocharian side already, and a change from PIE **o* to **ę* reminiscent of the one found in pre-PT **kwonə-/kwënə-* > pre-PS **węn* 'dog'.

For both the Samoyed and Tocharian words there exist alternative loan etymologies. For Nganasan *maagəə* specifically, a connection with Evk. *maŋa* 'strong' has been proposed (Wagner-Nagy 2019: 503).[113] However, this would separate the Nganasan word from its counterparts in Enets and Nenets, despite the regular phonological correspondence and identical semantics. A shared origin of Ng. der. *maagəə*, NeT der. *mey°ra-*, EnF *mui* 'sturdy, solid, strong' from PS **mȩjə̑* is clearly to be preferred, and this cannot be Evk. *maŋa* 'strong'.

TB *maiyya*, meanwhile, has been connected to Iranian **māyā-* (Av. *māiiā-* 'craft, capacity, strength') with a development of **āy* to **ayy* in the Old Steppe Iranian language that was in contact with pre-Proto-Tocharian (Bernard 2023: 58–60). Adams (2013: 508) derives the Tocharian from TB *maiwe* 'young' with addition of a suffix **-ya*, i.e., **moywya-* with loss of the medial **-w-*, but this would separate TB *maiyya* from the very similar Indo-Iranian words (see also Bernard's criticism of this etymology). TB *maiwe* 'young' itself is not borrowed from Iranian, but it is formally not as easy to connect to Samoyed **mȩjə̑* due to the cluster /jw/: such clusters occur in Proto-Uralic and Proto-Samoyed (e.g., PU **ojwa* > PS **åjwå*), so that the adoption of pre-PT **moywo-* or **mëywë-* as **mȩjə̑* rather than **mȩjwə̑* seems strange.

Semantically, 'young' is also not the same as 'sturdy, solid, strong', although in principle a shift could be imagined on the Samoyed side from 'young' to 'new, newly made' to 'sturdy, solid', as the opposite of 'old and almost falling apart'. An advantage of *maiwe* as a source for the Samoyed would be that it is also an adjective, rather than *maiyya*, a noun, and that it is with greater certainly inherited in Tocharian rather than borrowed from Iranian. Perhaps a connection of the Samoyed word with Iranian **māyā-* can be considered instead, adding to a small group of words of this kind (e.g., PS **jaə̑* 'flour' cf. PIr. **yawa-* 'grain'; Janhunen 1983: 123–124; Holopainen 2019: 105).

In conclusion, there exist similar words denoting some concept of 'strength' in Tocharian and Samoyed. The Tocharian noun TB *maiyya* 'power, might, strength' provides the closest phonological match to PS **mȩjə̑* 'sturdy, solid, strong', but the

[112] Adams (2013: 508) cites passages where it refers to 'bodily strength' or abstract "force". Similarly, the derived adjective *maiyyātstse* is used to describe a powerful 'ignorance' and 'hatred', as well as a 'powerful man'.

[113] Pointed out to me by Kaisla Kaheinen.

Tocharian etymology as a direct inheritance from Proto-Indo-European is highly uncertain. The adjective TB *maiwe* 'young' provides a worse phonological match, and a larger semantic difference between it and the Samoyed word requires a semantic change on the Samoyed side. A derivation of PS **mɐjə̑* from Tocharian may thus be considered as a possibility, but it is definitely not the best example of this category. Some connection between the Samoyed and Indo-Iranian words could be considered instead.

## 9.4 Discussion

The number of promising loanwords from pre-Proto-Tocharian into pre-Proto-Samoyed remains very low. At the same time, some words do show quite tantalizing correspondences (like **måŋkə̑*, **lə̑ŋkə̑*, **laŋkå-*, **mänäwjə*). A summary of the evaluations is provided in Table 9.1.

Table 9.1: A summary of the loanword evaluations.

| pre-Proto-Tocharian | Proto-Samoyed | Evaluation |
|---|---|---|
| **septəm* 'seven (7)' | **säjt³wə̑* 'seven (7)' | rejected |
| **wesa* 'gold' | ±**wäsa* 'metal, iron' | problematic |
| **mēnē* | ±**mänäwjə̑* 'full moon' | possible |
| **monko* 'lack, deficit' | **måŋkə̑* 'poverty, need' | plausible |
| **kwënəm* 'dog obl.sg.' | **wɐ̣n* 'dog' | plausible |
| **lonko* 'steep slope, cliff' | **lə̑ŋkə̑* 'steep bank; steep' | plausible |
| **lenk-* 'dangle, sway, hang' | **laŋkå-* 'dangle, hang' | possible |
| **der-* 'split → separate' | **tar-* 'divide, separate' | plausible |
| **newo* 'new' | **ńuə̑* 'child, young boy' | possible |
| **meywya* 'power, strength' or **maywe* 'young' | **mɐjə̑* 'sturdy, solid' | problematic problematic |

At first blush, the overall number of possible loans appears to contrast with those from early Indo-Iranian in the other Uralic languages, but the difference is not necessarily that significant at the level of individual branches. Holopainen lists 13 etymologies in the pre-Indo-Iranian category (Holopainen 2019: 336). However, if we check the distribution of these loanwords, it becomes clear that for instance Mansi and Khanty only have very few, one each in this category. This may be due to various factors that are difficult to test, including later lexical replacement. It is well-known that Samoyed has undergone some significant lexical replacement in even the basic vocabulary (e.g., Aikio 2002: 52), providing ample opportunity for secondary loss of any early loanwords. If there was only

one very old layer of borrowings from early Tocharian into early Samoyed, the number of reflexes is expected to be low. A lack of ample evidence for early loan contact is thus not a reliable indication that there was no early loan contact at all. We might imagine that retrievable early Tocharian loans from Samoyed behave according to a similar pattern as the earliest Indo-Iranian borrowings in Ob-Ugric.

That is not to say that the possible loans discussed in this chapter are (a) necessarily all accurately identified, as due caution is expressed where appropriate, or (b) incontrovertibly prove ancient contact between pre-Proto-Samoyed and pre-Proto-Tocharian; simply that the small number of etymologies is not an indictment of the proposed borrowings or of a contact hypothesis in and of itself. The limitations of this line of enquiry were already voiced by Janhunen (1983), as briefly addressed in the introduction to this paper as well. Furthermore, if the majority of phonological and grammatical contact phenomena are in line with influence on Tocharian through language shift from an early form of Samoyed, that particular variety Samoyed could logically not have been the direct ancestor of the modern Samoyed languages, so that limited borrowing from Tocharian might be expected. At least the new comparisons adduced here and the rehabilitation of the older suggestion of PS **måŋkə̑* with a closer nominal origin in Tocharian should provide an illustration of the possibility for further results in this early contact situation, and brings us a step closer towards a better understanding of the prehistoric interactions of the early Samoyeds with Indo-Europeans.

# 10 Conclusions

There are many linguistic features of Tocharian that show a correspondence with Uralic and Samoyed. I have discussed them in this dissertation in order to test and evaluate the hypothesis that Tocharian was influenced by Uralic or, more specifically, early Samoyed. In this conclusion, I will give a summary of the results of the investigations carried out in each chapter and answer the various research questions posed in the introduction.

## 10.1 Overview

Previous research on Tocharian-Uralic/Samoyed contact hypotheses has stressed on the Tocharian stop system (e.g., Kallio 2001, Bednarczuk 2015, Peyrot 2019a) and, with a specific connection to the Samoyed branch, the vowel system (Peyrot 2019a). The stop systems indeed correspond quite closely if we consider pre-Proto-Tocharian before phonemic palatalization. The pre-Proto-Tocharian phonological system at that point in the relative chronology also provides the closest match for the pre-Proto-Samoyed vowel system. However, the specificity correspondence between the pre-Proto-Tocharian vowel system and the pre-Proto-Samoyed vowel system may have been overestimated (Peyrot 2019a; differently, but still perhaps too optimistically Warries 2022), so that it cannot easily be decided how much value should be assigned to it in relation to a specifically Samoyed connection. The pre-Proto-Tocharian vowel system is also very similar to the Proto-Uralic vowel system. On the phonological comparison in more detail, see chapter 4.

The Tocharian accent or stress system probably passed through a stage of general word-initial accentuation, which corresponds to the stress placement of Uralic, although it is difficult to determine this with certainty due to later changes. Initial stress is typologically common, so that it does not constitute a highly specific match. For more details on the Tocharian accent, see chapter 5.

The agglutinative case system of Tocharian has frequently been connected with substrate interference (e.g., Krause 1951, K.H. Schmidt 1990, Bednarczuk 2015, Peyrot 2019a). Uncertainties around the age of the Tocharian secondary cases and differences in meaning make the comparison with the Samoyed case system insecure. This is discussed in chapter 6.

Tocharian agrees with Samoyed, and other eastern Uralic languages more generally, in the wide use of contextual orientation in the participles, a fact that had not been previously pointed out as a possible contact feature. The participial system, with its distinction between present, past, negative (privative) and necessitative participles (present gerundive) is also structurally reminiscent of the Samoyed participial system. See chapter 7 for more information.

The Tocharian use of pronoun suffixes to mark pronominal objects is similar to the Uralic objective conjugation, but the constructions in the two language groups differ in their origins and in the specifics of the types of syntactic structures that they could be used for. This may be understood as the result of the building blocks that were available in pre-Proto-Tocharian to make such a structure in the first place, so that it can plausibly still be linked with Uralic influence. The age of the objective conjugation in Uralic itself is a disputed problem, but I believe that it goes back to Proto-Uralic and probably finds its origins in a pre-Proto-Uralic stage. This issue is treated in chapter 8.

Finally, a handful of potential loanwords might support the connection with Samoyed specifically, but the evidence does not reach a critical mass (Janhunen 1973, Napol'skikh 2001, Kallio 2004, Warries fthc.). Proposed Tocharian loanwords in other Uralic languages or Proto-Uralic itself are too uncertain to use as evidence for prehistoric contact. An overview can be found in chapter 9.

## 10.2 Phonology

Tocharian phonology underwent major changes from Proto-Indo-European. The loss of distinctions related to voicing or aspiration, as well as vowel length, constitute significant innovations that resulted in a Uralic-looking phonological system in Tocharian. No other known language in the area shares this combination of features, with the possible exception of the still poorly understood pre-Proto-Yukaghir (see the discussions in 4.2.2 and 4.3.2). This makes early Uralic/Samoyed the most reliable source of substrate influence in this regard, according to our current knowledge.

In my discussion of Tocharian historical phonology, I treated some problematic points, like the timing of palatalization and umlaut, the merger of PIE **o* and **ē*, the reflexes of PIE **eh₂* and **ō*, and the developments of PIE **d*. I concluded that the rise of the palatalized vs. non-palatalized opposition in consonants can be connected with the merger of PIE **i* and the **ə* that arose from syllabic resonants and anaptyxis in zero-grades as **i*, and the merger of PIE **o* and **ē* as **e* (2.4). The merger of **i* and **u* as **ə* has to be dated later, since the **u* caused *u*-umlaut after palatalization had been completed, indicating that it must have remained distinct for a longer period of time. The merger of **o* and **ē* as **e* can be further supported by the fact that these vowels were affected in the same way by *u*-umlaut and *a*-umlaut, which can both be dated to the (pre-)Proto-Tocharian period as well (2.5.1 and 2.5.5).

The development of **ō* remains difficult to establish due to a lack of reliable examples. However, it is certain that it became PT **o* in the second syllable in a few examples, so that no general development to PT **a* can be posited (2.4.5). Whether the PT **o* from marginal **ō* and the PT **o* that resulted from umlaut need to be separated from the reflex of PIE **eh₂* is uncertain, due to the small group of examples showing contradictory

reflexes. I suggest that *$eh_2$ to PT *$å$ as opposed to *$o$ from umlaut and *$ō$ could be merged simply as PT *$o$ if the two examples that show a reflex *$a$ in Tocharian A are assumed to be the result of dissimilation from a preceding labial consonant *w-* or *pr-* (2.5.6.4). However, this remains tentative. A phoneme PT *$å$ can in any case be reconstructed as the reflex of umlauted *$a$ in the *o*-presents [class 4] and in TB *onolme* 'living being' (2.5.4). This *$å$ does regularly yield TB *o* and TA *a*, and thus should not be equated with at least the result of *u*-umlaut or *o*-umlaut of *$e$, as this normally becomes TAB *o* in the first syllable.

The consonant with the most varied and obscure development in Tocharian is PIE *$d$, with its various reflexes as PT *$t^s$, *$ś$, *$t$ and ∅. I concluded that the different reflexes can be interpreted in various ways depending on the assumed phonetic interpretation of PIE *$d$, and that the Tocharian data is insufficiently clear to allow us to infer the original state of *$d$ in the stage of Proto-Indo-European from which Tocharian descends. A relatively good candidate for an initial stage in the development is a [ð], as this is a natural intermediate step on the way to full loss (the reflex ∅), and lenition from [d] to [ð] can be blocked in an understandable way by a preceding nasal, so that the (devoiced) reflex *$t$ in the original cluster *$nd$ is accounted for. However, the development from [ð] to TAB $t^s$ remains difficult to understand, and I know no direct parallels for such a change. A glottalic interpretation of *$d$ could accord with the reflex ∅ by assuming debuccalization to a [ʔ], and perhaps an originally glottalized dental stop could have become TAB $t^s$ and *ś* by assuming assibilation from a potential (but also unverifiable) intermediate development to a [$t^h$] (for details, see 2.3.4.3–2.3.4.6).

For the relative chronology of Tocharian, the following conclusion is particularly relevant: at the time that palatalization took place, a vowel system without length distinctions can be reconstructed. This matches very well with the lack of vowel length distinctions in the early Samoyed phonological system, which also cannot be reconstructed with pervasive palatalization oppositions of the type that we see in Tocharian. This means that the phonological systems of pre-Proto-Tocharian and pre-Proto-Samoyed were structurally very similar at some point in the distant past. The pre-Proto-Samoyed system did not differ significantly from the Proto-Uralic system as far as the relevant features are concerned, so that this does not constitute a highly specific match between early Tocharian and Samoyed specifically.

In my discussion of Samoyed historical phonology, I concluded that the Proto-Samoyed reflex *$o$ rather than *$u$ in stems of the type PS *$CoC$ ~ FU *$CuCi$ can be most easily understood as an archaism in Samoyed, in accordance with Janhunen's (1981: 231) original formulation of this correspondence (3.6.3). The split development of PU *$a$-$a$ into PS *$å$-$å$ and *$a$-$ə̂$ should probably also be understood as some kind of archaism, since there does not seem to be any Samoyed-internal way to generate the two different reflexes, at least not with our current understanding of Samoyed historical phonology.

Zhivlov's (2014) suggested reconstruction of two different *a*-vowels in the second syllable in Proto-Uralic thus needs to be seriously considered and further researched (3.7.1).

A difficult point in the relative chronology of Samoyed is the split of PU **ę* into **ę* and **į*. The general distribution can be understood on the basis of the following vowel, with PS **ę* resulting from the original sequence PU **ę-i* and PS **į* from PU **ę-a*. However, there also appears to have been an interaction with coda nasals, which is clearly reflected in the developments of **lęmpi* 'swamp' and **lęnti* 'plain, valley', and perhaps in **joŋsi* 'bow' if this developed via **jęŋsi*. The absence of such a development of **δ´ęmi* 'bird cherry' to PS **jęm* rather than ***jįm* might indicate that the change from **ę* to **į* took place before apocope of **-i*, but paradigmatic alternations could always have restored the vowel, analogically restoring hypothetical **jįm* : gen.sg. **jęmân* to **jęm* : *jęmân*. This makes the dating of this vowel development relative to apocope uncertain (3.6.2).

In general, paradigmatic alternations resulting from some of the Samoyed sound changes should be borne in mind, as with the split development of roots of the shapes **CVli* and **CVji* to both **CVj* and **CVə* (Aikio 2012). The regular development is that **CV{l/j}i* became **CVə* when the final **-i* was preserved, and **CVj* in consonant stems or after apocope. The resulting alternations were levelled in most roots, but the original distribution can be seen in the verb PS **åâ- ~ *åj* 'be'. Other changes that are expected to have caused paradigmatic alternations include the shift from **o* to **å* in **CoCa* and **CoCCi* but not **CoCi*. The split vowel reflex of the verb PU **ćoji-* 'sound' and the derived noun **ćojma* 'sound' into PS **soj-* and **såjmå* respectively shows this. Aside from **såjmå*, other derivations and inflectional forms of **soj-* should regularly have undergone the same change, e.g., quasi-PU **ćojća-* (pst.), **ćojnta* (ipfv.ptc.), **ćojŋa-* (aor.) must have developed into **såjså-*, **såj(n)tå*, **såjŋa* by regular sound law, but were evidently restored based on forms like the imperative/connegative **ćoji-k* or the old prs.3sg. **ćoji*, where no development from **o* to **å* is expected. The forces of analogical regularization of paradigms should not be underestimated in the historical phonology of Samoyed (3.4.2).

## 10.3 Accent

The Tocharian accent, as can be deduced from Tocharian B only, does not conform to the Samoyed or Uralic stress on the first syllable. Rather, Tocharian B had a general second-syllable accent, albeit with many exceptional categories where the accent was fixed on the first syllable for various reasons. From the discussion of the diachronic development of the Tocharian accent in chapter 5, I concluded that earlier Tocharian would have had initial accent, and that the accent patterns found in Tocharian B are partially innovative. The pre-Proto-Tocharian initial accent is different from both the ancestral Proto-Indo-European mobile accent, and from later Tocharian B, but it does

accord with the first-syllable stress of early Uralic and Samoyed. It thus provides an additional phonological feature of pre-Proto-Tocharian that may have been taken over from early Uralic/Samoyed. Since first-syllable accentuation is quite common in languages around the world, it does not provide a particularly specific point of comparison, however, and other languages could also have caused this type of shift in early Tocharian.

## 10.4 Morphology

The morphological comparisons between Tocharian and Uralic/Samoyed are difficult to evaluate. The agglutinative case system of Tocharian offers a general match with Uralic/Samoyed, but it is not quite specific. The meanings of the Tocharian locative and perlative correspond at their core to the Samoyed locative and prolative, but they also have a wider functional range that is not paralleled. Especially the static use of the Tocharian perlative, and illative use of the Tocharian locative find no direct correspondence in Samoyed. It is unknown how much of this discrepancy may be due to later change in Tocharian, or whether it can be attributed to the result of the original meaning of the elements that were used to form the case suffixes.

Furthermore, the Tocharian agglutinative case system was clearly productive even after the Proto-Tocharian period, with diverging cases and case endings appearing in Tocharian A and B. This causes uncertainty over the true age of the Tocharian cases: they may (partially?) be old enough to result from language contact in the distant past, but some elements of the system could also be relatively too young for that. The Tocharian perlative case does receive its most appropriate counterpart in the Samoyed prolative, which is not reconstructed for Proto-Uralic. This may connect pre-Proto-Tocharian to pre-Proto-Samoyed specifically, rather than early Uralic more generally, but it is unknown how old exactly the Samoyed prolative case itself is. In principle it could be younger than the Tocharian perlative, so that the connection must remain tentative (6.4).

The Tocharian participial system also deviates from that of the Indo-European languages, as it contains a number of contextually oriented participles. This is most clearly seen in the preterite participle, which descends from an active Indo-European participle formation, but displays a flexible contextual range in both Tocharian languages. The privative (a negative participle) and the gerundives (necessitative or future participles) behave in a similar manner. The two present participle formations of Tocharian stand out in this context. One is an active participle that continues a Proto-Indo-European active voice participle. It commonly functions as an agent noun formation in Tocharian. The other descends from the Proto-Indo-European middle voice participle, but in Tocharian it has for the most part lost its connection with the middle voice. The common use of this participle as an adverbial form in Tocharian makes it

difficult to establish its behaviour as a true participle, but some examples of contextual orientation can be found (7.2.2).

The contextual orientation of most Tocharian participles aligns well with the common contextual orientation of Samoyed participles. The presence of negative and future participle formations also conforms to the standard Samoyed participial system (7.3). Since participle orientation of Uralic languages in the west is influenced by the orientation of surrounding Indo-European languages (Shagal 2018), it seems plausible that there could have been an obverse influence from Uralic/Samoyed participles in the east on Tocharian. However, it should be borne in mind that contextual participial systems are quite common in that part of the word, found for example in Ket, and in some Turkic, Tungusic, and Mongolic languages. This non-Indo-European feature of Tocharian can therefore not strictly be linked to Samoyed/Uralic specifically. It does, however, provide a connection with the Siberian linguistic area, and supports the other features that point to Uralic/Samoyed contact.

In verbal object marking, the Tocharian use of pronoun suffixes is similar to the Uralic objective conjugation, but the comparison is certainly not exact. The elements used to form the Tocharian object suffixes are clearly object pronouns that were suffixed to the finite verb form. The object suffix referring to a first-person singular object descends from a first-person singular clitic pronoun, and similarly with the second and third person singular and the plural.

In the Uralic/Samoyed objective conjugation, on the other hand, the objective suffixes derive from pronouns corresponding to the agent, leaving the presence of a topical object implied. In other words, the suffix referring to a first-person agent and a third person singular object shares phonological material with the first-person pronoun only. The origins and original nature of the Uralic objective conjugation are still debated. After considering the evidence, I concluded that the connection with the possessive declension is best supported (8.3.2.6). This connection is reflected most clearly in Samoyed, Mansi and Khanty, but also in traces in Mordvin. The connection to the possessive declension pertains not only to the shape of the 3sg. objective conjugation suffix *-*sA*, as is commonly acknowledged, but extends also to the marking of the number of the object with plural *-*n*-. As far as I can tell, this second connection between the objective conjugation and the possessive declension has received no attention in the discussion surrounding the origins of the objective conjugation, but I think that it clinches the argument against the hypotheses in which direct suffixation of object pronouns is assumed. The formation of the Uralic objective conjugation was thus very different to the creation of the Tocharian system of verbal pronoun suffixes. However, it is possible that the specific use of the pronoun suffixes in Tocharian as verbal suffixes rather than continued use as more general clitics was inspired by the Uralic/Samoyed objective conjugation.

## 10.5 Lexicon

The number of lexical comparanda between Samoyed and Tocharian remains very low. The most plausible borrowings from Tocharian into Samoyed are PS ±**mänäwjə̂* 'full moon' from pre-PT **mēnē-*, PS **måŋkə̂* 'need, poverty' from pre-PT **monko-* 'lack', and PS **tar-* 'divide, separate' from pre-PT **der-* 'split, separate', of which only the latter is a new proposal (see Warries fthc.). The etymology of PS **wẹ̑n* 'dog' as a borrowing from pre-PT **kwënə-*, the oblique stem of *ku-* 'dog' requires the simplification of **kʷ* to **w*, which has no parallel, but is otherwise plausible. PS **lə̂ŋkə̂* 'slope' is very similar to TB *leṅke* 'slope'. It has been suggested that TB *leṅke* might instead be derived from Iranian, so that the similarity between the Samoyed and Tocharian words is perhaps rather coincidental. However, the Iranian word itself does not have a good etymology, while it can be derived within Tocharian. PS **laŋkå-* 'hang', which is only reflected in Nenets, can be compared to pre-PT **lenk-* 'hang', but the very limited distribution of this word in Samoyed cannot inspire great confidence. The connection of PS **ńuə̂* 'child' with pre-PT *newo-* 'new' works quite well phonologically, but is semantically imperfect.

The connection between PS **säjt³wə̂* 'seven' and pre-PT **septəm* 'seven' is inexact, and the Samoyed word is rather to be derived from PU **ćäjćimä* 'seven'. PS **wäsa* 'metal' and pre-PT **wesa* 'gold' are extremely similar formally, and the difference in meaning may be acceptable if we assume a semantic widening or narrowing in one of the two languages. However, it is unattractive to completely separate PS **wäsa* from words for 'metal' in the other Uralic languages, such as SaaN *veaiki* 'copper', Fi. *vaski* 'copper, bronze', Hu. *vas* 'iron'. The internal correspondences are not entirely regular, pointing to a dispersal of this word among the Uralic branches after Proto-Uralic had split up, but a reconstruction **wäćkä* largely unifies the different forms. Crucially, the **-k-* of **wäćkä* cannot be derived from Tocharian. It is not present in Samoyed **wäsa*, but it would have been regularly lost in the cluster **-ćk-* in any case, so that it is impossible to argue that it was never there. A derivation of the Tocharian word from Samoyed would be phonologically possible, but only after **-ćk-* had been simplified. The full simplification to a single PS **-s-* seems to have been among the later changes in the relative chronology (see 3.9), so that a reverse direction for this loan etymology might be chronologically difficult. Such an interpretation is made further implausible due to the lack of any parallel cases.

All in all, while some connections look very attractive, the lexicon on its own does not provide a truly convincing indication that early Samoyed and early Tocharian were in contact with one another, as the small number of comparanda might be considered coincidental. It also remains possible that the contact between Tocharian and Uralic involved a para-Samoyed language that was generally identical in terms of phonology and morphology to pre-Proto-Samoyed but not the ancestor of Proto-Samoyed itself. If

this para-Samoyed was in contact with both Tocharian and pre-Proto-Samoyed, it is possible that just a few words were passed along from pre-Proto-Tocharian to pre-Proto-Samoyed via this intermediary, before para-Samoyed disappeared.

## 10.6 Evaluation

The results of this investigation are altogether mixed. On the one hand, a number of features that have been adduced as Uralicisms in Tocharian hold up well to closer scrutiny, especially in the phonology, and it is possible to identify more, such as in the participial system. On the other hand, there are some shared features that are widely distributed in the area, and could be attributed to influence from other languages as well, such the agglutinative case system.

Cumulatively, Uralic/Samoyed does provide the most coherent source for innovative Tocharian structural features. For instance, while Turkic also has agglutinative cases and while Ket has contextually oriented participles, these languages also have very different phonological systems when compared to Tocharian. The problem is that some other languages like pre-Proto-Yukaghir cannot be reliably reconstructed far enough back in time to test any Tocharian-Yukaghir contact hypothesis. Thus, if we try to understand Tocharian innovative features as the result of language contact, the hypothesis that the contact language was Uralic/Samoyed has the most explanatory power, and is to be preferred.

The connection with specifically the Samoyed branch of Uralic rests on a very narrow evidentiary basis according to the majority of parameters investigated here. The pre-Proto-Tocharian vowel system may be more in line with a vowel system that can be reconstructed for pre-Proto-Samoyed than with that of Proto-Uralic, but the differences are not very pronounced. In most other relevant aspects, Samoyed is like Proto-Uralic, so that no proper differentiation is possible. The prolative case and the necessitative participle provides a better connection with Samoyed specifically, and the few loanwords, if correctly identified, also do seem to provide a tether to Samoyed. However, in theory they could also stem from a para-Tocharian language that was spoken farther to the north or remained in Siberia when the ancestors of the Tocharians that ended up in the Tarim Basin had already migrated south.

A large number of the innovative typological features of Tocharian can thus be understood as the result of contact with early Uralic/Samoyed. This anchors pre-Proto-Tocharian to the forest area of south Siberia. Especially if the Samoyed connection is taken seriously, pre-Proto-Tocharian should accordingly be moved to the eastern part of south Siberia. Geographically, this matches the Afanasievo-Tocharian hypothesis very well. Chronologically, there are some uncertainties, however. If Proto-Uralic is dated around 2500 BCE, that means that there was no differentiated pre-Proto-Samoyed at the

time that the early Tocharians moved east with the Afanasievo Culture around 3100 BCE. Even if the early Tocharians only moved into the Dzungar basin around 2800, Tocharian-Samoyed contact specifically might have been impossible at such an early date.

However, these time estimates may be subject to change. If Samoyed is more conclusively shown on the basis of phonological and morphological arguments to be the first branch of Uralic to have split off, as per the traditional phylogeny, the dating of Proto-Uralic may shift farther back in time again to accommodate a period of shared Finno-Ugric innovations. A small core of Samoyed specific features that have been adduced seem difficult to entirely ignore in this regard, such as the presence of **o* in **CoCi*, associated with the lack of a merger of PS **tuj* 'fire' and **toj-* 'come', as opposed to **tuli* 'fire' and **tuli-* 'come' in the Finno-Ugric languages. The addition of the suffix **-la* to PU **ńoma* 'hare' could be another Finno-Ugric innovation, although this word is not attested in all branches. This topic requires further research before we can draw our conclusions.

Another possible solution for the chronological uncertainties would be to assume contact between pre-Proto-Tocharian and pre-Proto-Uralic instead. However, without adequate knowledge of pre-Proto-Uralic phonology and morphology, this can hardly be tested concretely at our current state of knowledge. The discussion surrounding the exact positions of the Proto-Uralic and pre-Proto-Uralic homelands would also interact with such a contact scenario.

A Tocharian-Uralic/Samoyed contact hypothesis can thus support the Afanasievo-Tocharian hypothesis and the associated placement of pre-Proto-Tocharian in space, but not necessarily in time. Further research is needed in that regard, and alternative archaeological scenarios with a later migration to the east should be explored if they can at some point be identified in the archaeological and genetic record on the basis of future discoveries.

# Etymological reference list 1: Tocharian

In the following numbered list, the words adduced and discussed in chapter 2 are given with references to the literature, and, where relevant, a short discussion. The main source of Tocharian B etymological information is Adams (2013), which contains references to the earlier literature. Especially Malzahn (2010) and Peyrot (2013) also provide a lot of information on verbs, which are cited according to the system of Peyrot (2013), rather than those employed by Malzahn and Adams. Tocharian A lacks a dedicated etymological dictionary. The Tocharian A forms are cited after the Dictionary and Thesaurus of Tocharian A by Carling and Pinault (DThTA).

The words are listed according to the Proto-Indo-European reconstruction, or the Proto-Indo-European root from which the Tocharian word is derived. As elsewhere in this dissertation, Proto-Indo-European accents are omitted because they cannot always be determined with certainty, and they are generally irrelevant for the topics under discussion. The alphabetical order is:

*bʰ d dʰ e ē ǵ ǵʰ g gʰ gʷ gʷʰ h₁ h₂ h₃ i ḱ k kʷ l m n o ō p r s t u*

Words do not start with the vowels **e*, **ē*, **o*, **ō*, or with the resonant **r*. Unlike some scholars, I do not write vowel colouring in the Proto-Indo-European reconstructions, so that, e.g., **h₂e* remains written as **h₂e* rather than becoming **h₂a*. I consider both the laryngeal colouring and syllabification of syllabic resonants to be post-Proto-Indo-European developments. In this list, the sign marking syllabic laryngeals **h̥₁*, **h̥₂*, **h̥₃* and syllabic resonants **l̥*, **m̥*, **n̥*, **r̥* and the sign for the glides **i̯* and **u̯* are omitted.

1. PIE **bʰeh₂ǵʰu-* > PT **poko*, obl.sg. **pokay* 'arm'
   TB *poko**, obl.sg. *pokai* | TA nom/obl.sg. *poke* (DThTA: 285b)
   Adams 2013: 434.

2. PIE **bʰreh₂tēr* > PT **protʸer* 'brother'
   TB *procer* | TA *pracar* (DThTA: 303b)
   Adams 2013: 454–455.

3. PIE *$b^hrǵ^h$-ro- > PT *pərk(ə)re 'long'
   TB *pärkare* /pərkə́re/ | TA *pärkär* (DThTA: 278a)
   Adams 2013: 399-400.

4. PIE *$deḱm$ > PT *śəkə 'ten'
   TB *śak* /śə́k/ | TA *śäk* (DThTA: 476a)
   Adams 2013: 674–675.

5. PIE *$demh_2$- > PT *$t^s$əma- 'grow, increase'
   TB *tsəma-* | TA *tsämā-* DThTA: 555b)
   Adams 2013: 804; Malzahn 2010: 985–987; Peyrot 2013: 843.

6. PIE *der- 'split' > PT *$t^s$ər- 'be separated, be apart; caus. separate'
   TB *tsəra-* | TA *tsärā-* 'separate' (DThTA: 556a)
   Adams 2013: 805; Malzahn 2010: 987–988; Peyrot 2013: 845.

7. PIE *$deuh_2$- > PT *$t^s$əwa- 'attach oneself, stick to (itr.)'
   TB *$t^s$əwa-* | TA *$t^s$äwā-* 'fit; obey' (DThTA: 558b)
   Adams 2013: 808–809; Malzahn 2010: 996–998; Peyrot 2013: 846.

8. PIE *deuk- > PT *$t^s$əwka- 'suck (out)'
   TB *tsəwka-* | TA *tsäwkā-* (DThTA: 558b)
   Adams 2013: 809; Malzahn 2001: 994–996; Peyrot 2013: 841.
   — The etymology *$d^heug^h$- 'produce (milk)', while not impossible, is semantically less fitting than a derivation from *deuk- 'pull (out)'; see Adams (2013: 809).

9. PIE *$dnǵ^hueh_2$ > PT *kəntwo 'tongue'
   TB *kantwo* /kə́ntwo/ | TA *käntu* (DThTA: 121a)
   Adams 2013: 147.
   — The dental and the velar were metathesised in Tocharian, with TB *kantwo* rather than **tankwo*, TA *käntu* rather than **tänku*.

10. PIE *doru-, *dreu- > PT *orə, pl. *arwa* 'wood'
    TB *or*, pl. *ārwa* | TA *or* (DThTA: 86b)
    Adams 2013: 127.

— The initial **d-* was lost by regular sound law in the oblique cases where it came into contact with the **r*, and it was subsequently removed in other forms by analogy. The singular TAB *or* shows *u*-umlaut of pre-PT **e* (PIE **o*) plural TB *ārwa* shows *a*-umlaut of pre-PT **e* (PIE **o*).

11. PIE **drem-* 'run' → ? **drem-wor* > PT *rəmer* 'quickly'
TB *ramer* | ? TA der. *ymār* (DThTA: 379b)
Adams 2013: 571–572; Pinault 2011: 168; Van Windekens 1976: 592; Winter 1962: 30.
— TB *ramer* could theoretically be from **drem-or* if this is simplified from **drem-wor*. Pinault (2011: 168) explains TA *ymār* as a local adverb formation from PT **rəmer* with prefixation of **yə*(*n*)- and pluralization in *-a*, yielding pre-TA **yrmārā* with umlaut; after regular apocope and dissimilation of the two *r*'s, this became TA *ymār*. In this way, TA *ymār* would not be a reflex of **h₁i-mōr*, as assumed by Van Windekens (1976: 592).

12. PIE **duei-* > PT **wəy-* 'be frightened'
TB *wəy-* | TA *wäyā-* (DThTA: 446a)
Adams 2013: 652; Malzahn 2010: 900–901; Peyrot 2013: 818.

13. PIE **dui-dḱmt* → ?**duih₁ḱmt* > PT **wʸikən* 'twenty (20)'
TB *ikäṃ* /yə́ykən/ | TA (analogical) *wiki* (DThTA: 443a)
Adams 2013: 66–67
— It is generally assumed that TB *ikäṃ*, TA *wiki* can be derived via pre-PT **wīkəmt* from a preform **duih₁kmt* with a development from PIE **ih₁* to pre-PT **ī* (Adams 2013: 66–67; Hackstein 2017: 1314), but the reconstruction of the**h₁* to the Proto-Indo-European level is uncertain. The development thus remains a tentative approximation that cannot strictly be used to show the development of true PIE **ih₁*. At the same time, there does not seem to be any other known way to get the TAB *i* than by assuming a development of **dui-dḱmt* via **duih₁kmt* at some stage.

14. PIE **dui-to-* > PT **wəte* 'second'
TB *wate* /wə́te/ | TA *wät* (DThTA: 433a)
Adams 2013: 625–626.

15. PIE **duoh₁, *duoi* > PT **wu, *wey* 'two (2)'
TB *wi* (m/f.) | TA *wu* (m.), *we* (f.) (DThTA: 446b)
Adams 2013: 651.

16. PIE **dʰegʷʰ-* > PT **tˢək-* 'burn (act. = tr.; mid. = intr.)'
TB *tsək-* | TA *tsäk-* (DThTA: 553b)
Adams 2013: 802; Malzahn 2010: 980–981; Peyrot 2013: 842.

17. PIE **dʰeigʰ-* 'form' > PT **tˢika-* 'form, shape'
TB *tsəyk-* | TA *tsāykā-* 'form tr.' (DThTA: 557b)
Adams 2013: 807; Malzahn 2010: 991–992; Peyrot 2013: 844.

18. PIE **dʰeih₂gʷ-* → *dʰih₂gʷ-* > PT **tˢaka-* 'pierce, bite'
TB *tsaka-* | TA *tsākā-* 'pierce' (DThTA: 552b)
Adams 2013: 800; LIV: 142; Malzahn 2010: 975–976; Peyrot 2013: 840.
— The laryngeal in this root is only reconstructed for the benefit on the Tocharian, and is not necessary for Latin *fīgō* 'drive in, insert; fasten' or Lithuanian *díegti* 'plant; sting', for which **dʰeigʷ-* would work too (De Vaan 2008: 219; Derksen 2015: 127). However, the other etymologies provided by Adams do not work phonologically for Tocharian: PIE **dē(n)ḱ-nh₂-* would have become ***śakna-*, **tēgʰ-nh₂-* should have become ***cakna-*, and the lengthened grades are difficult to account for, so that a connection with PIE **dʰeih₂gʷ-* remains the best option.

19. PIE ?**dʰengʰ-* > ? PT **tˢənk-* '(a)rise'
TB *tsənka-* | TA *tsänk-* (DThTA: 554b)
Adams 2013: 802–803; Malzahn 2010: 983–984; Peyrot 2013: 834.
— PT **tˢənk-* has also been connected with a PIE **dʰengʰ-* 'reach' (see Adams 2013: 802-803 with references). However, the supposed cognate Skt. *daghnóti* 'reach toward' is rather taken together with Gr. φθάνω 'anticipate, be ahead' from **dʰegʷʰh₂-* 'reach toward' in LIV (134-135), thus, crucially, without an internal nasal; see also Malzahn 2010: 983–984. Other supposed cognates OIr. *daingen* 'strong, firm', Russian (dial.) *djágnutь* 'get well, recover', *djáglyj* 'healthy, strong', and Latvian *dęñkts* 'strong, healthy, important' rather point to a root **dʰeng-* on account of the accentuation (Derksen 2008: 106), and do not fit with Tocharian semantically. The Proto-Indo-European antecedent of this Tocharian verb is thus very uncertain.

20. PIE *$d^{h}eub^{h}$- → *$d^{h}ub^{h}ro$- ‘deep’ > PT *$təpre$ ‘high’
TB *tapre* /təpre/ | TA *tpär* (DThTA: 221a)
Adams 2013: 296–297.
— This word is traditionally reconstructed as *$d^{h}ubro$-, but Kroonen (2011: 253) has shown that the root should be *$d^{h}eub^{h}$- on account of the Germanic evidence.

21. PIE *$d^{h}ugh_{2}tēr$ > PT *$təkat^{y}er$ ‘daughter’
TB *tkācer* /tkácer/ | TA *ckācar* (DThTA: 188b)
Adams 2013: 330–331.
— The initial *c*- of TA *ckācar* is the result of assimilation to the internal -*c*-.

22. PIE *$ǵenh_{1}$- ‘be born’ > PT *$kən$- ‘come about’
TB *kən*- | TA *kän*- (DThTA: 120a–121a)
Adams 2013: 169–170; Friis 2024: 373; Malzahn 2010: 568–570; Peyrot 2013: 731.
— Friis has adduced the optative TA *knitär* as an example where PIE **i* failed to palatalize the preceding consonant, potentially due to the intervening laryngeal: *$gnh_{1}$-$ih_{1}$- (Friis 2024: 373).

23. PIE *$ǵneh_{3}$- > *$kna$- ‘know’
TB *nana*- ‘appear : erscheinen’ | TA *knā*- (DThTA: 158a)
Adams 2013: 3; Malzahn 2010: 610–611, 678–679, Peyrot 2013: 440–442, 762.
— PT *$kna$- could in principle reflect either the *e*-grade of the root *$ǵneh_{3}$-, in which it would be one of the few examples of *$eh_{3}$ in Tocharian – and all examples are uncertain in some way – or it could reflect the zero-grade *$ǵnh_{3}$- with the much more commonly attested vocalization of laryngeals to PT **a*. The latter option seems more likely, although the former cannot be entirely excluded (see Peyrot 2013: 440–442 for a more detailed discussion).

24. PIE *$ǵomb^{h}o$- > PT *$keme$ ‘tooth’
TB *keme* | TA *kam* (DThTA: 95a)
Adams 2013: 208.

25. PIE $g^{(h)}ul$- > PT *$k^{w}əla$- ‘fail, weaken, recede’
TB *$k^{w}əla$*- | TA *$k^{w}älā$*- (DThTA: 144a)
Adams 2013: 256; Kim 1999: 153.

— The most probable of the proposed cognates are to be found in Baltic, e.g., Lith. *gulĩti* 'lie down', *guléti* 'lie, be ill' (see Adams 2013: 256, Kim 1999: 153; though the Tocharian connection is not mentioned by Derksen 2015: 193).

26. PIE *$*g^{w}eh_{3}$-u-* > PT **kew* 'cow'
TB *$ke_{u}$* | TA *ko* (DThTA: 151a)
Adams 2013: 201–202; Lubotsky 1990: 133-134; Beekes 2010: 232–233; Nielsen Whitehead 2018.

— The precise Proto-Indo-European reconstruction is debated. If we depart from *$*g^{w}eh_{3}$-u-*, one might theoretically expect *$*eh_{3}$* to yield a long vowel like pre-PT *$*\bar{o}$*, but the *e* in TB *$ke_{u}$* clearly reflects a pre-PT short **o*. The particular development of *$*h_{3}$* in combination with **e* in *$*g^{w}eh_{3}$-u-* is mirrored in TB *kaw-* 'kill' from PIE *$*keh_{2}u$-* 'hew, strike', with laryngeal colouring of the *$*eh_{2}$* to *$*ah_{2}$* but no further lengthening to *$*\bar{a}$*. The traditional reconstruction is *$*g^{w}ou$-*, but this foregoes the attractive connection with the verb *$*g^{w}eh_{3}$-* 'feed, tend', clashes with the short *a* in Rigvedic forms like dat.sg. *gáve* where lengthening of **o* to ***$\bar{a}$* should have occurred in accordance with Brugmann's Law, and leaving the accent of Greek βοῦς unexplained (Lubotsky 1990: 133–134; Beekes 2010: 232–233). Nielsen Whitehead reconstructs an original hysterodynamic paradigm with a nom.sg. *$*g^{w}h_{3}$-ēu(-s)*, acc.sg. *$*g^{w}h_{3}$-eu-m* and a consistent initial cluster *$*g^{w}h_{3}$-* (Nielsen Whitehead 2018). If such a cluster underwent delabialization and the laryngeal blocked palatalization (see 2.4.2), a reconstructed *$*g^{w}h_{3}$-ēu(-s)* might regularly have become TB *$ke_{u}$*. With a change from *$*h_{3}e$* via **o* to PT **e*, *$*g^{w}h_{3}$-eu-m* might also provide the basis for TB *$ke_{u}$*. There are thus many ways to interpret the development of 'cow' in Tocharian depending on the specific reconstruction taken as a point of departure, each involving theoretically possible or even plausible changes that are not directly paralleled elsewhere in Tocharian.

27. PIE *$*g^{w}enh_{2}$-* > PT *$*k^{y}$ənа* 'wife'
TB *śana* /śə́na/ | TA *śäṃ* (DThTA: 466b)
Adams 2013: 676–677.

28. PIE *$g^{w}ih_{3}u$- > PT *$k^{y}aw$- ‘live’
TB *śaw-*, *śay-* | TA *śāw-* (DThTA: 466a)
Adams 2013: 683–684; Malzahn 2010: 916–119; Peyrot 2013: 825.

29. PIE *$g^{w}m$-*sk*- > PT *$k^{w}$*əmnəsk*- ‘come’ (present stem)
TB *kənməsk*- | TA *$k^{w}$ämnäs*- (DThTA: 122a)
Adams 2013: 170–171; Malzahn 2010: 571–572; Peyrot 2013: 732.

30. PIE $g^{w}reh_{2}$- → *$g^{w}rh_{2}$-*mr*- > PT **kramər* ‘heaviness’, der.adj. *kramərtˢtˢe* ‘heavy’
TB *krāmär* /krámər/, der.adj. *kramartstse** /kramártˢtˢe/ | TA der.adj. *krāmärts* (DThTA: 161a)
Adams 2013: 230–231; Kroonen 2013: 312; Ringe 1996: 8.
— The precise derivation of this word from the root PIE *$g^{w}reh_{2}$- ‘be heavy’ is unclear, as there is no exactly corresponding formation elsewhere in Indo-European. The reconstruction with a zero-grade, *$g^{w}rh_{2}mr$-, is attractive because we know that a vocalized laryngeal became PT **a*. An An *o*-grade as in the reconstruction *$g^{w}roh_{2}$-*mr*- (Ringe 1996: 8) would be unexpected, and the development of **oH* in Tocharian is quite unclear (see subsection 2.4.5).

31. PIE *$g^{w}rh_{2}$-*uon*- > PT **kərwen$^{y}$e* ‘rock, stone’
TB *kärweñe**, pl. *kärweñi* /kərwéñi/ | TA *kärwaṃ** (DThTA: 125b)
Adams 2013: 176; Ringe *apud* Kim 1999: 151 fn. 26.
— The lack of an initial labiovelar in Tocharian A may be due to a dissimilation with the following *-w-* (Ringe *apud* Kim 1999: 151 fn. 26). The laryngeal was not vocalized in this position, see also 2.3.1.

32. PIE *$g^{wh}en$- ‘hit’ → *$g^{wh}n$-*sk*- > PT **kəsk*(*a*)- ‘scatter, strike apart’
TB *kəsk*(*a*)- ‘scatter’ | TA *käskā*- (DThTA: 134a)
Adams 2013: 189; LIV: 218–219; Malzahn 2010: 596; Peyrot 2013: 739.

33. PIE $h_{1}eg^{wh}$- → ±*$h_{1}e$-$h_{1}g^{wh}$- > PT **yok*- ‘drink’
TB *yok*- | TA *yok*- (DThTA: 373a)
Adams 2013: 551–552; Kim 2000: 153-155; Kloekhorst 2008: 236–237; Malzahn 2010: 810–811; Peyrot 2013: 801.
— Kim instead reconstructs a paradigm *$h_{1}ēh_{2}g^{wh}$- : *$h_{1}eh_{2}g^{wh}$- > PT **yek$^{w}$*- : **ok$^{w}$*-, from which **yok*- resulted by a crossing of the two forms, dismissing the possibility of secondary rounding from **e* to **o* due to *$k^{w}$ as required by the

reconstruction *$h_1eh_1g^{wh}$- (Kim 2000: 153-155). The Anatolian cognate can be better understood departing from *$h_1eg^{wh}$- : *$h_1g^{wh}$-, however, with no place for the *$h_2$ added by Kim (see Kloekhorst 2008: 236–237), so that a reduplicated reconstruction of *$h_1eh_1g^{wh}$- for Tocharian and subsequent rounding from *$yek^w$- to PT *$yok$- seems preferable to me.

34. PIE *$h_1ei$- → imp. *$h_1id^hi$ > PT imp. *$pəṣə$ 'go!' (*$p$- marks the imperative)
TB *paṣ* /páṣ/ | TA *piṣ*
Jasanoff 1987: 106–112.
— The TA vowel *i* does not correspond to TB *a* /á/, and would have to be analogical on the basis of the corresponding verb root *y*- 'go'.

35. PIE *$h_1ek\acute{u}o$- > PT *$yək^we$ 'horse'
TB *yakwe* /yákwe/ | TA *yuk* (DThTA: 370b)
Adams 2013: 528–529.

36. PIE *$h_1esh_2$-*r* > PT *$yəsar$ 'blood'
TB *yasar* /yásar/ | TA *ysār* (DThTA: 382b)
Adams 2013: 525; Beekes 2010: 366; Kloekhorst 2008: 256-260.
— Adams derives this from a "collective" *$h_1esh_2ōr$, but PT *$yəsar$ matches the "non-collective" *$h_1ésh_2r$ quite exactly, and it is the latter that has direct correspondences in other Indo-European languages, such as Hitt. *ēšḫar*, Gr. ἔαρ, etc. (Kloekhorst 2008: 256-260; Beekes 2010: 366)

37. PIE *$h_1i$-*tr*- → ?*$h_1itōr$ > PT *$yətar$ 'path, road, way'
TB der. *ytārye* /ytárye/ | TA *ytār* (DThTA: 376a)
Adams 2013: 559; Kortlandt 1988b: 84–85.
— The precise reconstruction of the suffix is difficult, since non-Tocharian cognates like Latin *iter* show a zero-grade. Such a reconstruction will not work for Tocharian: *$h_1itr$ should yield **$yətər$ with a /ə/ rather than /a/ in the second syllable. This /a/ needs to be accounted for in some way. Adams reconstructs a collective *$h_1itōr$, which is dubious due to the lack of other evidence for this formation. Kortlandt (1988b: 84–85) takes the *-*ar* as analogous after PT *$yəsar$ 'blood' (TB *yasar*, TA *ysār*), replacing expected **$yətər$. The reason for this analogy is rather unclear, however, as there is no other connection between the two words, aside from a superficial similarity.

38. PIE **h₁leng$^{wh}$-* > PT **lənk-* 'hang, dangle'
TB *ləṅk-* | TA *läṅk-* (DThTA: 399b)
Adams 2013: 598 (**leng-*); Malzahn 2010: 839–841; Peyrot 2013: 476, 808.

39. PIE **h₁leng$^{wh}$-* ?→ **h₁long$^{wh}$o-* > PT **lenke* 'valley, cleft'
TB *leṅke*
Adams 2013: 608.
— Might be derived from the verb for 'to hang' with semantics similar to German *Hang* 'slope, hill', *Abhang* 'slope'. Chams Bernard (p.c., 28-03-2024) instead suggests a loanword from Old Steppe Iranian **lænækæ*, as reflected in Ossetic Iron *lænk* and Digoron *læncæ* 'ravine', although this word itself does not have an etymology.

40. PIE **h₁leng$^{wh}$-* → **h₁lng$^{(w)h}$u-* → ** h₁lng$^{(w)h}$u-tio* > PT **lənk$^{w}$tse* 'easy, light (not heavy)'
TB *laṅk$_{u}$tse* /lə́nk$^{w}$tse/
Adams 2013: 590–591.
— To be derived from a *u*-stem **h₁l̥ng$^{(w)h}$u-* also seen in Gr. ἐλαχύς 'small' and Skt. *laghú-* 'light', *raghú-* 'swift' rather than directly from the root **h₁leng$^{wh}$-*. This accounts for the labiovelar reflex.

41. PIE **h₁neḱ-* → **h₁nḱ-* > PT **enk-* 'seize, take'
TB *eṅk-* | TA *eṃtsa-* (DThTA: 70a)
Adams 2013: 81–82; Malzahn 2010: 538–539; Peyrot 2013: 725.

42. PIE **h₁reg$^{w}$-* → **h₁rg$^{w}$-mō-* or **h₁rg$^{(w)}$-u-mō-* > PT **orkəmo* 'dark; darkness'
TB *orkamo* /orkə́mo/ | TA *orkäm* (DThTA: 86b)
Adams 2013: 128.
— The derivation of PT **orkəmo* via **h₁rg$^{(w)}$-u-mō-* rather than directly from **h₁rg$^{w}$-mō-* was suggested to me by Professor Lubotsky (p.c., 15-10-2024). It would account for the vowel TAB *o* with *u*-umlaut instead of by *o*-umlaut caused by PIE **ō*, which has no good parallels. See the discussion at the end of 2.5.3.

43. PIE **h₁reg$^{w}$-* → **h₁rg$^{w}$-ont-* > PT **erkent* 'black'
TB m.obl.sg. *erkent* | TA m.obl.sg. *arkant* (DThTA: 15b)
Adams 2013: 101.

44. PIE *$h_1rud^h$-ro- > PT *rətre 'red'
TB *ratre* /rə́tre/ | TA *rtär* (DThTA: 394b)
Adams 2013: 570–571.

45. PIE *$h_2eǵ$- > PT *ak- 'lead, guide, drive'
TB *ak-* | TA *āk-* (DThTA: 23a)
Adams 2013: 38–39; Malzahn 2010: 519–520; Peyrot 2013: 720.

46. PIE *$h_2eḱu$- > PT der. *$ak^wətˢe$ 'sharp'
TB *akwatse* /ak$^w$ə́tˢe/
Adams 2013: 5; Pinault 2008: 445.

47. PIE *$h_2enh_1$- → *$h_2enh_1$-ske/o- > PT *anask- 'inhale'
TB *anask-*
Adams 2013: 12–13; Malzahn 2010: 524; Peyrot 2013: 721.

48. PIE *$h_2enh_1$- 'breathe' > *ana- > der. *ana-elme* > PT *ånålme- 'living being'
TB *onolme*
Adams 2013: 121; Peyrot 2010: 72.

49. PIE *$h_2erǵ$- → *$h_2erǵu$- > PT *$ark^wi$ 'white'
TB *ārkwi* | TA *ārki* (DTHTA: 43b)
Adams 2013: 52–53.

50. PIE *$h_2nt$-$b^hoh_1$ > PT *antpu- 'both'
TA f. *āmpuk* (DThTA: 39a)
Hilmarsson 1989b: 56-58; Ringe 1996: 89
— The feminine gender of TA *āmpuk* is unexpected with its derivation from PIE *$h_2(e)ntb^hoh_1$, see Hilmarsson (1989b: 56-58). The masculine form is TA *āmpi*. Cf. TB *antapi* /antəpəy/ ~ *āntpi* /ántpəy/

51. PIE *$h_2og$-$eh_2$-, or *$h_2eug$- → ?*$h_2ugeh_2$ > PT *oko 'fruit; result, effect'
TB *oko*
Adams 2013: 115; Del Tomba 202: 190.
— TA *oko* 'id.' is borrowed from Tocharian B (DThTA: 78b). The derivation from the root PIE *$h_2eug$- 'grow' to *$h_2ugeh_2$- is proposed by Del Tomba (2020: 190),

while *$h_2ogeh_2$ has parallels in Lith. *úoga* 'berry', OCS *agoda* 'fruit', etc., and more distantly Goth. *akran* 'fruit', Eng. *acorn*, etc. (see Adams 2013: 115). Both PIE *$h_2ogeh_2$ and ?*$h_2ugeh_2$ would likely give the same result in Tocharian, as we find both *o*-umlaut of pre-PT **e* (from PIE **o*) and of **u*, cf. quasi-PIE **someh$_2$m* > TB *somo* 'one:F.OBL.SG' and PIE **uksōn* > TB *okso* 'ox'.

52. PIE *$h_2$ster- > PT der. **ścərəye* 'star'
TB *ścirye* | TA *śre**, nom.pl. *śreñ* (DThTA: 482b)
Adams 2013: 701.

53. PIE *$h_2$ueh$_1$nto- > PT **w$^y$ente* 'wind'
TB *yente* | TA *want* (DThTA: 423b)
Adams 2013: 546.

54. PIE *$h_3$emso- > PT **anse* 'shoulder'
TB *āntse* /ánt$^s$e/ | TA *es* (DThTA: 74a)
Adams 2013: 46; Beekes 2010: 1679–1680; Kloekhorst 2008: 178.
— Initial *$h_2$ could be reconstructed as one way to account for the Tocharian **a-* (Kloekhorst 2008: 178), but that would require an unexpected *o*-grade in Greek ὦμος. *$h_3$ would cause a development to **o* from the *e*-grade *$h_3$emso-, and is therefore reconstructed by Beekes (2010: 1679–1680). The development of initial *$h_3$e- in Tocharian is uncertain due to a lack of reliable examples, and thus could in principle be assumed to be *a-* just like *$h_2$e-; cf. *$h_3$ewis* 'sheep' for the same issue. However, laryngeal colouring of PIE **e* to pre-PT **o* does seem to be found in PIE *$g^weh_3u$- to TB *ke$_u$* 'cow' (q.v.) and in *$h_3$eros- to TB *ere* 'form, appearance' (q.v.), if this does not reflect *$h_3$oro- instead, so that it difficult to decide either way. Adams assumes **Hōmso-*, with the very uncertain development from **ō* to **a* (see subsection 2.4.5). Both the precise reconstruction and the phonological development of this word are too uncertain to use it in further argumentations.

55. PIE *$h_3$er- 'rise up' → *$h_3$er- or *$h_3$or- > PT **er-* 'evoke, cause'
TB *er-* | TA *ar-* (DThTA: 14a)
Adams 2013: 98–99 (**h$_1$or-*); Kloekhorst 2008: 200; Malzahn 2010: 541–542; Peyrot: 2013: 726.

— Tocharian has a causative meaning of the Indo-European root. The ablaut grade from which PT **er-* is derived is unclear, as this could in principle reflect either **$h_3$er-* or **$h_3$or-*. Adams argues that the Hittite cognate precludes a reconstruction of **$h_3$*, and thus reconstructs **$h_1$or-* instead, but see Kloekhorst (2008: 200) in favour of **$h_3$*.

56. PIE **$h_3$er-* → **$h_3$er-os-* or **$h_3$or-o-* > PT **ere* 'form, appearance'
TB *ere*
Adams 2013: 99; Del Tomba 2020: 84.

— The *s*-stem PIE **$h_3$eros-* derived from the verb **$h_3$er-* is paralleled in Greek ὄρος 'mountain' and Hittite *ḫaršar*, *ḫaršn-* 'head; person; front; beginning'. For Tocharian, a thematic **$h_3$oro-* from the same root would phonologically have the same outcome. If initial **$h_3$e-* is taken to become TB *a-*, as appears to be the case in TB *āntse* /ánt$^{s}$e/ from **$h_3$emso-* (q.v.) and TB *ā$_u$w* 'ewe' from **$h_3$eui-* (q.v.), the reconstruction **$h_3$eros-* would not be possible for TB *ere* and thus **$h_3$oro-* should be preferred. The validity of the change **$h_3$e-* to TB *a-* remains uncertain, however. The thematic inflection of TB *ere* with an obl.pl. *ereṃ* can be secondary and does not provide conclusive proof for a reconstruction **$h_3$oro-* over **$h_3$eros-* (Del Tomba 2020: 84). The reconstruction **$h_3$eros-* has better parallels in the other Indo-European languages.

57. PIE **$h_3$eui-* 'sheep' > PT **awə*, pl. *awəyə* 'ewe'
TB *ā$_u$w* /áw/, pl. *awi* /awə́y/
Adams 2013: 38; Beekes 2010: 1061; De Vaan 2008: 437–438; Kloekhorst 2008: 338; Lubotsky 1990: 130.

— Alternatively reconstructed with initial **$h_2$-* specifically to account for the Tocharian **a-* (e.g., Adams 2013: 38). The argument that Lycian **χawa-* 'sheep' would also require **$h_2$* has been addressed and rejected by Kloekhorst (2008: 338). Initial **$h_3$* would accord well with the reflexes in the other Indo-European languages, especially the initial *h-* of Armenian *hoviw* 'shepherd' from **$h_3$eui-peh$_2$-* (Lubotsky 1990: 130). The general assumption is that **$h_3$e-* yields Tocharian **o-* rather than **a-*, but there are only very few possible examples of that development, and none seem to be completely reliable. The *e*-grade expected in an *i*-stem would not yield the *o*-vocalism of Latin *ovis* 'sheep', Greek ὄϊς 'sheep' with **$h_2$* instead of **$h_3$* (see De Vaan 2008: 437–438, Beekes

2010: 1061), and an *o*-grade could not regularly yield Hitt. *ḫāu̯i-* 'sheep' (*$*h_3o$* > Hitt. /ʔā/; see Kloekhorst 2008: 75), Skt. *ávi-* (with Brugmann's law, *$*h_3oui$-* should have regularly become Skt. ***ā́vi-*; see Lubotsky 1990: 130). A reconstruction *$*h_3eui$-* thus best accounts for all reflexes except the Tocharian one, where it would require the assumption that *$*h_3e$-* became PT **a-*. On this same issue, see the discussion under PIE *$*h_3emso$-* 'shoulder' above.

58. PIE *$*h_3ok^w$-s-* > PT **ek* 'eye'
TB *ek** | TA *ak* (DThTA: 1b)
Adams 2013: 78–79.

59. PIE *$*ḱleu$-* → *$*ḱleu-mō$-* > PT *$*kl^yomo$* 'noble'
TB *kłyomo* | TA *klyom* (DThTA: 171a)
Adams 2013: 250.

60. PIE *$*ḱmtom$* (*$*h₁ḱmtom$/*$*dḱmtom$*)> PT **kənte* 'hundred (100)'
TB *kante* /kə́nte/ | TA *känt* (DThTA: 121a)
Adams 2013: 146–147.

61. PIE *$*ḱuō$*, acc. *$ḱuonm$* > PT *$*k^{(w)}u$*, obl. *$*k^wenə$* 'dog'
TB *ku, kweṃ* | TA *ku*, obl. *koṃ* (DTHTA: 136b)
Adams 2013: 190.

62. PIE *$*kedh_2$-*, prs. *$*kdnh_2$-/*$*kndh_2$-* > PT **kəta-*, prs. **kənta-* 'strew'
TB *kəta-, kənta* (>> *kətna-*) | TA *kätā-, känā-* (DThTA: 119a)
Adams 2013: 166–167; Malzahn 2010: 564–565; Peyrot 2013: 730.

63. PIE *$*keh_2u$-* 'strike' > PT **kaw-* 'kill'
TB *kaw-* | TA *kāw-*
Adams 2013: 222; Malzahn 2010: 606-607; Peyrot 2013: 728.

64. PIE **kʷekʷlo-* 'wheel; circle' > PT **kʷəkʷ(ə)le* 'cart, wagon, chariot'
TB *kokale* | TA *kukäl* (DThTA: 137a)
Adams 2013: 214.

65. PIE **kʷel-* 'turn' → **kʷolo-* > PT **kele* 'navel'
TB *kele*
Adams 2013: 211.

66. PIE **kʷetuores* > PT **kʸətwerə* 'four (4)'
TB *śtwer* | TA *śtwar* (DThTA: 478b)
Adams 2013: 703.

67. PIE **kʷi-so* > PT **kʷəse* 'who'
TB *kᵤse* | TA *kus* (DThTA: 146a)
Adams 2013: 199–201.

68. PIE **kʷo-* → ±**kʷo-tos* > PT **kete* 'whose, to whom, for whom'
TB *kete*
Adams 2013: 203-204.

69. PIE **kʷreih₂-* → **kʷrih₂-wor* > PT **kʷəryår* 'commerce, trade'
TB *karyor* /kə́ryor/ | TA *kuryar* (DThTA: 143a)
Adams 2013: 153.
— See the discussion in subsection 2.3.3.2.

70. PIE **kʷrih₂-*, prs. **kʷrinh₂-* > PT **kərya-*, prs. **kərna-* 'buy'
TB *kərya-*, prs. *kərna-*
Adams 2013: 174–175; Malzahn 2010: 577; Peyrot 2013: 734.

71. PIE **leuk-* > PT der. **ləkʷtse* 'bright, light'
TB *lakᵤtse* /lə́kʷtˢe/
Adams 2013: 589.

72. PIE **leuk-* → **leuk-o(s)-* > PT **lʸuke* 'shine, splendour'
TB *lyuke*
Adams 2013: 616.

73. PIE **limn-* > PT **lʸəmə* 'lake'
TB *lyam* /ĺəm/ | TA *lyäm* (DThTA: 410b)
Adams 2013: 614.

74. PIE ? **lug-lo-* > PT **ləkle* 'pain, sorrow'
TB *lakle* /ləkle/
Adams 2013: 589–590.
— There is no exact corresponding formation in other Indo-European languages, but Gr. λευγαλέος 'wretched, unhappy' and λυγρός 'id.' are close. If assimilation of **r* to **l* in Tocharian is assumed, a PIE **lugró-* could potentially have yielded both TB *lakle* and Gr. λυγρός (cf. Adams 2013: 589–590), but this assimilation in Tocharian would be without parallels and thus uncertain.

75. PIE **luHs*, pl. **luHs-h₂* > PT obl.sg. *luwa*, pl. **lwasa* 'animal'
TB *luwo*, obl.sg. *luwa*, pl. *lwāsa* | TA *lu*
Del Tomba 2019: 131–133; differently Adams 2013: 606–607.
— Del Tomba (2020: 131–133) has argued that the nominative singular *luwo* is an analogical creation replacing expected nom.-obl.sg. *luwa* on the basis of the names of other animals that show nom.sg. *-o* ~ obl.sg. *-a*. The meaning 'louse' found in the Germanic **lūs* and Celtic **luwā* probably developed from the meaning 'animal' of the Tocharian. Cf. Irish *míol* 'animal, insect, creature; louse' from Old Irish *míl* 'id.' next to Scottish Gaelic *mial* 'louse'. In Scottish Gaelic the archaic meaning 'animal' is preserved only in certain cases such as the name of a sea monster, *mial-mhòr a' chuain* lit. 'great *mial* of the sea', where *mial* is not a 'louse'.

76. PIE **medʰ-u-* > PT **mʸətə* 'honey'
TB *mit*
Adams 2013: 494.

77. PIE **meh₁n-s-* → *meh₁n-ē(n)-* > PT **mʸenʸe* 'moon, month'
TB *meñe* | TA *mañ* (DThTA: 329b)
Adams 2013: 503.

78. PIE *-*mh₁no-* > PT *-*mane* present participle suffix
TB *-mane* | TA *-māṃ*
Klingenschmitt 1975.

79. PIE **neḱ-* → *neḱu-*, gen.sg. *nḱuos* > PT **enkʷe* 'man'
TB *eṅkwe* | TA *oṅk* (DThTA: 80a)
Adams 2013: 82–83.

80. PIE **neud-* → **nud-sk-* > PT **nətk-* 'hold off, push away'
TB *nətka-* | TA *nätk-*
Malzahn 2010: 682–683; Adams 2013: 356–357; Peyrot 2013: 764.

81. PIE **-nti* > PT **-nʸtʸə* prs.act.3pl. verbal ending
TA *-ñc*
Malzahn 2010: 35–36; Peyrot 2013: 415
— Both Malzahn (2010: 35–36) and Peyrot (2013: 415) with ample further references.

82. PIE **nokʷt-* → **nokʷt-u-* > PT **noktə* 'night'
TA adv. *nokte* 'last evening', adv. *noktiṃ* 'in the evening' (DThTA: 257a, 257b)
Pinault 1990: 184–190; Pinault 2008: 424.

83. PIE **nokʷt-* → **nokʷt-eu-io-* > PT **nektʸəwʸe* 'at night, last night'
TB *nekciye* /nekcíye/ | TA *nakcu* (DThTA: 239a)
Adams 2013: 363; Pinault 1990: 184–190; Pinault 2008: 424.

84. PIE **oḱtō* > PT *oktə* 'eight : acht (8)', der. **oktukə* 'eight (80)'
TB *okt* | TA *okät*, *oktuk* (DThTA: 77a, 79b)
Adams 2013: 115-116; Ringe 1996: 89.
— The base numeral TB *okt* lost the expected final *-u* under the influence of *ṣukt* 'seven' or its ancestor PT **sʸəptə* (TA *ṣpät*), but it remained in the Tocharian A decad *oktuk* 'eighty' (Ringe 1996: 89).

85. PIE **peh₂-nt-* → **peh₂-nt-es* > PT m.pl. **ponʸtʸə* (**pånʸtʸə*) 'all'
TB *poñc* | TA *poñś* (DThTA: 286b)
Adams 2013: 432–433.

86. PIE **peh₂-nt-* → **peh₂-nt-ih₂* > PT **pontˢa-* (?**påntˢa*) 'all'
TA f.obl.sg. *pontsāṃ* (DThTA: 286b)
Pinault 2008: 524.
— The feminine displays the development of PIE **-ntih₂*, which this word is meant to illustrate in the main text. The TA *o* can probably only result from PIE **eh₂* as in *poke* from **bʰeh₂ǵʰu-* 'arm' above.

87. PIE **pekʷ-* > PT **pək-* 'cook, ripen (act. = tr.; mid. = intr.)'
TB *pək-* | TA *päk-* (DThTA: 275a)
Adams 2013: 393–394; Malzahn 2010: 700–701; Peyrot 2013: 770.

88. PIE **pekʷ-* → **pekʷ-l* > PT **pʸəkʷəl*, pl. *pʸəkʷəla* 'year'
TB *pikul*, pl. *pikwala* | TA *pᵤkäl* (~*pkul*), pl. *puklā* (DThTA: 288b)
Adams 2013: 410–411; Pinault 2008: 446.

89. PIE **preK-* → **prK-ske/o-* → pre-PT **proskeh₂* > PT **prosko* (?**pråskå*) 'fear'
TB *prosko*, der. *proskiye* | TA der. *praski* (DThTA: 308a)
Adams 2013: 455.

90. PIE **penkʷto-* > PT **pʸənkte* 'fifth'
TB *piṅkte* | TA *pänt* (DTHTA: 277a)
Adams 2013: 411.

91. PIE **ph₂tēr* > PT **pacer* 'father'
TB *pācer* /pácer/ | TA *pācar* (DThTA: 268a)
Adams 2013: 390.

92. PIE **plh₁-no-* 'full' > PT der. **pəllewə* 'full moon'
TB *pälleᵤ** /pəllέw/, obl.sg. *pällent* /pəllént/
Adams 2013: 406.
— The laryngeal was evidently lost, see also 2.3.1.

93. PIE **prHuo-* > PT **pərwe* 'first'
TB *parwe* /pə́rwe | TA der. *pärwat* 'first, oldest' (DThTA: 279a)
Adams 2013: 383–384.

94. PIE **puh₂-r* > PT **pəwar* 'fire'
TB *puwar* /pə́war/ | TA *por*
Adams 2013: 421–422; Winter 1965a: 192–193.

95. PIE **seh₂i-* > PT **soy-* (?**såy-*) 'be sated'
TB *soy-*, prt.ptc. *sosoyu* | ? → TA prt.ptc. *sasyu* with *säy*(*n*)- in 96.
Adams 2013: 770–771; Malzahn 2010: 955; Peyrot 2013: 833 fn. 949, 835.

96. PIE **seh₂i-* → **sh₂i-nu-* > PT **səyn-* 'satiate'
TB *səyn-* | TA *säy*(*n*)-, prt.ptc. *sasyu* (DThTA: 522a–b)
Adams 2013: 757; Friis 2024: 372–373; Hackstein 1995: 299–300; Malzahn 2010: 945–956; Peyrot 2013: 833.
— The lack of palatalization may be due to the laryngeal blocking it in the sequence **sh₂i*, but it could also have been removed due to analogy with the related verb PT **soy-* 'be sated' (see Friis 2024: 372–373.).

97. PIE **sem-* → f. **someh₂-* > PT **somo* (?**såmå*) 'one (1), some' (f.obl.sg.)'
TB *somo* | TA *ṣom* (DThTA: 511b)
Adams 2013: 721–722; Hilmarsson 1986: 33; Pinault 2008: 437–438.

98. PIE **septm* > PT **sʸəptə* 'seven (7)'
TB *ṣukt* | TA *ṣpät* (DThTA: 501b)
Adams 2013: 720.
— The consonant cluster in Tocharian B was influenced by *okt* 'eight' (Adams 2013: 720).

99. PIE **skHi-eh₂* > PT **skəyo* 'shadow'
TB *skiyo* /skə́yo/
Adams 2013: 773; Beekes 2010: 1351; Del Tomba 2020: 144–146; Ringe 1996: 18–19.
— The development of the word for 'shadow' in Tocharian is very difficult. See the discussions by Ringe (1996: 18–19) and Del Tomba (2020: 144–146), and cf. subsection 2.4.2.

100. PIE **sneh₁-ru*, pl. **sneh₁-ur-h₂* 'sinew' > PT **sʸnʸorə*, pl. *sʸnawra* 'sinew'
TB *ṣñor*, pl. *ṣnaura*
Adams 2013: 729; Del Tomba 2020: 87–88.

— The nominative is from pre-PIE **sneh₁-ur* with metathesis of final *-ru-* (Del Tomba 2020: 87-88). The original order **-ur-* is preserved in the plural *ṣnaura* (ibid.).

101. PIE **sokʷo-* 'sap' > ? PT **sekʷe* 'pus'
TB *sekwe* | TA *saku**, ins.sg. *sakuyo* (DThTA: 506a)
Adams 2013: 764; Beekes 2010: 1093.
— The etymological connection with Gr. ὀπός 'plant juice' (secondary loss of expected initial *h-*, see Beekes 2010: 1093), Lith. *sokaĩ* (m.pl.) 'resin', OCS *sokъ* 'sap, juice of fruits', Alb. *gjak* 'blood', though semantically not quite exact, nevertheless seems reasonable.

102. PIE **solh₂uo-* > PT **selwe* 'whole', der. **solme* 'completely'
TB der. *solme* | TA *salu* (DThTA: 511a)
Hilmarsson 1986: 32; Adams 2013: 771
— TA *salu* clearly has not undergone *u*-umlaut of **e* (PIE **o*), apparently because the word was **selwe* at the time, with a consonantal **-w-* rather than umlauting **-u* (cf. TA *maku* from **mekwa*, TB *mekwa* 'nails'). TB *solme* did undergo *u*-umlaut, presumably from a derivation that looked like **selu-me-* at the time, whence **solume* to **soləme* and eventually PT **solme* with syncope, and ultimately TB *solme* without further changes.

103. PIE **spend-* > PT **spənta-* 'trust'
TB *spənta-* | TA *späntā-* (DThTA: 543a)
Adams 2013: 785–786; Malzahn 2010: 967–968; Peyrot 2013: 838.

104. PIE **splend-* > PT **planta-* 'be pleased'
TB *planta-* | TA *plantā-* (DThTA 315a)
Adams 2013: 459; Malzahn 2010: 741–742; Peyrot 2013: 781.

105. PIE **stembʰH-* > PT **stəma-* 'stand' (sbj./prt. stem, suppletive with prs. *kəĺ-*)
TB *stəma-* | TA *ṣtämā-* (DThTA: 131a)
Adams 2013: 184–186; LIV: 595; Kroonen 2013: 473; Peyrot 2013: 547.
— Adams prefers a connection with **stem-* (or rather, **stemH-*, cf. Kroonen 2013: 473), as found in Germanic (e.g., ON *stemma* 'stem, stop, dam up') and Baltic (e.g., Lith. *stùmti* 'shove, push'), but the preterite TB *ścama* /ścə́ma/

corresponds closely to Ved. *ástambhīt*, from **(h₁e-)stemb$^{h}$H-t*, so that this connection is to be preferred (see Peyrot 2013: 547, LIV: 595).

106. PIE **sueh₂d-* → **suh₂d-ro-* > **sware* 'sweet'
TB *swāre* /swáre/ | TA *swār* (DThTA: 547b)
Adams 2013: 795–796.

107. PIE **suid-ie-* > PT **səya-* 'sweat'
TB *səya-* | TA *syā* (DThTA )
Adams 2013: 756; Malzahn 2010: 947; Peyrot 2013: 832.

108. PIE **teh₂-ns* > PT **tons* (?**tåns*)
TB (arch.) *toṃ* | TA *tos-äṃ*
Del Tomba 2018: 347; 2020: 216–220

109. PIE **tek$^{w}$-* → **tek$^{w}$os-* > PT **t$^{y}$ake* 'river'
TB *cake* /cə́ke/
Adams 2013: 267.

110. PIE **temp-* > PT **t$^{y}$əmp-* 'be able to'
TB *cəmp-* | TA *cämp-* (DThTA: 184a)
Adams 2013: 272–273.
— The semantic connection with PIE **temp-* 'stretch' is not entirely compelling, but possible; cf. German *reichen* 'reach; suffice, be enough' from an earlier meaning 'stretch' as more obvious in Dutch *reiken* 'reach, extend, stretch out'. From 'suffice, be enough', a development along the lines of 'I am enough (for ...)' to 'I am able to do (...)' can be assumed.

111. PIE **-ti* > ? PT **-s$^{y}$ə* prs.3sg. verbal ending
TA -(*ä*)*ṣ*
Jasanoff 1987: 110–111; Malzahn 2010: 32-33; differently Peyrot 2013: 413, differently again Pinault 2008: 620.

112. PIE **-ti* > ? PT **-s$^{y}$ə*
TA -(*ä*)*ṣ* (ablative ending)
Jasanoff 1987: 109–110.

113. PIE **tuHom* > PT **təwe* 'you (sg.)'
TB *tuwe* /tə́we/ ~ *twe* | TA *tu* (DThTA: 216b)
Adams 2013: 319–320.

114. PIE **ueǵʰ-no-* > PT **wʸəkne* 'way, manner'
TB *yakne* | TA *wkäṃ* (DTHTA: 448b)
Adams 2013: 518.

115. PIE **ueh₂st-u-* > PT **wostə* (?**wåstə*) 'house'
TB *ost* | TA *waṣt*
Adams 2013: 134; Beekes 2010: 158.
— A zero-grade **uh₂st-* (generalized in Greek ἄστυ; Beekes 2010: 158) should have yielded PT ***wastə*, whence TB ***wāst* /wást/ and TA ***wāṣt*, since there is no known *u*-umlaut of **a* (from **h₂* or **h₂e*) to **o*. Tocharian must thus reflect the full-grade **ueh₂st-*. The vocalism of TA *waṣt* cannot be explained as the result of *a*-umlaut in the plural **wostwa* (*pace* Adams 2013: 134), since all other certain cases of *a*-umlaut affect **e* rather than **o*.

116. PIE **ueiḱ-os-* > PT **wʸike* 'place'
TB *ike*, pl. *ykenta*
Adams 2013: 67.

117. PIE **uelh₂-* → **ulh₂-ntih₂* > PT **lantˢa* 'queen'
TB *lāntsa* | TA *lānts* (DThTA: 397a)
Adams 2013: 594.

118. PIE **uelh₂-* → **ulh₂-ōnts*, gen. *ulh₂-ntos* > PT **wəlo*, gen. *lante* 'king'
TB *walo* /wə́lo/, gen. *lānte* /lánte/ | TA *wäl*, gen. *lānt* (DThTA: 439b)
Adams 2013: 631; Lubotsky 1994.

119. PIE **uerǵ-* → **uerǵ-unt-ōn* > PT **wʸərkʷənto* 'wheel'
TB *yerkwanto**, obl. *yerkwantai* | TA *wärkänt* (DThTA: 436a)
Adams 2013: 547–548; Kim 1999: 168.
— The *ye-* of Tocharian B does not match the *wä-* of Tocharian A and may be secondary, perhaps analogous to the *ye-* of *yerter* 'wheelrim felloe' (Kim 1999: 168; Adams 2013: 547–548).

120. PIE **ues-r* 'spring' → ? **uesōr-* > der. PT **wʸəsare* 'wheat; grain'
TB *ysāre* | TA *wsār* (DThTA: 456b)
Adams 2013: 567–568; Peyrot 2018b: 251–252.

121. PIE **uid-uon-* > PT **wəwe* 'learnèd, skillful'
TB *uwe*
Adams 2013: 76–77.

122. PIE **uih₁ro-* 'man' > PT **wʸire* 'young, juvenile'
TA *wir* (DThTA: 446a)
Friis 2024: 22-24; Hackstein 2017: 1314.
— The semantic match between Tocharian *wir* and the other Indo-European forms is inexact, since elsewhere it means 'man, hero' rather than 'young, juvenile'. For a discussion, see Friis (2024: 22–24).

123. PIE **uiso-* > PT **wəse* 'poison'
TB *wase* /wə́se/ | TA *wäs* (DThTA: 442b)
Adams 2013: 634.

124. PIE **uksōn-* > PT **okso* 'ox'
TB *okso* | TA *opäs**, pl. *opsi* (DThTA: 83b)
Adams 2013: 117.

125. PIE **ulkʷo-* > PT **wəlkʷe* 'wolf(?)'
TB *walkwe**, pl. *walkwi*
Adams 2013: 632.
— The meaning of the Tocharian is not quite certain. PK NS 30 a2: /// *ma ywārcä walkwi ramtä wīyäskeṃ m(ñ)cu(ṣkeṃ)* /// 'amid ... they frighten the princes like *walkwi*' with *walkwi* plausibly meaning 'wolves' (Adams 2013: 632). The preservation of the labiovelar next to PIE **o* is rare, but potentially paralleled in PIE **sokʷos* > PT **sekʷe*, q.v.

126. PIE **uod-r* → *ud-r-*, der.adj. *ud-r-io-* > PT **wər* 'water', der.adj. *wərəye* 'dew; watery'
TB *war* /wə́r/, der.adj. *wriye** | TA *wär*, der.adj. *wri* (DThTA: 435a)
Adams 2013: 627–628, 672.

127. PIE **uōkʷs, uokʷ-* > PT **wek* 'voice'

TB *wek* | TA *wak* (DThTA: 422a)

Adams 2013: 660; Beekes 2010: 1138.

— The Proto-Indo-European paradigm was nom.sg. **uōkʷs*, elsewhere *uokʷ-*: Sanskrit nom.sg. *vā́k*, acc.sg. *vā́cam*, Latin nom.sg. *vōx*, acc.sg. *vōcem* with a long vowel next to Greek acc.sg. ὄπα with a short vowel; the Greek nominative is not attested (see Beekes 2010: 1138 for details). The long vowel in Sanskrit *vā́cam* can derive from members of the paradigm with the shape **uokʷV* and the change from **o* to **ā* in open syllables (Brugmann's Law), so that it can be derived from an alternating paradigm **uōkʷs, uokʷ-* regularly. The long vowel in Latin *vōcem* can be due to analogy with the nominative. The vowel **e* in PT **wek* must derive from the PIE short **o*, meaning that the vocalism of **uokʷ-*, as found in the Greek acc.sg. ὄπα, was generalized in Tocharian (Adams 2013: 660).

# Etymological reference list 2: Samoyed

Below follows a numbered list of the Samoyed etymologies cited in chapter 3, with references to various Samoyed dictionaries, wordlists, and articles in which the etymologies are proposed or discussed. Translations of the Samoyed words are provided in English and Russian for the Proto-Samoyed head words, to aid in the retrieval of these words in current and any future dictionaries of Samoyed languages that use Russian as the matrix language.

- **Proto-Uralic** reconstructions are based on Janhunen (1981), Uralisches Etymologisches Wörterbuch (UEW), Sammallahti (1988), and various other sources which are cited for specific entries.
- **Proto-Samoyed** forms are generally reconstructed on the basis of Samojedischer Wortschatz (SW) by Janhunen (1977) or Matorische Sprache (MS) by Helimski (1997), which differ only in some details, with the updated interpretation of the development of PU **ä* and **e* as per Helimski (2005). If I deviate from their reconstruction or other reconstructions in the literature, the note will explain why. Some words in the individual Samoyed languages that are not found in more recent dictionaries are cited according to SW, whereby the abbreviations for authors of the original sources are kept; relevant here are only (C) for Castrén, (L) for Lehtisalo, (M) for Mikola and (T) for Tereščenko (see SW for the exact references).
- **Nganasan** forms are generally cited after the Nganasan-Russian dictionary of Kosterkina, Momde and Ždanova (2004) with a transliteration from the Cyrillic to the Roman alphabet. The abbreviation "NgSl." is here used to refer to this dictionary. Whenever another source has a relevant Nganasan word missing from this dictionary, that source is cited instead.
- **Tundra Enets** forms are provided on the basis of Helimski's unpublished Tundra Enets dictionary, here abbreviated as "EnSl.".
- **Forest Enets** forms are taken from Olesya Khanina and Andrey Shluinsky's unpublished Forest Enets wordlist, here abbreviated as "EnWl.".
- **Tundra Nenets** is cited based on the dictionary of Tereščenko (1965), abbreviated as "T65" with the phonological transcription of Salminen (1998).
- **Forest Nenets** is phonologized similarly to Tundra Nenets after the rules of Salminen's system (Salminen 2007), and as a direct transcription from Cyrillic of

the relevant entries in the Forest Nenets-Russian dictionary by Barmič & Vello (2002), here abbreviated "NeSl.", in parentheses. The entries in Barmič & Vello are in some ways phonologically underdifferentiated when there are no minimal pairs, so that some information would be lost if only those entries are copied. For example, <a> may represent /a/ or /ă/.

- **Selkup** is cited after Alatalo's Sölkupisches Wörterbuch (SkW; 2004), which provides "Common Selkup" headwords on the basis of the attested variation of the different dialects. Asterisks are employed to mark roots or stems that only exist in derivations, but other "Common Selkup" forms are given without asterisks as in the dictionary. Underlined segments indicate uncertainty. The number cited along with these forms refers not to the page number, but to the entry number in the dictionary.
- **Kamas** words are taken over from Kamassz Szótár (KSz) by Tamás Janurik (2021). As with Selkup, the associated numbers correspond to the entries, not to the pages.
- **Mator** is provided on the basis of Matorische Sprache (MS) by Helimski (1997). Helimski made a phonological and morphological analysis of all attested Mator material and provided Proto-Samoyed reconstructions where appropriate. I refer to the entry numbers of the Samoyed words in Helimski's wordlist, rather than to the page numbers.

The order in which the words are presented in the alphabetical order below. Of these, the consonants **γ*, **δ*, **ŋ* and **r* only occur word-internally, never word-initially:

*a ä ć č γ δ δ́ e ę i į j k l m n ń ŋ o p r s t u ü w*

Some words show a discrepancy between Finno-Ugric **u* and Samoyed **o* (see subsection 3.6.3). They are here reconstructed with **o* in Proto-Uralic, but the reconstructions with **u* are provided as well for ease of interpretation. Whenever the entry with **u* would have been placed differently in the relative order of the list, an unnumbered entry with a reference is provided. The general phonological developments from Proto-Uralic to Samoyed are discussed in chapter 3, with further references to the relevant literature), and will not be repeated here unless they are in some detail directly relevant for a specific entry.

1. PU **aćkal* > PS **asVl-* ‘step : шагать’

   Sk. *āsəl-* (SkW 230) | Km. *azə?-* (KSz 0077)

   Aikio 2002: 40–41; 2006a: 30; UEW: 19; Zhivlov 2023: 158.

   — The quality of the second-syllable vowel strictly speaking cannot be seen from either Selkup or Kamas, the only Samoyed languages with a reflex of this word. Expected to be PS **asə̂l-* with a **ə̂* rather than **å*/*a* on the basis of the development of the first-syllable PU **a* to PS **a* rather than **å* (cf. 3.7.2).

2. PU **aŋa-* ‘open’ > PS der. **ńaŋə̂-* ‘take off (clothes) : снять (одежду)’

   NeT der. *nyaəcy°* ‘id.’ (T65) | NeF *nyaŋăqsy°* (NeSl. *njaŋa’š*) | Sk. adj. *ńaŋə* ‘naked : nackt’ (SkW 1675)

   Aikio 2002: 50, 2020: 20–21; SW: 106; UEW: 11.

   — The initial PS **ń* is a secondary prothetic nasal. This **ń-* would arise secondarily before word-initial vowels in Nenets, but not in Selkup, which thus suggests that it has to be reconstructed. However, the vowel in Selkup is also irregular, seeing as Sk. *a* normally reflects PS **ə̂* rather than PS **a*, which is expected to become Sk. *ā*. The same meaning as in Samoyed is found in Mansi and Khanty (see Aikio 2020: 20).

3. PU **aŋi* > PS **aŋ* ‘mouth : рот’

   Ng. *ŋaŋ* (NgSl.) | EnT *e?* (EnSl.) | NeT *nyah* (T65) | NeF *nya°ŋ syĭ* lit. ‘opening of the mouth’ (*nya°ŋ* from gen.sg. **aŋə̂n*) (NeSl. *njaŋ ši*) | Sk. *āŋ* (SkW 215) | Km. *aŋ* (KSz 0049) | Mt. *äŋ* (MS 76)

   SW: 20 (**äŋ*); UEW: 11.

4. PU **aŋti* > PS **aŋtə̂* ‘blade : лезвие’

   Ng. *ŋaťə* (NgSl.) | EnT *edo* (EnSl.) | NeT *nyant°* (T65) | NeF der. e.g. *nyantułampyosy°* ‘sharpen a knife : точить нож’ (NeSl. *njantułampjoš*) | Sk. *āŋtə* (SkW 217) | Km. *aŋ* (KSz 0050) | Mt. *ändä* (MS 74)

   SW: 20–21 (**äŋtə̂*); UEW: 12–13.

5. PU **äjmä* > PS **äjmä* ‘needle : игла’

   Ng. *ńejmï* (NgSl.) | EnT *ee* (EnSl.) | EnF *nɛɛ* (EnWl.) | NeT *nyíbya* (T65) | Km. *ńimi* (KSz 0753) | Mt. *ime* (? *īme*) (MS 342)

   Sammallahti 1988: 536; SW: 22 (**ejmä* ~ **nejmä*); UEW: 22.

— With a prothetic initial nasal in most Samoyed languages, entirely regular in Nganasan and Tundra Nenets, but not in Kamas and Forest Enets. Tundra Enets and Mator show the expected original vowel-initial form.

6. PU **ćaδa-* > PS **sårå-* 'rain : идти (о дожде)', der.n. **sårö* (?**såråw*) 'rain : дождь'
Ng. *soruďa*, der.n. *sor″a* (NgSl.) | EnT *sara-*, der.n. *sare* (EnSl.) | EnF der.n. *sare* 'rain' | NeT der.n. *saryo* (T65) | NeF *xałasy*°, der.adj. *xałyoj*° (NeSl. *xāłaš*, *xałjoj*) | Sk. *suârə̲-* (SkW 2677a.) | Km. der. *surno* 'rain : Regen : дождь' (KSz 0985) | Mt. der.n. *sörüh* (? *sirüh* ~ *sürüh*) (MS 908)
SW: 135–136; Sammallahti 1988: 540.

7. PU **ćala-* > PS **ćålâ-* 'flash : блестеть'
Ng. der. noun *solə* 'brightness : блеск' (NgSl.) | NeT der. *sal°wə*(*sy°*) (T65) | Mt. *salə-*, der. *saləmər-* (MS 842, 843)
Aikio 2002: 27–29; 2006a: 30.
— The change from PU **a-a* to PS **å-â* in this word is unusual and quite unique: expected would be PS **ćålä-* (cf. PU **kala* > PS **kålä* 'fish') or PS **ćalâ-* (as in PU **kama* > PS **kamâ* 'scale').

8. PU **ćaŋka-* > PS **såŋkV-* 'stick into : воткнуть', noun *såŋkâ* 'penis : пенис'
EnT noun *sago* (EnSl.) | EnF noun *sago* (EnWl.) | NeT noun *saŋk°* (L) | Sk. *suâŋkə̲-* (SkW 2625)
Aikio 2006a: 24.

9. PU **ćäjćimä*/*ćäććimä* > PS **säjt³wə* 'seven : семь (7)'
Ng. *śajbə* (NgSl.) | EnT *seʔo* (EnSl.) | EnF *sɛʔo* (EnWl.) | NeT *syíqw°* (T65) | NeF *syeqw°* (NeSl. *še″w*) | Sk. *sēlću* (SkW 2741) | Km. *sejʔbü* (KSz 0951) | Mt. *kejʔbə* ~ ?*kejʔbü* (MS 451)
Aikio 2020: 109–111; 2022a: 25; SW: 139.
— The development from PU **m* to PS **w* is irregular.

10. PU **ćęrki* > PS **sęr* 'ice; white; salt : лёд; белый; соль'
Ng. *sir* (NgSl.) | EnT *siʔ* (EnSl.) | EnF *siʔ* (EnWl.) | NeT *ser* (T65) | NeF *xĕł* (NeSl. *xeł*) | Sk. *sęrə* (SkW 2697) | Km. der.? *sĭri* (KSz 0968) | Mt. der. *sęrę̆t* (MS 874)
Aikio 2020: 126; SW 138.

— The consonant-final forms of the northern Samoyed languages should be taken as the regular reflex, indicating PS **sęr*. The Selkup reflex also derives from this, as it should have become ***sīrə* rather than *sęrə* if derived from a PS **sęrâ* (see Helimski 2007 on the sound laws involved).

11. PU **ćilmä* > PS **säjmä* 'eye : глаз'
Ng. *śejmï* (NgSl.) | EnT *sei* (EnSl.) | EnF *sej* (EnWl.) | NeT *sæw*° (T65) | NeF *xäm*° (NeSl. *xëm*) | Sk. *saji* (SkW 2562) | Km. *sima* (KSz 0957) | Mt. *sīme* (MS 886)
Sammallahti 1988: 540; SW: 132; UEW: 479.

12. PU **ćoδ´a* > PS **såjå-* 'wage war : вести войну'
Ng. der. *souru* 'warrior : воин' (NgSl.) | NeT der. *sayu*° 'enemy army : войско противника', *sayuw*° 'army : войско' (T65)
Aikio 2002: 27; 2015: 61.

13. PU **ćoji-* > PS **soj-* 'be heard, be audible : быть слышным, слышаться'
EnT der. *sôδoru* 'loud sound : сильный шум' (EnSl.) | EnF *sɔ-* (EnWl.) | NeT *sosy*° (T65) | NeF *xosy*° (NeSl. *xoš*)
Aikio 2020: 132; UEW: 482–283.

14. PU **ćojma* > PS **såjmå* 'sound'
Ng. *sojmü* (Aikio 2015a: 16 fn. 5) | Sk. *sūmə* (SkW 2483)
Aikio 2015: 16 fn. 5; 2020: 133

15. PU **ćoma* > PS **somå* 'good : хороший'
Ng. der. *sumugu* 'success : удача' (NgSl.) | EnT der. *sojδa* (EnSl.) | EnF der. *sɔjza* (EnWl.) | NeT *səwa* (T65) | NeF *xoma* (NeSl. *xoma*) | Sk. *soma* (Skw 2469)
Aikio 2020: 135; Salminen 2023: 380–385; SW: 132–133 (**sâmå*).
— Tundra Nenets and Forest Enets seem to reflect PS **sâmå*, but Nganasan, Forest Nenets and Selkup can only reflect PS **somå*, which is in line with other Uralic cognates (see Aikio 2020: 135, Salminen 2023: 380–385).

16. PU **ćowi* (**ćuwi*) > PS **soj* (or **soə̑j*?) 'throat : горло'
EnT *sôô* (EnSl.) | EnF *so* (EnWl.) | NeT *syo* (T65) | NeF *syŏ* (NeSl. *šo*) | Sk. *sōl'* (SkW 2568) | Km. *soj* (KSz 0974) | Mt. *soj* ?*sōj* (MS 897)
Aikio 2020: 145; Sammallahti 1988: 540 (**śoxi*; SW: 142).
— Theoretically, a reconstruction PS **soə̑j* would also be possible. This might even be implied by the long vowel reflexes in Tundra Enets *sôô*, Selkup *sōl'* and potentially Mator *soj* ?*sōj* (thus given in MS, attested as ⟨soóschto⟩, analysable as /sosto/ or /sōsto/ with poss.3sg. *-to* and a regular development from *-j-* to *-s-* before the *-t-*). Nenets, Forest Enets, Selkup and Kamas would not show a clear reflex of a vowel sequence, and there is no Nganasan cognate to either confirm or disprove it. The Finno-Ugric languages reflect a preform **ćuwi* rather than **ćowi*.

— PU **ćuwi* 'throat' → see **ćowi*.

17. PU **ćüδ́i* > PS **sijə* 'coal : уголь'
EnT in compound *tuuśio* (EnSl.) | Sk. *sīćə*/*śīćə* (SkW 2556) | Km. der.? *si?* (KSz 0956) | Mt. *ki* (? *kī*, ? *kiji*) (MS 479)
Sammallahti 1988: 540; SW: 140; UEW: 477.

18. PU **ćülki-kći-* > PS **siəsə-* 'spit : плевать', der. noun *siəsö* (?**siəsəw*) 'foam : пена'
EnT *śiosi* (EnSl.) | Sk. *süəsə̲-* (SkW 2659) | Km. *suzə-*, der. noun *suzu* (KSz 0986)
Aikio 2020: 148; Sammallahti 1988: 549 (**śülki-*/*śüδ́ki-*); UEW: 479–480.

19. PU **ćüńćä* > PS **sünsə* 'breast : грудь'
Ng. *śinsə* (NgSl.) | EnT *śuḋo* (EnSl.) | EnF *šuzu* (EnWl.) | ? cf. NeT *syunc°* 'youngest, dearest child : последный, младший ребёнок' | Mt. *künžü* (MS 595)
Aikio 2020: 149–150; Sammallahti 1988: 540; SW: 144; UEW: 480.
— The retention of PU **ü* as PS **ü* rather than **i* is surprising, as is the reduction of the second-syllable PU **-ä* to PS **-ə*, but the comparison otherwise seems too close to ignore (and cf. the development of PU **peksä-* 'soften a hide', where **-ä* was reduced as well and even lost); see also the discussion by Aikio (2020: 149–150)

20. PU **čaŋa-* > PS **cåŋå-* ‘rub, wear out (tr.) : натереть’ intr. **cåŋo-* (?**cåŋåw-*) ‘wear out (intr.) : износиться’
NeT *taŋa*(*sy*°), intr. *taŋo*(*sy*°) (T65) | NeF der.tr. *taŋałasy*°, der.intr. *taŋałăsy*° (NeSl. *taŋałaš*, *taŋałăš*) | Km. *taŋo-* (KSz 1150)
Aikio 2002: 11–12; SW 151.

21. PU **čečä* > PS **cecä* ‘uncle (mother’s younger brother) : дядя (младший брат матери)
Ng. der. *tïtïďa* (NgSl.) | NeT *tyidya* (T65) | Sk. **čičä* (SkW)
Sammallahti 1988: 536; SW: 33 (**cicä*).

22. PU **čočki* (**čučki*) > ? PS **cocə̂* ‘pole : шест’
Sk. *čōčə* (SkW 1060)
Aikio 2013: 163–164.
— The vowel **o* instead of **u* is unexpected, but the correspondence with PSaa. **ćoskę* ‘block of wood’, MdE *čočko* ‘log’ is otherwise regular (Aikio 2013: 163–164).

23. PU **δ´ęmi* > PS **jęm* ‘bird cherry : черёмуха обыкновенная’
Sk. *ćęmə*, *ćį̄mə-* | Km. *lem* (KSz 0533)
Aikio 2012: 228; Sammallahti (1988: 536 (**ďɨxmɨ*); UEW: 65–66.
— The vowel-final base *ćęmə* as given by SkW is secondary: PS ***jęmə̂* would have become ***ćį̄mə*, as in the genitive *ćį̄mən* from **jęmə̂n* (cf. Helimski 2007 on the relevant sound developments).

24. PU **δ´ümä* > PS **jimä* ‘glue : клей’
Ng. *ďïmi* (NgSl.) | EnT *ďee* (EnSl.) | EnF *ďïi* (EnWl.) | NeT *yibya* (T65) | NeF *jĭmya* (NeSl. *djimja*) | Sk. *ćįmə* (SkW 1469) | Km. *nĭme* (KSz 0695) | Mt. **ńime*, *nime* (MS 755)
Sammallahti 1988: 537; SW: 45; UEW: 66.

25. PU **elä-* > PS **elä-* ‘live : жить’
Ng. *ńilïďi* (NgSl.) | EnT *ire-* (EnSl.) | EnF *ďiri-* (EnWl.) | NeT *yilye*(*sy*°) (T65) | NeF *jĭłyesy*° (NeSl. *djiłješ*) | Sk. **ilə-* (SkW 344) | Km. der. *ťili* ‘lively : lebendig : живой’ (KSz 1371) | Mt. *ilə-* (MS 340)
Sammallahti 1988: 536 (**ilä-*); SW: 27 (**ilä-* ~ **jilä-*); UEW: 73.

— Forest Enets, Nenets and Kamas reflect a pre-form with a secondary initial palatal glide **j-*, thus seemingly reflecting PS **jelä-*. On account of other Uralic cognates without the initial **j-* and the lack of **j-* in the Nganasan and Selkup reflexes **elä-* is proper Proto-Samoyed form as well, with later prothesis of **j-* in Enets, Nenets and Kamas; see Sammallahti (1988: 536) and Zhivlov (2023: 167) rather than the alternating Proto-Samoyed reconstructions by SW and Aikio (2020: 43).

26. PU **ɐćiw-* > PS **ɐso-* (?**ɐsə̂w-*) 'camp : встать чумом'
EnT *usu-* (EnSl.) | EnF *usu-* (EnWl.) | NeT *ŋeso-* (T65)
Aikio 2012: 241; UEW: 18–19.

27. PU **ɐδi* 'year' > PS der. **ɐrö* 'autumn : осень'
NeT *ŋeryo* (T65) | Sk. *ɐra* (SkW 277) | Km. *ere* (KSz 0210) | Mt. *örö*(*h*) (MS 819)
Aikio 2012: 233–234; Sammallahti 1988: 552 (Proto-Finno-Permic **ooti*); SW 22 (**ɐrə̂jə̂j* ?**ɐrö*); UEW: 335–336 (**oδe*).

28. PU **ɐkta-* > PS **ɟtå-* 'hang up : повесить', der.intr. **ɟtö-* (?**ɟtåw-*) 'hang : висеть'
Ng. *ŋitiďi* (NgSl.) | EnT *iδa-* (EnSl.) | EnF *iδa-* (EnWl.) | NeT *ŋida*(*sy*°), *ŋidyo*(*sy*°) (T65) | NeF der.intr. *ŋĭtyosy*° (NeSl. *ŋïčoš*) | Sk. **ɟt*(*t*)*ə-*, itr. *ɟt*(*t*)*i-* (SkW 92) | Km. *edə-* (intr. *edü-*?) (KSz 199) | Mt. *idə-* (MS 335 **ɟta-*)
Aikio 2020: 51 (**ɟtå-*; **ɟtö-*); SW: 25–26 (**ɟtɜ-*; ?**ɟtå-*); Sammallahti 1988: 536.

29. PU **ɐla* > PS **ɟl*(*ə̂*) 'under; space underneath : вниз, под'
Ng. *ŋilə* (NgSl.) | EnT *iro-* (EnSl.) | EnF *iro-* (EnWl.) | NeT *ŋilə-* (T65) | NeF *ŋĭťă-* (NeSl. *ŋïťa-*) | Sk. **ɟl-* 'under : unter', *ɟlə* 'underside : das Untere' (SkW 322) | Km. *jil* (KSz 0259)
Sammallahti 1988: 536; SW: 24; UEW: 6.

30. PU **ɐna-ɐppi* > PS **ɟnə̂pə̂* 'father-in-law : тесть'
Ng. (C) *˜inaba* (/ŋinəbə/) (SW) | EnT *inobo* (EnSl.) | NeT *ŋinəb*° (T65) | NeF (L) *ŋɟnnap* (/ŋĭnăp°/) (SW) | Km. *ömbə* (KSz 0784)
Sammallahti 1988: 536; SW: 24–25; UEW: 9.

31. PU **ẹńa* > PS **i̯ńə̂* ‘tame : ручной’
NeT *ŋiy°* (T65) | Sk. **ī̯ńə* (Aikio 2020)
Aikio 2020: 55; Sammallahti 1988: 536; SW: 25.

32. PU **ẹpti* > PS **ẹptə̂* ‘hair (of the head) : волосы’
Ng. *ŋabtə* (NgSl.) | EnT *ito* (EnSl.) | EnF *itu* (EnWl.) | NeT *ŋebt°* (T65) | NeF *ŋept°* (NeSl. *ŋept*) | Sk. *ōptə* (SkW 24) | Km. *eʔbdə* (KSz 197) | Mt. *ẹbtẹ* (MS 246)
Sammallahti 1988: 536; SW: 21; UEW: 14–15.

33. PU **imi-* > PS **ӭm-mä*(-) ‘suckle; breast : быть грудным; женская грудь’
NeT verb *ŋəmya*(*sy°*), noun *ŋəmya* (T65)
Aikio 2020: 59–60; UEW 82–83; Sammallahti 1988: 536.

34. PU **ipsä-* > PS **ӭptä-* ‘smell (v.) : пахнуть’
NeT *ŋəbtye*(*sy°*) (T65)
Aikio 2020: 60–61; Sammallahti 1988: 536; SW 16; UEW: 83–84 (**ipз* ~ **ipз-zз* ~ **ipз-śз*).

35. PU **ipsi* > PS **ӭptӭ* ‘smell (n.) : запах’
Ng. (C) *˜obta* /ŋobtə/ (SW) | EnT irreg. *opto* (expected ***oto*) (EnSl.) | EnF *ɔtu* (EnWl.) | NeT *ŋəbt°* (T65) | Sk. *aptə* (SkW 20)
Aikio 2020: 60–61; Sammallahti 1988: 536; SW: 16; UEW: 83–84 (**ipз* ~ **ipз-zз* ~ **ipз-śз*).
— The development from the initial PU **i* to PS **ӭ* is not phonologically regular, but can be understood as the result of analogy with the verb PU **iptä-*, where it is expected (see Aikio 2020: 60–61).

36. PU **itä-* ‘appear’ > ? PS **ə̂tə̂-* ‘see : видеть’, der.refl. **ӭti-* ~ **ə̂tu-* ‘be visible : быть видным’, ? der.trsl. ±**ӭtӭjm-* ‘become visible, appear : явиться, появиться’
Ng. *ŋətəďa*, der.refl. *ŋəðüśa*, der.trsl. *ŋətumsi̯*, der.n. *ŋəðu* ‘physical appearance; colour : внешний вид; цвет’ | EnT der.refl. *oδi-*, der.trsl. *oδim-* (EnSl.) | EnF der.refl. *ozi-* (EnWl.) | NeT der.refl. *ŋədyə*(*sy°*) ‘be visible : быть видным’, der.trsl. *ŋədyimcy°* (T65) | Sk. *atə̱-*, der.refl. *atu-*, der.trsl. *atom-* (SkW 66) | Km. *ĭdə-* (KSz 0243) | Mt. ?*adə-*, der.trsl. *adəm-* (MS 7, 8)
Helimski *apud* Aikio 2002: 49; Aikio 2020: 65; SW: 16; UEW: 85.

— The reconstruction of the base verb is quite difficult. Ng. *ŋətəďa* 'meet; find; check; attempt; encounter : увидеться; найти; проверить; попробовать; встретить' and *ŋəðüśa* 'see, watch, observe : видеть, следить, наблюдать' are both back vocalic while only *ŋətumsi̯* 'appear; be born : появиться, показаться; родиться' is front vocalic, as would be expected from (a derivative of) front-vocalic PU **itä-*. Ng. *ŋəðüśa* would be expected from a PS **ât̮u-*, while NeT *ŋədyə-* 'be visible : быть видным' seems to require **äti-*. The meaning 'see' rather than 'appear', presumed as the base **ât̮ə-* in SW, is not found in the Uralic cognates, and the phonological derivation of this back vocalic form from front vocalic PU **itä-* is unclear to me.

37. PU **jasi* > PS **jat* 'early winter : ранняя весна'
Sk. *ćāt*(*ə̲*) (SkW 1475)
Aikio 2015: 52.
— Presumably, the final vowel is given in SkW in parentheses as a possibility because the dialects in which this word is attested would not have preserved it either way; this makes the reconstruction of the Proto-Samoyed word as a consonant stem uncertain, even though it is expected based on the regular change from Proto-Uralic words of the type **CVCi* to PS **CVC*.

38. PU **jäwji* > PS **jüjə* 'lichen (on trees) : ягель'
Ng. *ďïə* (NgSl.) | NeT *yuy*° (Aikio 2006)
Aikio 2006a: 12–14; UEW: 96.

39. PU **jɐ̯γi-* > PS ders. **ɐ̯r-*, **ɐ̯kə̂l-* 'drink : выпить'
EnT *ixoro-* (EnSl.) | EnF *ixuru-* (EnWl.) | NeT *ŋercy*°, *ŋexəlcy*° (T65) | NeF *ŋełsy*° (NeSl. *ŋełьš*) | Sk. *ɐ̄r-* (SkW 276) | Mt. *ɐ̯r-* (MS 249)
Aikio 2002: 38–40; SW: 21–22.

40. PU **joŋsi* > PS **j̮ŋtə̂* 'bow : лук'
Ng. *ďïntə* (NgSl.) | EnT *ido* (EnSl.) | EnF *idu* (EnWl.) | NeT *ŋin*° (T65) | NeF *ŋĭn*° (NeSl. *ŋïn*) | Sk. **j̮ntə* (SkW) | Km. *jinə* (KmSz.) | Mt. *mindi* (analogical **w-*, which was consequently assimilated to *m-*; see MS 678 for details)
Sammallahti 1988: 537; SW: 25 (**j̮ntə̂*).
— Nganasan *ďïntə* alone seems to have preserved the initial PU **j-*, a potential archaism of this language, since there do not seem to be any examples where an initial PS **j̮* developed a secondary **j-* in Nganasan. The initial *j-* in Kamas

*jinə* is not original, as it should otherwise have become a ***t'-* (cf. PU **juka* > PS **jə̂kå* > Km. *t'aga* 'river' and PU **e̯la* > PS **i̯l(ə̂)* > Km. *jil* 'below').

41. PU **juka* > PS **jə̂kå* 'river : река'
EnT *d'axa* (EnSl.) | EnF *d'ɔxa* (EnWl.) | NeT *yəxa* (T65) | NeF *jăxa* (NeSl. *djăxă*) | Sk. *ćaqə* 'hole filled with water' (SkW 1530??) | Km. *t'aga* (KmSz.) | Mt. *čaga* (MS 169)
Janhunen 1981: 223; Sammallahti 1988: 537; SW: 35.

42. PU **jupta-* > PS **jə̂ptə̂-* 'say, tell : говорить, рассказывать'
Ng. der. *d'ebtəðasa* (NgSl.) | Sk. *ćaptə̱-*, der. *ćaptä* 'tale : Märchen' (SkW 1443, 1444)
Helimski *apud* Aikio 2002: 48; SW: 35.

43. PU **juri-* > PS **jür-* 'get lost : теряться'
EnT *d́uro-* (EnSl.) | EnF der. *d'urota-* 'forget' (EnWl.) NeT *yurə(sy°)* 'forget' | NeF der. *jułipyosy°* (NeSl. *djułïpjoš*) | Sk. der.? *üru-* (SkW 285) | Km. *t'ür-* (KSz 1413)
Aikio 2002: 46–48.
— The final **-i* should have been lost, so that the final *-ə̂* of the northern Samoyed forms is likely to be secondarily introduced by analogy. Kamas *t'ür-* then reflects the expected outcome.

44. PU **kaća-* > PS der. **kåso* 'gift : подарок'
EnT *kasu* (EnSl.) | NeT *xaso* (T65) | NeF (L) *kāssū* (/kasu/) (SW) | Sk. *qosə* (SkW 2190)
Sammallahti 1988: 538; SW: 61 (**kåsə̂j*); UEW: 111.

45. PU **kaδ´a-* > PS **kåjä-* 'leave : оставить', itr. **kåjo-* (?**kåjäw-*) 'stay : остаться'
Ng. tr. *koid'a*, itr. *kousa* (NgSl.) | EnT tr. *kae-*, itr. *kai-* (EnSl.) | EnF tr. *kaji-* ~ *kaja-*, itr. *kaji-* ~ *kaja-* (EnWl.) | NeT tr. *xaye(sy°)*, itr. *xayo(sy°)* (T65) | NeF tr. *kajesy°*, itr. *kajosy°* (NeSl. *kadješ*, *kadjoš*) | Sk. *quə̂ćə-*, itr. *quə̂ći-* (SkW 2057) | Km. itr. *kojo-* (KSz 0405) | Mt. *kojo-* (MS 531)
Sammallahti 1988: 538; SW: 58; UEW: 115–116.

46. PU **kaja* > PS **kåjå* 'sun : солнце'
Ng. *kou* (NgSl.) | EnT *kaja* (EnSl.) | EnF *kaja* (EnWl.) | NeT der. *xayera(sy°)* 'become sunny', *xayer°q* 'sun' (T65) | Sk. *quə̂ćə* 'heat : Hitze' (SkW 2054) | Km. *kuja* (KSz 0474) | Mt. *kaja* (MS 389)
Sammallahti 1988: 538; SW: 58.

47. PU **kaji* > PS **kåə̂* ‘slender object : тонкий предмет’
Sk. *qū* (Aikio 2013)
Aikio 2013: 166–167.

48. PU **kajši-*, noun **kajšo* > PS **kåjtə̂*(-) ‘be sick, sickness, der. noun **kåjto* ‘sickness’
Ng. *koťə*, *koťəďa* (NgSl.) | EnF *kaťe* (SB) | NeT *xædo* (T65) | Sk. **qūtə-* (SkW 1970)
Aikio 2014a: 3–5; SW: 58–59.

49. PU **kajwa* > PS **kajwå* ‘spade : лопата’
Ng. *kajbu* (NgSl.) | EnT *sea* (EnSl.) | EnF *sɛɛ* (EnWl.) | NeT *syíwa* (T65) | NeF *syewa* (NeSl. *šewa*) | Km. *ko* (KSz 0395)
Aikio 2002: 41–42; SW: 63.

50. PU **kala* > PS **kålä* ‘fish : рыба’
Ng. *kolï* (NgSl.) | EnT *kare* (EnSl.) | EnF *kare* (EnWl.) | NeT *xalya* (T65) | NeF *kałya* (NeSl. *kāłja*) | Sk. *quə̂lə* (SkW 2362) | Km. *kola* (KSz 0407) | Mt. *kälä* (MS 461)
Sammallahti 1988: 538; SW: 59; UEW 119.

51. PU **kali-* > PS **kåə̂-* ‘die : умереть’, der.caus. **kå-tå-* ‘kill : убить’
Ng. *kuoďa*, der.caus. *kotuďa* (NgSl.) | EnT *kaa-*, der.caus. *kaδa-* (EnSl.) | EnF *ka-*, der.caus. *kaza-* (EnWl.) | NeT *xasy*°, der.caus. *xada*(*sy*°) (T65) | NeF *kasy*°, der.caus. *katasy*° (NeSl. *kāš*, *katāš*) | Sk. *qū-*, der.caus. *quə̂t-* (SkW 1759, 1901) | Km. *kü-*, *kut-* (KSz 0504, 0501) | Mt. *kā-*, der.caus. *kadə-* (MS 367, 374)
Aikio 2015: 52; Sammallahti 538 (**kåxli-*); SW: 56–57; UEW: 173.
— The causative **kåtå-* appears to be made on the basis of **kåə̂-* after the intervocalic **-l-* had disappeared, so that **kå-* could be taken as the new “consonant stem” from which to derive a causative. I think that this is paralleled in the derivative PS **ńånsə̂-* ‘lick’ next to underived PS **ńåə̂-* ‘lick’ from PU **ńali-* (entry 116 below). Selkup and Kamas both analogically removed the final vowel of PS **kåtå-* and reflect **kåt-* instead.

52. PU **kama* > PS **kamə̂* ‘scale : чешуя’
NeT *syaw*° (T65) | NeF *syam*° (NeSl. *šam*) | Sk. *kām~kāmə* (SkW 1822)
Aikio 2015: 55; SW 63; UEW: 121–122.

53. PU **kani-* > PS **kån-* 'go : уйти', der.caus. **kåntå-* 'carry : увести'
Ng. der. *konïďa,* tr. *kontuďa* (NgSl.) | EnT der. *kane-*, tr. *kada-* (EnSl.) | EnF der. *kańe-*, *kada-* (EnWl.) | NeT tr. *xana*(*sy*°) | NeF tr. *kanasy*° (NeSl. *kanāš*) | Sk. *quân-* (SkW 2002) | Km. *kan-* (KSz 0306) | Mt. tr. *kandə-* (MS 415)
Janhunen 1981: 231; Sammallahti 1988: 538; SW: 59–60; UEW 124.
— Only the causative **kåntå-* is found in the other branches of Uralic, so that PS **kån-* represents a valuable archaism of the Samoyed branch. The vowel reflex PS **å* is phonologically unexpected in underived **kån-*, but regular in **kåntå-*.

54. PU **kari* > PS **kar* 'skin, surface, bark : кожа, поверхность, корочка'
NeT *syar* | NeF *syał* (NeSl. *šał*)
Aikio 2012: 233.

55. PU **käli* > PS **käj* (~ ?**kää-*) 'tongue : язык'
Ng. der. *śiəďə* (NgSl.) | ? EnT der. *siʔoro* (EnSl.) | ? EnF der. *šiʔɔru* (EnWl.) | NeF *syĕ* (NeSl. *še*) | Sk. *śē* (SkW 2420) | Km. der. *šekə* (KSz 1050) | Mt. *kEj* (? *kāj*, ? *kēj*, ? *kij*; MS 450 **keəj*)
Aikio 2012: 228; Sammallahti 1988: 538; SW 66.
— The precise Proto-Samoyed reconstruction is difficult, cf. Aikio (2012: 228). There may have been a paradigmatic alternation between **käj* and **kää* that was resolved in different ways in the different Samoyed languages (cf. PU **woli-* > PS **åj-* ~ **åå-* 'be' in entry 198 below). Nganasan *śiəďə* could reflect **kää*, with a derivational suffix **-jə*, while Mator should go back to **käj* or **kää-j*. Forest Nenets *syĕ* and Selkup *śē* can reflect PS **käj* (cf. the vowel development of PS **säjswä* 'seven' to NeF *syeqw*°, Sk. *sēľću*), but it is unclear how **ää* would have developed in Selkup: expected would be **ē* or perhaps **ī*. Nganasan could perhaps also reflect a derivation in **-jə* from **käj*, via **käj-ə-jə* with loss of **-j-* between the first and second syllable.

56. PU **käri-* > PS **kär-* 'dress, put on (clothes) : надеть'
Ng. *śerəďi* (NgSl.) | EnT *seri-* (EnSl.) | EnF *sɛro-* (EnWl.) | NeT *syerə*(*sy*°) (T65) | NeF *syełăsy*° (NeSl. *šełaš*) | Sk. *śēr-* (SkW 2701) | Km. *šer-* (KSz 1054) | Mt. *ker-* (MS 470)
Aikio 2002: 18–20; SW 68.
— The northern Samoyed languages have a vowel stem, which is probably secondary as a result of analogy, since the consonant stem reflected in Selkup, Kamas and Mator is phonologically expected (cf. section 3.3).

57. PU **kätki-* > PS **kätə-* 'wrap up : привязать'
Ng. *śetəďi* 'load : погрузить' (NgSl.) | NeT *syedə*(*sy*°) (T65) | Mt. *kEdə-* (MS 447)
Aikio 2002: 20; 2006a: 29.

58. PU **käwδi* > (in derivations) PS **kürə* 'band (for tightening) : лента'
Ng. der. *kirəməsï* 'tighten : затянуть' (NgSl.) | EnT *śurobo-* 'tighten (with a strap) : стянуть (шнурком)' (EnSl.) | NeT *syur*°*q* 'waistband for trousers : объём пояса' (T65)
Aikio 2006: 19–20.

59. PU **keti* > PS **ket* 'shape, appearance; figure : образ, изображение' (also a pronoun stem)
Ng. *sï?* (NgSl.) | EnT *śi?* (EnSl.) | EnF *ši?-* (pronoun stem) (EnWl.) | NeT *syiq* (T65) | Sk. *śi-* (pronoun stem) (SkW 2428)
Aikio 2006a: 17–19.

60. PU **kęčči-* > PS **kęcə̂-* 'bad smell, stench : вонь'
SkTaz *qętį*, SkKet *qęči-* (Aikio 2014: 6)
Aikio 2014a: 6.

61. PU **koćki-* > PS **kåsə̂-* 'be dry, dry out (itr.) : высохнуть'
Ng. *kosəďa* (NgSl.) | EnT *kaso-* (EnSl.) | EnF *kasu-* (EnWl.) | NeT *xasə*(*sy*°) (T65) | NeF *kasăsy*° (NeSl. *kasaš*) | Sk. **kūsə* 'shallow' (SkW 2202) | Km. *ko?-* (**kås-*; KSz 0397)
Sammallahti 1988: 537; SW: 60–61.

62. PU **koδ´ka* > PS **kåjkə̂* 'spirit, god; idol : идол, шайтан'
Ng. *kojkə* (NgSl.) | EnT *kaxa* (EnSl.) | EnF *kixu* (EnWl.) | NeT *xæx*° (T65) | NeF *käx*° (NeSl. *këxë*) | Sk. der.? *qāka* (SkW 2097)
Aikio 2002: 13–15 ; SW: 51.
— Selkup has an unexpected *ā* rather than *ū̆*, as if from PS **kakə̂*. Perhaps it was affected by the second-syllable vowel *-a*, which looks to be a derivational suffix (expected would be *-ə*).

63. PU **kojra* > PS **korå* 'male, male reindeer : бык, пороз (самец оленя)'
Ng. *kuru* (NgSl.) | EnT *kôra* (EnSl.) | EnF (!) *kɔra* (EnWl.) | NeT *xora* (T65) | NeF *koła* (NeSl. *koła*) | Sk. *qorə* (SkW 2259) | Km. *kora* (KSz 0430)

SW: 74; UEW: 168–169; Zhivlov 2023: 162.

— The vowel development from PU **kojra* to PS **korå* rather than **kårå* is unusual, see subsection 3.6.3.

64. PU **koki-* > PS **ko-* 'see, find : найти'
EnT *ku-* (EnSl.) | EnF *ko-* (EnWl.) | NeT *xosy°* (T65) | NeF *kosy°* (NeSl. *kōš*) | Sk. *qo-* (SkW 1749) | Km. *ku-* (KSz 0459) | Mt. *ko-* (MS 513)
SW: 72; UEW: 171.

65. PU **kompa* > PS **kåmpå* 'wave : волна'
Ng. *koŋhu* (NgSl.) | EnT *kaba* (EnSl.) | EnF *kaba* (EnWl.) | NeT *xampa* | NeF *kampa* (NeSl. *kampa*) | Sk. *qōmpə* (SkW 1851)
Sammallahti 1988: 537; SW: 59; UEW: 203.

66. PU **konta* > PS **kåntâ*(-) 'hunt, kill; sacrificial animal'
Ng. *kontə* (NgSl.) | EnT *kado* (EnSl.) | EnF der. *kadta-* 'kill hunting : убить на охоте' | NeT *xan°* (T65) | NeF *kan°* (NeSl. *kān*) | Mt. *kandəgat* 'pugnacious, eager to fight : raufsüchtig, rauflustig' (MS 416)
Aikio 2006a: 15–17.

67. PU **kopa* > PS **kopå* 'skin : кожа, кора'
Ng. *kuhu* (NgSl.) | EnT *kôba* (EnSl.) | EnF *koba* (EnWl.) | NeT *xoba* (T65) | NeF *kopa* (NeSl. *kōpa*) | Sk. **qopə* (SkW 1794) | Km. *kuba* (KSz 0463) | Mt. *koha*, *koho* (MS 520)
Sammallahti 1988: 537; SW: 73; UEW: 180–181.

— The vowel development from PU **kopa* to PS **kopå* rather than ***kåpå* is unusual, see subsection 3.6.3.

68. PU **korpi-* 'blaze' > PS **kårpâ* 'northern lights : северное сияние', der. *kårpâr-* 'blaze'
EnT *kabo* (EnSl.) | EnF *kabu* (EnWl.) | NeT *xarp°*, der. *xarpərcy°* (T65)
Aikio 2002: 15–16.

— A loan etymology for this Samoyed word has also been suggested, from Proto-Tungusic **garpa-* 'radiate light, shine', **garpa* 'ray, beam' (Helimski *apud* Aikio 2006a: 29). The similarities may also be accidental.

69. PU **kosi-* (**kusi-*) > PS **kot*(-) ‘cough : кашель, кашлять’, der.augm. *kotâjr-* ‘cough : кашлять’

Ng. *kuʔ*, der.v. *kutəδuśa* (NgSl.) | EnT *kuʔ*, der. *kuδur-* (EnSl.) | NeT *xoq* (T65) | NeF *kot°* (NeSl. *kot*) | Sk. **qot*, verb **qotə-*, der. **qotar-* (SkW 1915) | Km. *kuʔ-* (KSz 0462) | Mt. der. *kodor-* (MS 517)

Sammallahti 1988: 537; SW: 74; UEW: 223.

— The Finno-Ugric languages reflect a preform **kusi-* rather than **kosi-*.

70. PU **koska* > PS **kåtå* ‘grandmother : бабушка’

Ng. *kotu* (NgSl.) | EnT *kaδaʔ* ‘mother’s sister’ (EnSl.) | EnF *kaza* (EnWl.) | NeT *xada* (T65) | NeF *kata* (NeSl. *kāta*) | ? Sk. **kotə̲* ‘old’ (SkW 1918) | Mt. **kada* → Tuvan *kada* ‘aunt, grandmother, mother-in-law’ (MS 378)

Sammallahti 1988: 537; SW: 62.

71. PU **kuδa* > PS **kârâ* ‘morning’

Sk. *qarə* (SkW 2217) | Km. der. *karəlʹɛn* ‘in the morning : morgen : утром’

Aikio 2002: 42–43 (PU **kuδi*, PS **kâr*); 2022: 27 (PU **kuda*); UEW: 193.

— Aikio first reconstructed PU **kuδi* (2002), but more recently **kuδa* (2022a: 27), which fits better with the Samoyed vowel developments.

72. PU **kuji-* > ? PS **kijčV-* ‘lie : лежать’

Sk. der. *kučal-* ‘lie down : sich legen’ (SkW 1943) | Mt. *kistə-* (MS 504)

UEW: 197; Zhivlov 2023: 164.

— Zhivlov reconstructs PS **kijtV-*, but, to my knowledge, a development from **-jt-* to Sk *-č-* would be unparalleled; cf. the development in PS **såjtâ-* ‘sew’ > Sk. *sït-* (SW 134–135). The vowel correspondence between Selkup and Mator does not seem to be regular either, so that this etymology should be treated with great caution.

73. PU **kulki-* > ? PS **kəj-* ‘go : пойти’

NeT *xœsy°* (aor. *xəya*) | NeF *käsy°* (aor. *kǎja*; NeSl. *këš*, aor. *kadja*) | Mt. *kaj-* (MS 384)

Aikio 2002: 26; SW: 51.

— The connection of **kəj-* with PU **kulki-* is posited by Aikio (2002: 26). This reflex of PU **kulki-* competes with the following entry. In principle, we could imagine a paradigmatic split where **kulki-* regularly becomes pre-PS **kujâ-*, whence the vowel stem in forms like the connegative **kujâ-k* develops into PS **kuâ-* while the consonant stem in forms like the aorist develops via **kuj-ŋå*

into PS **kâjŋå* (> NeT *xəya*). A stem **kâj-* could have been extracted from the latter. However, as far as I can tell, it is technically unclear whether PS **kəj-* was originally front vocalic or back vocalic.

74. PU **kulki-* > ? PS **kuâ-* 'float downstream : поплыть (по течению)'
NeT der. *xúqla*(*sy*°), der. *xúw*° 'driftwood' (T65) | Sk. der. *qūrə-* (SkW 2284)
SW: 76; UEW: 198; Zhivlov 2023: 164.
— This Samoyed reflex of PU **kulki-* competes with the preceding entry, q.v.

75. PU **kuma-* > PS **kâmâ-* 'fall : повалиться, развалиться'
Ng. *kəməďa* (NgSl. кэмэдя II) | EnT der. *kooxa* 'fallen tree' | NeT *xəwəsy*° (T65) | NeF *kămăsy*° (NeSl. *kamaš*) | Sk. **qam-* (SkW 1823) | Mt. der. *kamaga* 'fallen tree' (MS 405)
Sammallahti 1988: 537; SW: 52.

76. PU **kuńa-* > PS **kâńâ-* 'be closed, close (the eyes) : закрыть глаза'
Ng. der. *kəńədusa* (Keheinen 2023: 80) | EnT *końi-* (EnSl.) | NeT *xəyəsy*° (T65) | Sk. *qāńə-* (SkW 2074) | Km. *kaj-* | Mt. **kańə-*, *kajə-* (MS 420)
Sammallahti 1988: 537; SW: 53–54.
— The vowel of Sk. *qāńə-* is irregular; expected would be ***a*. The retention of **-ń-* in Nganasan, rather than a development to **-j-*, is also unexpected, and I have no explanation for it.

77. PU **kuńći* > PS **kunsâ* 'urine : моча'
Ng. (M) *kunsə* (SW) | Sk. *küśə*, verb *küśə-* ~ *kǖśə-* (SkW 2212) | Km. *künzə* (KSz 0512) | Mt. *kunžʒ* (MS 596)
Sammallahti 1988: 537; SW: 77; UEW: 210.

78. PU **kupsa-* > PS **kâptå-* 'extinguish, put out : тушить, погасить', itr.der. ±**kâptö-* (?**kâptåw-*) 'go out, die away : погаснуть'
Ng. tr. *kəbtuďa*, itr. *kəbtᵘasa* (NgSl.) | EnT tr. *kota-*, itr. *koti-* (EnSl.) | EnF *kɔta* (EnWl.) | NeT tr. *xəbta*(*sy*°), itr. *xəbtyo*(*sy*°) (T65) | NeF tr. *kăptasy*°, itr. *kăptyosy*° (NeSl. *kăptăš*, *kapčoš*) | Sk. tr. *qaptə-*, itr. *qapti-* (SkW 1789) | Km. *kuʔbdə-*, itr.der. *kuʔbdō-* (KSz 0465) | Mt. tr. *kaptə-* (MS 372)
Janhunen 1981: 223; Sammallahti 1988: 537; SW: 54.

79. PU **kura* > PS **kârå* 'bend, curve : изгиб, извилина'
Ng. *kəru* 'манера, привычка : manner, habit, custom' (NgSl.) | EnT *kora* (EnSl.) | NeT *xəra* (T65) | Sk. der. **qaru* (SkW 2225) | Km. *kara-* (KSz 0318)
Aikio 2006: 14–15; SW: 55; UEW: 220.

80. PU **kura* > ? der. PS **kâru* (?**kârâw*) 'knife : нож'
EnT *koru* | EnF *kɔru* (EnWl.) | NeT *xər°* (T65) | NeF *kăt°* (NeSl. *kaŧ*) | Mt. *kuru*(*h*) (MS 611)
Sammallahti 1988: 537; SW: 54.

81. PU **kurki* (**kẹ̈rki*) > PS **kârö* (?**kârâw*) 'crane : журавль'
EnT *kori* (*xario* borrowed from NeT; EnSl.) | NeT *xəryo* (T65) | NeF *kăqłyu* (unexpected *-qł-*; NeSl. *ka"ľju*) | Sk. *qara* (SkW 2223) | Km. *kuro*, *kuruʔjo* (KSz 0498) | Mt. *körüh* ~ *köröh* (MS 557 **kərö*)
SW: 54 (**kârâ*, ? **kârâjâj*)
— This is in origin an onomatopoeia of the sound that cranes make in flight, a trill with a rather abrupt start, so ±"*ʔrrr*". Correspondences and formations may be irregular and partially independent; cf. also Ng. *kokərï* 'crane' (NgSl.).

— PU **kusi-* 'cough' → see **kosi-*.

82. PU **kuwakka* > PS **kuâkå-nå* 'long ago : давно, раньше'
Ng. *kuogunu* (NgSl.)
Aikio 2012: 244.

83. PU ±**kuwi* > PS **kįj* 'moon, month : луна, месяц'
Ng. der. *kiťəδəə* (NgSl.) | Km. *ki* (KSz 0379) | Mt. *kistit* ~ *kistiʔn* (MS 507)
Sammallahti 1988: 537 (**kɨxi*); SW: 69; UEW: 211–212.
— The medial consonant is difficult to reconstruct in this Proto-Uralic word. PS **kįj* could theoretically reflect a derivative in **-j* rather than a change from **-w-* to **-j*. Cf. PU **ćowi* > PS **soj* (curiously enough, these two words are structurally very similar). The Samoyed vowel **į* differs from the vowel **u* reflected in the other Uralic languages.

84. PU **külki* > ? PS ?**käj* 'side : сторона', der. **käjwə* 'id.'

Ng. *kəi* (NgSl.) | EnT *koo* (EnSl.) | EnF *ke* (EnWl.) | NeT der. *xæw°* (T65) | NeF der. *käw°* (NeSl. *këw*) | Sk. der.? *qə̄* (SkW 1768)

Salminen 2023: 377–380; SW: 57–58.

— This etymology is proposed by Salmine 2023: 377–380. The development from PU **ü* to PS **ə* is unexpected, but may be paralleled in PU **küńil* 'tear' > PS **käńəli*. Ng. *kəi* instead of ***kəj* probably does not directly reflect PS **käj*, but rather seems to be a derivative like ?**käjə-j*. A derived body part 'rib : ребро' is also shared widely in Samoyed with the same derivational suffix **-tə(jə)*: Ng. *kəiðəə*, EnT *kooδi*, EnF *kiuze*, NeT *xæw°di°*, NeF *käw°ti°*, Sk. *qə̄tə*, Km. *kot*, Mt. *kajbədV*.

85. PU **künti* > PS **küntə* 'smoke, vapour : дым, пар'

Ng. *kintə* (NgSl.) | EnT *śudo* (EnSl.) | EnF *šudu* (EnWl.) | NeT *syun°*, der. *syinyo* 'fog : Nebel : туман' (T65) | NeF der. *syŭnyu* 'fog : Nebel : туман' (NeSl. *šunju*) | ? Sk. *s'ümtə* (SkW 2487) | Km. *šüńü* (KSz 1098) | Mt. *kündü* (MS 593)

Sammallahti 1988: 537; SW: 79 (**küntə̂*); UEW: 158 (**kintɜ*, **küntɜ*).

— The preservation of PU **ü* as PS **ü* in this word is unexpected, cf. 3.6.1.

86. PU **küńil* > PS ±**käńəlį* 'tear : слеза'

Ng. *kəəlï* poss.3sg. *-tï* (NgSl.) | EnT *koori* (EnSl.) | EnF *kɔjilu* (EnWl.) | NeT *xəyəl°* (T65) | Sk. **qańi* (SkW 2072) | Km. *kajəl* (KSz 0285) | Mt. *kań(ə)li*, *kaj(ə)li* (MS 421 **kəńəli*)

Sammallahti 1988: 537; SW: 53.

— The reconstruction of this word in Proto-Uralic is difficult, but the Proto-Samoyed form should be front vocalic PS **käńəl-* on account of the front vowel harmony of the Nganasan reflex, making the reconstruction PU **kuńili* (Janhunen 1981: 255, Sammallahti 1988: 537) difficult (see Salminen 2023: 379). Selkup is missing the **-l* for some unexplained reason (cf. SW).

87. PU **lapta* > PS **jåptå* 'thin : тонкий'

Ng. der. *ďobtəgəə* (seemingly **jåptə̂*; NgSl.) | EnT *ďata* (EnSl.) | EnF *ďata* (EnWl.) | NeT *yabta* (T65) | NeF *japta* (NeSl. *djāpta*) | Sk. **ćōptə* (SkW 1450) | Mt. der. *čabtəmbuj* (MS 163)

SW: 38; UEW: 238; Zhivlov 2014: 125.

88. PU **läćä-* > PS **jäsä-* 'cover the tent : покрыть чум'
EnT *ćese-* (EnSl.) | NeT *yesye*(*sy*°) (T65) | Sk. *ćęsə̯-* (SkW 1578)
Helimski *apud* Aikio 2002: 49.

89. PU **lämi* > PS **jämVjə* 'broth : суп'
Ng. *ďeməə* 'porridge made of blood, meat, flour and water' (Aikio 2012: 231) | NeT *yewey*° 'soup, broth'
Aikio 2012: 231.
— Ng. *ďeməə* suggests PS **jäməjə* while NeT *yewey*° suggests PS **jämäjə*.

90. PU **lämpi* > PS der.? **jämpV* 'clothes : одежды'
Ng. der.? *ďeŋhʲa* | EnT der. *ćebiδe?* | NeT der. *yimpit*°*q* | Sk. der. *ćęmpət-* (SkW 1472)
Aikio 2002: 13; SW: 42–43.

91. PU **läsi-* > PS **jät*(*ə̈*)- 'by, near, for : к, на, о'
Ng. *ďatə-*, lat. *ďatəŋ* (NgSl.) | EnT lat. *ćeδo?* (EnSl.) | EnF *ďez* (EnWl.) | NeT *yeq*, lat. *yed*°*h* (T65) | Sk. *ćātə* (SkW 1477)
Helimski *apud* Aikio 2002: 48; SW: 44.
— Sk. *ćātə* as if from **jatə* instead; PS **jätə* should have become Sk. ***ćītə*, while PS **jät* should have become Sk. ***ćet*.

92. PU ?**lewi-* > ? PS **jeə̈-* or **jew-* 'shoot : стрелять'
Sk. *ćə̄-* (SkW 1434)
Aikio 2002: 53; Salminen 2023: 392.
— NeT *yenyer-* 'id.' is connected to the Selkup verb by Alatalo (SkW), but not mentioned by Salminen (2023: 392); the connection is indeed phonologically difficult, as PU/PS **e* does not become Nenets *e*, but rather *i*. The Proto-Uralic reconstruction is given by Zhivlov as **lexi-*rather than **lewi-*, which would only be reconcilable with the Selkup verb if this developed from a vowel sequence PS **jeə̈-*. A vowel sequence is found as a reflex of PU **Vγi* only once in PU **suγi-* 'row' (entry 157 below), and once potentially as the result of PU **Vwi* in PS **tiə̈* from PU **siwi* 'year ring' (entry 155 below). As far as I can see, in the word for 'shoot', the nature of the medial consonant cannot be reliably established as either PU **-γ-* or **-w-* based on the other Uralic cognates. However, even PU **lewi-* would perhaps be expected to develop to PS **ji-*, whence Sk. ***ći-*, as in the derivatives from PU **sewi-* 'eat', PS **ti-* (entry 151 below), this etymology

remains very uncertain. If PS **eə̈* develops in Selkup like PS **iə̈*, Sk. *ӱ* would be the regular reflex of a hypothetical PS **jeə̈-*.

93. PU **lę̑δi-* > PS **lę̑r-* 'be frightened : испугаться', der. **lę̑rö-* (?**lę̑rə̑w-*)
NeT der. *leryo*(*sy*°) | NeF *ťiťyosy*° (NeSl. *ťiťjoš*) | Sk. der. *lę̑ri-* (SkW 2895), ? *nį̑r*(*a̲k*)*ka̲-* 'fürchten' not the same acc. to SkW (1411) | ? Km. *nere-* (KSz 0687) | Mt. *nę̑r-* (?) (MS 747)
Aikio 2014b: 85–86; SW: 83.
— NeT *leryo-*, NeF *ťiťyo-* and Sk. *lę̑ri-* clearly belong together, but it is unclear to me whether the *n*-initial forms in Selkup, Kamas and Mator really belong here, or if they can be the result of an irregular change from **l-* to *n-*.

94. PU **lę̑mpi* > PS **lį̑mpə̑* 'swamp : болот', **lį̑mpå-* 'be swampy : быть болотистым', **lį̑mpə̑-* 'sink into a swamp : повалиться в болото'
Ng. *ľüŋhə* 'boggy place' (Aikio 2014b: 86) | EnT *lubo* (EnSl.) | EnF *lubu* | NeT *limpa*(*sy*°), *limpə*(*sy*°), der. *limpəd*° 'swamp : болото' (T65) | NeF der. *ťĭmpăt*° 'swamp, bog : болото, топь' (NeSl. *ľimpat*) | Sk. der.? *lį̑mpä* ~ *lį̑mpi* (SkW 2803)
Aikio 2014c: 86; SW: 83; UEW: 235 (**lampe*).

95. PU **lę̑nti* > PS **lį̑ntə̑* 'plain, valley : равнина, долина'
Ng. *ľintə* (NgSl.)
Aikio 2014b: 86; Sammallahti 1988: 552 (Proto-Finno-Permic); UEW: 235–236.

96. PU **lę̑pći* > PS **lę̑psə̑* ~ *ję̑psə̑* 'cradle : люлька'
Ng. *labsə* (NgSl.) | EnT *ĺiťo* (EnSl.) | EnF *l'ičʉ* (EnWl.) | NeT *yebc*° (T65) | Sk. *ćopsə* (SkW 1452) | Km. *ťepsə* | Mt. *čə̄psę* ~ *sə̄bsę* (MS 205 **ĺę̑*(*ə*)*psə*)
Aikio 2002: 48; SW: 41 (**ję̑psə̑*, ?**ję̑ə̑psə̑* ~ **lę̑ə̑psə̑*); UEW: 260.
— The vacillation of initial **l-* or **j-* in this word is unique and therefore puzzling. Salminen (2023: 384) argues for a reconstruction PS **ĺę̑psə̑* (so, too, Helimski MS), with an initial palatal **ĺ* reconstructed up to the Proto-Uralic stage (cf. also UEW). Since the Proto-Uralic phoneme **ĺ* would be quite unique to this word, alternative explanations should also be considered, such as an assimilation in palatality to the originally palatal PS **s* from PU **ć*, which could have taken place within early Samoyed; i.e., (pre-)PS **lę̑pśə̑* > **ĺę̑pśə̑* vel sim. The resulting PS **ĺę̑psə̑* could have developed with **ĺ* to **j* in all branches except for Nganasan and Enets, as posited by Salminen and Helimski.

97. PU **lęwi* > PS **lę* 'bone : кость'
Ng. der.? (**lę-jə̂*) *laa* 'ring : кольцо', der. (**lę-tə̂jə̂*) *latəə* 'bone : кость' (Kaheinen 2023: 167; NgSl.) | EnT der. *Ĺiδi* (EnSl.) | EnF der. *lizi* (EnWl.) | NeT *lí, le-* (T65) | NeF *łĭ, łe-* (NeSl. *lĭ̆, le-*) | Sk. *lį̄, lę-* (SkW 2761) | Km. *le* (KSz 0529) | Mt. *lE* (? *lę*) (MS 618)
Sammallahti 1988: 538 (**lɨxɨ*); SW: 82; UEW: 254–255 (**luwe*); Zhivlov 2023: 161 (**lęwi*)
— Nenets and Selkup show a shared distribution of long **i̯* in the nominative singular and **ę* in the other cases, but it is unclear how this might best be explained. Perhaps a reconstruction such as nom.sg. **lęw* > **lį̄* next to acc.sg. **lę(w)m* = **lęm* could be attempted, but there are no direct parallels.

98. PU **lomi* (**lumi*) > PS **jom*(-) 'snow (v. and n.) : снег, идти (о снеге)'
Ng. *d'um* (NgSl.) | NeT der. *yomcyah* (T65) | Sk. **ćom-* (SkW 1459)
Sammallahti 1988: 538; SW: 46.
— The Finno-Ugric languages reflect a preform **lumi* rather than **lomi*.

99. PU ?**luča* > PS **jə̂cə̂* 'block of wood : доска'
Ng. *d'etə* (Aikio 2006: 29) | EnT *d'oδo* (EnSl.) | EnF *d'ɔzu* (EnWl.) | NeT *yəd°* (T65) | NeF (L)*jᵊà̀t̀* (*jăt°*) (SW) | Sk. *ćačə* (SkW 1474)
Aikio 2002: 12–13; 2006a: 29; SW: 36 (**jə̂t¹ə̂*).
— A reconstruction PU **luča* is needed to unite Fi. *luta* 'tool for removing birch bark' with the Samoyed words, but requires Saami forms like SaaN *lohti* 'wedge' to be borrowed from Proto-Finnic, since *lohti* could only reflect PU **luta* (Aikio 2006: 29).

— PU **lumi* 'snow' → see PU **lomi*.

100. PU **lupsa* > PS **jə̂ptå* 'dew : роса'
Ng. der. *d'ebtua* (NgSl.) | EnT *d'ota* (EnSl.) | NeT *yəbta* (T65) | NeF *jăpta* (NeSl. *djăpta*) | Sk. der. *ćaptu* (SkW) | Km. *t'eʔbda* (KSz 1338) | Mt. чипталъ (-лъ unclear; MS 208)
Janhunen 1981: 223; Sammallahti 1988: 538; SW: 35–36; UEW: 261.

101. PU **meni-* > PS **men-* 'go : идти, ехать'
Ng. *mïnsï* (NgSl.) | EnT der. *midir-* 'carry : нести' (EnSl.) | EnF der. *midir* 'carry : нести' | NeT *myincy°* (T65) | NeF *myĭnsy°* (NeSl. *minьš*) | Km. *mĭn-* (KSz 0592)
Sammallahti 1988: 538; SW: 94; UEW: 272.

102. PU **meγi-* > PS **mi-* 'give, sell : дать, продать'
Ng. der. *miśi* (NgSl.) | EnT der. *mis-* (EnSl.) | EnF der. *mis-* | NeT der. *myicy°* (T65) | NeF *myĭqsy°* (NeSl. *mi'š*) | Sk. *mi-* (SkW 735) | Km. *mĭ-* (KSz 0588) | Mt. *mi-* (MS 667)
Aikio 2012: 248; Sammallahti 1988: 538; SW: 94; UEW: 275 (**miγe-*).

103. PU **mẹ̑ksa* > PS **mi̯tâ* 'liver : печень'
Ng. *mitə* (NgSl.) | EnT *muδo* (EnSl.) | EnF *muzu* (EnWl.) | NeT *mid°* (T65) | NeF *mĭt°* (NeSl. *mït*) | Sk. *mī̯tə* (SkW 757) | Km. *mĭt* (KSz 0604)
Sammallahti 1988: 538; SW: 93–94; UEW: 264 (**maksa*).

104. PU **minä* > PS **mən* 1sg. personal pronoun 'I : я'
Ng. der.? *mənə* (NgSl.) | EnT der. *moʹdi* (EnSl.) | EnF *modʹi* der. (EnWl.) | NeT der. *məny°* | NeF der. *mănу°* (NeSl. *manь*) | Sk. *man* (SkW 771) | Km. *man* (KSz 0550) | Mt. *mən* (Ms 627)
Janhunen 2013: 214; SW: 86.

105. PU **moćki-* > PS **måsâ-* 'wash, smear : вымыть, смазать'
EnT *masu-* (EnSl.) | EnF *masu-* ~ *məsu* (EnWl.) | NeT *masə*(*sy°*) (T65) | NeF *masăsy°* (NeSl. *masaš*) | Sk. *mūsə-* / *müsə-* (SkW 860) | Km. *buzə-* (KSz 0170) | Mt. *masə-* (MS 642)
Sammallahti 1988: 538; SW: 89; UEW: 289 (**muśke-*, **mośke-*).

106. PU **muδ´a* > PS **mâjå* 'earth, land, terrain : земля, страна, край'
Ng. *mou* (NgSl.) | Sk. *maćə* 'Wald(rücken)' (SkW 799) | Km. *mĭja* 'mountain' | Mt. *bijä* 'mountain : Berg' (MS 124)
Aikio 2002: 22–23; SW: 85.

107. PU **muja-* > PS **mâjâ-* 'become happy : обрадоваться'
NeT *məyə*(*sy°*) (T65) | NeF *măjăsy°* (NeSl. *mădjăš*)
Aikio 2002: 22.

108. PU **muka* > PS **mâkå* 'back : спина'
Ng. *məku* (NgSl.) | EnT *maxa* (EnSl.) | EnF *maxa* (EnWl.) | NeT *məxa* (T65) | NeF *măxa* (NeSl. *maxa*) | Sk. *moqə*, der. *moqal*/*moqol* 'id.' (SkW 818) | Km. der. *begəl* (KSz 0120) | Mt. *baga* (MS 90)
Janhunen 1981: 223; Sammallahti 1988: 538; SW: 85.

109. PU **muna* > PS **mânå* 'egg : яйцо'
Ng. *mənu* (NgSl.) | EnT *mona* (EnSl.) | EnF *mɔna* (EnWl.) | Sk. *manə* 'penis' (SkW 772) | Km. der. *munoj* (KSz 0626)
Janhunen 1981: 223; Sammallahti 1988: 538; SW: 86; UEW: 285–286.
— For the semantics of the Selkup word, cf. the Finnish cognate *muna* 'egg; penis, testes'.

110. PU **mura* > ? der. PS **mârVŋkå* 'cloudberry : морошка'
Ng. *murïŋga* (NgSl.) | EnT *moraga* (EnSl.) | EnF *mɔroga* | NeT der. *məraŋka* | NeF *măłaŋka* (NeSl. *małaŋka*)
Janhunen 1981: 223-224; Sammallahti 1988: 538; UEW: 287.
— The vowels of Nganasan *murïŋga* do not regularly correspond to Enets and Nenets forms; expected would be Ng. ***məruŋga*, cf. PS **mânå* 'egg' > Ng. *mənu*.

111. PU ±**nataw* > PS **nåto* (?**nåtåw*) 'brother-in-law : деверь'
NeT *nado* (T65) | Km. *nado* (KSz 0644)
Sammallahti 1988: 539 (**nåtiw*); SW: 98 (**nåt¹ə̂* ? **nåt¹ə̂j*); UEW: 299–300.

112. PU **nimi* > PS **nim* 'name : имя'
Ng. *ńim* (NgSl.) | EnT *ńi?* (EnSl.) | EnF *ńi?* (EnWl.) | NeT *nyum* (T65) | NeF *nyŭm* (NeF *nim*) | Sk. *nim* ~ *ńim* (SkW 1349) | Km. *nim* (KSz 0691) | Mt. *Nim* ~ *nüm* (MS 756)
Sammallahti 1988: 538; SW: 102; UEW: 305 (**nime*).

113. PU **nokki* > PS der. **nåku* 'neck : затылок'
Sk. *nuku* (SkW 1385)
UEW: 328–329; Zhivlov 2016: 299.

114. PU **nusi-* > PS **nât-* 'scrape : выскоблить'
EnT (analogical) *nos-* (EnSl.) | EnF (analogical) *nɔs-* (EnWl.) | NeT *nəcy°* (T65) | Sk. **nat-* (SkW 1351) | Km. *nă?-* (KSz 0676)
Sammallahti 1988: 538; SW 97; UEW: 309.
— The reduction of PU **u* to PS **ə̂* before PU **-i* rather than before **-a* seems irregular here, but it may have been extracted and generalized from common forms like the verbal noun PU **nusma* > PS **nâtmå* and the aorist stem, quasi-PU **nusŋa-* > PS **nâtŋå-* (NeT *nəqŋa-*, etc.).

115. PU **nüδi* > PS **nir* 'handle : рукоятка'
Ng. *ńir* (NgSl.) | EnT *ńi?* (EnSl.) | EnF *ńi?* (EnWl.) | NeT *nyir* (T65) | NeF *nyił* (NeSl. *nił*) | Sk. *nir* (SkW 1417) | Km. der. *ńirže* (KSz 0754)
Sammallahti 1988: 538; SW: 102–103; UEW: 304 (**niδe*, **nüδe*).

116. PU **ńali-* > PS **ńåâ-* 'lick : лизать', der. **ńå-nsâ-* 'id.'
Ng. der. *ńonsəďa* (NgSl.) | EnT *naǵo-* (EnSl.) | EnF *naso-* ~ *nazo-* (EnWl.) | NeT *nyancə*(*sy*°) | NeF *nyansăsy*° (NeSl. *njansăš*) | Sk. *ńū-* (SkW 1627) | Km. *nü-* (KSz 0723)
Sammallahti 1988: 539 (**ńåxli-*); SW: 105; UEW: 321.
— Nganasan *ńonsə-* has to be derived from a stem *ńå-*, while the other cognates, especially Selkup and Kamas, suggest the more expected reconstruction PS **ńåâ-*. It would be preferable to derive the Nganasan, Enets and Nenets forms from the same ancestral form, however, so that the absence of the vowel sequence needs to be accounted for. I find it probable that **ńå-nsâ-* was simply based on a new "consonant stem" of the root **ńåâ* after disappearance of the **-l-*, cf. PS **kå-tå-* 'kill' rather than ***kåâ-tå-* as the causative of PS **kåâ-* 'die' (see the entry of PU **kali-* 'die'). At least the vowels of underived Sk. *ńū-* and Km. *nü-* are entirely parallel to those of PS **kåâ-* from PU **kali-* 'die'.

117. PU **ńęli* > PS ±**ńęj* 'arrow : стрела'
NeT compound *tu*°*nyi* 'rifle, shotgun : ружьё, винтовка' | Sk. *ńęj̲ə̲* (SkW 1663) | Km. *ńie* 'arrow : Pfeil', *ńa* 'bullet : Kugel' (KSz 0734) | Mt. *ńej*, *nej* (MS 729)
Aikio 2012: 228; Sammallahti 1988: 539 (**ńɨxli-*); SW: 108.

118. PU **ńęri* > PS **ńęr* 'cartilage : хрящ'
Ng. *nir* | EnT *ni?* (EnSl.) | EnF *ni?* (EnWl.) | NeT *nyer* (T65) | NeF der. *nyelăqku* (NeSl. *njelaqku*) | Sk. *ńi̯r* (SkW 1707)
Aikio 2006: 20–21; SW: 108; UEW: 324.

119. PU **ńimi-* > PS **ńim-* 'suck : сосать', n.der. *ńimsä* 'breast, milk : молоко'
Ng. der. *ńimibtiśi* (NgSl.) | EnF der. *nuutagu-* (EnSl.) | NeT der. *nyimnye*(*sy*°) (T65) | NeF der. *nyimnyapyosy*° (NeSl. *nimnjapjoš*) | Sk. *ńim-*, n.der. *ńipsə* (SkW 1638) | Km. der. *ńimer-* (KSz 0752) | Mt. der. **ńümnə-*, *nümnə-* (MS 786)
Aikio 2002: 23–25; SW: 110.

120. PU **ńoγi-* (**ńuγi-*) > PS **ńo-* (or **ńoə̑-*?) ‘pursue : преследовать’
NeT der. *nyoda*(*sy*°) (T65) | NeT *nyotasy*° (NeSl.) | Sk. *ńō-*, der. *ńōtə̲-* (SkW 1625) | Mt. **ńodə-*, *nodə* (MS 766)
Sammallahti 1988: 539 (**ńoxɨ-*); SW: 111; UEW: 323–324 (**ńoŋδa-*, **ńowδa-*).
— None of the attested Samoyed descendants reliably show a difference between **o* and **oə̑*, so that there is no way to be absolutely sure which should be reconstructed. Most words of the structure PU **CVγi* end up as PS **CV*, but PU **suγi-* does seem to have become PS **tuə̑-* rather than **tu-* (q.v.). The Finno-Ugric languages may reflect a preform **ńuγi-* rather than **ńoγi-*.

121. PU **ńoma* > PS **ńåmå* ‘hare : заяц’
Ng. *ńomu* (NgSl.) | EnT *naba* (EnSl.) | EnF *n’aba* (EnWl.) | NeT *nyawa* (T65) | Sk. (der.?) *ńoma* (SkW 1633)
Sammallahti 1988: 539; SW: 106.
— Selkup (Tym) *ńoma* should probably be a derivation, since PS **-å* usually becomes *-ə* in Selkup. The Finno-Ugric languages reflect a derived **ńoma-la* ‘hare’ instead of the underived PS **ńåmå*. An etymological connection with the verb PU **ńoγi-* ‘pursue’ (Janhunen 1981: 242) is possible if the Samoyed **o* vowel of **ńo-* ‘pursue’ is more archaic than the **u* of FU **ńuγi-* (against this connection, see Zhivlov 2023: 162, who reconstructs a development of PU **ńuγi-* to PS **ńo-* rather than of PU **ńoγi-* to PS **ńo-*, FU **ńuγi-*; cf. the discussion in subsection 3.6.3). The etymologized reconstruction PU “**ńoγma*” is not expected to develop differently from **ńoma*.

122. PU **oδ´a* > PS **åjå* ‘meat, body : мясо, тело’
EnT *aja* (EnSl.) | EnF *aja* (EnWl.) | NeT *ŋaya* (T65) | NeF *ŋaja* (NeSl. *ŋadja*) | Sk. *uə̑ćə* (SkW 162) | Km. *uja* (KmSz.) | Mt. der. *ojobtu* (MS 799)
Aikio 2006: 11–12; SW: 17.

123. PU **ojwa* > PS **åjwå* ‘head : голова’
Ng. *ŋojbu* (NgSl.) | EnT der. *aburi* (EnSl.) | EnF *εba* (EnWl.) | NeT *ŋœwa* (T65) | NeF *ŋăwa* (NeSl. *ŋajwa*) | Mt. *ajba* (MtSpr.)
Sammallahti 1988: 536; SW: 17; UEW: 336–337.

124. PU **pala-* > PS **pålä-* ‘swallow’
NeT *palye-* | Sk. *pōlə-* (SkW 693)
Sammallahti 1988: 540; SW: 116; UEW: 350.

125. PU **pata* > PS der.v. **patə̂-* ‘put something in a pot, in water : положить на котёл, в воду’, der. ?**patü-* ‘id.’

Ng. *h$^{u}$atiďa* ‘dip, plunge (pfv.) : окунуть’, *h$^{u}$atətəsa* ‘id. (ipfv.)’ (NgSl.) | EnT *peδi-* (EnSl.) | NeT *pyadə*(*sy*°), intr. *pyadə*(*sy*°) ‘be put in the water : быть опущенным в воду’ (T65) | NeF *pyatăsy*° (NeSl. *pjātăš*) | Sk. *pōt-* (SkW 431) | Km. *padə-* ‘insert; feed : hineinstecken, füttern : воткнуть; кормить’ (KSz 0798)

Aikio 2015: 55; Sammallahti 1998: 548 (FU); SW: 118; UEW 358 (FU).

— The Uralic connection is with the noun PU **pata* ‘pot’, found in a number of Finno-Ugric languages, but the use of the Samoyed verb also seems to be more general than simply inserting things into pots specifically. Sk. *pōt-* is also used for laying nets in the water, and the next entry in SkW 432 *pōti* ‘ice hole for a drawnet : Wuhne für das Zugnetz’ seems to be a likely derivative on the basis of this semantic link. Loss of the more general meaning found in Samoyed could in principle be a shared innovation in the Finno-Ugric languages relevant for the phylogeny. As for the formal reconstruction in Samoyed, Ng. *h$^{u}$ati-*, EnT *peδi-*, NeT *pyadə-*, NeF *pyată-* might all reflect PS **patü-*, but the Nenets forms may also go back to PS **patə̂-*, **patu-* or **patį̆-*. Sk. *pōt-* rather seems to derive from a form **pat-*, with the frequent analogical removal of final schwa (and a development from **a* to *ō* rather than to ***ā* due to the initial *p-*). Km. *padə-* can certainly derive from PS **patə̂-*, which is also necessary for derived Ng. *h$^{u}$atətə-*, and which forms the basis for a secondary consonant stem in Sk. *pōt-* as well.

126. PU **pawi* > PS **pa* ‘tree : дерево’

Ng. der. (**pa-jə̂*) *h$^{u}$aa* (NgSl.) | EnT *pe* (EnSl.) | EnF *pɛ* (EnWl.)| NeT *pya* (T65) | NeF *pyă* (NeSl. *pja*) | Sk. *pō* (SkW) | Km. *pa* (KmSz.) | *hä* (MS 267)

Sammallahti 1988: 539; SW: 117 (**pä*); UEW: 410–411.

127. PU **päli* > PS **pää* (~ **päj* ?) ‘edge, outside : остриё’

Ng. der.? *h$^{i}$ai* (NgSl.) | EnT *pio* (EnSl.) | EnF *pe* (EnWl.) | NeT *pyí* | NeF der. *pyinyaŋ* ‘out : наружу’ (NeSl. *pīnjaŋ*) | Sk. *pō* | MS loc. *hegen* (MS 276)

Aikio 2012: 239; Janhunen 1981: 241; SW: 124 (**piə̂*).

— Ng. *h$^{i}$ai* could reflect PS **päj*, or perhaps a schwa-final form like **päjə* or derivative **päjə-j*, while the other Samoyed languages can go back to PS **pää* (especially EnT *pio* and NeF *pyinyaŋ* rather than ***pyenyaŋ* point to this reconstruction in favour of **päj*). Selkup *pō* could probably reflect PS **päj* with

rounding from expected ***pē* due to the initial labial consonant, but it is unclear if PS **pää* would not develop in the same way in this language.

128. PU **pälkä* > ? PS ±**peä* 'thumb : большой палец'
Ng. der. *hïaďə* (NgSl.) | EnT der. *piiťu* (EnSl.) | NeT der. *pyíkcya* (T65) | Km. der. *pidə* (KSz 0840) | ? Mt. der. *hegəbti* (MS 275)
Aikio 2020: 95; SW: 123.
— The Proto-Samoyed reconstruction is difficult, and the precise connection and relation with PU **pälkä* is not at all straightforward.

129. PU ?**pä/iŋi* > PS **päŋ* 'hollow hand : ладонь'
Ng. *hʲaŋ* | EnT *peo* (EnSl.) | NeT *pyeh* (T65) | NeF *pyěŋ* (NeSl. *pjeŋ*) | Sk. *pīŋ* (SkW 597) | Km. *pĭŋ* (KSz 0853) | Mt. *höŋ* (MS 309)
Janhunen 1981: 257–258; Sammallahti 1988: 539.
— Sk. *pīŋ* is due to pre-schwa lengthening of **ä* to Proto-Selkup **ī*, see Helimski (2007: 129); EnT *peo* similarly reflects a schwa-final variant or case form (ibid.). The vowel correspondence with the cognate Finnish *pivo* 'hand, palm' (and other Finnic forms) is not regular.

130. PU **peɣi-* > PS **pi-* 'cook, ripen (by cooking) : свариться', der.caus. PS **pirä-* 'cook (tr.) : сварить'
Ng. der.caus. *hiriďi* (NgSl.) | EnT *pi-*, der.caus. *pire-* | EnF *pi-*, der.caus. *piri-* (EnWl.) | NeT *pyisy°*, der.caus. *pyirye*(*sy°*) (T65) | NeF *pyĭsy°*, der.caus. *pyĭťyesy°* (NeSl. *piš*, *piťjēš*) | Sk. der.caus. *pirə̲-* (SkW 680) | Km. *pü-*, der.caus. *pürə-* (KSz 0875)
Sammallahti 1988: 539; SW: 123–124; UEW: 368.

131. PU **peksä-* > PS **pet-* 'hit, soften a hide : размять шкуру'
Ng. *hït-* (Aikio 2015: 14) | NeT der. *pyidyelcy°*, *pyidyercy°* (T65) | Sk. *pit-*, *pitə-* (SkW 462) | Km. *püʔ-* (KSz 0877)
Aikio 2002: 54 fn. 8; 2014a: 14; SW: 126; UEW: 368–369.
— The reduction of PU **-ä* to **-ə* or ∅ in Samoyed is very infrequent and not properly described or understood.

132. PU **peli-* > PS **pej-* ‘be afraid : испугаться’
Ng. der. *hüümsï,* der. *hiľïʔśi* ‘fear : бояться’ (NgSl.) | EnT der. *piiro-*, *piis-* (EnSl.) | EnF der. *pizi-* ‘frighten : испугать’ | NeT der. *pyílcy°*, der. *pyínə*(*sy°*) (T65) | NeF *pyinăsy°* (NeSl. *pinăš*) | Sk. *pī-* (SkW 393b.) | Km. der. *pim-* (KSz 0844) | Mt. der. *himər-* (? *hīmər-*) (MS 300)
Sammallahti 1988: 539; SW: 125; UEW: 370.

133. PU **pęni-* > PS **pęn-* ‘put : положить’
Ng. *h$^{u}$ansa* (NgSl.) | EnT *pun-* (EnSl.) | EnF *pun-* (EnWl.) | NeT *pency°* (T65) | NeF *pensy°* (NeSl. *реньš*) | Sk. *pęn-* (SkW 491) | Km. *pen-* (KSz 0829) | Mt. *hęn-* (MS 288)
Sammallahti 1988: 539; SW: 118; UEW: 353–354 (**pane-*).

134. PU **piδi* > PS der. **pirə* ‘height : высота’
Ng. *hirə* (NgSl.) | EnT *piro* (EnSl.) | EnF postp. *piro* (EnWl.) | NeT *pyir°* (T65) | NeF der. *pyĭłăt°* (NeSl. *piłat*) | Sk. postp. *pīr*, *pīrə* (SkW 677)
Sammallahti 1988: 539; SW: 125.

135. PU **piδkä* > PS **pirkä* ‘high : высокий’
EnT *píde* (EnSl.) | EnF *pise~pize* (EnWl.) | NeT *pyircya* (T65) | NeF *pyĭłsya* (NeSl. *piłьša*) | Sk. *pirkə* (SkW 977) | Km. *pĭrže* (KSz 0855) | Mt. *hirge* (MS 305)
SW: 125.

136. PU ?**pijri* ‘circle, around’ > ? PS **pijr* ‘inner bark of birch : камбиальный слой берёзы’
NeT *pyír*
Salminen 2023: 388–390.

137. PU **piksi* > PS **pütə* ‘net string, cord : нитка’
NeT *pyud°* (L 390 *pʹūð$^{\text{ɐ}}$*) | Sk. *pūtə* (SkW 458)
Sammallahti 1988: 539; UEW: 380.

— With an irregular development from **i* to **ü*, possibly due to the initial **p-*; this is also seen with **wi* to **(w)ü* in PU ±**wijitti* > PS **wüət* ‘ten (10)’ and PU **wiγi-* > PS **ü-* ‘drag’ (s.vv.), but I know of no further examples of **pi* to **pü*; rather, the other, reliable example of PU **pi* yield PS **pi* in Samoyed derivatives from the root **piδi* ‘long’. Not in SW; Selkup cognate absent in UEW, SkW gives the connection with Nenets.

138. PU **ponči* > PS **pâncə̂* 'hem, lower edge : нижний край'
Ng. der. *hontəə* | EnT *padi* (EnSl.) | EnF *pade* (EnWl.) | NeT *pan°* (T65) | NeF *pan°* (NeSl. *pan*) | Sk. der. *pōnčar* (SkW 486) | Km. *pandar* (KSz 0804) | Mt. ? *handa* (MS 258)
Aikio 2006a: 23–24; SW: 116.

139. PU **pori-* (**puri-*) > PS **por-* 'eat : съесть', der.vn. **porså* 'fish meal : варка'
Ng. der.vn. *hursu* (NgSl.) | EnT der.vn. *podá* (EnSl.) | NeT der.vn. *porca* (T65) | Sk. der.vn. *porsə* (SkW 647, derived there from Even *норочō* instead) | Mt. *hor-* (MS 316)
Sammallahti 1988: 539; SW: 127–128; UEW: 405–406 (**pure-*).
— PS **porså* is a derivative of the same type as PS **âmså* 'food' from **âm-* 'eat' (SW 15), and it has a wide distribution from Nganasan to Selkup. The underived verb in Mator **hor-* confirms the Proto-Samoyed reflex of the verb PU **pori-*. The Finno-Ugric languages reflect a preform **puri-* rather than **pori-*.

140. PU **poski* > PS **påtə̂* 'cheek : щека'
Ng. der. *xotuə* (NgSl.) | En. (C) der.? *fáru*, irreg.? *faede* ~ *paede* (SW) | NeT der. *padu* (T65) | Sk. der. *pūtəl* | Km. der. *puʔma* (KSz 0869) | Mt. der. *hoʔlo* (MS 306)
Sammallahti 1988: 539; SW: 117; UEW: 396.

141. PU **pučki* > PS **pucə̂* 'core : сердцевина'
Ng. der. *hütəðə* 'torso, body, figure : туловище, тело, фигура' (NgSl.) | EnT der. *puδodo* 'torso, body : туловище, тело' (EnSl.) | EnF der. *puzuj* 'spinal cord' (EnWl.) | NeT *pud°* 'that which gives balance, steadiness', *pyah pud°* 'core of a tree : сердцевина дерева' | Sk. *pūčə* (SkW 442) | Km. *püt* (KSz 0898) | Mt. *hudu-* 'self' : selbst, sich selbst' (MS 321)
Aikio 2014a: 11–15; Sammallahti 1988: 539; SW: 129–130.

142. PU **puδ´a-* > PS **pâjä-* 'chop wood : нарубить двор'
Ng. *hoiďi*, der. *hoiʔśi* (NgSl.) | EnT der. *petur-* (EnSl.) | EnF der. *pɛturu-* (EnWl.) | NeT der. *pæ°cy°* (T65) | NeF der. *pǎjǎqsy°* (NeSl. *padja'š*) | Sk. der. *paćal-* (SkW 499) | Mt. der. *häjäl-* (MS 277)
Aikio 2006a: 22–23; SW: 112; UEW: 389–390 (**poδ´ɜ-*).

143. PU ?**pujɨ* > PS **puə̑* ‘the backside : сзади, назад’
Ng. *huə*, adv. (loc.) *huənu* (NgSl.) | EnT der.adv. (lat.) *puoʔ* (EnSl.) | EnF adv. (loc.) *pɔna*, der.adj. *pɔnuj* (EnWl.) | NeT adv. (loc.) *púna*, (abl.) *púd°* (T65) | NeF adv. (loc.) *puna*, (abl.) *put°* (NeSl. *puna*, *put*) | Sk. adv. *pūn* (SkW 379) | Km. der. *puʔmo* (KSz 0873) | Mt. adv. (loc.) *huna* (? *hūna*) (MS 325)
Aikio 2012: 247; SW: 129 (**puə̑*, ? **puwə̑*); UEW: 401.

144. PU **puna-* > PS **pə̑n-* ‘braid, weave : заплести’
Ng. *honsa* (NgSl.) | EnF der. *pɔguru-* (EnWl.) | NeT der. *pəŋkəlcy°* | NeF der. *pă̈ŋkă̈lsy°* (NeSl. *paŋkałьš*) | Sk. *pan-* / *pač-* (SkW 473) | Km. *pün-* (KSz 0888)
Sammallahti 1988: 539; SW: 133; UEW: 402–403.
— The final **-a* is unexpectedly lost in Samoyed, most likely because it is of the type that became PS **-ə̑*, which could be eliminated by analogy (see subsection 3.7.1).

145. PU **pura* > der. PS **pə̑råjŋ* ‘drill : сверло’
EnT der. *poĺiďu* (?) ‘awl : шило’ (EnSl.) | EnF *pɔľiďu* (EnWl.) | NeT *pəreh* | NeT *pă̈łiŋ* (NeSl. *paɬüŋ*) | Sk. *paräŋ* / *pariŋ* ‘icepick : Eishaue’ (SkW 629) | Km. *paraŋ* (KSz 0808)
Sammallahti 1988: 539; SW: 114; UEW: 405.

— PU **puri-* ‘bite’ → see **pori-*.

146. PU **purki* > PS **pur-* (in derivations) ‘smoke : дым’
NeT der. *pur°q* (T65) | ? Sk. der. *purqə* (SkW 661)
Aikio 2002: 25–27; SW: 131.
— The Selkup might be a loan from Khanty instead, KhE *pŏrkį* ‘id.’ (see SkW, Aikio 2006: 30).

147. PU **puwa-* > PS **puə̑-* ‘blow : дуть’
Ng. der. *hüarsa*, der. *hüoləsa*, der. *hüoltəsa*, etc. (NgSl.) | EnT der. *puaro-*, ? der. *pueso-* (EnSl.) | EnF ? der. *pɔsa~pesa-* (EnWl.) | NeT der. *púcy°*, der. *púqlyo*(*sy°*), ? der. *posə*(*sy°*) (T65) | NeF ? der. *pusipyosy°* (NeSl. *pusïpyoš*) | Sk. *pū-*, der. *pūl-* ‘swell : schwellen’ (SkW 381, 700) | Km. der. *püʔ-* (KSz 0877) | Mt. der. *hal-* (MS 255)
Aikio 2012: 244; SW: 128–129; UEW: 411.

148. PU **sala-* > PS **tålä-* ‘steal : украсть’, der.augm. *tåleŋr-* (?**tåläjr-*) ‘steal (ipfv.) : воровать, красть’
Ng. *tolïďa*, der.augm. *tolarsa* (NgSl.) | EnT *tare-*, der.augm. *tarir-* (EnSl.) | EnF *tari-* (EnWl.) | NeT *talye*(*sy*°), der.augm. *taler-* (T65) | NeF *tałyesy*° (NeSl. *tałješ*) | Sk. *tuâlə-*, der.augm. *tuâlir-* (Sk 1306) | Km. *tol'er-* (KSz 1232) | Mt. *tälər-* (MS 975)
Sammallahti 1988: 540; SW: 150–151; UEW: 430–431.

149. PU **sarka* > PS **tårkå* ‘fork, bifurcation, branch : отросток, ветв, ответвление’
EnT *taga* (EnSl.) | NeT *tarka* (T65) | NeF *tałka* (NeSl. *tałka*)
Sammallahti 1988: 540; SW: 152.

150. PU **säji* > PS **täj* (~ **täə* ?) ‘pus : гной’, der.v. ‘rot : сгнить’
Ng. *ťüm-* | EnT *ćüm-* (EnSl.) | NeT *tyím-* | Sk. *tē*, verb *tē-* (SkW 950) | Km. verb *te?-* (KSz 1291) | Mt. *ti* (?) (MS 1007)
Sammallahti 1988: 540 (**sexji*); SW: 161 (**ti* ~ **te*).
— Cf. PU **käli* > PS **käj* (~ **käə* ?) ‘tongue’. The Mator reflex “*ki*” ‘tongue’ next to “*gei*” and “*káschtä*” ±|käj-tä| ‘his/her tongue’ (MS 1007) may point to a parallel reconstruction of these two words. The vowel reflex of Selkup *tē* is also identical with that of *śē* ‘tongue’. Ng. *ťüm-*, EnT *ćüm-* and NeT *tyím* probably reflect **täjm-* rather than **täəm-*, but **täəjm-* might also be a possibility.

151. PU **sewi-* ‘eat’ > der. PS **timä* ‘tooth : зуб’
Ng. *ťimi* | EnT *ćü* (EnSl.) | EnF *čü* (EnWl.) | NeT *tyibya* (T65) | NeF *tyĭm*° (NeSl. *čim*) | Sk. *timə* (SkW 1037) | Km. *time* (KSz 1198) | Mt. *time* (MS 1091)
Aikio: 2002: 34–36.
— This etymology is called into question by Salminen (2023: 394), since, according to him, only PU **seɣi-* should yield PS **ti-*, while PU **sewi-* should become PS ***tew-*. However, other instances of **Vwi* normally behave very much like sequences of **Vɣi* in Samoyed as well (cf. subsection 3.4.3) so that a development from PU **sewi-* to PS **ti-* seems plausible enough (as in PU **meɣi-* > PS **mi-* ‘give, sell’). The reflexes within Samoyed are furthermore regular, with the exception of the second-syllable vowel of Forest Nenets *tyĭm*° rather than expected ***tyĭmya*. There, we can assume the same analogical development as took place in PU **ćilmä* ‘eye’ > PS **səjmä* >> NeF *xäm*°, NeT *sæw*° rather than expected NeF ***xămya*, NeT ***sæbya*, only this time Tundra Nenets *tyibya* did not undergo the same change (not NeT ***tyiw*° to regularly correspond to NeF *tyĭm*°).

152. PU **sewi-* ‘eat’ > der. PS **titü* (?**titäw*) ‘bait : приманка’
Ng. *ťütü* (Aikio 2021: 173) | EnF *čiδi* (Aikio 2021: 173) | Sk. *tütu* (SkW 1089)
Aikio 2021: 173.
— This recently proposed etymology by Aikio strengthens the derivation of PS **timä* from PU **sewi-*, since it corroborates the reflex **ti-* of this root.

153. PU **sęksa* > der. PS **tįtåjŋ* ‘Siberian pine : кедр’ (Pinus sibirica)
EnT *tiδii* ‘log without branches : чистое бревно без сучьев’ (EnSl.) | NeT *tideh* (T65) | NeF *tĭtiŋ* (NeSl. *tĭtī̆ŋ*) | Sk. *tįtəŋ* (SkW 1075) | Km. *tedəŋ* (KSz 1180) | Mt. *tidEŋ* ~ *tüdEŋ* (MS 1012 **tįteŋ*)
Sammallahti 1988: 540; SW: 160; UEW: 445–446.

154. PU **sęni* > PS **cęn* ‘sinew : жила’
Ng. *taŋ* (NgSl.) | EnT *tiʔ* (EnSl.) | EnF *tiʔ* (EnWl.) | NeT *teh* (T65) | NeF *tĕŋ* (NeSl. *teŋ*) | Sk. **čęn, čī̆nə-* (SkW: 1122) | Km. *ten* (KmSz.) | Mt. *tęn* (MtSpr: 1004)
Aikio 2012: 231; Sammallahti 1988: 548 (PFU); SW: 32–33.
— The development from PU **s-* to Sk. *č-* is irregular; expected would be ***t-* (cf. Aikio 2012: 231). It is technically impossible to tell for certain if the irregularity occurred in Selkup or in earlier Samoyed, since PS **t* and **c* merge in all other Samoyed languages. However, in all other cases of Selkup *č*, this is reconstructed for Proto-Samoyed, so that the internal logic of the reconstruction demands PS **cęn* rather than PS **tęn* in the absence of good arguments in favour of the latter reconstruction.

155. PU **si*/*äwi* > PS **tiə* ‘year ring (of tree) : годичное кольцо на дереве’
Ng. (C) *tĭ̂* (SW) | EnT *ťii* (EnSl.) | NeT *tyí* | NeF *tyĭ* (NeSl. *čĭ*) | Sk. *tū̆* ‘fibre (in tree) : Faser (im Baum)’ (SkW 952)
Aikio 2012: 244; Sammallahti 1988: 540 (**süxi*).
— The Proto-Uralic reconstruction is made uncertain due to SaaS *sïeve* ‘line, streak’ indicating *säwi* rather than *siwi*, but for the **i* of Samoyed PS **tiə*, PU **siwi* seems to work quite well (see Aikio 2012: 244). A vowel sequence like **iə* resulting from PU **Vwi* is unusual, however, as **Vwi* normally yields a simple vowel **V* (cf. subsection 3.4.3).

156. PU **soski-* > PS der. **tåtu-* (?**tåtâw-*) ‘chew : жевать’
Sk. *tutu̲-* (SkW 1068) | Km. *tudo-* (KSz 1256)
Sammallahti 1988: 540; Zhivlov 2016: 299

— The word family PS **tât³wå-* 'chew' to Ng. *tubuďa* (NgSl.), EnT *tôʔa-* (EnSl.), NeT *təqwa*(*sy*°), Mt. *toʔbə-* (MS 1039) looks superficially similar, and is hesitantly compared by Helimski (MS), but proper reconciliation with the Selkup and Kamas forms from PS **tåtu-* is difficult. The vowels and the internal consonants do not match.

157. PU **suγi-* > PS **tuə̂-* 'row : грести'

Ng. der.instr. *tuobsa* (NgSl.) | Sk. *tū-* (SkW 941) | Km. der. *tuʔ-* (KSz 1254)

Sammallahti 1988: 540; SW: 166; UEW: 449–450.

— The vowel sequence is necessary for the Nganasan instrument noun, which should have been ***tubsa* if derived from a stem **tu-*; Sk. *tū-* also accords better with a vowel sequence than with a simple vowel, cf. PS **ju-* 'melt : Schmelzen' > Sk. *ću-* 'id' (SkW 1427).

158. PU **suδ´a* > PS **tə̂jå* 'finger : палец'

Mt. *taja* (MS 941)

Janhunen 1981: 224; Sammallahti 1988: 540 (**suwδ´å*); UEW: 449.

— The hypothetical connection with SaaN *čuvdi*, etc., requires a PU **ćuwδ´a*, while the Ugric cognates and Samoyed (Mator) require initial **s-* and do not show any **-w-* (see although the latter may in principle have been lost regularly, there are no parallels to confirm or disprove such a loss).

159. PU **suksi* > PS **tutə̂* 'ski : лыжа'

Ng. *tutə* (NgSl.) | EnT *tuδo* (EnSl.)

Sammallahti 1988: 540; UEW: 450.

— Sk. *tāt-* 'gehen, Ski laufen' (SkW 1043) is connected by UEW, but this would rather have to go back to a pre-form **tat-* and cannot regularly be derived from PS **tutə̂*.

160. PU **suŋi* > PS **tə̂ŋ* 'summer : лето'

Ng. *təŋə* (NgSl.) | EnT *too* (EnSl.) | EnF *tɔɔ* (EnWl.) | NeT *təh* (T65) | NeF *tăŋ* (NeSl. *taŋ*) | Sk. *taŋə* (SkW 1211) | Km. *taŋa* (KSz 1146) | Mt. der.adv. *taŋan* (? *taŋən*) (MS 962)

Sammallahti 1988: 540; SW: 148; UEW: 454.

— Nenets does not have a final **-ə̂* in this word, as expected due to apocope; in the other branches, a **-ə̂* was probably analogically introduced to the base word from forms like the adverbial **tə̂ŋə̂n*, adjectival **tə̂ŋə̂jə̂*, or other such derivations. Enets *toʔ* (from **tə̂ŋ*) seems to be attested in older material (see SW).

161. PU **süli* > PS **tij* ‘fathom : сажень’, der. **tijmä* ‘id.’
Ng. (C) *t'ĭmi* (SW) | EnT *ćiee* (EnSl.) | EnF *čibi* (M 305) | NeT *tyíbya* (T65) | NeF *tyimya* (NeSl. *čimja*) | Sk. *tī* (SkW 962) | Mt. *timE* (? *tīmE*) (MS 1020)
Sammallahti 1988: 540; SW: 163; UEW: 444.

162. PU **šelki-* > PS **tej-* (or **teə-* ?) ‘fly (pfv.) : полететь, улететь’, der.ipfv. **tej-r-* ‘fly (ipfv.) : летать, лететь’
Ng. (der.?) *təuď'i*, der.ipfv. *təirśi* (NeSl.) | EnT *ćio-*, der.ipfv. *ćiir-* (EnSl.) EnF *či-*, der.ipfv. *čir-* (EnWl.) | NeT *tyísy°*, der.ipfv. *tyírcy°* (T65) | NeF *tyisy°*, der.ipfv. *tyiłsy°* (NeSl. *čiš*, *čiłьš*) | Sk. *tī-* (SkW 965) | Mt. *ti-* (? *tī-*) (MS 1009)
Aikio 2002: 26; Sammallahti 1988: 550 (PFU); SW: 161–162 (**ti-* ? **tij-*).
— PS **tej-* (in the older reconstruction, **tij-*) is normally given as the Proto-Samoyed stem, but EnT *ćio-* may instead point to **teə-*; this is expected to become NeT *tyí-*, NeF *tyi-* in the same way as **tej-*, and the other languages are not entirely clear either. The vowels in the Nganasan forms are difficult to understand. While *təirśi* ‘fly (ipfv.)’ may reflect PS **tej-r-* (or **teə-jr-*?), the second-syllable vowel of the perfective *təuď'i* ‘fly (pfv.)’ is unaccounted for; cf. Kaheinen (2023: 172–173). Since *təuď'i* is disyllabic, it might match EnT *ćio-*, but it is unclear how exactly.

163. PU **šęra-* > PS **tįrå-* ‘dry out (intr.) : высохнуть’
EnT *tira-* (EnSl.) | EnF *tira-* (EnWl.) | NeT *tira(sy°)* | NeF *tĭłasy°* (NeSl. *tĭ̆łaš*) | Sk. *tįr̲ə̲-* ? (SkW 1299, meaning not entirely certain) | Mt. der.adj. *tiri*, der.caus. *t(i)ribtə-* (MS 1028, 1029)
SW: 160; UEW: 502; Zhivlov 2023: 160.

164. PU **taka-* > PS **takə̂-* adv., postp. ‘the back : задняя часть’
Ng. lat. *tagə*, loc. *takənu*, abl. *takəðə* (NgSl.) | EnT loc. *taxane*, abl. *taxaδ* (EnSl.) | EnF *taxa* (EnWl.) | NeT lat. *tyax°h*, loc. *tyax°na*, abl. *tyaxəd°* (T65) | Sk. *tāk-* (SkW 1155) | Km. *tak-* (KSz 1121)
SW: 154; UEW: 506–506.
— The expected form is PS **takə̂* with the final vowel, so the consonant-final especially seen in Selkup is probably analogical.

165. PU **talwa-* > PS **tåjwå-* 'reach : добраться'
Ng. *tojbusa* (NgSl.) | EnT *taa-* (EnSl.) | EnF *tɔɔ-* (EnWl.) | NeT *tæwə*(*sy*°) (T65) | NeF *tăwăsy*° (NeSl. *tëwaš*) | Km. *tu-* (KSz 1253) | Mt. *tajbə-* (MS 942)
Aikio 2002: 29–30; SW: 146.

166. PU ?**täγδi* or **täwδi* > PS **tärə* 'contents; full : содержимое; полный'
Ng. *ťerə* (NgSl.) | EnT *tero* (EnSl.) | EnF *tɛro* (EnWl.) | NeT *tyer*° (T65) | NeF *tyeł*° (NeSl. *čeł*) | Sk. *tīr* (SkW 1295)
Aikio 2002: 31–34; Sammallahti 1988: 550–551 (FU); SW: 158 (**terə̂* ~ **tirə̂*); UEW: 518; Zhivlov 2023: 167.
— Salminen (2023: 392) argues for a Proto-Samoyed reconstruction with retained **-w-* to account for the Selkup long vowel reflex *ī*, but Helimski has shown that PS **tärə* (SW notation **terə*) regularly yields Proto-Selkup **tīrə* (Helimski 2007: 129). Aikio's reconstruction PU **täwδi* (Aikio 2002: 31–24) would show a different vowel development from PU **käwδi* 'rope' > PS **kürə* and also PU **jäwji* 'lichen' > PS **jüjə*, so that Zhivlov's reconstruction as **täγδi*, with **γ* rather than **w* may be preferable (Zhivlov 2023: 167).

167. PU **täwiw* > PS ?**tä*/*ew* 'lung : лёгкое'
Ng. der. *ťeińďə* (NgSl.) | EnF der. *čiji* (EnWl.) | NeT der. *tyiwak*° (T65) | Sk. *tū̄* (SkW 956) | Km. *tu* (KSz 1252) | Mt. *tüh* (MS 1081)
Sammallahti 1988: 541; SW: 164.
— The vowel is difficult to reconstruct. Ng. *ťeińďə* should probably reflect PS **ä* rather than **e*, since **e* normally becomes Ng. *ï* or *i*, but **e* fits better with the *i* in Enets and Nenets. The other reflexes are more ambiguous, with rounding effects of the **w* having changed the vowel quality.

168. PU **tinä* > PS **tän* 'you (2sg.) : ты'
Ng. der.? *tənə* (NgSl.) | EnT der. *toďi* (EnSl.) | Sk. *tan* (SkW 1100) | Km. *tan* (KSz 1136) | Mt. *tən* (MS 928)
Janhunen 2013: 214; SW: 147.

169. PU **toγi-* > PS **tå-* (~ **tə̂-*) 'bring, give : давать', der. **tə̂tå-* 'id.'
Ng. imp. (T) *tə?* (SW), der. *tətuďa*, imp. *təðu?* (NgSl.) | EnT *ta-*, der. *teδa-* (and *toδa-*) (EnSl.) | EnF *taa-*, der. *tɔza-* (EnWl.) | NeT *tasy*°, fut. *tətə-* (T65) | NeF *tasy*° (NeSl. *tāš*) | Sk. der. *tat-* (SkW 1042) | ? Km. *det-* | Mt. ? *ta-* (MS 930)

Sammallahti 1988: 550; SW: 145; UEW: 529–530.

— Strikingly, the reflex PS **tå-* from PU **toγi-* is different from PS **ńo-* 'pursue' from PU **ńoγi-*. PU **o* normally becomes PS **å* in stems of the type **CoCa* or **CoCCi*, not in the type **CoCi* as here. To understand this, we could potentially consider forms like the aorist stem PS **tå-ŋå-* (NeT *ta°*) or the imperfective participle PS **tå-ntå* (NeT *tana*), which can be derived regularly from the consonant stem as in **toγ-ŋå* and **toγ-ntå* with disappearance of the **γ* and a shift from PU **o* to PS **å*. The reflex **tə̂-* in the derivative **tə̂tå-* has a wide distribution in Samoyed as well, but there is no known parallel for the development from PU **o* to PS **ə̂*. If there was a variant PU **tuγi-*, the **ə̂* in PS **tə̂tå-* could be explained as an effect of a following suffixal **-å* in **tə̂tå-*, as argued by Peyrot (2019b: 193), but then the vowel difference between a hypothetical **tuγ-tå-* on the one hand and **toγ-ntå* (etc.) on the other would still need to be accounted for. The other evidence supports **toγi-* (e.g., SaaN *duohkat* 'sell' rather than ***duhkat*), so that Peyrot's explanation for the vocalism of the derivative **tətå-* cannot be confirmed.

170. PU **toli-* (**tuli*) > PS **toj-* ~ *tu-* 'come : прийти, приехать'

Ng. *tujśa*, cneg. *tuʔ* (NgSl.) | EnT *tô-*, cneg. *tuʔ* (EnSl.) | EnF *to-* (EnWl.) | NeT *tosy°*, cneg. *tuq* (T65) | NeF *tosy°*, cneg. *tŭq* (NeSl. *toš*, *tu"*) | Sk. *tū̄-* (SkW 958) | Km. *šo-* (KSz 1070) | Mt. *toj* (MS 1047)

Sammallahti 1988: 540; SW: 164; UEW: 535.

— The connegative/imperative form of Tundra Enets and both Tundra and Forest Nenets seems to reflect PS **tu-k*, with a different vowel than the one found in the normal stem **toj-* (Nganasan *tuʔ* is ambiguous and can reflect either PS **tuk* or **tok*). To my knowledge, this vowel discrepancy is not explained, and there seems to be no productive way to obtain it within Samoyed. The Finno-Ugric languages reflect a preform **tuli-* rather than **toli-*. If departing from PU **tuli-k* in the connegative/imperative, PS **tuə̂k* should be expected, whence NeT ***túq*, NeF ***tuq* with a long vowel and Ng. ***tuoʔ* with a vowel sequence. Thus, I see no regular way to derive PS **tu-k*.

171. PU **tolwa* > PS **tajwå* 'nail, wedge : гвоздь'

EnT *tea* (EnSl.) | NeT *tyíwa* | NeF *tyewa* (L ± *t'śĭèββaə̂* SW) | Mt. *täjbä* (MS 975)

Aikio 2002: 54 fn. 8; SW: 154.

172. PU **tora-* > PS der. **tåro-* 'wrestle : бороться'
Ng. der. *torasa*, der. *torarsa* (NgSl.) | EnT der. *tarur-* (EnSl.) | EnF *tarur-* (EnWl.) | NeT der. *tarocy°* (T65) | Mt. *toro-* (MS 1064)
Samallahti 1988: 540; SW: 152; UEW: 531.

173. PU **totki* > PS der. **tåtu* 'crucian carp : карась'
Sk. *tutu* (SkW 1066)
UEW: 532; Zhivlov 2016: 299.

174. PU **towi* (**tuwi*) > PS **to* 'lake : озеро'
Ng. der. *turku* | EnT *tu* ? (EnSl.) | EnF *to* (EnWl.) | NeT *to* (T65) | NeF *tŏ* (NeSl. *to*) | Sk. *tō~tū*, *to-* (SkW 936) | Km. *tu* (KSz 1251) | Mt. *toh* (MS 1043)
Gusev 2008; Sammallahti 1988: 540 (**toxɨ*); SW: 164; UEW: 533.
— The Finno-Ugric languages reflect a preform **tuwi* rather than **towi*.

175. PU **tuli* > PS **tuj* 'fire : огонь'
Ng. *tuj* (NgSl.) | EnT *tu* (EnSl.) | EnF *tu* (EnWl.) | NeT *tu* (T65) | NeF *tŭ* (NeSl. *tu*) | Sk. *tū* (SkW 954) | Km. *šü* (KSz 1088) | Mt. *tuj* (MS 1085)
Sammallahti 1988: 540; SW: 166; UEW: 535.

— PU **tuli-* 'come' → see **toli-*.

176. PU **tulka* > ? PS **tuå* 'feather, wing : перо, крыло'
Ng. *ťüo* (NgSl.) | EnT *tua* (EnSl.) | EnF *tɔ* (EnWl.) | NeT *to* (T65) | NeF *tŏ* (NeSl. *to*) | Sk. *tū* (SkW 940) | Mt. *tua*, *tuga* (? *toga*) (MS 1070)
Aikio 2012: 229; Sammallahti 1988: 540; SW: 166; Zhivlov 2023: 164.

177. PU **tumti-* > PS **tumtə̂-* 'know, recognize : узнать'
Ng. *tumtəďa* (NgSl.) | EnT *tudo-* (EnSl.) | EnF *tudu-* (EnWl.) | NeT *tumtə*(*sy°*) (T65) | Km. *tĭmnə-* (KSz 1207) | Mt. *tumdə-* (MS 1091)
Sammallahti 1988: 541; SW: 167; UEW: 536.

178. PU **tunta-* > PS **tə̂ntå* 'teach: учить', der. *tə̂ntö-* (?**tə̂ntåw-*) 'learn : учиться'
EnF *tadda-* (M 2433) | NeT der. *tənara*(*sy°*) 'punish, teach : наказать, проучить', der. *tənyo*(*sy°*) 'calm down : угомониться' (T65) | Sk. **tan*(*t*)*ə̲-* (SkW 1105)
Aikio 2002: 44–45; SW: 147.

— PU **tɯwi* 'lake' → see **towi*.

179. PU **tütki-* > PS **titə-* 'open : открыть'
Sk. *tītə-* (SkW 1096) | Km. *šiʔdə-* (KSz 1060)
Aikio 2006a: 26.

180. PU **uji-* (?**owji-*) > PS **uə̂-* 'swim : плыть'
NeT der. *ŋoxolcy°* (T65) | NeF *ŋusy°* (Salminen 2012: 349), der. *ŋuxułsy°* (L *ŋūxur'ś*) | Sk. *ū-*, der. *ūqəl-* (SkW 4) | Mt. *u-* (? *ū-*) (MS 1108)
Aikio 2020: 82 (PU ?**owji-*; PS **uə̂-*); Salminen (2012: 349–350; PS **uj-*); Sammallahti 1988: 536 (PU **uxi-*, PS **u-*); SW: 29.
— As Aikio (2020) points out, Salminen's (2012) reconstruction as PS **uj-* does not accord with the Selkup vowel, which should have become Sk. ***ūː*; PS **uə̂* regularly yields NeT *ú* and Sk. *ū*, rendering PS **uə̂* the more fitting reconstruction.

181. PU **ukti* > PS **utə̂* 'way, track, footprint : путь, след'
Ng. *ŋutə* (NgSl.) | EnT der. *uδi* (EnSl.) | EnF *uʔ* (EnWl.) | NeT *ŋuq* (T65) | NeF *ŋut°* (NeSl. *ŋut*)
Aikio 2012: 230; 2020: 86; SW: 30 (**uə̂t*); UEW: 546 (**utka*).
— Ng. *ŋutə* and NeF *ŋut°* point to PS **utə̂* while EnF *uʔ* and NeT *ŋuq* point to PS **ut*; PS **utə̂* is the expected form on account of the Proto-Uralic cluster **-kt-*, after which PU **-i* is normally preserved (cf. section 3.3).

182. PU **ulki* > PS **uj* 'pole : шест'
Ng. *ŋüj* (NgSl.) | EnT *ŋu* (EnSl.) | EnF der. *ŋuzu* (EnWl.) | NeT *ŋu* (T65) | NeF *ŋŭ* (NeSl. *ŋu*) | Sk. *ūˑ* (SkW 8)
Aikio 2002: 26; Sammallahti 1988: 536; SW: 29–30; UEW: 543.

183. PU **üli* > PS **i-* 'on; space above, top : на; вершина'
Ng. *ńi* (NgSl.) | EnT *ńiʔ* (EnSl.) | EnF *n'i* (EnWl.) | NeT e.g., lat. *nyih*, abl. *nyid°* | NeF e.g., loc. *nyĭnya*, abl. *nyĭt°* (NeSl. *ninja*, *nĭt*) | Sk. *i-* (SkW 10) | Km. *ni* (KSz 0690) | ? Mt. ?der. "нем-" ? *ńim-*, *nim-* 'raise : hochnehmen, heben' (MS 734)
Sammallahti 1988: 536; SW: 26.
— With the initial nasal in all Samoyed languages except Selkup seems to be taken over from the genitive **-n* (SW 26). The reduction to PS **i* rather than

more expected **ij* or **iə* is probably because this is a postposition, a function word, which are susceptible to phonological reductions.

184. PU ?**ülkä* > ? PS **ij* 'son : сын'
? NeT der. *nyícy*° 'calve : отелиться' | Sk. *ī* (SkW 13)
Aikio: 2020: 95.
— The phonological development from PU **ülkä* (PSaa. **ęlkē* 'son', Fi. *ylkä* 'bridegroom') to PS **ij* would require a development as though it were **ülki*, like in **ulki* 'pole' to PS **uj*. This could be paralleled in **pälkä* (q.v.) but remains poorly understood (see Aikio 2020: 95).

185. PU **üwä* > PS **(n/j)iə* 'belt : пояс'
Ng. der. *ŋiəďə* | EnT *niojo* (EnSl.) | EnF *n'eju* (EnWl.) | NeT *nyí* (T65) | NeF *nyǐ* (NeSl. *ni*) | Sk. *ćī̆* (SkW 1433) | Km. *t'i* (KSz 1361) | Mt. *Ni* (MS 750)
Aikio 2012: 230; Sammallahti 1988: 536 (?**üxji*); SW: 102; UEW: 90 (**jäje*).
— Ng. *ŋiəďə* should derive from **iə* rather than from **niə* or **jiə*. The initial **n-* (in Enets and Mator) and **j-* (in Selkup and Kamas) are both prothetic elements. In Nenets the initial nasal would be automatic in any case.

186. PU **wajŋi* > PS **wajŋ* 'breath; soul : дыхание, душа'
Ng. der. *baťüʔ*, der. *baťəďa* 'breathe : дышать' (NgSl.) | EnT ? *beu*, der. *beduʔo* (EnSl.) | EnF *beduʔ* (EnWl.) | NeT der. *yínt°q* (T65) | NeF der. *wyent°* (NeSl. *wjent*) | Sk. *kuâj(i)* (SkW 2078) | Km. *majə* (KSz 541) | Mt. der. *möjüh* (? *mojuh*) (MS 684)
Sammallahti 1988: 541; SW: 173; UEW: 552–553.

187. PU **wala* > PS **wålä* 'song : песня'
Ng. *bəlï* (NgSl.) | EnT *bare* (EnSl.) | EnF *bare* (EnWl.) | ? NeT der.? *wal°q*
Aikio 2006: 26–27; UEW: 812.

188. ? PU **wali-* > PS **wåâ-* 'speak, promise : говорить, обещать', der.n. *wåâtå* 'word : слово'
Ng. *buoďa*, der.n. *buodu* (NgSl.) | EnT *baa-* (EnSl.) | EnF der.n. *baza* (EnWl.) | NeT *wasy°*, der.n. *wada* (T65) | NeF der.n. *wata* (NeSl. *wāta*) | Mt. der.n. *bāda* (MS 89)
SW: 171.
— A hypothetical Proto-Uralic verb corresponding to the noun PU **wala* > PS **wålä* 'song' with the same stem-final vowel alternation between **-i* and **-a* in verbs and nouns as discussed briefly by Aikio (2002: 54 fn. 10). The

reconstruction to PU **wali-* is based on this likely relationship to the noun **wala* 'song', which is found in other Uralic languages as well. The phonological development would be entirely parallel to that of PU **kali-* > PS **kåâ-* 'die'. Not the same word as PU **wali-* 'carve', which is unknown from Samoyed.

189. PU ±**wančaw* > PS **vånco* (**våncåw*) 'root : корень'
Ng. *bəntu* (NgSl.) | EnT *badu* (EnSl.) | EnF *badu* (EnWl.) | NeT *wano* (T65) | NeF *wanu* (NeSl. *wanu*) | Sk. *končə* (SkW 2023) | Km. *mona* (KSz 0614) | Mt. *mondo*(*h*) (MS 689)
Sammallahti 1988: 541 (**vånčå*); SW: 171 (**vånc3* ? **våncâj*); UEW: 548–549.

190. PU **wara* > PS **vårå* 'mountain, ridge : гора, скала'
Ng. *bəru* (NgSl.) | Sk. *kūrə* (SkW 2288) | ? Km. *bor* (KSz 0146)
Aikio 2006a: 27–28; Sammallahti 1988: 551 (PFU).

191. PU **watka-* > ? PS **våt-* 'debark (a tree) : обдирать'
NeT *wacy*°, der. *watə*(*sy*°) (T65)
Aikio 2015b: 56; UEW: 561.

— The combination of loss of the final vowel and the development of the first-syllable vowel is aberrant: expected would be PS ***vatâ-* (NeT ***yadə*(*sy*°)), secondarily **vat-* (NeT ***yacy*°) or PS ***våtå-* (NeT ***wada*(*sy*°)), based on the behaviour of other, structurally similar words. A reconstruction PU **wotka-* would be better for Nenets, as this could yield PS **våt*(*â*) and thus NeT *wacy*° similar to PU **woča* to PS **våc*(*â*), NeT *waq* 'enclosure', but this does not work for the other Uralic cognates like MdE *vatka-* 'debark'. Other possible reconstructions based on the Nenets form are PS **våk-*/**vås-*/**våc-*, which, combined with the unexpected vocalism, makes the Uralic etymology less than compelling, despite the good semantic match and general phonological similarity.

192. PU **weɣi-* > ? PS **i-* 'take : брать'
Sk. *ī-* (SkW 14) | Km. *ī-* (KSz 0229)
Aikio 2013: 171.

— The initial **w-* was apparently unstable, and disappeared in a few cases, cf. PU **woli-* 'be' > PS **åâ-*, **åj-* and PU **woδ´i* 'new' > PS **oj-*. This is similar to the disappearance of the other glide **j-* in PU **jɛ̨ki-* 'drink' to PS **ɛ̨-* and in PU **joŋsi* 'bow' everywhere except in Nganasan.

193. PU **weti* > PS **wet* 'water : вода'
Ng. *bi?* (NgSl.) | EnT *bu?* (EnSl.) | EnF *bu?* (EnWl.) | NeT *yiq* (T65) | NeF *wyǐq* (NeSl. *wi"*) | Sk. *üt, ūtə-* (SkW 107) | Km. *bü* (KSz 0172) | Mt. *bü* (MS 143)
Sammallahti 1988: 541; SW: 176; UEW: 570.

194. PU ±?**wijit(t)i* > PS **wüät* 'ten : десять (10)'
Ng. *büi?* (NgSl.) | EnT *biu?* (EnSl.) | EnF *biu?* ~ *bi?* (EnWl.) | NeT *yúq* (T65) | NeF *juq* (NeSl. *dju"*) | Sk. *kōt* (SkW 1965) | Km. *ńe?n* (KSz 0187) | Mt. *čūt~čū?n* (MS 245)
Aikio 2012: 241–242; Sammallahti 1988: 541; SW: 177; UEW: 577.
— In NeF ***wyuq* would be expected rather than *juq*. The exact Proto-Uralic reconstruction is unclear, cf. Aikio (2012: 241–242).

195. PU **wiγi-* > ? PS **ü-* 'drag : тащить', der. *ükəl-* 'id.'
NeT der. *nyuxəlye(sy°)* (T65) | Sk. *ū-*, der. *ūqəl-* (SkW 9)
Helimski *apud* Aikio 2002: 49; SW: 30–31.

196. PU **woča* > PS **wåc(ə̂)* 'enclosure : загон'
Ng. (T) *bə?* (SW) | EnT *ba?* (EnSl.) | NeT *waq* (T65) | NeF *wat°* (NeSl. *wat*) | Sk. der. *kuə̂čʊ* 'fishing weir; tributary : Damm, Wehr; Zufluß' (SkW 1903)
Aikio 2015b: 62; Sammallahti 1988: 541; SW: 171; UEW: 577–578.
— Cf. also Sk. *quə̂ččə* 'town : Stadt' (SkW 1912), which is given as the Selkup reflex by UEW. However, the initial *q-* is at least unexpected (normally PS **w-* to Sk. *k-* with possible assimilation to a following *-q-*, of which this word has none), and so is the geminate *-čč-*. The second-syllable vowel is clearly lost in Nganasan, Tundra Enets and Tundra Nenets, perhaps due to analogical removal from expected PS **wåcə̂*. Forest Nenets *wat°* could reflect this expected pre-form, but it may also have (re-)obtained a second-syllable schwa secondarily as the result of an analogical change in the opposite direction.

197. PU **woδ´i* (or **wuδ´i*) > PS **oj-* 'new : новый'
SkTaz der.adv. *oćiŋ* 'again, anew' (Zhivlov 2023: 166)
Sammallahti 1988: 551; UEW: 587; Zhivlov 2023: 166.
— The Finno-Ugric language reflect a preform **wuδ´i* rather than **woδ´i*. The initial **w-* was apparently unstable, and disappeared in a few cases, cf. PU **woli-* 'be' > PS **åə̂-*, **åj-* and PU **weγi-* 'take' > PS **i-*. This is similar to the

disappearance of the other glide **j-* in PU **jęki-* 'drink' to PS **ę-* and in PU **joŋsi* 'bow' everywhere except in Nganasan.

198. PU **woli-* 'be' > PS **åə̂-* ~ **åj-* (?**aj*) 'be : быть'

Ng. (**åə̂-*) imp.2sg. *ŋuo?* | EnT *a-* (EnSl.) | EnF (**åj-*) *ɛ-*, (**åə̂-*) *ŋa-* (EnWl.) | NeT (**åj-*) *ŋæsy°*, (**åə̂-*) *ŋa-* (T65) | NeF *ŋäsy°* (NeSl. *ŋëš*) | Sk. (?**aj-*) **ē-* (SkW) | ? Km. (?**aj*) *i-* (KmSz.) | Mt. (?**aj-*) *äj-* (MS 59)

Janhunen 2020: 134–135; SW: 16–17; UEW: 580.

— The initial **w-* was lost irregularly. This verb shows an alternation between reflexes of PS **åə̂* and **åj* in Nenets and in Forest Enets (the two sequences regularly merge in Tundra Enets); cf. Janhunen (2020: 134–135). I think that PS **åə̂* represents the original vowel stem of **woli-*, while PS **åj-* reflects the consonant stem **wol-*, used before suffixes longer than a single consonant. The southern Samoyed languages seem to reflect a vowel **a* rather than **å*, an issue that is found in various words and still requires a proper explanation. Kamas *i-* is put here as well in SW, but this may rather reflect PS **(j)i-* 'be', as in Ng. *iśa* (suppletive with imp.2sg. *ŋuo?*), NeT *yib°q* 'if it is' (see Janhunen 2020).

199. PU **wolka* > PS **wajk* 'shoulder : плечо', der. *wajkkə̂* 'neck : шея'

Ng. der. *bakəðəə* (NgSl.) | EnT der. *beko* (EnSl.) | EnF der. *bɛko* (EnWl.) | NeT der. *yík°* (T65) | NeF der. *wyeqk°* (NeSl. *wje"k*) | Sk. *quə̂q* 'shoulder' (SkW 2106) | Km. der. *bajkə* (0091) | Mt. der. *bäjkä* (?) K ~ *böjkö* (? *büjkü*) M (MS 107)

Sammallahti 1988: 551 (PFU); SW 173; UEW: 581.

— The loss of PU **-a* may have occurred via a reduction to **-ə̂*, cf., e.g., PU **woča* 'enclosure' > PS **wåc*. The development of PU **-lk-* to PS **-jk-* seems to be unique to this word, cf. PU **ulki*, PU **ülkä*, PU **šelki-*, PU **tolka*, where the cluster develops like **-j-*. A hypothetical PU **wolkka* cannot account for this, since such a reconstruction is incompatible with cognates such as SaaN *oalgi* (rather than ***oalki*) and Fi. *olka* (rather than ***olkka*).

200. PU **woŋki* > PS **wåŋkə̂* 'hole, cave : яма'

Ng. *baŋkə* 'dugout : землянка' (NgSl.) | EnT *bago* (EnSl.) | EnF *bago* (EnWl.) | NeT *waŋk°* (T65) | NeF *wāŋk°* (NeSl. *waŋk*)

Sammallahti 1988: 551; SW: 171–172; UEW: 583; Zhivlov 2023: 163.

— PU **wuδ´i* 'new' → see **woδ´i*.

201. PU *(*w*)*uwa* > PS **wuə̑* ‘current : течение’
Ng. *buo* (NgSl.)
Aikio 2012: 244.
— The Saami cognates such as SaaN *avvit* ‘flow’ do not show the initial **w*- reflected in the Nganasan *b*-, meaning that it was either irregularly lost in Saami or irregularly added in Samoyed/Nganasan.

# Bibliography

### Bibliographical abbreviations

CEToM = *A Comprehensive Edition of Tocharian Manuscripts* (*CEToM*). Created and maintained by Melanie Malzahn, Martin Braun, Hannes A. Fellner, and Bernhard Koller. https://cetom.univie.ac.at/

DThTA = Carling & Pinault 2023.

EnSl. = Евгений Хелимский. Энецкий словарь. [Eugene Helimski. Enets dictionary.] Manuscript.

EnWl. = Olesya Khanina & Andrey Shluinsky. Forest Enets wordlist. Manuscript.

INEL Selkup Corpus = Maria Brykina, Svetlana Orlova & Beáta Wagner-Nagy. 2021. *INEL Selkup Corpus. Version 2.0*. Publication date 2021-12-31. Archived at Universität Hamburg. https://hdl.handle.net/11022/0000-0007-F4D9-1. In: The INEL corpora of indigenous Northern Eurasian languages. https://hdl.handle.net/11022/0000-0007-F45A-1.

KSz = Janurik 2021.

LIV = Rix & Kümmel 2001.

MS = Helimski 1997.

NeSl. = Barmič & Vello 2002.

NgSl. = Kosterkina, Momde & Ždanova 2001.

SkW = Alatalo 2004.

SW = Janhunen 1977.

SWTF I = Waldschmidt & Bechert. 1973.

T65 = Tereščenko 1965.

UEW = Rédei 1988–1991.

### References

Abaffy, Erzsébet E. 2001. On the origin of the Hungarian verbal number and person markers: some controversial issues. *Acta Linguistica Hungarica* 48, 321–335.

Abondolo, Daniel. 1998. Introduction. In Daniel Abondolo (ed.), *The Uralic Languages*. London-New York: Routledge, 1–42.

Adams, Douglas Q. 1983. Studies in Tocharian Vocabulary II: Words pertaining to the Lower Limbs in Tocharian B. *Journal of the American Oriental Society* 103, 611–614.

2000. Some observations of peoples, places, and languages in the Tarim Basin in the first millennium AD. *Tocharian and Indo-European Studies* 9, 1–28.

2013. *A Dictionary of Tocharian B. Revised and Greatly Enlarged.* Amsterdam-New York: Rodopi.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2002. New and Old Samoyed Etymologies. *Finnisch-Ugrische Forschungen* 57, 9–57.

2006a. New and Old Samoyed Etymologies (Part 2). *Finnisch-Ugrische Forschungen* 59, 5–34.

2006b. On Germanic-Saami contacts and Saami prehistory. *Suomalais-Ugrilaisen Seuran Aikakauskirja* 91, 9–55.

2012. On Finnic long vowels, Samoyed vowel sequences and Proto-Uralic **x*. In Tiina Hyytiäinen et al. (eds.), *Per Urales ad Orientem: Iter polyphonicum multilingue. Festschrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012*. Helsinki: Suomalais-Ugrilainen Seura, 227–250.

2013. Studies in Uralic Etymology I: Saami Etymologies. *Linguistica Uralica* 3/2013, 161–174.

2014a. Studies in Uralic Etymology II: Finnic Etymologies. *Linguistica Uralica* 1/2014, 1–19.

2014b. Studies in Uralic Etymology III: Mari Etymologies. *Linguistica Uralica* 2/2014, 81–93.

2014c. On the reconstruction of Proto-Mari Vocalism. *Journal of Language Relationship* 11, 125–157

2014d. The Uralic-Yukaghir lexical correspondences: genetic inheritance, language contact or chance resemblance? *Finnisch-Ugrische Forschungen* 62, 7–76.

2015. 'Finnic 'secondary *e*-stems' and Proto-Uralic vocalism. *Suomalais-Ugrilaisen Seuran Aikakauskirja* 95, 25–66.

2020. Uralic Etymological Dictionary (draft version of entries A-Ć). https://www.academia.edu/41659514/URALIC_ETYMOLOGICAL_DICTIONARY_draft_version_of_entries_A_Ć_

2021. Studies in Uralic Etymology V: Permic Etymologies. *Linguistica Uralica* 3/2021, 161–179.

2022a. Proto-Uralic. In Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 3–27.

2022b. **-kkAs*: a Proto-Uralic inclinative suffix. In Kaisla Kaheinen et al. (eds.), *Hämeenmaalta Jamalille. Kirja Tapani Salmiselle 07.04.2022*. Helsinki: University of Helsinki, 15–33.

Aikio, Ante & Jussi Ylikoski. 2016. The origin of the Finnic *l*-cases. *Fenno-Ugrica Suecana Nova Series* 15, 59–158.

Akao, Miyu. 2020. Internal and External Factors behind the Development of the Tocharian Secondary Cases. In David M. Goldstein, Stephanie W. Jamison and Brent Vine (eds.), *Proceedings of the 31st Annual UCLA Indo-European Conference. Los Angeles, November 8th and 9th, 2019.* Tübingen: Buske, 1–16.

Alatalo, Jarmo. 2004. *Sölkupisches Wörterbuch aus Aufzeichnungen von Kai Donner, U.T. Sirelius und Jarmo Alatalo*. Helsinki: Suomalais-Ugrilainen Seura.

Alhoniemi, Alho. 1996. Zur verwendung der Objektkonjugation im Mordwin-ischen. In Lars-Gunnar Larsson (ed.), *Lapponica et Uralica. 100 Jahre finnisch-ugrischer Unterricht an der Universität Uppsala: Vorträge am Jubiläumssymposium 20.-23. April 1994*. Uppsala: Acta Universitatis Upsaliensis, 145–152.

Allentoft, M.E. et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522, 167–172.

Anderson, Gregory D.S. 2006. Towards a Typology of the Siberian Linguistic Area. In Yaron Matras et al. (eds.), *Linguistic Areas: convergence in Historical and Typological Perspective*. Basingstoke-New York: Palgrave Macmillan, 266–300.

Anikin & Helimski 2007. = А. Е. Аникин & Е. А. Хелимский 2007. *Самодийско-тунгусо-маньчжурские лексические связи*. [Samoyed-Tungus-Manchu lexical connections.] Москва: Языки славянской культуры.

Anthony, David. 2007. *The horse, the wheel and language*. Princeton: Princeton University Press.

2013. Two IE phylogenies, three PIE migrations, and four kinds of steppe pastoralism. *The Journal of Language Relationship* 9, 1–21.

Anthony, David & Don Ringe. 2015. The Indo-European Homeland from Linguistic and Archaeological Perspectives. *Annual Review of Linguistics*, 1, 199–219. https://doi.org/10.1146/annurev-linguist-030514-124812

Authier, Gilles. 2012. New strategies for relative clauses in Azeri and Apsheron Tat. In Volker Gast and Holger Diessel (eds.), *Clause linkage in cross-linguistic perspective: Data-driven approaches to cross-clausal syntax*. Berlin-Boston: De Gruyter, 225–252.

Bakró-Nagy, Marianne, Katalin Sipőcz & Elena Skribnik. 2022. North Mansi. In Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 537–564.

Barmič & Vello 2002. = М. Я. Бармич & И. А. Вэлло. 2002. *Словарь ненецко-русский и русско-ненецкий (лесной диалект)*. [Nenets-Russian and Russian-Nenets dictionary (forest dialect).] Санкт-Петербург: Просвещение.

Bateman, Nicoleta. 2007. *A Crosslinguistic Investigation of Palatalization*. PhD-dissertation, University of California, San Diego.

2011. On the Typology of Palatalization. *Language and Linguistics Compass* 5/8, 588–602.

Bednarczuk, Leszek. 2015. Non-Indo-European features of the Tocharian dialects. In Elżbieta Mańczak-Wohlfeld and Barbara Podolak (eds.), *Words and Dictionaries*. Kraków: Jagiellonian University Press, 55–67.

Beekes, Robert. 2010. *Etymological Dictionary of Greek*. Leiden-Boston: Brill.

2011. *Comparative Indo-European Linguistics: An Introduction*. Second edition. Amsterdam-Philadelphia: John Benjamins.

Belyaev, Oleg. 2010. Evolution of Case in Ossetic. *Iran and the Caucasus* 14, 287–322.

Bereczki, Gábor. 1988. Geschichte der wolgafinnischen Sprachen. In Denis Sinor (ed.), *The Uralic Languages: Description, History and Foreign Influences*. Leiden: Brill, 314–351.

Bergsland, Knut. 1973. The Mordvinian Conjugation as a Problem of Recon-struction. In Erkki Itkonen and Alho Alhoniemi (eds.), *Commentationes Fenno-Ugricae in honorem Erkki Itkonen sexagenarii die XXVI mensis aprilis anno MCMLXXIII*. Helsinki: Suomalais-Ugrilainen Seura, 44–52.

Bernard, Chams. 2023. *Like Dust on the Silk Road. An investigation of Iranian loanwords in Tocharian, with a focus on the oldest ones*. PhD-dissertation, Leiden University.

Bernhard, Franz. 1958. *Die Nominalkomposition im Tocharischen*. PhD-dissertation, Göttingen.

Bernhardt, Mariann. 2020. The variation of conjugation types in Erzya perception verbs. *Finnisch-Ugrische Forschungen* 27, 21–48.

Bjørn, Rasmus G. 2022. Indo-European loanwords and exchange in Bronze Age Central and East Asia: Six new perspectives on prehistoric exchange in the Eastern Steppe Zone. *Evolutionary Human Sciences* 4, e23. https://doi.org/10.1017/ehs.2022.16

Broomhead, J.W. 1962. *A textual edition of the British Hoernle, Stein and Weber Kuchean manuscripts. With transliteration, translation, grammatical commentary and vocabulary*. PhD-dissertation, Cambridge: Trinity College.

Brosman, Paul W., Jr. 2010. The *nt*-participles and the verbal adjectives in *-to*. *Indogermanische Forschungen* 115, 22–34.

Budzisch, Josefina. 2021. *Definitheit im Selkupischen*. Szeged: University of Szeged.

Burkova, Svetlana. 2022. Nenets. In Marianne Bakró-Nagy, Johanna Laakso and Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 674–708.

Burlak, Svetlana. 2000. *Историческая фонетика тохарских языков* [The historical phonetics of the Tocharian languages.] Москва: Российская Академия Наук, Институт Востоковедения.

Burlak, Svetlana & Ilya B. Itkin. 2003. A sound change that never happened: The fate of Proto-Tocharian**o* (В *o*) in Tocharian A. *Tocharian and Indo-European Studies* 10, 17–35.

Carling, Gerd. 1999. The Tocharian inflected adverbials and adpositions in relation to the case system. *Tocharian and Indo-European Studies* 8, 95–110.

2000. *Die Funktionen der lokalen Kasus im Tocharischen*. Berlin: De Gruyter.

2005. Proto-Tocharian, Common Tocharian, and Tocharian - on the value of linguistic connections in a reconstructed language. In Karlene Jones-Bley et al. (eds.), *Proceedings of the sixteenth annual UCLA Indo-European conference*. Washington, C.D.: Institute for the Study of Man, 47–71.

2006. Notes on the Syntactic Functions of the Tocharian Clitic Pronouns. In Gerd Carling (ed.), *$^{GIŠ.ḪUR}$gul-za-at-ta-ra: Festschrift for Folke Josephson*. Göteberg: Meijerbergs institut för svensk etymologisk forskning, 32–45.

2012. Development of form and function in a case system with layers: Tocharian and Romani compared. *Tocharian and Indo-European Studies* 13, 57–76.

Carling, Gerd & Georges-Jean Pinault. 2023. *Dictionary and thesaurus of Tocharian A*. Wiesbaden: Harrassowitz.

Čeremisina 1991 = М. И. Черемисина. 1991. *Селькупский глагол: формы и их синтаксические функции*. [The Selkup verb: forms and their syntactic functions.] Novosibirsk: Novosibirsk State University.

Childebayeva, Ainash et al. 2024. Bronze age Northern Eurasian genetics in the context of development of metallurgy and Siberian ancestry. *Communications Biology* 7.

Collinder, Björn. 1940. *Jukagirisch und Uralisch*. Uppsala: Almqvist & Wiksell.

Cooreman, Ann M. 1987. *Transitivity and discourse continuity in Chamorro narratives*. Berlin-New York: De Gruyter.

Coppock, Elizabeth & Stephen Wechsler. 2010. Less-travelled paths from pronoun to agreement: the case of the Uralic objective conjugations. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG10 Conference*. Standord: CSLI Publications, 186–206.

Couvreur, Walter. 1947. *Hoofdzaken van de Tochaarse klank- en vormleer*. Leuven: Beheer van Philologische Studiën.

Cowgill, Warren. 1967. Ablaut, Accent and Umlaut in the Tocharian Subjunctive. In Walter W. Arndt et al. (eds.), *Studies in Historical Linguistics in Honor of George Sherman Lane*. Chapel Hill: University of North Carolina Press, 171–181.

Csepregi, Márta. 2023. Khanty. In Daniel Abondolo & Riitta-Liisa Valijärvi (eds.), *The Uralic Languages*. Second Edition. London-New York: Routledge, 703–752

Csúcs, Sándor. 2005. *Die Rekonstruktion der permischen Grundsprache*. Budapest: Akadémiai Kiadó.

Däbritz, Chris Lasse. 2017. Zur Markierung von Numerus an Nomina in den samojedischen Sprachen – Synchronie und mögliche Implikationen für die Rekonstruktion des Protosamojedischen und des Protouralischen. *Finnisch-Ugrische Mitteilungen* 41, 63–101.

Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge: Cambridge University Press.

Damgaard, P. de B. et al. 2018. 137 ancient human genomes from across the Eurasian steppes. *Nature* 557, 269–374.

Del Tomba, Alessandro. 2018. On the pronominal feminine plural in Tocharian, *Indogermanische Forschungen* 123, 339-356.

2020. *The Tocharian Gender System: A Diachronic Study*. PhD-dissertation, Leiden University.

Derksen, Rick. 2008. *Etymological Dictionary of the Slavic Inherited Lexicon*. Leiden-Boston: Brill.

2015. *Etymological Dictionary of the Baltic Inherited Lexicon*. Leiden-Boston: Brill.

De Vaan, Michiel. 2008. *Etymological Dictionary of Latin and the other Italic Languages*. Leiden-Boston: Brill.

Dragoni, Federico. 2023. *Watañi lāntaṃ: Khotanese and Tumshuqese Loanwords in Tocharian*. Wiesbaden: Reichert.

É. Kiss, Katalin. 2010/11. A magyar tárgyas és alanyi igeragozás kialakulásának szintaktikai hátteréről. [On the syntactic background of the evolution of the Hungarian

objective and subjective conjugations.] *Nyelvtudományi Közlemények* 107, 131–146.

2013. From Proto-Hungarian SOV to Old Hungarian Top Foc V X. *Diachronica* 30, 202–231.

Erdal, Marcel. 2004. *A Grammar of Old Turkic*. Leiden: Brill.

Eyþórsson, Þórhallur. 1993. Proto-Indo-European origins of Tocharian accent: Class V subjunctives in Tocharian B. *Tocharian and Indo-European Studies* 6, 43–94.

Fellner, Hannes. 2005. *Die Vertretung der Labiovelare im Tocharischen*. MA-thesis, Universität Wien.

2013. *Studies in Tocharian adjective formation.* PhD-dissertation, Harvard University.

2014. Tocharian special agents: the *nt*-participles. *Tocharian and Indo-European Studies* 15, 53–67.

Fellner, Hannes. & Laura Grestenberger. 2018. Die Reflexe der **-nt-* und **-mh,no-* Partizipien im Hethitischen und Tocharischen. In Elisabeth Rieken (ed.), *100 Jahre Entzifferung des Hethitischen. Morphosyntaktische Kategorien in Sprachgeschichte und Forschung. Akten der Arbeitstagung der Indogermanische Gesellschaft vom 21. bis 23. September 2015 in Marburg.* Wiesbaden: Reichert, 63–82.

Filchenko, Andrey Yury. 2007. *A Grammar of Eastern Khanty*. PhD-dissertation, Houston, Texas.

Friis, Louise S. 2021. Tocharian B agent nouns in *-ntsa* and their origin. *Indo-European Linguistics* 9, 1–25.

2024. *Studies in Tocharian verbal morphology relevant to the cladistic position of Tocharian in Indo-European*. PhD-dissertation, Leiden University.

Gamkrelidze & Ivanov 1972 = Т. В. Гамкрелидзе, В. В. Иванов 1972. Лингвистическая типология и реконструкция системы индоевропейских смычных. [Linguistic typology and reconstruction of the system of Indo-European stops.] In C.B. Bernstein et al. (eds.), *Конференция по сравнительно-исторической грамматике индоевропейских языков (12-14 декабря), предварительные материалы*. Moskva: Nauka, 15–18.

Georg, Stefan. 2023. Connections between Uralic and other language families. In Daniel Abondolo and Riitta-Liisa Valijärvi (eds.), *The Uralic Languages*. Second edition. London-New York: Routledge, 176–209.

Georg, Stefan & Uwe Seefloth. 2019. Uralo-Eskimo? In E. V. Golovko et al. (eds.), *Лингвистика и другие важные вещи. Памяти Александра Павловича Володина (1935–2017)*. Sankt-Peterburg: RAN, 436–451.

Gippert, Jost. 1987. Zu den sekundären Kasusaffixen des Tocharischen. *Tocharian and Indo-European Studies* 1, 22–39.

Givón, Talmy. 1994. The pragmatics of de-transitive voice: Functional and typological aspects of inversion. In Talmy Givón (ed.), *Voice and Inversion*. Amsterdam-Philadelphia: John Benjamins, 3–44.

Grestenberger, Laura. 2020. The diachrony of participles in the (pre)history of Greek and Hittite. Losing and gaining functional structure. *Diachronica* 37(2), 215–263.

Griffith, Aaron. 2011. Old Irish pronouns: agreement affixes vs. clitic arguments. In Andrew Carnie (ed.), *Formal approaches to Celtic linguistics*. Newcastle upon Tyne: Cambridge Scholars Publishing, 65–93.

Grünthal, Riho et al. 2022. Drastic demographic events triggered the Uralic spread. *Diachronica* 39, 490–524.

Goldstein, David. 2016. Variation versus Change: Clausal Clitics between Homer and Herodotus. *Indo-European Linguistics* 4, 53–97.

Gusev 2008 = Валентин Ю. Гусев. 2008. Ауслаутное *-w* в самодийских языках: нганасанско-маторская изоглосса. [Auslaut *-w* in the Samoyed languages: a Nganasan-Mator isogloss.] *Материалы 2-й между-народной конференции по самодистике*. Санкт-Петербург: Нестор-История, 114–128.

Hackstein, Olav. 1995. *Untersuchungen zu den sigmatischen Präsensstamm-bildungen des Tocharischen*. Göttingen: Vandenhoeck & Ruprecht.

2003. Zur Entwicklung von Modalität in Verbaladjektiven. In Eva Tichy, Dagmar S. Wodtko & Britta Irslinger (eds.), *Indogermanisches Nomen. Derivation, Flexion und Ablaut. Akten der Arbeitstagung der Indo-germanischen Gesellschaft, Freiburg, 19. bis 22. September 2001*. Bremen: Hempen, 51–66.

2017. The phonology of Tocharian. In Jared Klein et al. (eds.), *Handbook of comparative and historical Indo-European linguistics*. Volume 2. Berlin: De Gruyter, 1304–1335.

Häkkinen, Jaakko. 2012. Early contacts between Uralic and Yukaghir. In Tiina Hyytiäinen et al. (eds.), *Per Urales ad Orientem. Iter polyphonicum multilingue. Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012*. Helsinki: Suomalais-Ugrilainen Seura, 91–101.

2023. On locating Proto-Uralic. *Finnisch-Ugrische Forschungen* 68, 43–100.

Häkkinen, Kaisa. 1984. Wäre es schon an der Zeit, den Stammbaum zu fällen? Theorien über die gegenseitigen Verwandtschaftsbeziehungen der finnish-ugrischen Sprachen. *Ural-Altaische Jahrbücher*, Neue Folge, Band 4, 1–24.

Hale, Mark R. 1987. *Studies in the comparative syntax of the oldest Indo-Iranian languages*. PhD-dissertation, Harvard University.

Hall, T. A. 2000. Typological generalizations concerning secondary palatalization. *Lingua* 110, 1–25.

Hall, T. A. & Silke Hamann. 2006. Towards a typology of stop assibilation. *Linguistics* 44, 1196–1236.

Hall, T. A. & Silke Hamann & Marzena Zygis. 2006. The phonetic motivation for phonological stop assibilation. *Journal of the International Phonetic Association* 36, 59–81.

Hamari, Arja & Rigina Ajanki. 2022. Mordvin (Erzya and Moksha). In Marianne Bakró-Nagy, Johanna Laakso and Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 392–431.

Haspelmath, Martin. 1994. Passive Participles across Languages. In Paul J. Hopper and Barbara A. Fox (eds.), *Voice: Form and Function*. Amsterdam-Philadelphia: John Benjamins, 151–177.

Havas, Ferenc. 2004 Objective conjugation and medialisation. *Acta Linguistica Hungarica* 51, 95–141.

Helimski, Eugene. 1983. *The Language of the First Selkup Books*. Szeged: University of Szeged.

1986. = Евгений Хелимский. 1986. Etymologica 1–48: Материалы по этимологии маторско-тайгийско-карагасского языка. [Etymologica 1–48: Materials on etymologies of the Mator-Taigi-Karagass language.] *Nyelvtudományi Közlemények* 88, 119–143.

1992/93. = Евгений Хелимский. 1992/93. Etymologica 49–79: 'Материалы по этимологии маторско-тайгийско-карагасского языка.' [Etymologica 49–79: Materials on etymologies of the Mator-Taigi-Karagass language.] *Nyelvtudományi Közlemények* 93, 101–123.

1993. = Евгений Хелимский. 1993. Прасамодийские **ə̑* и **ə̈*: прауральские источнки и нганасанские рефлексы. [Proto-Samoyed **ə̑* and **ə̈*: Proto-Uralic sources and Nganasan reflexes.] In Marianne Bakró-Nagy & Enikő Szíj, *Hajdú Péter 70 Éves*. Budapest: MTA Nyelvtudományi Intézet, 125–133.

1997. *Die Matorische Sprache*. Szeged: University of Szeged.

1998. Nganasan. In Daniel Abondolo (ed.), *The Uralic Languages*. London-New York: Routledge, 480–515.
2005. The 13th Proto-Samoyedic vowel. In Wagner-Nagy Beáta (ed.), *Mikola-konferencia 2004*. Szeged: University of Szeged, 27–39.
2007. = Евгений Хелимский. 2007. Продление гласных перед шва в селькупском языке как фонетический закон. [Pre-schwa vowel lengthening in Selkup as a sound law.] *Linguistica Uralica* 43, 124–133.

Hilmarsson, Jörundur. 1986. *Studies in Tocharian Phonology, Morphology and Etymology with special emphasis on the* o*-vocalism*. PhD-dissertation, Leiden University.
1989a. Rounding and exceptions from rounding in East Tocharian. *Indogerman-ische Forschungen* 94, 101–134.
1989b. *The dual forms of nouns and pronouns in Tocharian*. Reykjavík: Málvísindastofnun Háskóla Íslands.
1991. *The Nasal Prefixes in Tocharian: A Study in Word Formation.* Reykjavík: Málvísindastofnun Háskóla Íslands.
1993. Development of labiovelars (and tectals plus *u̯*) in initial position in Tocharian (an overview). *Die Sprache* 35, 176–86.

Hoffner, Harry A. & H. Craig Melchert. 2008. *A grammar of the Hittite language. Part 1, Reference grammar*. Winona Lake, Indiana: Eisenbrauns.

Holopainen, Sampsa. 2019. *Indo-Iranian borrowings in Uralic. Critical overview of the sound substitutions and distribution criterion*. PhD-dissertation, University of Helsinki.
2021. On the question of substitution of palatovelars in Indo-European loanwords in Uralic. *Suomalais-Ugrilaisen Seuran Aikakauskirja* 98, 197–233.
2023. Hunting for the Uralic accusative(s). Review of: László Honti, Az ősi uráli tárgyrag története és vesszőfutása: Accusatum et expulsum. *Finnisch-Ugrische Forschungen* 68, 237–244.

Honti, László. 1976. Az osztják személyjelölő suffixumok történeti áttekintése. [A historical overview of the Ostyak personal suffixes.] *Nyelvtudományi Közlemények* 78, 71–119.
1995. Zur Morphotaktik und Morphosyntax der uralischen/finnisch-ugrischen Grundsprache. In Heikki Leskinen et al. (eds.), *Congressus Octavus Internationalis Fenno-Ugristarum. Jyväskylä 10.–15.8.1995. Pars I: Orationes plenariae et conspectus quinquennales*. Jyväskylä: Moderatores, 53–82.
1998. ObUgrian. In Daniel Abondolo (ed.), *The Uralic Languages*. London-New York: Routledge, 327–357.

1998/99. Gondolatok a mordvin tárgyas igeragozás uráli alapnyelvi hátteréről. [Thoughts on the Proto-Uralic background of the Mordvin objective conjugation.] *Nyelvtudományi közlemények* 96, 106–119.

Hopper, P.J. 1973. Glottalized and murmured occlusives in Indo-European. *Glossa* 7, 141–166.

Hualde, José Ignacio. 2005. Quasi-phonemic contrasts in Spanish. In Vineeta Chand et al. (eds.), *WCCFL 23: Proceedings of the 23rd West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Press, 374–398.

Huard, Athanaric. 2022. *Recherches sur les textes de méditation en tokharien*. PhD-dissertation, École Pratique des Hautes Études, Paris.

Imamura, Satoshi. 2022. A corpus analysis of two by-passives in Japanese from the viewpoint of information structure. *Lingua* 272.

Itkin, Ilya B. 2011. Review of: Gerd Carling & Georges-Jean Pinault, A Dictionary and Thesaurus of Tocharian A. Volume 1: Letters a-j. *Tocharian and Indo-European Studies* 12, 243–254.

Ivanov, V. V. 1985. Tocharian and Ugrian. In U. Pieper and G. Stickel (eds.), *Studia Linguistica. Diachronica et Synchronica*. Berlin: De Gruyter, 411–419.

Jakob, Anthony. 2023. *A history of East Baltic through language contact*. PhD-dissertation, Leiden University.

Jalava, Lotta. 2022. *Tundraa myöten, joen mukana. Prosekutiivisijan alkuperästä samojedilais-itämerensuomalaisin vertailuin*. [Across the tundra, along the river. On the origin of the prosecutive case with a Samoyed-Finnic comparison.] In Kaisla Kaheinen et al. (eds.), *Hämeenmaalta Jamalille. Kirja Tapani Salmiselle 07.04.2022*. Helsinki: University of Helsinki, 145–158.

Janda, Gwen Eva, Johanna Laakso & Helle Metslang. 2022. Person marking. In Marianne Bakró-Nagy, Johanna Laakso and Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 894–903.

Janhunen, Juha. 1976. Adalékok az északi-szamojéd hangtörténethez. Vokalizmus: az első szótagi magánhangzók. [Additions to the sound history of Northern Samoyed. Vocalism: the first-syllable vowels.] *Néprajz és Nyelvtudomány* 19–20, 165–188.

1977. *Samojedischer Wortschatz. Gemeinsamojedische Etymologien*. Helsinki: Suomalais-Ugrilainen Seura.

1981. Uralilaisen kantakielen sanastosta. [On the vocabulary of the Uralic protolanguage.] *Suomalais-Ugrilaisen Seuran Aikakauskirja* 77, 219–274.

1982. On the structure of Proto-Uralic. *Finnisch-Ugrische Forschungen*, 44, 23–42.

1983. On early Indo-European-Samoyed contacts. In Juha Janhunen et al. (eds.), *Symposium Saeculare Societatis Fenno-Ugricae*. Helsinki: Suomalais-Ugrilainen Seura, 115–127.

1989. Samojedin predestinatiivisen deklinaation alkuperästä. [On the origin of the Samoyed predestinative declination.] *Suomalais-Ugrilaisen Seuran Aikakauskirja*. 82, 298–301.

1991. = Юха Янхунен. 1991. Нганасаны и распад прасамодийской языковой общности. [The Nganasans and the disintegration of the Proto-Samoyed speech community.] In Семинар "Проблемы происхождения народов уральской языковой семьи". Izhevsk, 16–19.

1998. Samoyedic. In Daniel Abondolo (ed.), *The Uralic Languages*. London-New York: Routledge, 457–479.

2000. Reconstructing pre-Proto-Uralic typology spanning the millennia of linguistic evolution. *Congressus Nonus Internationalis Fenno-Ugristarum: pars I*, 59–76.

2002. The Nenets imperative sentence and its background. *Finnisch-Ugrische Mitteilungen* Band 24/25, 71–85.

2007. The primary laryngeal in Uralic and beyond. In Jussi Ylikoski and Ante Aikio (eds.), *Sámit, sánit, sátnehámit. Riepmočála Pekka Sammallahtii miessemánu 21. beaivve 2007.* Helsinki: Suomalais-Ugrilainen Seura, 203–227.

2009. Proto-Uralic—what, where, and when? In Jussi Ylikoski (ed.), *The Quasquicentennial of the Finno-Ugrian Society*. Helsinki: Suomalais-Ugrilainen Seura, 57–78.

2013. Personal pronouns in Core Altaic. In Martine Robbeets and Hubert Cuyckens (eds.), *Shared Grammaticalization: With special focus on the Transeurasian languages*. Amsterdam: John Benjamins, 212–226.

2020. Issues of comparative Uralic and Altaic Studies (6). Uralic copulas and their analogues in other Eurasian languages. In *Ёмас сымыӈ нэ̄квe вӧртур э̄тност самын натум – Scripta miscellanea in honorem Ulla-Maija Forsberg*. Helsinki: Suomalais-Ugrilainen Seura, 129–140.

2022. = Юха Янхунен. 2022. Великое прошлое малых народов (на примере самодийцев). [The great past of small peoples (the case of the Samoyeds).] *Археология Евразийских Степей* No. 2, 283–289.

Janurik, Tamás. 2021. *Kamassz Szótár*. [Kamas Dictionary.] Székesfehérvár: Budenz Alkotóház.

Jasanoff, Jay H. 1987. Some irregular imperatives in Tocharian. In Calvert Watkins (ed.), *Studies in memory of Warren Cowgill (1929–1985). Papers from the Fourth East*

*Coast Indo-European Conference, Cornell University, June 6–9, 1985*. Berlin-New York: De Gruyter, 92–112.

2013. The Tocharian subjunctive and preterite in *-*a*-. In Adam I. Cooper et al. (eds.), *Multi Nominis Grammaticus. Studies in Classical and Indo-European Linguistics in honor of Alan J. Nussbaum on the occasion of his sixty-fifth birthday*. Ann Arbor: Beech Stave Press, 105–120

2015. The Tocharian B accent. In Melanie Malzahn et al. (eds.), *Tocharian texts in context. International conference on Tocharian manuscripts and Silk Road culture held June 26–28, 2013 in Vienna*. Bremen: Hempen, 87–98.

Kaheinen, Kaisla. 2023. *Etymologia ex silentio: Nganasanin äännehistoria ja kielikontaktit*. [Etymologia ex silentio: Nganasan historical phonology and language contact.] PhD-dissertation, University of Helsinki.

Kallio, Petri. 2001. Phonetic Uralisms in Indo-European? In Christian Carpelan et al. (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations.* Helsinki: Suomalais-Ugrilainen Seura, 221–234.

2004. Tocharian loanwords in Samoyed? In Irma Hyvärinen et al. (eds.), *Etymologie, Entlehnungen und Entwicklungen: Festschrift für Jorma Koivulehto zum 70. Geburtstag*. Helsinki: Suomalais-Ugrilainen Seura, 129–137.

2006. Suomen kantakielten absoluuttista kronologiaa [On the absolute chronology of the Finnic protolanguages]. *Virittäjä* 110, 2–25.

2012. The non-initial-syllable vowel reductions from Proto-Uralic to Proto-Finnic. In Tiina Hyytiäinen et al. (eds.), *Per Urales ad Orientem. Iter polyphonicum multilingue. Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012*. Helsinki: Suomalais-Ugrilainen Seura, 163–175.

Kazakevič, Olga. 2022. Selkup. In Marianne Bakró-Nagy, Johanna Laakso and Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 777–816.

Kenesei, István & Krisztina Szécsényi. 2022. Hungarian. In Marianne Bakró-Nagy, Johanna Laakso and Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 636–658.

Keresztes, László. 1998. Mansi. In Daniel Abondolo (ed.), *The Uralic Languages*. London-New York: Routledge, 387–427.

1999. *Development of Mordvin definite conjugation*. Helsinki: Suomalais-Ugrilainen Seura.

Khanina, Olesya & Andrey Shluinsky 2015. = О. В. Ханина & А. Б. Шлуинский. 2015. Прямой объект в энецком языке: объектное согласование глагола

[Direct object in Enets: verbal object cross-reference]. In *Типология морфосинтаксических параметров: Материалы международной конференции "Типология морфо-синтаксических параметров"* Volume 2. Moscow: MPGU, 392-410.
http://mpgu.su/wp-content/uploads/2017/01/hanina_shluinskij.pdf

2023. Forest and Tundra Enets. In Daniel Abondolo & Riitta–Liisa Valijärvi (eds.), *The Uralic Languages. Second Edition*. London, New York: Routledge, 793–852.

2024. Uralic object agreement on the verb: Enets evidence for a topic-marking strategy. In S. Björklöf et al. (eds.), *Itämeren kieliapajilta Volgan verkoille: Pühendusteos Riho Grünthalile 22. mail 2024*. Helsinki: Suomalais-Ugrilaisen Seura, 333–358.

Kim, Ronald I. 1999. The development of labiovelars in Tocharian: A closer look. *Tocharian and Indo-European Studies* 8, 139-187.

2000. 'To drink' in Anatolian, Tocharian and Proto-Indo-European. *Historische Sprachforschung* 113, 151–170.

2003. Uncovering the Prehistory of the Tocharian Preterite II. *Historische Sprachforschung* 116, 190–233.

2007a. The Duke of York Comes to Xinjiang: Ablaut, Analogy and Epenthesis in Tocharian B Nasal Presents. *Historische Sprachforschung* 120, 66–104.

2007b. Vowel Weakening in Tocharian A Preterite Participles and Abstract Nouns. *Kyoto University Linguistic Research* 26, 1–30.

2013. The Indo-European, Anatolian, and Tocharian 'secondary' cases in typological perspective. In Adam I Cooper, Jeremy Rau and Michael Weiss (eds.), *Multi Nominis Grammaticus. Studies in Classical and Indo-European Linguistics in honor of Alan J. Nussbaum on the occasion of his sixty-fifth birthday*. Ann Arbor: Beech Stave Press, 121–142.

2018. *The Dual in Tocharian, From Typology to Auslautgesetz*. Dettelbach: J.H. Röll.

Klingenschmitt, Gert. 1975. Tocharisch und Urindogermanisch. In Helmut Rix (ed.), *Flexion und Wortbildung. Akten der V. Fachtagung der Indo-germanischen Gesellschaft, Regensburg, 9–14. September 1973.* Wiesbaden: Reichert, 148–163.

Kloekhorst, Alwin. 2008. *Etymological Dictionary of the Hittite Inherited Lexicon*. Leiden-Boston: Brill.

2016. The Anatolian stop system and the Indo-Hittite hypothesis. *Indogermanische Forschungen* 121, 213-247.

2018. Anatolian evidence suggests that the Indo-European laryngeals $*h_2$ and $*h_3$ were uvular stops, *Indo-European Linguistics* 6, 69–94.

Klumpp, Gerson. 2002. *Konverbkonstruktionen im Kamassischen*. Wiesbaden: Harrassowitz.

2022. Kamas. In Marianne Bakró-Nagy, Johanna Laakso and Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 817–843.

Klumpp, Gerson & Josefina Budzisch. 2023. In Daniel Abondolo and Riitta-Liisa Valijärvi (eds.), *The Uralic Languages*. Second Edition. London-New York: Routledge, 897–938.

Koller, Bernhard. 2018. On the internal structure of the Tocharian A comitative suffix. In David M. Goldstein, Stephanie W. Jamison and Brent Vine (eds.), *Proceedings of the 28th Annual UCLA Indo-European Conference*. Bremen: Hempen, 167–184.

Kölver, Bernhard. 1965. *Der Gebrauch der sekundären Kasus im Tocharischen*. PhD-dissertation, Universität Frankfurt am Main.

Koponen, Eino. 2022. Saami. In Marianne Bakró-Nagy, Johanna Laakso and Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 103–112.

Korhonen, Mikko. 1967. *Die Konjugation im Lappischen. Morphologisch-historische Untersuchung. I. Die finiten Formkategorien*. Helsinki: Suomalais-Ugrilainen Seura.

1981. *Johdatus lapin kielen historiaan*. [An introduction to the history of the Saami language.] Helsinki: Suomalaisen Kirjallisuuden Seura.

1988. The Lapp Language. In Denis Sinor (ed.), *The Uralic Languages: Description, History and Foreign Influences*. Leiden-New York: Brill, 41–57.

Kortlandt, Frederik H.H. 1985. Proto-Indo-European Glottalic Stops: the Comparative Evidence. *Folia linguistica historica* 19, 183–202.

1988b. On the development of PIE final syllables in Tocharian. *Tocharian and Indo-European Studies* 2, 80-88.

Körtvély, Erika. 2005. *Verb conjugation in Tundra Nenets*. Szeged: Goldpress Nyomda.

Kosterkina, Momde & Ždanova 2001. = Н.Т. Костеркина & А.Ч. Момде & Т.Ю. Жданова. 2001. *Словарь нганасанско-русский и русско-нганасанский*. [Nganasan-Russian and Russian-Nganasan dictionary.] Санкт-Петербург: Просвещение.

Kövesi, Magda A. 1973. Zu den “umstrittenen” Fragen der objektiven Konjugation in den ugrischen Sprachen. *Finnisch-Ugrische Forschungen* 40, 96–106.

Krause, Wolfgang. 1951. Zur Frage nach dem nichtindogermanischen Substrat des Tocharischen. *Zeitschrift für vergleichende Sprachforschung* 69, 185–203.

1952. *Westtocharische Grammatik 1: Das Verbum*. Heidelberg: Winter.

Krause, Wolfgang & Werner Thomas. 1960. *Tocharisches Elementarbuch. Band I. Grammatik*. Heidelberg: Winter.

Kroonen, Guus. 2011. False exceptions to Winter's Law: On the effects of Kluge's Law on the Proto-Germanic consonantism. In Tijmen Pronk and Rick Derksen (eds.), *Accent matters: Papers on Balto-Slavic accentology*. Amsterdam: Rodopi, 251–61.

2013. *Etymological Dictionary of Proto-Germanic*. Leiden-Boston: Brill.

Kroonen, Guus et al. 2022. Indo-European cereal terminology suggests a North-west Pontic homeland for the core Indo-European languages. *PLOS ONE* 17(10). doi:10.1371/journal.pone.0275744

Kroonen, Guus, Gojko Barjamovic & Michaël Peyrot. 2018. Linguistic supplement to Damgaard et al. 2018. Early Indo-European languages, Anatolian, Tocharian and Indo-Iranian. doi.org/10.5281/zenodo.1240524.

Kulikov, Leonid. 2012. The Proto-Indo-European Case System and its Reflexes in a Diachronic typological Perspective: Evidence for the Linguistic Prehistory of Eurasia. *Rivista degli studi orientali*. Nuovo serie 84, 289–309.

Kulonen, Ulla-Maija. 1989. *The passive in Ob-Ugrian*. Helsinki: Suomalais-Ugrilainen Seura.

1993. *Johdatus unkarin kielen historiaan*. [An introduction to the history of the Hungarian language.] Helsinki: Suomalaisen Kirjallisuuden Seura.

Kümmel, Martin Joachim. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert Verlag.

2009. The range of Tocharian *a*-umlaut. In Thomas Olander and Jens Elmegård Rasmussen (eds.), *Internal reconstruction in Indo-European: Methods, results and problems, Section papers from the XVIth International Conference on Historical Linguistics held at the University of Copenhagen*. Copenhagen: Museum Tusculanum, 171–179.

2012. Typology and reconstruction: The consonants and vowels of Proto-Indo-European. In Benedicte Nielsen Whitehead et al. (eds.), *The sounds of Indo-European: Proceedings of the conference on Indo-European linguistics held at Copenhagen University 16–19 April 2009*. Copenhagen: Museum Tusculanum, 292–329.

2020. The development of the perfect within IE verbal systems. In Robert Crellin and Thomas Jügel (eds.), *Perfects in Indo-European Languages and Beyond*. Amsterdam: John Benjamins, 15–47.

Künnap, Ago. 1971. *System und Ursprung der kamassischen Flexionssuffixe*. Helsinki: Suomalais-Ugrilainen Seura.

1999. *Kamass*. München: LINCOM.

2004. On the evidence of the verbal 3rd person suffix **-sV* in Uralic. *Linguistica Uralica* 40, 81–86.

Kuryłowicz, Jerzy. 1970–72. Gli aggettivi in *-l-* e il perfetto slavo. *Ricerche slavistiche* 17–18, 321–328.

Kuritsyna, Anna V. 2022. Tocharische Präsenspartizipien im attributiven Gebrauch. Presentation on 15-09-2022 for the Deutsche Orientalistentag held from 12 to 17 September 2022 at the Freie Universität Berlin.

2023. = Анна В. Курицына. 2023. *Синтаксис тохарских языков в сопоставительном аспекте*. [Syntax of Tocharian languages in comparative aspect.] PhD-dissertation, Moskow.

Kuznecova, Helimski & Gruškina = А. И. Кузнецова, Е. А. Хелимский & Е. В. Грушкина. 1980. *Очерки по селькупскому языку. Тазовский диалект. Том I*. [Essays on the Selkup Language. Taz Dialect. Volume I.] Moscow: Moscow University.

Laakso, Johanna. 2022a. Finnic. In Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 240–253.

2022b. Finnish, Meänkieli, and Kven. In Marianne Bakró-Nagy, Johanna Laakso and Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 254–268.

Lehmann, Christian. 1984. *Der Relativsatz*. Tübingen: Gunter Narr.

Leisiö, Larisa. 2006. Passive in Nganasan. In Werner Abraham and Larisa Leisiö (eds.), *Passivization and Typology: Form and Function*. Amsterdam: John Benjamins, 213–230.

Li, Chunxiang et al. 2010. Evidence that a West-East admixed population lived in the Tarim Basin as early as the Bronze Age. *BMC Biology* 8, 15.

Liimola, Matti. 1963. *Zur historischen Formenlehre des Wogulischen. I. Flexion der Nomina*. Helsinki: Suomalais-Ugrilainen Seura.

1973. Noch einmal über das *l* der objektiven Konjugation des Wogulischen. In Erkki Itkonen and Alho Alhoniemi (eds.), *Commentationes Fenno-Ugricae in honorem Erkki Itkonen sexagenarii die XXVI mensis aprilis anno MCMLXXIII.* Helsinki: Suomalais-Ugrilainen Seura, 196–204.

Lowe, John J. 2015. *Participles in Rigvedic Sanskrit: the syntax and semantics of adjectival verb forms*. Oxford: Oxford University Press.

Lubotsky, Alexander M. 1990. La loi de Brugmann et *$H_3$e-. *La Reconstruction des Laryngales*. Paris: Les Belles Lettres, 129–136.

1994. The original paradigm of the Tocharian word for 'king'. In B. Schlerath (ed.), *Tocharisch. Akten der Fachtagung der Indogermanischen Gesellschaft. Berlin, September 1990*. Reykjavík: Málvísindastofnun Háskóla Íslands, 66–72.

2001. The Indo-Iranian Substratum. In Christian Carpelan, Asko Parpola and Petteri Koskikallio (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations*. Helsinki: Suomalais-Ugrilainen Seura, 301–317.

Lubotsky, Alexander & Starostin Sergei. 2003. Turkic and Chinese Loan Words in Tocharian. In B.L. Bauer and G.-J. Pinault (eds.). *Language in time and space: A Festschrift for Werner Winter on the occasion of his 80$^{th}$ birthday*. Berlin-New York: De Gruyter.

Lundquist, Jesse & Anthony D. Yates. 2018. The morphology of Proto-Indo-European. In Jared Klein et al. (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*. Volume 3. Berlin: De Gruyter, 2079–2195.

Lyutikova, Ekaterina & Andrei Sideltsev. 2020. On the Syntax of Argument Clitics in Hittite. *Transactions of the Philological Society*, 118, 29–78.

Macdonell, Arthur. 1927. *A Sanskrit Grammar for Students*. 3$^{rd}$ ed. London: Oxford University Press.

Mallory, James P. 1989. *In Search of the Indo-Europeans: Language, Archaeology and Myth*. London: Thames & Hudson

2015 *The Problem of Tocharian Origins: An Archaeological Perspective*. Philadelphia: University of Pennsylvania.

Mallory, James P. & Victor H. Mair. 2000. *The Tarim Mummies: Ancient China and the Mystery of the Earliest Peoples from the West*. London: Thames & Hudson.

Malzahn, Melanie. 2010. *The Tocharian Verbal System*. Leiden: Brill.

2011. Die tocharischen Präsens- und Konjunktivstämme auf suffixales *-sk-* und eine innertocharische Vokalschwächungsregel. In Thomas Krisch and Thomas Lidner (eds.), *Indogermanistik und Linguistik im Dialog, Akten der XIII. Fachtagung der Indogermanischen Gesellschaft vom 21. bis 27. September 2008 in Salzburg*. Wiesbaden: Reichert, 361–366.

2014. Tocharian A *śorki* 'fear' and two other TA scary words. *Tocharian and Indo-European Studies* 15, 87–94.

2016. The second one to branch off? The Tocharian lexicon revisited. In Bjanr Simmelkjær Sandgaard Hansen et al. (eds.), *Etymology and the European lexicon. Proceedings of the 14th Fachtagung der Indogermanischen Gesellschaft, 17-22 September 2012, Copenhagen*. Wiesbaden: Reichert, 281–292.

Marcantonio, Angela. 1985. On the definite vs. indefinite conjugation in Hungarian: a typological and diachronic analysis. *Acta Linguistica Academiae Scientiarum Hungaricae* 35, 267–297.

Marchenko, Z.V. et al. 2017. Radiocarbon chronology of complexes with Seima-Turbino type objects (Bronze Age) in southwestern Siberia. *Radiocarbon* 59, 1381–1397.

Marggraf, Wolf-Jürgen. 1970. *Untersuchungen zum akzent in Tocharisch B*. PhD-dissertation, Christian-Albrechts-Universität Kiel.

Maslova, Elena S. 2003. *Tundra Yukaghir*. München: LINCOM.

2008. Review of: Irina Nikolaeva. A historical dictionary of Yukaghir. *Studies in Language* 32, 254–264.

Matasović, Ranko. 2009. *Etymological Dictionary of Proto-Celtic*. Leiden: Brill.

Matras, Yaron. 2009. *Language Contact*. Cambridge: Cambridge University Press.

Mayrhofer, Manfred. 1992. *Etymologisches Wörterbuch des Altindoarischen*. Band 1. Heidelberg: Carl Winter.

Meillet, Antoine & Joseph Vendryes. 1948. *Traité de grammaire comparée des langues classiques.* 2e édition revue et augmentée par J. Vendryes. Paris: Librairie Ancienne Honoré Champion.

Melchert, H. Craig. 2002. Review of: G. Carling, Die Functionen der lokalen Kasus im Tocharischen. *Journal of the American Oriental Society* 122, 107–108.

Meunier, Fanny. 2017. Quelques éléments de syntaxe des adjectifs verbaux (gérondifs) et participes en tokharien. In Claire Le Feuvre, Daniel Petit and 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, 221–232.

Meyer, Robin. 2017. *Iranian-Armenian language contact in and before the 5th century CE*. PhD-dissertation, University of Oxford.

Mikola, Tibor. 1988. Geschichte der samojedischen Sprachen. In Denis Sinor (ed.), *The Uralic Languages: Description, History and Foreign Influences*. Leiden-New York: Brill, 219–263.

1995. *Morphologisches Wörterbuch des Enzischen*. Szeged: University of Szeged.
2004. *Studien zur Geschichte der samojedischen Sprachen*. Szeged: University of Szeged.

Morgunova, N. L. & O. S. Khokhlova. 2013. Chronology and periodization of the pit-grave culture in the region between the Vogla and Ural Rivers based on radiocarbon dating and paleopedological research. *Radiocarbon* 55, 1286–1296.

Mus, Nikolett. 2023. Nenets. In Daniel Abondolo and Riitta-Liisa Valijärvi (eds.), *The Uralic Languages*. Second Edition. London-New York: Routledge, 853–896.

Narisimhan, Vagheesh et al. 2019. The Formation of Human Populations in South and Central Asia. *Science* 365(6457).

Napol'skikh, Vladimir. 2001. Tocharisch-uralische Berührungen: Sprache und Archäologie. In Christian Carpelan, Asko Parpola and Petteri Koskikallio (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations*. Helsinki: Suomalais-Ugrilainen Seura, 367–383.

Nielsen Whitehead, Benedicte. 2018. PIE *$g^{w}h_{3}$-*éu̯*- 'cow'. In Lucien van Beek et al. (eds.), *Farnah: Indo-Iranian and Indo-European Studies in Honor of Sasha Lubotsky*. Ann Arbor-New York: Beech Stave Press, 217–231.

Nikolaeva, Irina. 1999a. *Ostyak*. München: LINCOM.
1999b. Object agreement, grammatical relations, and information structure. *Studies in Language 23*, 341–386.
2001. Secondary topic as a relation in information structure. *Linguistics* 39, 1–49.
2006. *A Historical Dictionary of Yukaghir*. Berlin-New York: De Gruyter.
2014. *A Grammar of Tundra Nenets*. Berlin-Boston: De Gruyter.
2015. On the expession of TAM on nouns: evidence from Tundra Nenets. *Lingua* 166, 99–126.
2020. *Yukaghir morphology in a historical and comparative perspective.* München: LINCOM.

Ning, Chao et al. 2019. Ancient genomes reveal Yamnaya-related ancestry and a potential source of Indo-European speakers in Iron Age Tianshan. *Current Biology* 29 2526–2532.

Oettinger, Norbert. 2013/14. Die Indo-Hittite-Hypothese aus heutiger Sicht. *Münchener Studien zur Sprachwissenschaft* 67/2, 149–176.

Olander, Thomas. 2020. To **b* or not to **b*, Proto-Indo-European **b* in a phylo-genetic perspective. *Historische Sprachforschung* 133, 182–208.

Olsen, Birgit Anette. 1999. *The noun in Biblical Armenian. Origin and word formation – with special emphasis on Indo-European heritage*. Berlin: De Gruyter.

Onishi, Teigo. 2022. *Pronominal clitics in Tocharian: a study in the morphology-syntax interface*. PhD-dissertation, University of California Los Angeles.

Pakendorf, Brigitte. 2010. Contact and Siberian languages. In Raymond Hickey (ed.), *The Handbook of Language Contact*. Malden: Wiley-Blackwell, 714–737.

2012. Patterns of relativization in North Asia: Towards a refined typology of prenominal participial relative clauses. In Volker Gast and Holger Diessel (eds.), *Clause linkage in cross-linguistic perspective: Data-driven approaches to cross-clausal syntax*. Berlin-Boston: De Gruyter, 253–284.

Papp, István. 1968. *Unkarin kielen historia*. [The History of the Hungarian Language.] Helsinki: Suomalaisen Kirjallisuuden Seura.

Parpola, Asko. 2012. Formation of the Indo-European and Uralic (Finno-Ugric) language families in the light of archaeology: Revised and integrated "total" correlations. In Riho Grünthal and Petri Kallio (eds.), *A linguistic map of prehistoric northern Europe*. Helsinki: Suomalais-Ugrilainen Seura, 119–184.

Peyrot, Michaël. 2007. Review of: Haruyuki Saito, Das Partizipium Präteriti im Tocharischen. *Bibliotheca Orientalis* 64, 798–801.

2008. *Variation and change in Tocharian B*. Amsterdam-New York: Rodopi.

2010. On the formation of the Tocharian preterite participle. *Historische Sprachforschung* 121, 69–83.

2012a. The Tocharian A match of the Tocharian B obl.sg. *-ai*. *Tocharian and Indo-European Studies* 13, 181–220.

2012b. Tocharian 'eat' and the strong imperfect in Tocharian A. In Olav Hackstein and Ronald I. Kim (eds.), *Linguistic developments along the Silk Road: Archaism and innovation in Tocharian.* Wien: Österreichische Akademie der Wissenschaften, 85–119.

2013. *The Tocharian Subjunctive*. Leiden: Brill.

2017. On the part of speech and the syntax of the Tocharian present participle. 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, 327–341.

2018a. Interrogative stems in Hittite and Tocharian. *Indogermanische For-schungen* 123, 65–90.

2018b. Tocharian agricultural terminology: Between inheritance and language contact. In Guus Kroonen et al. (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*). Washington DC: Institute for the Study of Man, 242–277.

2019a. The deviant typological profile of the Tocharian branch of Indo-European may be due to Uralic substrate influence. *Indo-European Linguistics* 7, 72–121.

2019b. Indo-Uralic, Indo-Anatolian, Indo-Tocharian. In Alwin Kloekhorst and Tijmen Pronk (eds.), *The Precursors of Proto-Indo-European: The Indo-Anatolian and Indo-Uralic hypotheses*. Leiden-Boston: Brill, 186–202.

2022a. Tocharian. In Thomas Olander (ed.), *The Indo-European language family – A phylogenetic perspective*. Cambridge: Cambridge University, 83–101.

2022b. Notes on Tocharian A *o*(*k*), A *oram* and B *sorromp* 'down', B *oṣno*, B *nanāmo* 'recognising', B *pāwe*, and B †*səwm-* 'trickle'. *Tocharian and Indo-European Studies* 21, 163–177.

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.

Piispanen, Peter S. 2013. The Uralic-Yukaghiric Connection Revisited: Sound Correspondences of Geminate Clusters. *Suomalais-Ugrilaisen Seuran Aikakauskirja* 94, 165–197.

Pinault, Georges-Jean. 1990. Notes sur les manuscrits de Maitreyasamiti. *Tocharian and Indo-European Studies* 4, 119–202.

2005. Tocharian 'woman'. In Nikolai N. Kazansky (ed.), *Hṛdā́ mánasā. Studies presented to Professor Leonard G. Herzenborg on the occasion of his 70th birthday*. Saint Petersburg: Nauka, 201–214.

2008. *Chrestomathie tokharienne*. Leuven-Paris: Peeters.

2009. On the formation of the Tocharian demonstratives. In Elisabeth Rieken and Paul Widmer (eds.), *Pragmatische Kategorien. Form, Funktion und Diachronie. Akten der Arbeitstagung der Indogermanischen Gesellschaft vom 24. bis 26. September 2007 in Marburg*. Wiesbaden: Reichert, 221–245.

2010. Le pronom d'ipséité en tokharien. In Injoo Choi-Jonin, Marc Duval & Olivier Soutet (eds.), *Typologie et comparatisme. Hommages offerts à Alain Lemaréchal*. Leuven: Peeters, 351–365.

2011. Let us now praise famous gems. *Tocharian and Indo-European Studies* 12, 155–220.

2012. Tocharian *-nt-* participles and agent nouns. In Olav Hackstein and Ronald I. Kim (eds.), *Linguistic developments along the Silk Road: Archaism and innovation in Tocharian*. Wien: Österreichische Akademie der Wissenschaften, 179–204.

Pinault, Georges-Jean & Michaël Peyrot. fthc. A dramatic "Life of the Buddha" in Tocharian B. Edition of manuscript PK AS 12 of the Pelliot Collection.

Poliakov, Andrey V. et al. 2019. A review of the radiocarbon dates for the Afanasyevo Culture (Central Asia): shifting towards the "shorter" chronology. *Radiocarbon* 61 (1), 243–263.

Pronk, Tijmen. 2009. Reflexes of the deletion and insertion of Proto-Tocharian **ä* in Tocharian B. *Tocharian and Indo-European Studies* 11, 69–113.

2018. Language contact and prosodic change in Slavic and Baltic. *Diachronica* 35:4, 552–580.

Pystynen, Juho. 2014. *Close vowel reduction in Samoyedic*. Blog post. https://protouralic.wordpress.com/2014/09/22/close-vowel-reduction-in-samoyedic/.

2017. **ü > *i, *ü in Samoyedic*. Blog post. https://protouralic.wordpress.com/2017/10/07/u-i-u-in-samoyedic/

Rau, Jeremy. 2017. The derivational history of the perfect participle active. In Claire Le Feuvre, Daniel Petit and 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, 377–389.

Raun, Alo 1988a. The Mordvin Language. In Denis Sinor (ed.), *The Uralic Languages: Description, History and Foreign Influences*. Leiden-New Yok: Brill, 96–110.

1988b. Proto-Uralic comparative historical morphosyntax. In Denis Sinor (ed.), *The Uralic Languages: Description, History and Foreign Influences*. Leiden: Brill, 555–571.

Rédei, Károly. 1988. Geschichte der permischen Sprachen. In Denis Sinor (ed.), *The Uralic Languages. Description, history and foreign influences*. Leiden: Brill, 351–394.

1988–91. *Uralisches Etymologisches Wörterbuch*. Budapest: Akadémiai Kiadó.

1989. A finnugor igeragozásról, különös tekintettel a magyar igei személyragok eredetére. [About the Finno-Ugric conjugation, with special attention for the origin of the Hungarian verbal conjugations.] *Nyelvtudományi közlemények* 90, 143–160.

1998/99. Gondolatok az uráli tárgyas igeragozásról. [Thoughts about the Uralic definite conjugation.] *Nyelvtudományi közlemények* 96, 124–128.

1999. Zu den uralisch-jukagirischen Sprachkontakten. *Finnisch-Ugrische Forschungen* 55, 1–58.

Riese, Timothy. 1998. Permian. In Daniel Abondolo (ed.), *The Uralic Languages*. London: Routledge, 249–275.

Rießler, Michael. 2016. *Adjective attribution*. Berlin: Language Science Press.

Ringe, Donald A. 1987. On the prehistory of Tocharian B accent. In C. Watkins (ed.), *Studies in Memory of Warren Cowgill (1929–1985). Papers From the Fourth East Coast Indo-European Conference. Cornell University, June 6–9, 1985*. Berlin-New York: De Gruyter, 254–269.

1989. Tocharian B *ausu*, *auṣu*, *aultsu*. *Tocharian and Indo-European Studies* 3, 35–50.

1990. The Tocharian active s-preterite: a classical sigmatic aorist. *Münchener Studien zur Sprachwissenschaft* 51, 183–242.

1991. Evidence for the position of Tocharian in the Indo-European family? *Sprache* 34, 59–123.

1996. *On the Chronology of Sound Changes in Tocharian*. Volume 1. New Haven, Connecticut: American Oriental Society.

Rix, Helmut & Martin Kümmel. 2001. *Lexikon der indogermanischen Verben: die Wurzeln und ihre Primärstammbildungen.* 2nd ed. Wiesbaden: Reichert.

Róna-Tas, András. 1998. The Reconstruction of Proto-Turkic and the Genetic Question. In Lars Johanson and Éva Á. Csató (eds.), *The Turkic Languages*. London-New York: Routledge, 67–125.

Saag, Lehti et al. 2019. The Arrival of Siberian Ancestry Connecting the Eastern Baltic to Uralic Speakers further East. *Current Biology* 29, 1701–1711.

Saarikivi, Janne. 2022. The divergence of Proto-Uralic and its offspring. In Marianne Bakró-Nagy, Johanna Laakso and Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 28–58.

Saarinen, Sirkka. 2022. Mari. In Marianne Bakró-Nagy, Johanna Laakso and Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 432–470.

Saito, Haruyuki. 2006. *Das Partizipium Präteriti im Tocharischen*. Wiesbaden: Harrassowitz.

Salminen, Tapani. 1996. A commentary address to László Honti's "Zur Morphotaktik und Morphosyntax der uralischen/finnisch-ugrischen Grundsprache". *Congressus octavus Internationalis Fenno-Ugristarum, Jyväskylä 10–15.8 1995. Pars VIII*, 25–27.

1997. *Tundra Nenets Inflection*. Helsinki: Suomalais-Ugrilainen Seura.

1998. *A morphological dictionary of Tundra Nenets*. Helsinki: Suomalais-Ugrilainen Seura.

2002. Problems in the taxonomy of the Uralic languages in the light of modern comparative studies. In А.Е. Кибрик (ed.), *Лингвистический беспредел: сборник статей к 70-летию А.И. Кузнецовой*. Москва: Издательство Московского университета, 44–55.

2007. Notes on Forest Nenets Phonology. In Jussi Ylikoski and Ante Aikio (eds.), *Sámit, sánit, sátnehámit. Riepmočála Pekka Sammallahtii miessemánu 21. beaivve 2007.* Helsinki: Suomalais-Ugrilainen Seura, 349–372.

2012. Traces of Proto-Samoyed vowel contrasts in Nenets. In Tiina Hyytiäinen et al. (eds.), *Per Urales ad Orientem. Iter polyphonicum multilingue. Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012*. Helsinki: Suomalais-Ugrilainen Seura, 339–358.

2014. Suomalais-samojedilaisia muotovertailuja. [Finnic-Samoyed morpho-logical comparisons] In Nobufumi Inaba et al. (eds.), *Juuret marin murteissa, lehvat yltää Uraliin. Juhlakirja Sirkka Saarisen 60-vuotispäiväksi 21.12.2014.* Helsinki: Suomalais-Ugrilainen Seura, 289–300.

2023. Suomalais-samojedilaisia sanavertailuja. [Finnic-Samoyed lexical comparisons] In Sampsa Holopainen et al. (eds.), *Elämä ja Etymologia. Janne Saarikiven 50-vuotisjuhlakirja*. Helsinki: Suomalais-Ugrilainen Seura, 377–399.

Sammallahti, Pekka. 1974. *Material from Forest Nenets*. Helsinki: Suomalais-Ugrilainen Seura.

1988. Historical phonology of the Uralic languages with special reference to Samoyed, Ugric and Permic. In Denis Sinor (ed.), *The Uralic Languages. Description, history and foreign influences*. Leiden: Brill, 478–554.

1990. On the relationships of the Uralic languages. In László Jakab et al. (eds.), *Congressus Septimus Internationalis Fenno-Ugristarum, 3A, Sessiones Sectionum – Dissertationes Linguistica*. Debrecen: Kossuth Lajos tudományegyetem, 7–12.

1998. Saamic. In Daniel Abondolo (ed.), *The Uralic Languages*. London-New York: Routledge, 43–95.

Schmidt, Karl Horst. 1990. The Postulated Pre-Indo-European Substrates in Insular Celtic and Tocharian. In T.L. Markey & A.C. Greppin (eds.), *When Worlds Collide: the Indo-Europeans and the Pre-Indo-Europeans*. Ann Arbor: Karoma, 179–202.

Schmidt, Klaus T. 1985. Beiträge zur Kenntnis der tocharischen Verbalmorpho-logie. In Bernfried Schlerath (ed.), *Grammatische Kategorien, Funktion und Geschichte, Akten der VII. Fachtagung der Indogermanischen Gesellschaft, Berlin, 20.–25. Februar 1983.* Wiesbaden: Reichert, 424–434.

1992. Archaismen des Tocharischen und ihre Bedeutung für Fragen der Rekonstruktion und der Ausgliederung. In Robert S.P. Beekes, Alexander Lubotsky and Jos Wietenberg (eds.), *Rekonstruktion und relative Chronologie.*

*Akten der VIII. Fachtagung der Indogermanischen Gesellschaft, Leiden, 31. August - 4. September 1987*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck, 101–114.

1997. Zu einigen Archaismen unter den tocharischen Präteritalbildungen. *Tocharian and Indo-European Studies* 7, 255–261.

Seržant, Ilja A. 2020. Paradigmatisation of the perfect and resultative in Tocharian. In Robert Crellin & Thomas Jügel (eds.), *Perfects in Indo-European Languages and Beyond*. Amsterdam: John Benjamins, 215–244.

Shagal, Ksenia. 2018. Participial systems in Uralic languages: an overview. *Journal of Estonian and Finno-Ugric linguistics* 9 (1), 55–84.

2019. *Participles – A Typological Study*. Berlin-Boston: De Gruyter.

Sieg, Emil. 1944. *Übersetzungen aus dem Tocharischen I.* Berlin: Akademie der Wissenschaften.

Siegl, Florian. 2013. *Materials on Forest Enets, an Indigenous Language of Northern Siberia*. Helsinki: Suomalais-Ugrilainen Seura.

Simon, Zsolt. 2020. Urindogermanische Lehnwörter in den uralischen und finno-ugrischen Grundsprachen: Eine Fata Morgana? *Indogermanische Forschungen* 125, 239–265.

Sipos, Mária. 2022. North Khanty. In Marianne Bakró-Nagy, Johanna Laakso and Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 582–607.

Skribnik, Elena. 2001. Pragmatic structuring in Northern Mansi. In Tõnu Seilenthal (ed.), *Congressus Nonus Internationalis Fenno-ugristarum. 7.-13.8.2000 Tartu. Pars VI. Dissertationes sectionum: Linguistica III*. Tartu: Tartu Ülikool, 222–239.

Sorokina 2010. = И. П. Сорокина. 2010. *Энецкий язык*. [The Enets Language.] Санкт-Петербург: Наука.

Sorokina & Bolina 2001. = И. П. Сорокина & Д. С. Болина. 2001. *Словарь энецко-русский и русско-энецкий*. [Enets-Russian and Russian-Enets dictionary.] Санкт-Петербург: Просвещение

Starostin 1982. = Сергей А. Старостин. 1982. Праенсейская реконструкция и внешние связи енисейских языков. [Proto-Yeniseian reconstruction and external connections of the Yeniseian languages.] In: Е. А. Алексеенко et al. (eds.), *Кетский сборник. Антропология, этнография, мифология, лингвистика – Studia Ketica. Physical anthrpology, ethnography, mythology, linguistics*. Leningrad: NAUKA, 144–237.

Sturtevant, Edgar H. 1933. *A Comparative Grammar of the Hittite Language*. Philadelphia: Linguistic Society of America/University of Pennsylvania.

Tambets, Kristina et al. 2018. Genes reveal traces of common recent demographic history of most of the Uralic speaking populations. *Genome Biology* 19.

Tereščenko 1965 = Н. М. Терещенко. 1965. *Ненецко-русский словарь*. [Nenets-Russian dictionary.] Москва: Советская Энциклопедия.

1973. = Н. М. Терещенко. 1973. *Синтаксис самодийских языков*. [The syntax of the Samoyed languages.] Ленинград: Наука.

Thomas, Werner. 1952. *Die tocharischen Verbaladjektive auf* -l. Berlin: Akademie-Verlag.

1957. *Der Gebrauch der Vergangenheitstempora im Tocharischen*. Wiesbaden: Harrassowitz.

1994. Zur Frage der nichtindogermanischen Komponente im Tocharischen. In Roland Bielmeier & Reinhard Stempel (eds.), *Indogermanica et Caucasica. Festschrift für Karl Horst Schmidt zum 65. Geburtstag*. Berlin-New York: De Gruyter, 222–233.

Thomason, Sarah G. 2001. *Language Contact*. Edinburgh University Press.

Thomason, Sarah G. & Terrence Kaufman 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkely-Los Angeles: University of California Press.

Thurneysen, Rudolph. 1946. *A Grammar of Old Irish.* Revised and enlarged edition, translated from the German by D. A. Binchy and Osborn Bergin. Dublin: The Dublin Institute for Advanced Studies.

Thorsø, Rasmus & Andrew Wigman et al. 2023. Word Mining: Metal Names and the Indo-European Dispersal. In Kristian Kristiansen, Guus Kroonen and Eske Willerslev (eds.), *The Indo-European Puzzle Revisited. Integrating Archaeology, Genetics and Linguistics.* Cambridge: Cambridge University Press, 105–126.

Trosterud, Trond. 2006. *Homonymy in the Uralic Two-Argument Agreement Paradigms*. Helsinki: Suomalais-Ugrilainen Seura.

Þórhallsdóttir, Guðrún 1988. Tocharian contraction across *-w-*. *Tocharian and Indo-European Studies* 2, 184–210.

Urmanchieva 2023 = Анна Ю. Урманчиева. 2023. Реконструкция глагольного спряжения в самодийских языках: лично-числовые показатели индикатива и форманты типов спряжения. [Reconstruction of verbal conjugation in Samoyedic languages: Personal inflexions of the indicative and markers of conjugation types.] *Acta Linguistica Petropolitana* 19 part 1, 256–300.

Vajda, Edward. 2022. The Dene-Yeniseian Hypothesis. In Michael Fortescue & Edward Vajda, *Mid-Holocene Language Connections between Asia and North America*. Leiden-Boston: Brill, 235–480.

Van Windekens, Albert J. 1976. *Le tokharien confronté avec les autres langues indo-européennes. Vol. I: La phonétique et le vocabulaire.* Louvain: Centre International de Dialectologie Générale.

1979. *Le tokharien confronté avec les autres langues indo-européennes. Vol. II, 1: La morphologie nominale.* Louvain: Centre International de Dialectologie Générale.

Viitso, Tiit-Rein. 1998. Fennic. In Daniel Abondolo (ed.), *The Uralic Languages*. London:-New York Routledge, 96–114.

Virtanen, Susanna. 2014. Pragmatic direct object marking in Eastern Mansi. *Linguistics* 52, 391–413.

Viti, Carlotta. 2016. Null objects and clitics in some early Indo-European languages. In Rosemarie Lühr (ed.), *Idiosynkrasie. Neue Wege ihrer Beschreibung*. Wiesbaden: Reichert, 13–32.

Wackernagel, Jacob. 1892. Über ein Gesetz der indogemanischen Wortstellung. *Indogermanische Forschungen* 1, 333–436.

Wagner-Nagy, Beáta. 2019. *A Grammar of Nganasan*. Leiden-Boston: Brill.

Wagner-Nagy, Beáta & Sándor Szeverényi. 2022. Samoyedic. In Marianne Bakró-Nagy, Johanna Laakso and Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 659–673.

Waldschmidt, Ernst & Heinz Bechert. 1973. *Sanskrit-Wörterbuch der buddhistischen Texte aus den Turfan-Funden und der kanonischen Literatur der Sarvastivada-Schule, I.* Göttingen: Vandenhoeck und Ruprecht.

Warries, Abel R. 2022. Towards a new comparison of the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems. *Indo-European Linguistics* 10, 169–213.

fthc. An investigation of Tocharian loanwords in Samoyed.

Winter, Werner. 1962. Die Vertretung indogermanischer Dentale im Tocharischen. *Indogermanische Forschungen* 67, 16–35.

1965a. Tocharian evidence. In Werner Winter (ed.), *Evidence for laryngeals*. The Hague: De Gruyter, 190–211.

1977. Internal structure and external relationship of two verbal paradigms: Tocharian B *weñ-*, A *weñ* 'SAY'. *Journal of Indo-European Studies* 5, 133–159.

1990. The importance of fine points in spelling: Deletion of accented vowels in Tocharian B. In Jasek Ficiak (ed.), *Historical Linguistics and Philology*. Berlin-New York: De Gruyter, 371–91.
1993. The development of underlying accented "schwa" before dentals in Tocharian B. *Tocharian and Indo-European Studies*, 197–205.
1994a. Syncope in Tocharian A. In Benedicte Nielsen and Jens E. Rasmussen (eds.), *In honorem Holger Pedersen: Kolloquium der Indogermanischen Gesellschaft vom 26. bis 28. März 1993 in Kopenhagen*. Wiesbaden: Reichert. 401–415.
1994b. Zum tocharischen Verb. In Bernfried Schlerath (ed.), *Tocharisch, Akten der Fachtagung der Indogermanischen Gesellschaft, Berlin, September 1990.* Reykjavík: Málvísindastofnun Háskóla Íslands, 284–309.
1997. Lexical archaisms in the Tocharian languages. In Hans Henrich Hock (ed.), *Historical, Indo-European and Lexicographical studies. A Festschrift for Ladislav Zgusta on the occasion of his 70th birthday*. Berlin-New York: de Gruyter, 183–193

Ylikoski, Jussi. 2011. A survey of the origin of directional case suffixes in European Uralic. In Seppo Kittilä et al. (eds.), *Case, Animacy and Semantic Roles*. Amsterdam: John Benjamins, 235–280.
2016. The origins of the western Uralic *s*-cases revisited: historiographical, functional-typological and Samoyedic perspectives. *Finnisch-Ugrische Forschungen* 63, 6–78.
2017. On the tracks of the Proto-Uralic suffix **-ksi*: a new but old perspective on the origin of the Mari lative. *Suomalais-Ugrilaisen Seuran Aikakauskirja* 96, 369–419.
2018. The so-called relation forms of nouns in South Saami: A byproduct or remnant of Uralic **-mpV*? *Finnisch-Ugrische Forschungen* 64, 6–71.

Zaicz, Gábor. 1998. Mordva. In Daniel Abondolo (ed.), *The Uralic Languages*. London-New York: Routledge, 184–218.

Zhang, Fan et al. 2021. The genomic origins of the Bronze Age Tarim Basin mummies. *Nature* 559, 256–261.

Zhivlov, Mikhail 2014. Studies in Uralic Vocalism III. *Journal of Language Relationship* 12, 113–148.
2018. = Михаил Живлов. 2018. Прасамодийский **s* – фонология и фонетика. [Proto-Samoyed **s* – phonology and phonetics.] Семинар по сравнительно-исторической фонетике самодийских языков (Москва). Conference presentation. Powerpoint available at: https://www.academia.edu/38048065/Прасамодийский_s_фонология_и_фонетика.

2022. Pre-Proto-Yukaghir Consonant Clusters. *International Journal of Eurasian Linguistics* 4, 41–59.

2023. Reconstruction of Proto-Uralic. In Daniel Abondolo and Riitta–Liisa Valijärvi (eds.), *The Uralic Languages*. *Second Edition*. London-New York: Routledge, 117–175.

## Samenvatting

Dit proefschrift betreft twee afzonderlijke taalgroepen, het Tochaars en het Samojeeds, en hun mogelijke prehistorische contacten. Het Tochaars is een uitgestorven tak van het Indo-Europees die tot ongeveer duizend jaar geleden werd gesproken het Tarimbekken in het huidige Xinjiang, Noordwest-China. Hoe het Tochaars in die regio is gekomen vanuit het oorspronkelijke Indo-Europese taalgebied op de Pontisch-Kaspische Steppe is onzeker. Op basis van de meest recente archeologische en genetische inzichten ligt een migratie via zuidelijk Siberië voor de hand. Daar werd in het verre verleden al een voorstadium van het Samojeeds gesproken, en de voorouders van de Tocharen hadden dus onderweg naar het Tarim Bekken mogelijk contact met de sprekers van dit pre-Samojeeds.

Archeologie en genetica geven verder geen uitsluitsel over dit mogelijke contact, maar de vergelijkende taalwetenschap kan wel inzichten bieden. Voor een aantal vanuit het Indo-Europees gezien afwijkende aspecten van het Tochaars is voorgesteld dat ze voortkomen uit contact met talen zoals het vroege Samojeeds of het Oeraals, de bredere taalfamilie waartoe het Samojeeds behoort. In dit proefschrift worden verschillende aspecten van het Tochaars vanuit dit oogpunt nader onderzocht om de hypothese over prehistorisch contact te kunnen beoordelen, en zo de hypothesen over de migratieroute van de Tocharen taalkundig te toetsen.

De relatieve chronologie van de diachrone ontwikkelingen in de klanken van het Tochaars en het Samojeeds worden eerst elk afzonderlijk in detail behandeld, en vervolgens worden de data bij elkaar gebracht in een vergelijking van de klanksystemen. De conclusie is dat het gehele pre-Tochaarse klanksysteem, de medeklinkers én de klinkers, in het verleden erg leek op dat van het vroege Samojeeds (hoofdstuk 2–4).

De bespreking van de ontwikkeling van het Tochaarse klanksysteem wordt voortgezet met de plaatsing van de klemtoon. Het Tochaars had, voor zover zichtbaar, de klemtoon op de tweede lettergreep (in drielettergrepige woorden). Hiermee wijkt het af van het Samojeeds, waar de klemtoon vanouds vast op de eerste lettergreep valt. Aan de hand van ontwikkelingen in de lettergreepstructuur in het pre-Tochaars wordt beargumenteerd dat een eerder stadium van het Tochaars de klemtoon eerst wel op de eerste lettergreep had, net als het Samojeeds (hoofdstuk 5).

Het Tochaarse naamvalsysteem is grotendeels vernieuwd ten opzichte van de Indo-Europese voorouder. Vooral de Tochaarse perlatief doet erg denken aan de Samojeedse prolatief, maar het is onmogelijk om met zekerheid te bepalen hoe oud de Samojeedse prolatief zelf is. De Tochaarse en Samojeedse naamvalsystemen lijken op elkaar wat betreft hun algehele structuur, maar in de details zijn er ook verschillen. Het blijft onzeker in hoeverre we hierbij te maken hebben met latere ontwikkelingen, en in

hoeverre de kern van het Tochaarse systeem kan zijn ontstaan door Samojeedse invloeden (hoofdstuk 6).

In dit proefschrift wordt ook een nieuwe overeenkomst tussen het Tochaars en het Samojeeds voorgesteld, namelijk in het systeem van de deelwoorden. Het Tochaars kent meerdere soorten deelwoorden, waarvan er slechts een kan worden omschreven als een actief deelwoord. De andere zijn contextueel, d.w.z. hun interpretatie als actief of passief hangt af van de rest van de zin en is niet inherent aan het deelwoord zelf. Contextuele deelwoorden komen veel voor in het Samojeeds, en ook in andere talen van Siberië. Wat de Tochaarse deelwoorden verder verbindt met het Samojeeds is dat het Tochaars naast deelwoorden voor de tegenwoordige tijd en de verleden tijd ook een negatief deelwoord en een deelwoord van noodzaak heeft. Diezelfde vier soorten deelwoorden kenmerken ook het Samojeedse systeem (hoofdstuk 7).

Ook het Tochaarse finiete werkwoord vertoont een markante overeenkomst met het Samojeedse werkwoord. In alle oude Indo-Europese talen worden werkwoorden vervoegd voor persoon en getal van het subject, maar in het Tochaars bestaat daarnaast nog een reeks achtervoegsels waarmee ook het object gemarkeerd wordt. Dit is vergelijkbaar met de zogenoemde objectieve werkwoordconjugatie van verschillende Oeraalse talen, waaronder het Samojeeds. Binnen de Oeralistiek is de interpretatie en de reconstructie van de objectieve conjugatie een omstreden onderwerp, en er bestaat onenigheid over de oorsprong en de ouderdom ervan. Hier wordt beargumenteerd dat de objectieve conjugatie zeker zo oud is als het Proto-Oeraals, wat betekent dat de objectieve conjugatie ook al bestond in het vroege Samojeeds en niet pas later is ontstaan. Het Tochaarse systeem van objectmarkering kan dus zijn ontstaan onder invloed van de pre-Samojeedse objectieve conjugatie (hoofdstuk 8).

Mogelijke woorden die vanuit het Tochaars in het Samojeeds zijn geleend worden als laatste besproken. Het gaat om zeer weinig voorbeelden, maar deze vormen toch een welkome ondersteuning van de grammaticale aspecten van de contacthypothese (hoofdstuk 9).

Leenwoorden bieden bovendien een van de weinige aspecten van de vergelijking die specifiek aansluiten op het Samojeeds, en niet ook op het Oeraals in bredere zin. De taalkundige kenmerken van het Tochaars passen bij een migratieroute via zuidelijk Siberië en contact met vroeg Samojeeds, maar er is nog vervolgonderzoek nodig naar de vroege voorgeschiedenis van ook de andere talen van de regio om een complexer en genuanceerder beeld te krijgen.

## Curriculum Vitae

Abel Warries was born in 1997 in Amsterdam. During his time at the Barlaeus Gymnasium in Amsterdam, he developed an interest in various languages, especially Scottish Gaelic, North Saami and Japanese. After graduating in 2015, he went to study the BA Linguistics in Leiden. There, he took Tocharian as an elective in the second year, but was initially most interested in the Celtic languages, attending Scottish Gaelic summer courses at Sabhal Mòr Ostaig on the Isle of Skye in 2016 and 2017. He wrote his BA thesis on the phylogeny of the Celtic languages in 2018, under the supervision of Tijmen Pronk. Warries then stayed in Leiden for the MA Linguistics, and as he developed a stronger interest in Saami and the other Uralic languages, he started investigating theories on the ancient contacts between the Indo-European and Uralic language families for term papers. He wrote an MA thesis on the Tocharian and Samoyed case systems under the supervision of Michaël Peyrot in 2019.

Directly following this, Warries joined Michaël Peyrot's ERC-funded project *The Tocharian Trek* at the Leiden University Centre for Linguistics to work on this dissertation. In the course of the years, he presented at conferences such as the 13$^{th}$ Finno-Ugric Congress in Vienna in 2022, went for a two-month research stay in Helsinki and taught an introductory Scottish Gaelic course at the Leiden Summer School of Languages and Linguistics in 2023. Starting in 2019, Abel also became a volunteer at the Naturalis Natural History Museum in Leiden, educating visitors about the museum collection and biology more broadly, and he is currently employed there.