UNIVERSITY OF CALIFORNIA

Los Angeles

Studies in Tocharian Phonology above the Word-Level

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Indo-European Studies

by

Bernhard Koller

2015

ABSTRACT OF THE DISSERTATION

Studies in Tocharian Phonology above the Word-Level

by

Bernhard Koller
Doctor of Philosophy in Indo-European Studies
University of California, Los Angeles, 2015
Professor H. Craig Melchert, Chair

The present work is a collection of studies on the Tocharian languages that focuses on the phonological properties of units larger than the word. The first chapter involves a study of the segmental properties of external sandhi in Tocharian A couched in the framework of Optimality Theory. A philological study involving the Tocharian brāhmī alphabet reveals that there exists a correlation between external sandhi and the orthographic strategy used to render a word-final consonant. The second chapter provides an analysis of the prosodic relationship between clitics and their hosts, showing that these elements form a prosodic constituent intermediate between the word and the phrase. This is the case both in Tocharian A as well as in Tocharian B. The final chapter addresses two different aspects of the internal structure of Tocharian Wh-words. First, the prosodic analysis employed for clitics and their hosts is extended to Wh-words in Tocharian B, making it possible to account for their otherwise aberrant accentuation as well as some of their segmental properties. The second part of the chapter investigates the nature of indefinites in Tocharian A, showing that, contrary to the descriptions in most handbooks, the attested forms do not belong to a single paradigm. Instead, we are dealing with two separate sets of Wh-words doubling as indefinites, a fact that has been obscured by a phonological process that renders one set identical to demonstrative pronouns.

The dissertation of Bernhard Koller is approved.

Stephanie W. Jamison

Dominique Sportiche

Brent Vine

H. Craig Melchert, Committee Chair

University of California, Los Angeles

2015

To Anna

Contents

List of Tables

Symbols and Abbreviations

Symbols Used in Transcriptions of Tocharian Examples

=	vowel deleted in sandhi
()	lost text that has been restored (either from context or based on parallel texts)
{}	an assumed scribal error that has been corrected by the editor
–	missing akṣara
·	missing part of akṣara
#	verse boundary
;	major colon boundary
,	minor colon boundary
•	punctuation mark used in Tocharian manuscripts, usually to mark clause boundaries
:	punctuation mark used in Tocharian manuscripts, usually to mark verse boundaries
///	point at which the manuscript leaf breaks off
[...]	irrelevant part of an example sentence that has been omitted

Abbreviations

^{A}	archaic text
C	consonant
CEToM	A Comprehensive Edition of Tocharian Manuscripts
ClG	clitic group
CS	Chicano Spanish
IDP	The International Dunhuang Project
^{L}	late text
OT	Optimality Theory
PIE	Proto-Indo-European
PPhr	phonological phrase
PT	Proto-Tocharian
PWd	phonological word
R	sonorant
σ	syllable
TA	Tocharian A
TB	Tocharian B
TITUS	Thesaurus Indogermanischer Text- und Sprachmaterialien
V	vowel

Grammatical Features

abl	ablative
abstr	abstract noun derived from a verb
adj	derived adjective
all	allative
conj	conjunction
dem	demonstrative
emph	emphatic particle
gen	genitive
ger	gerund
impf	imperfect
indef	indefinite
inf	infinitive
ins	instrumental
int	interrogative
loc	locative
mp	m-participle (unspecified for voice)
nom	nominative
obl	oblique
opt	optative
per	perlative
pl	plural
PN	personal name
pp	preterite participle
prev	preverb
pro	pronoun
prt	preterite
ptcl	particle
rel	relative
sg	singular

Acknowledgements

First, I would like to thank my teachers and the members of my dissertation committee, Craig Melchert, Stephanie Jamison, Dominique Sportiche, and Brent Vine, not only for their valuable advice on the dissertation itself but for their support and encouragement all throughout my graduate studies.

I am deeply grateful to Melanie Malzahn, without whom, it is fair to say, this dissertation would not exist. It was she who first introduced me to the study of Tocharian at the University of Vienna and who sparked my interest in doing my own research in the field. She later gave me the opportunity to collaborate on the project *A Comprehensive Edition of Tocharian Manuscripts*, thereby granting me unlimited access to the electronic language corpus, which this dissertation relies on heavily. She further took the time to read drafts of each chapter and to provide copious helpful suggestions.

The present work owes much to Hannes Fellner, who was always willing to listen to my ideas and random thoughts about the Tocharian data and who also read and commented on drafts of chapter 2 and 4. Jessica DeLisi generously agreed to read and provide comments and corrections on chapter 3. I would also like to thank Mattyas Huggard, David Goldstein, Laura Grestenberger, Dieter Gunkel, and Kie Zuraw for valuable advice and discussion on various aspects of this dissertation, as well as the participants of the 33rd East Coast Indo-European Conference for their input on some of the material in chapter 2. Thomas Graf gave me some pointers on using the Vim text editor more efficiently, which facilitated the writing process considerably. The Berlin-Brandenburgische Akademie der Wissenschaften

kindly gave me permission to use photographs of manuscript leaves within their collection.

I am grateful for the generous financial support from the late A. Richard Diebold. In addition to his providing funds to the Program in Indo-European Studies at UCLA in general, I was fortunate enough to receive both summer and dissertation year funding as part of my Diebold fellowship.

I would like to thank Tanya Kim, the Student Affairs Officer of the Dodd Humanities Group at UCLA, for her considerable expertise and patience in helping with with all the bureaucratic difficulties of finishing my dissertation outside of the US.

Here I would also like to express my gratitude to the outstanding teachers to whom I owe many of the skills that allowed me to succeed as a graduate student. From my time at the University of Vienna these include Martin Peters, Hans Martin Prinzhorn, and Markus Pöchtrager. At UCLA, I would especially like to thank Bruce Hayes and Russ Schuh, not only for teaching me much about linguistics but also how to teach it.

I would like to thank my parents for all their support throughout my entire education, as well as my friends Jürgen Greimel, Michael Lackner, and Katharina Simma for making a generally stressful time much easier to bear. Finally, I would like to thank the most important person in my life, my wife Anna Pagé, who spent a tremendous amount of time reading and rereading every part of this dissertation. Her endless loving support contributed to this work in more ways than I can count.

Vita

2008	M. A. Indo-European Studies, University of Vienna
2008-2015	Ph.D. student, Program in Indo-European Studies, UCLA
2009, 2010	Chancellor's Prize Summer Mentorship Award, UCLA
2010	Graduate Certificate in Linguistics, UCLA
2009-2013	Teaching Assistant for the Department of Linguistics, UCLA
2010-2015	Diebold Fellow, Program in Indo-European Studies, UCLA
2013	Candidate in Philosophy, Program in Indo-European Studies, UCLA
2014-present	Collaborator on the project "A Comprehensive Edition of Tocharian Manuscripts" at the University of Vienna

Publications and Presentations

Koller, Bernhard. (2015). "On the Relationship between Indefinites and Wh-Words in Tocharian A." Presented at the 34rd East Coast Indo-European Conference, June 4-7. University of Vienna.

–. (2014). "On the Relationship between Virāma Spelling and Tocharian Prosody." Presented at the 33rd East Coast Indo-European Conference, June 6-8. Virginia Tech

–. (2014). "Question Particles in Tocharian A." Presented at the Kyoto-UCLA Student Workshop in Indo-European. March 24-26. Kyoto University.

–. (2013). "Hittite *pai-* 'come' and *uwa-* 'go' as Restructuring Verbs." In Leonid Kulikov and Nikolaos Lavidas (eds). "Proto-Indo-European Syntax and its Development." *Journal of Historical Linguistics* 3.1. 77-97. Forthcoming reprint of volume: John Benjamins, 2015 (Benjamins Current Topics 75)

–. (2013). "On the Status of the Particle *ne* in Tocharian A Clause Structure." Presented at the 25th Annual West Coast Indo-European Conference. October 25-26. UCLA

–. (2013). "Internal and External Sandhi as a Diagnostic for the Affixal/Clitic status of Function Morphemes in Tocharian A." Presented at Tocharian Texts in Context. International Conference on Tocharian Manuscripts and Silk Road Culture. June 26-28. University of Vienna

–. (2011). "Hittite 'come' and 'go' as Restructuring Verbs." Presented at the XX International Symposium on Theoretical and Applied Linguistics. April 1-3. Thessaloniki, Greece.

–. (2008). "Schwa-Zero Alternations and Their Implications for Verbal Root Structure in Tocharian A." Presented at the 20th Annual West Coast Indo-European Conference. October 31-November 1. UCLA.

–. (2007). Co-authored with Anna-Maria Adaktylos, Hannes A Fellner, Melanie Malzahn, Katharina Simma, and Raimund Staudinger. "A Concordance of the Unedited Tocharian Texts of the Berlin Turfan Collection."Melanie Malzahn (ed.). *Instrumenta Tocharica.* Heidelberg. Universitätsverlag Winter. 39-78.

Chapter 1

Introduction

1.1 The Tocharian Languages

The name Tocharian refers to two Indo-European languages attested in manuscripts discovered during the late nineteenth and early twentieth century along the northern Silk Road in the Tarim Basin.[1] The manuscripts are dated to the period between the fifth and twelfth centuries AD,[2] and are written in two distinct but closely related languages, commonly referred to as Tocharian A (or East Tocharian) and Tocharian B (or West Tocharian). The vast majority of the texts contains various genres of Buddhist literature, with the exception of a relatively small number of secular texts such as administrative records.

1.1.1 Language Variation

While the attested Tocharian A (TA) texts are linguistically relatively homogeneous, Tocharian B (TB) texts exhibit a significant amount of variation. Winter (1955) determines that the

[1] A very detailed account of the various expeditions that led to the discovery of Tocharian, as well as of other previously unknown languages, can be found in Fellner (2007).

[2] For a recent discussion of the age of Tocharian manuscripts based on paleographical evidence, see Malzahn (2007b).

TB texts can be classified into three major linguistic categories (with a number of subcategories), which he explains as being due to dialectal variation. Stumpf (1990), in his detailed study on language variation in TB, reinterprets Winter's findings, arguing that the three different categories represent different diachronic stages of the language. Stumpf's theory is adopted and further refined by Peyrot (2008) and is now the predominant view on language variation in TB among Tocharian scholars. In the present work, I discuss differences between the different varieties of TB only where it is relevant for the analysis.

The following sections contain brief overviews of the Tocharian segmental inventories, TB accent, and Tocharian meter, as these issues are relevant throughout the dissertation. Other properties of Tocharian phonology will be introduced as they become necessary for the discussion.

1.1.2 Tocharian Segmental Inventory

The following tables provide a list of the phonological segments that can be clearly identified for both languages based on the Tocharian writing system, although the two languages differ with respect to the exact status of some of the segments within their respective phonological systems (on which see below). Each segment in the tables is given in IPA notation. In cases where it is not identical, I also provide the rendering of a segment in the traditional transcription system for Tocharian, which I will use to render all Tocharian example forms in this work.[3]

[3]For a more detailed overview of the consonant and vowel inventory of the Tocharian languages see e.g. Pinault (2008: 415-20) and Ringe (1996: xix-xxv). The table in (T:1.1) does not include segments that exclusively occur in Sanskrit loanwords.

Table 1.1: Consonant inventory

	Labial	Dental	Alveo-Pal.	Palatal	Velar	Labio-Velar
Stops	p	t			k	k^w (kw/k_u/$_u$k)
Affricates		ʦ (ts)		cȼ (c)		
Fricatives		s	ʃ (ṣ)	ȼ (ś)		
Nasals	m	n		ɲ (ñ)	ŋ (ṅ)	
Liquids		l r		ʎ (ly/ll)		
Glides				j (y)		w

Table 1.2: Vowel inventory

	Front	Central	Back/Round
High	i (i, ī)	ɨ (ä)	u (u, ū)
Mid	e	ə (a)	o
Low		a (ā)	

In addition to the vowels in (T:1.2), TB has a set of falling diphthongs involving a non-high vowel as first member followed by *i* or *u*. One major difference between the vowel systems of TA and TB concerns the central vowels. In TB *ä*, *a* and *ā* are realizations of the two underlying segments /ə/ and /a/ (see the next section for details). In TA, the segments *a* and *ā* are always contrastive, while *ä* is an epenthetic vowel, whose placement is fully predictable. Accordingly, I take TA *ä* as not being present in the underlying representation of words, following Jasanoff (1987), Kim (2007, 2012), and Koller (2008).

1.1.3 Accent in TB

While we know very little about the accent system of TA, TB is more informative in this regard. Starting with classical TB,[4] it is possible to identify the location of the accent of a

[4]In archaic texts, the alternations in question are not yet present or are at least not orthographically realized in any consistent way. For details on this issue see Krause (1952: 1-4) and more recently Stumpf (1990: 76-78) and Peyrot (2008: 33-41).

subset of words based on a specific type of vowel alternation. The two underlying segments /a/ and /ə/ receive different realizations depending on whether or not they are accented.

Table 1.3: TB central vowels

Underlying Segment	Unaccented	Accented
/ə/	[ɨ] *ä*	[ə] *a*
/a/	[ə] *a*	[a] *ā*

The difference between *ä*, *a*, and *ā* is reflected in the brāhmī alphabet through the use of different characters and vowel diacritics. The vowel *ä* is always unaccented, *ā* is always accented, and the accentual status of *a* has to be determined based on the vowel it alternates with. The specific grouping of the segments in (T:1.3) goes back to the observation by Krause (1952: 10-16) that *ā* and *a* in the first syllable of a disyllabic word alternate with *a* and *ä*, respectively, when the same stem occurs in a word of more than two syllables due to suffixation.

Krause specifically describes this pattern as involving the weakening of the initial vowel in words of more than two syllables. This interpretation of the data is challenged by Marggraf (1970). Marggraf argues that the stems that show the kind of vowel alternation observed by Krause are underlyingly accented on the second (i.e. final) syllable. He then establishes the rule that where the lexical accent on a morpheme would end up on the final syllable, it is retracted onto the preceding syllable instead (cf. Marggraf 1970: 16 and also Winter 1990: 372). As a complicating factor, a number of morphemes end in an underlying sequence of /Cə/. In TB a final /ə/ generally does not surface.[5] Yet, an underlying word-final /ə/ counts as an additional syllable peak with respect to the structural description of Marggraf's rule, resulting in surface forms with accent on the final syllable.

[5]More precisely, the final /ə/ is deleted in prose but can be realized as *ä* or *o* in metrical contexts (for details see Malzahn 2012a).

Table 1.4: TB accent placement

	1) retraction	2) in-situ	Underlying stem	Translation
a)	*k̲ante*	*k̲änte-nma-sa*.PER.PL	/kənté/ + /nma/ + /sa/	'hundred'
b)	*m̲āk̲a*	*m̲ak̲ā-ntso*.GEN.PL	/maká/ + /ntsə/	'much/many'
c)	*m̲āk̲a*	*m̲ak̲ā-ṃts*.GEN.PL	/maká/ + /ntsə/	'much/many'
d)	*eneṅk̲a*	*eneṅk̲ā-ññeṃ*.ADJ	/enenká/ + /ññen/	'inside'

In column 1) of (T:1.4) the lexical accent has been retracted from its underlying position in order to avoid accent on the word-final syllable. In column 2) the same stem is followed by an unaccented suffix and therefore the accent can remain in its original position. Note that the forms in row b) and c) involve the same underlying representations and the only difference is that the final /ə/ in the genitive plural morpheme /ntsə/ is realized as *o* in b) and as Ø in c), yet both forms fail to undergo accent retraction.

A clear advantage of Marggraf's account over Krause's is that it correctly captures the rare cases in which we find vowel alternations that do not involve the first two syllables (as in T:1.4d). According to Marggraf's model, we simply need to assume that *eneṅka* bears underlying accent on the final syllable, which must be retracted if the stem is realized in isolation. Krause's description, according to which words are either accented on the first or the second syllable, has nothing to say about these cases. Therefore, for the present study I adopt Marggraf's system for TB accent.

1.1.4 Basic Properties of Tocharian Meter

Many of the Buddhist texts attested in Tocharian are entirely metrical, such as the Buddhastotras,[6] and in addition a large number of the prose texts are interspersed with metrical passages. In what follows, I provide only a very brief overview of those properties of Tocharian meter that are relevant to the present discussion. Metrical texts in both Tocharian

[6]Buddhastotras are a genre of praise poetry to the Buddha.

languages are subject to the same basic principles. The largest metrical unit is the stanza, each with a fixed number of verses (usually four). Each verse contains a predetermined number of cola separated by caesurae, i.e. obligatory word-boundaries, and the cola themselves are made up of a fixed number of syllables. Tocharian meter is therefore primarily syllable-counting and there is no evidence for any sensitivity to syllable-quantity. As an illustration of this organizing principle, I give an example of one very common type of stanza. It contains four verses, each of which is made up of three cola with five, four and three syllables, respectively.[7]

(1.1) *Strophe with 4 x 12 syllables in TA (A 1 b6)*

	5 σ		4 σ		3 σ
verse a	māski kätkāläṃ	;	ktäṅkeñc tsraṣiñ	,	sāmuddrä
verse b	traidhātuk saṃs(ār	;	tsra)ṣṣuneyo	,	ktäṅkeñc kraṃś
verse c	kälpnāntär toriṃ	;	puttiśparäṃ	,	wärṣṣältse
verse d	mā =pärmāt tsru yärm	;	yātal ya(tsi)	,	tsraṣṣune

In (1.1), all four verses have the same length and structure, which is not the case for all stanza types. What is crucial, however, is that for each individual verse, the number and length of the cola is dictated by the type of stanza. The very strict nature of this system makes it straightforward to identify cases in which a meter has been violated by the poet.

Another important aspect of Tocharian meter is the distinction between major and minor cola, first introduced by von Gabain and Winter (1958: 34) and Winter (1959). In (1.1), each verse contains two major cola, separated by a semicolon, while the second major colon can be further broken down into two minor cola, separated by a comma. The main criterion for distinguishing between these two levels of metrical constituents is caesura violability. Caesurae between minor cola are violated more frequently than those between major cola.

[7]The Tocharian scribes generally marked verse boundaries with a special symbol in the brāhmī alphabet. They did not, however, explicitly indicate caesurae. The caesura markers ; and , are a property of the transcription only.

However, this distinction does not have an impact on the argumentation within the present study and I use the term *colon* simply to refer to any metrical unit that can only be further divided into syllables.

1.2 The Data

1.2.1 CEToM

Almost all of the data used for the present study is taken from the database "A Comprehensive Edition of Tocharian Manuscripts" (CEToM). CEToM is a project undertaken at the University of Vienna under the direction of Melanie Malzahn. Its goal is to provide editions of all available Tocharian manuscript leaves, including transliterations, transcriptions, and metrical analyses, as well as both linguistic and philological commentaries. The project further aims at the compilation of a complete electronic dictionary for both Tocharian languages. At the time of the writing of this dissertation, the project is still a work in progress and at this stage contains only a subset of published TB manuscripts. However, all major published TA texts are already included in the database in transcribed (albeit not necessarily translated) form. Most of the generalizations about the data made in this dissertation are based on electronic searches of the currently[8] available CEToM corpus.

1.2.2 Presentation of Examples

The studies within this dissertation use two different types of examples. The first type involves individual words or short phrases that are embedded in the main text of the dissertation or listed in tables, e.g. the TA form *kāsu* 'good'. In appendix A I provide a word index that contains each attested form in the text with up to four manuscript leaves where it

[8]Most of the used data is based on a copy of the database retrieved on May 21, 2015.

is attested.[9] All examples in the index are listed in the form in which they are transcribed in CEToM (with minor differences in word segmentation, on which see below). In cases where I cite a restoration from the secondary literature that is currently not used in CEToM, a footnote indicates how the form in question is listed in the index.

The second type are glossed sentences, which are set off from my main text. Both the transcriptions and the translations of these examples are taken from CEToM, unless otherwise indicated, and this includes restorations in parentheses. For the most part I am following the transcription practices used by CEToM, the only major exception being that I insert a space before certain particles where the traditional transcription practice does not, such as the TA relative particle *ne*. Following the translation, each example further contains the siglum of the manuscript leaf containing it, as well as the line of the leaf on which the example starts, e.g. a sentence spanning lines a3-a5 is simply labelled as a3. Examples that are identifiable as metrical are marked accordingly, including the specific meter if it can be determined. The interlinear glosses are mostly based on lexical and morphological information taken from CEToM, Adams (1999), and Carling et al. (2009), as well as the word index in Krause and Thomas (1964). Regarding language variation, most TB example sentences are taken from classical texts. In cases where a sentence has been taken from an archaic or late text, the manuscript siglum is marked with $^{\text{A}}$ and $^{\text{L}}$, respectively. Both the exact format of the given sigla, as well as the identification of a given leave as archaic or late are taken directly from CEToM.

1.3 Goals of this Dissertation

The present dissertation is a collection of studies on phonological phenomena in Tocharian involving units larger than the word. A major focus of this work therefore falls on two

[9] Hypothetical/reconstructed forms discussed in the text (and marked with an asterisk) are not included in the word index.

aspects of Tocharian phonology: a) external sandhi and b) the phonological interaction between clitics and their hosts. The latter, especially, is an area of Tocharian phonology that has so far only been studied sporadically and unsystematically. One of the main goals of this work is to establish a concrete analysis of the constituency between clitics and their hosts using findings from the field of Prosodic Phonology, which holds that different types of phonological constituents can be subject to different types of phonological phenomena. I show that by identifying and by distinguishing between different types of phonological constituents, we gain new insights into problems of Tocharian phonology that have so far escaped satisfactory analyses.

1.4 Overview of the Dissertation

Chapter 2 presents a two-part study of TA external sandhi, beginning with a detailed discussion of its segmental properties. I am taking as my starting point the study on Tocharian sandhi in Stumpf (1971b) and I provide a reanalysis of the available TA data couched in Optimality Theory. The second part of the chapter is both philological and linguistic in nature and deals with the way external sandhi is reflected in the use of the brāhmī alphabet. I show that the use of pre-vocalic virāma spelling in TA directly corresponds to the absence of external sandhi. Based on this observation, and in conjunction with the analysis of TA sandhi proposed in the first part of the chapter, I argue that pre-vocalic virāma spelling has a phonetic correlate in TA, i.e. the insertion of a glottal stop.

Chapter 3 explores the phonological relationship between clitics and their hosts within the framework of Prosodic Phonology. The first part of the chapter focuses on TA. Using the findings from the chapter on TA sandhi, as well as evidence from meter, phonotactic constraints, and the phonology of compounds, I argue that clitics and their hosts form a prosodic constituent that is distinct from both the word and the phrase. The second part of the chapter addresses clitics in TB. I show that the same kind of analysis that I employ

for TA clitics is also applicable to TB, even though the available phonological evidence is different. Unlike TA, TB provides insight into the accentual behavior of clitics, whereas external sandhi in TB is of little use in establishing their prosodic constituency.

Chapter 4 consists of two studies of the internal structure of Tocharian Wh-words. It has long been observed that most Tocharian Wh-words can be decomposed into an interrogative and a pronominal element, at least diachronically. In the first part of the chapter, I argue that by treating the interrogative part of Wh-words as clitics, following the specific prosodic analysis proposed in chapter 3, we can solve a long-standing problem in both the synchronic and diachronic phonology of TB, namely the apparently exceptional accentuation of the Wh-words *mäkté* 'how' and *kᵤsé* 'who, what'. The second part of the chapter provides a new analysis of TA indefinites. Most handbooks on Tocharian grammar treat the attested TA indefinites as forming a single paradigm, part of which is identical to that of the Wh-word *kus* 'who, what', while the remaining part seems to overlap with the demonstrative *saṃ*. I argue that this analysis is incorrect and that the indefinite forms of *kus* should simply be viewed as Wh-words doubling as indefinites, thereby following the original analysis by Sieg et al. (1931), which predates the standard view. For the remaining indefinites, I propose an analysis according to which they are part of a different Wh-paradigm. Their homophony with demonstratives simply arises due to a phonological process that deletes the interrogative particle.

Chapter 5 provides an overview of the results of the present study.

Chapter 2

On Tocharian A Sandhi and its Rendering in the Brāhmī Alphabet

2.1 Introduction

The aim of the present chapter is to provide a unified analysis of TA external sandhi[1] targeting word-final segments that precede word-initial vowels, henceforth X#V-sequences. Taking as my starting point Peter Stumpf's thorough study of vocalic sandhi in Tocharian (Stumpf 1971b), I demonstrate that all of these processes conspire to satisfy a single phonological constraint, namely a ban on onsetless syllables. I further show that the same constraint is responsible for a sandhi rule not discussed by Stumpf, namely gemination of word-final consonants. The second part of the chapter focuses on TA orthography, demonstrating how the same phonological constraint can help us understand the use of virāma spellings within the brāhmī alphabet.

[1] In what follows I will follow the common practice in linguistics of using the terms *sandhi* and *external sandhi* interchangeably.

2.2 External Sandhi in TA

I begin by giving a summary of Stumpf's generalizations on V#V-sandhi in TA, illustrated with examples.[2]

2.2.1 Stumpf on V#V-Sequences

Stumpf (1971b: 118) makes the following generalization regarding V_1#V_2-sequences.[3] V_1 undergoes one of two changes. It is deleted if it is [+back -round][4] or turned into a glide otherwise. The following tables provide examples for each of the two changes.

Table 2.1: V_1-Deletion

a)	*śla* 'with'	+ *oko* 'fruit'	→	*śl= oko*
b)	*tkanā* 'on the earth'	+ *ākāśā* 'in the air'	→	*tkan= ākāśā*

Example b) illustrates that when V_1 and V_2 are identical, sandhi yields a single instance of that vowel. By analogy to a) and similar cases in TB, Stumpf treats this as V_1-deletion.[5]

[2]While Stumpf covered both Tocharian languages in his study, I only discuss TB where it may improve our understanding of the TA data.

[3]Stumpf's study includes exhaustive lists of all V_1#V_2 and R#V-sequences undergoing sandhi attested within the Berlin collection. The only sizeable addition to the TA corpus since then was the Maitreyasamiti-Nāṭaka manuscript from Yanqi (YQ) published in Ji et al. (1998). The present study takes into account both Stumpf's collection of data as well as the YQ-manuscript.

[4]Stumpf uses the terms *palatal* (for *e* and *i*) and *labial* (for *u* and *o*). Throughout this dissertation I consider both central and back vowels to carry the feature [+back].

[5]Below I treat these cases as an instance of vowel coalescence instead.

Table 2.2: Glide Formation

a) *praski* 'fear'	+ *arä̱s* 'evoke'	→	*prasky arä̱s*
b) *eṃtsu* 'seized'	+ *ā̱sträṃ* 'clean'	→	*eṃtsw ā̱sträṃ*
c) *sne* 'without'	+ *āk* 'end'	→	*sny āk*
d) *malto* 'first'	+ *āksisam* 'I proclaim'	→	*maltw āksisam*

Stumpf further allows for three types of exceptions to his generalization. First, when the first word in the sequence is the negation *mā*, V-deletion targets V_2 instead of V_1.

Table 2.3: V_2-Deletion after *mā*

a) *mā* 'not'	+ *appärmāt* 'despised'	→	*mā =pärmāt*
b) *mā* 'not'	+ *aryu* 'long'	→	*mā =ryu*

Second, V_2-deletion also takes place when the second element in the V_1#V_2-sequence is an enclitic.

Table 2.4: V_2-Deletion in enclitic

a)	*tuṅkyo* 'love'	+ *aśśi* ptcl.	→	*tuṅkyo =śśi*
b)	*tā* 'where'	+ *aśśi* ptcl.	→	*tā =śśi*

Third, the rule reducing two identical vowels to a single one takes precedence over glide formation in cases of two round or front vowels. As in (T:2.1b) Stumpf treats these cases as instances of V_1 deletion.

Table 2.5: V_1-Deletion where V_1=V_2

b)	*wṣeññe* 'resting place'	+ *empele* 'terrible'	→	*wṣeññ= empele*
c)	*amokyo* 'art'	+ *opśi* 'adept'	→	*amoky= opśi*
d)	*āñmaṣi* 'self'	+ *ime* 'idea'	→	*āñmaṣ= ime*
e)	*pe nu* ptcl.	+ *ucchiṣṭ* 'unclean'	→	*pe n= ūcchiṣt*

Finally, there is an additional complication in that word-final *o* in TA before a non-identical vowel has two possible outcomes. While *maltw āksisam* (in T:2.2d) exhibits the expected glide formation, all the remaining cases show V_1-deletion instead.[6]

Table 2.6: Deletion of word-final *o*

a)	*poryo* 'with fire'	+ *āriñc* 'heart'	→	*pory= āriñc*
b)	*ṣokyo* 'truly'	+ *āṣträṃ* 'pure'	→	*ṣoky= āṣträṃ*

What all of the attested cases involving V_1-deletion of *o* have in common is that they either involve a noun ending in the instrumental morpheme *-yo* (4x)[7] or the adverb *ṣokyo* 'truly' (5x). This distribution led Stumpf to the conclusion that the final *o* in these cases was [-round] and therefore not subject to glide formation. As diachronic evidence for the exceptional phonological status of TA *-yo*, Stumpf (1971b: 132-3) adduces the potentially cognate particle TB *wai* 'and'. If the two forms are indeed cognate, the argument goes, the irregular corrrespondence TA *o* - TB *ai* (instead of expected *au*) indicates the [-round] status of the TA vowel. There are at least two problems with this hypothesis. First, it is not clear how it would apply to the adverb *ṣokyo* unless we are, in fact, dealing with the instrumental form of an otherwise unattested noun. A more serious problem is that this analysis requires us to posit an underlying segment that would only be attested in a single morpheme.[8] Stumpf also provides an alternative hypothesis based on a suggestion by Werner Thomas, according to which *o* lost its rounding due to assimilation to the preceding glide.

[6] A potential second case involving glide formation of *o* is *św empele* (A 51 a6). The authors of the CEToM edition (Gerd Carling, Georges-Jean Pinault, and Melanie Malzahn) derive *św* from *śo* while Stumpf derives it from *śu*.

[7] 3 attestations listed by Stumpf + 1 from the YQ-manuscript.

[8] One way to avoid this issue is to treat the behavior of *-yo* in external sandhi as a morphological idiosyncrasy. Within Stumpf's model, such an analysis is independently required for the negation *mā*, which triggers exceptional V_1-deletion.

While there is no independent evicence for such a phonological rule, it avoids the issues raised by the etymological approach.

In the next section, I provide an alternative account for external sandhi affecting V#V-sequences, including the behavior of word-final *o*.

2.2.2 An OT-Analysis

In this section I propose a unified analysis for a number of TA sandhi effects couched in Optimality Theory (Prince and Smolensky [1993] 2004). Within OT the output of an underlying phonological representation is selected from a candidate set based on violable constraints. OT constraints come in two types, markedness constraints and faithfulness constraints. Markedness constraints penalize output representations that exhibit certain phonological properties, while faithfulness constraints require the output form to deviate from the underlying representation as little as possible. Which constraints can be violated by an output form depends on the specific constraint ranking in a given language. One of the main advantages of this model is that it can straightforwardly capture conspiracies, in which a number of distinct phonological processes appear to be directed towards a common goal, i.e. to avoid outputs with some specific marked phonological property.[9] The OT framework is therefore particularly well-suited to capture TA sandhi in X#V-sequences, as all of the individual sandhi phenomena can be shown to be directed towards the same goal, i.e. to supply a vowel-initial syllable with an onset.

Before introducing the analysis, I must first address a complication regarding the data. Due to the relatively small size of the TA corpus, only a subset of all possible vowel combinations is attested as undergoing sandhi. The following represents a list of all types of

[9]The notion of a phonological conspiracy goes back to Kisseberth (1970).

attested $V_1\#V_2$-pairs in TA affected by sandhi.[10,11]

Table 2.7: Sandhi outcomes in $V_1\#V_2$-pairs

$V_2\rightarrow$		a	ā	e	o	i	u
$V_1\downarrow$	a	-	ā	-	o	-	-
	ā	ā	ā	-	-	-	-
	e	ya, e	yā, (ā)	e	yo, (o)	-	-
	o	o	wā, ā	we	o	-	-
	i	ya, (a)	yā, (ā)	ye, (e)	yo	i[12]	yu
	u	-	wā	-	wo	-	u

Any concrete analysis of TA sandhi will necessarily make some predictions about the content of some of the empty cells in (T:2.7)[13] based on the behavior of the attested cases and typological considerations, as well as comparison with TB. Therefore the present analysis must necessarily have a preliminary status and may well be falsified as additional TA manuscripts become available.

In what follows I argue that all sandhi effects in TA involving X#V-sequences are due to the constraint Onset, a markedness constraint requiring every syllable to have an onset (Prince and Smolensky [1993] 2004: 20). Within the constraint ranking that results in external sandhi,[14] Onset is undominated and can therefore never be violated. Thus each

[10]A similar overview encompassing both Tocharian languages can be found in Krause and Thomas (1960: 72-73).

[11]I excluded cases that Stumpf plausibly argued to be due to scribal errors. The outcomes in parentheses are argued by Stumpf to be mere orthographic variants. I return to those cases below.

[12]Two out of three attestations of sandhi involving i#i are spelled with <ī>, while the only attestation involving u#u is spelled with <ū>. As there is no independent evidence for contrastive vowel length in Tocharian this is most likely a purely orthographic phenomenon. It is possible that the scribes were orthographically influenced by Sanskrit, where two identical high vowels do yield a long vowel in sandhi.

[13]It is also possible in principle that some of the gaps are not accidental and the respective vowel combinations simply did not undergo sandhi.

[14]Since all sandhi effects have to be considered optional to some degree, we have to assume that X#V-

different kind of X#V-sandhi effect can be interpreted as a form of resyllabification[15] in order to avoid the realization of a word-initial vowel without an onset.

A high ranking of ONSET can be independently motivated by examining the behavior of VV-sequences word-internally. Word-internal hiatus is mostly prohibited[16] in Tocharian A. Wherever hiatus would arise through affixation, we get various phonological and morpho-phonological alternations conspiring to produce surface forms without hiatus.

The examples in (T:2.8) illustrate a word-internal version of vowel deletion. In the first row of each table, stem-final *ā* disappears whenever it is followed by a vowel-initial suffix. The relevant segments are underlined.

Table 2.8: Vowel deletion

Preterit Middle of *wärp^{ā}-* 'receive'			Preterit Singular Active of *läm$^{(ā)}$-* 'sit'		
/wrpā+e/	→	*wärpe* 1sg.	/lymā+ā/	→	*lymā* 1sg.
/wrpā+te/	→	*wärpāte* 2sg.	/lymā+ṣt/	→	*lymāṣt* 2sg.
/wrpā+t/	→	*wärpāt* 3sg.	/lya+lymā+n/	→	*lyalymāṃ* 3sg.caus-obj.

The tables in (T:2.9) show two instances of consonant epenthesis. The perlative suffix *ā* is attached to two different nominal stems ending in a vowel, triggering the epenthesis of a glide in one case and of a velar stop in the other.

sequences without sandhi are due to a different constraint ranking. I address this issue in more detail in the second part of this chapter.

[15]Resyllabification along the lines discussed here was already assumed by Stumpf himself. Specifically, Stumpf uses the term "Sandhisilbe" (p. 121) for a syllable spanning a word boundary due to sandhi.

[16]Bernhard (1958) lists a number of compounds involving hiatus at the compound boundary. While some of these cases could also be analyzed as phrases, a small number of examples appear to be genuine instances of compound-internal hiatus. I return to these in chapter 3.2.3.2.

Table 2.9: Consonant epenthesis

knānmune 'knowledge'			*lwā* 'animals'		
/knānmune+ṣy/	→	*knānmuneṣi* adj.	/lwā+śśy/	→	*lwāśśi* gen.pl.
/knānmune+ā/	→	*knānmuneyā* per.sg.	/lwā+ā/	→	*lwākā* per.sg.

Finally, as in external sandhi, high vowels alternate with glides, depending on whether or not they are followed by a consonantal segment.

Table 2.10: High vowel alternations

ārkiśoṣi 'world'			*riṣaki* 'sage'		
/ārkiśoṣy/	→	*ārkiśoṣi* nom/obl.sg.	/ryṣaky/	→	*riṣaki* nom/obl.sg.
/ārkiśoṣy+ an/	→	*ārkiśoṣy-aṃ* loc.sg.	/ryṣaky+āp /	→	*riṣaky-āp* gen.sg.

Even with omission of the phonological or morphophonological details, these examples illustrate that word-internal onsetless syllables are avoided in Tocharian A and that cases arising through morpheme concatenation undergo some form of repair.

Having established the importance of Onset for TA phonology, we now turn to the next question from an OT-viewpoint, namely which constraints can or cannot be violated in order to avoid onsetless syllables.

2.2.2.1 Glide Formation and Vowel Coalescence

I adopt the constraint set and ranking used by Baković (2007) to analyze vocalic sandhi in Chicano Spanish (CS). The vocalic sandhi effects in CS strongly resemble those of TA, both with regard to front and round vowels, as well as sequences of identical vowels. Considering the small size of the TA dataset, it is useful to illustrate how the available examples match up with similar cases from a living language, in which each individual phenomenon can be securely backed up with additional data. Therefore, I am giving the relevant CS examples

from Baković (2007: 2-3).[17] The comparison with a living language further serves to illustrate that, even though vocalic sandhi mainly occurs in metrical contexts within the Tocharian text corpus, we are dealing with a linguistically natural and typologically unremarkable phenomenon.

Just like in TA, a V#V-sequence involving two identical vowels in CS yields a single instance of that vowel.

Table 2.11: Coalescence

Chicano Spanish			Tocharian		
lo odio	→	[loðjo]	**prutko oki*	→	*prutkoki*
mi hijo	→	[mixo]	*āñmaṣi īme*	→	*āñmaṣime*
era así	→	[erasi]	**ywārckā āriñcaṃ*	→	*ywārckāriñcaṃ*
se escapó	→	[sescapó]	**wṣeññe empele*	→	*wṣeññ= empele*
tu uniforme	→	[tuniforme]	**penu ucchiṣt*	→	*penūcchiṣt*

Both languages turn mid and high vowels into high glides when followed by a non-identical vowel.

Table 2.12: Glide Formation: high vowels

Chicano Spanish			Tocharian		
mi ultima	→	[mjultima]	**lwāśi upage*	→	*lwāśyupage*
mi hebra	→	[mjeβra]	**ṣñi entsyo*	→	*ṣñyentsyo*
mi obra	→	[mjoβra]	**ñäkci opleṃ*	→	*ñäkcyopleṃ*
mi árbol	→	[mjaɾβol]	**tri āpāyṣiṃ*	→	*tryāpāyṣiṃ*
tu hijo	→	[twixo]	–	→	–
tu epoca	→	[twepoka]	TB **nauṣu ecce*	→	TB *nauṣwecce*
su Homero	→	[swomeɾo]	*lu oki*	→	*lwoki*
tu alma	→	[twalma]	**wät nu ālak*	→	*wät nwālak*

[17] In order to save space I am leaving out the translations of the examples. All that matters for our current purposes are the vowel alternations. For illustration purposes, I list input forms to sandhi phenomena even if the individual words are not attested in isolation. All the output forms are attested and are listed in appendix A.

Table 2.13: Glide formation: mid vowels

Chicano Spanish			Tocharian		
me urge	→	[mjurxe]	–		
pague ocho	→	[paɣjoʧso]	**lāntune opyāc*	→	*lāntunyopyāc*
porque aveces	→	[porkjaβeses]	*sne āk*	→	*snyāk*
tengo hipo	→	[teŋgwipo]	TB **po ike*	→	TB *pwīke*
como Eva	→	[komweβa]	TB **oku empelye*	→	TB *okwempelye*
lo habla	→	[lwaβla]	**maltu āksisam*	→	*maltwāksisam*

If no TA example is available, I provide a TB example from Stumpf's collection wherever possible. The TB examples, in conjunction with the close parallels between TA and CS, indicate that the gaps in attestation are likely accidental.

I now review the OT-constraints and phonological assumptions employed by Baković (2007) and demonstrate their application to the TA data.

Glide Formation

First, I assume that glide formation in TA results in rising diphthongs, as it does in CS, i.e. the glide and the following vowel occupy a single mora. This analysis receives support from cases where glide formation results in initial clusters that appear to violate TA phonotactics.

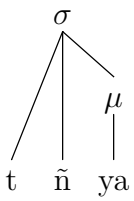
Table 2.14: Glide formation after C-clusters

pñi	+ *eluneyo*	→	*pñyeluneyo*	'with giving of merit'
tri	+ *āsrapäntu*	→	*tryāsrapäntu*	'three evils'
ytsi	+ *onu*	→	*ytsyonu*	'begin to go'
wtsi	+ *oky*	→	*wtsyoky*	'like a parasol'
tñi	+ *āriñc*	→	*tñyāriñc*	'your heart'

TA allows a maximum of two consonants in word-initial position, or three if the first

consonant is a sibilant.[18] All apparent exceptions to this rule involve cases of glide formation. If we assume, however, that the glide is located in the same mora as the following vowel, all of the examples in (T:2.14) can be assimilated to the cases with two initial consonants.

(2.1) *The initial syllable of* tñy āriñc



I therefore adopt Baković's specific version of the constraint ONSET, which treats glides preceding a vowel within the same mora as possible syllable onsets.[19] The observation that mid vowels turn into high glides can be captured by a markedness constraint HIG (Casali 1997: 516), which penalizes glides that are [-high]. This constraint can be considered undominated in TA, since non-high glides do not occur. Through the formation of rising diphthongs, glide formation violates the markedness constraint ONE-TO-ONE,[20] which requires a one-to-one correspondence between moras and segmental melody.

Vowel Coalescence

I further follow Baković in treating the examples involving identical vowels (as in T:2.11)

[18]For details on the special status of sC-clusters crosslinguistically see e.g. Kaye (1992).

[19]ONSET (Definition from Baković 2007: 6).

> Let σ be some syllable dominating some mora μ, which in turn dominates some segmental melody ξ. Assign a violation if μ is not preceded by another mora μ and μ is not preceded by another segmental melody ξ.

[20]ONE-TO-ONE (Definition from Baković 2007: 4).

> Let μ be some mora and ξ some segmental melody such that μ dominates ξ. Assign a violation if either (i) μ also dominates some other segmental melody ξ or (ii) ξ is also dominated by some other mora μ.

as vowel coalescence, *pace* Stumpf, who analyzes them as cases of deletion.[21] The constraint violated by vowel coalescence is UNIFORMITY,[22] which requires every input segment to correspond to a separate segment in the output. The markedness constraint NO-LONG[23] penalizes diphthongs with identical segmental content (such as *yi*, *wu*, *iy* and *uw*), as well as long vowels and cases of hiatus involving identical vowels. NO-LONG captures the observation that identical high vowels undergo coalescence instead of glide formation. Since TA does not exhibit sequences such as y#i or w#u, NO-LONG can be considered undominated as well.[24]

Finally, the analysis operates with two common types of faithfulness constraints. The first type is MAX-V,[25] which penalizes vowel deletion. Any vowel that is present in the input but has no correspondent in the output causes a violation. The second type is a family of constraints called IDENT. A constraint IDENT(F) assigns a violation to a candidate for each segment whose value for F differs from that of its counterpart in the input.

Let us now consider how the constraint set laid out above accounts for the difference between V#V-sequences undergoing glide formation and those undergoing coalescence. I begin with an example of coalescence involving two high vowels.

[21]Stumpf (1971b: 100 fn. 7) specifically argues that there is no reason to distinguish between deletion and coalescence. The purpose of this distinction in the present analysis will become clear below.

[22]For a more technical definition, see McCarthy and Prince (1995: 123).

[23]NO-LONG (Definition from Baković 2007: 5).

Let ζ_1 and ζ_2 be two adjacent segmental melodies or moras. Assign a violation if $\zeta_1 = \zeta_2$ (i.e., if ζ_1 and ζ_2 are featurally identical segmental melodies, or if ζ_1 and ζ_2 are moras that dominate the same segmental melody).

[24]It should be noted, however, that both *yi* and *wu* are attested word-internally. I am making the standard assumption that forms undergoing external sandhi have already been syllabified on the word-level (see e.g. Nespor and Vogel 1986 [2007]: 68-72 on the distinction between syllabification on the word-level and resyllabification above the word) and that the constraint rankings on the lexical and the postlexical level are not necessarily identical (Kiparsky 2000).

[25]See McCarthy and Prince (1995: 16) for details on MAX constraints.

(2.2) ONSET, NO-LONG, MAX-V >> UNIF

/āñmaṣi$_1$ i$_2$me/	ONSET	NO-LONG	MAX-V	UNIF
☞ a. *āñmaṣi$_{12}$me*				*
b. āñmaṣi$_1$ me			*!	
c. āñmaṣyime		*!		
d. āñmaṣiime	*!	*!		

The winning candidate is selected by ranking UNIFORMITY below ONSET, NO-LONG and MAX-V. The ranking thereby renders coalescence less offensive than hiatus, glide formation or vowel deletion. Compare this with a case of glide formation.

(2.3) ONSET, NO-LONG, MAX-V >> IDENT(HI), ONE-TO-ONE >> UNIF

/malto$_1$ ā$_2$ksisam/	ONSET	IDENT(LO)	MAX(V)	IDENT(HI)	1-TO-1	UNIF
☞ a. *maltwāksisam*				*	*	
b. malto$_1$ksisam			*!			
c. maltā$_{12}$ksisam		*!				*
d. malto$_{12}$ksisam		*!				*
e. maltoāksisam	*!					

ONE-TO-ONE, the constraint violated by glide formation, has to be ranked below MAX(V) in order to rule out vowel deletion. The only difference between glide formation of high vowels and mid vowels is that the latter also violates the faithfulness constraint IDENT(HI) in addition to the markedness constraint ONE-TO-ONE. The candidate involving vowel coalescence is ruled out by an undominated faithfulness constraint IDENT(LO), since the resulting vowel necessarily deviates from one of the two input vowels in the feature [low]. The undominated nature of IDENT(LO), in combination with HIG, further captures the fact that identical low vowels can only undergo coalescence, and not glide formation. IDENT(BK), another undominated faithfulness constraint, ensures that V#V-sequences in which the vowels differ in the feature [back] do not undergo coalescence either, as illustrated by (2.4).

(2.4) IDENT(BK) >> IDENT(HI), ONE-TO-ONE

/lāntune$_1$ o$_2$pyāc/	IDENT(BK)	IDENT(HI)	1-TO-1	UNIF
☞ a. *lāntunyopyāc*		*	*	
b. lāntuno$_{12}$pyāc	*!			*

Finally, cases of V#V with identical mid vowels illustrate why it is necessary to distinguish between vowel deletion and coalescence.

(2.5) *Coalescence wins over deletion*

/triko$_1$ o$_2$ki/	ONSET	MAX(V)	IDENT(HI)	1-TO-1	UNIF
☞ a. *triko$_{12}$ki*					*
b. trikwoki			*!	*!	
c. triko$_1$ki		*!			
d. trikooki	*!				

The example in (2.3) requires a ranking of both IDENT(HI) and ONE-TO-ONE below MAX-V in order to prevent deletion from applying instead of glide formation. Even though the candidate involving coalescence and the candidate involving deletion are homophonous, the form *trikoki* cannot reflect candidate c). This is because the highest ranked constraint violated by c) is ranked higher than the highest ranked constraint violated by the candidate with glide formation, rendering the latter more harmonic. The winning candidate must therefore be the result of coalescence, and it is rendered the optimal candidate by ranking UNIFORMITY below IDENT(HI) and ONE-TO-ONE.

Thus, drawing a distinction between vowel deletion and coalescence accounts for the observation that sequences of identical mid vowels do not undergo glide formation. This is in spite of the fact that the sequences *ye* and *wo* are clearly well-formed as the output of external sandhi, as they are attested as the result of the inputs /i#e/ and /u#o/.

At this point the direct parallels between CS and TA end. Thus far we have not seen a single example that would constitute a bona fide case of vowel deletion in TA under the

current analysis. The next section addresses how such cases can be accounted for with an extended version of the constraint set above.

2.2.2.2 Vowel Deletion

Having adopted a coalescence analysis for sandhi involving identical vowels, we can modify Stumpf's generalization regarding vowel deletion thusly:

(2.6) *Stumpf's generalization (disregarding vowel coalescence)*

In a sequence $V_1\#V_2$, where $V_1 \neq V_2$ and V_1 is [-round +back], V_1 undergoes deletion unless

1. V_1 is an *o* preceded by *y*, which behaves like [-round] vowels.
2. V_2 undergoes deletion instead of V_1 because
   (a) V_2 is part of an enclitic.
   (b) V_1 is part of the negation *mā*.

Even though it is typologically unproblematic for a language to delete V_1 before content words and V_2 with function words,[26] the TA data is not as straightforward as the generalization above suggests. In what follows I propose an alternative analysis that is equally compatible with the attested data but also accounts for the apparently idiosyncratic behavior of the negation *mā*.

The following table summarizes all the attested types of cases of V#V involving vowel deletion.

[26] One such language is Email (Benue-Congo), as discussed in Casali (1997: 511-4).

Table 2.15: Complete list of attested V#V-sequences involving deletion

	V_1	V_2		Output
a)	o	ā	→	ā
b)	a	ā	→	ā
c)	a	o	→	o
d)	ā	a	→	ā
e)	o	a	→	o

The table shows that cases of vowel deletion in TA can be grouped into two types: example a) involves deletion of word-final *o*, while b)-e) show deletion of *a* regardless of whether it is V_1 or V_2.

I begin with the cases in which word-final *o* undergoes deletion instead of glide formation. The following is a complete list of attested examples.[27]

Table 2.16: All attestations of deletion of final *o*

a)	*āsley= āsleyo*	'?'	h)	*ṣoky= ākā(l*	'truly a wish'
b)	*wsāluy= āmpi*	'both with ... clothes'	i)	*ṣoky= ākā ///*	'truly a wish'
c)	*tämy= ānant*	'Therefore, Ānanda ...'	j)	*ṣoky= āṣträṃ*	'truly poor'
d)	*tuṅky= āriñc*	'heart with love'			
e)	*pory= āriñc*	'heart with fire'			
f)	*yatsy= āmpi*	'both with ... skin'			

While Stumpf simply treats *o* in *yo* as [-labial], I propose an alternative solution that is indirectly connected to phonotactics. The forms in c)-j) all end in at least two consonants. This situation is highly reminiscent of a kind of variation we find with the pronominal forms *ṣñi* 'own' and *tñi* 'you' (gen. sg.).[28] In both of these pronouns, the glide can be omitted in

[27]Example a) is the only case from the YQ-manuscript, while all the other attestations are taken from Stumpf's data collection. One apparent example of V_2 deletion is *ṣokyo =kāl* (A 252 b7), which Stumpf convincingly deemed a scribal error. The previous line of the same leaf contains the expected sandhi outcome of the same two words.

[28]See Stumpf (1971b: 109).

sandhi.

Table 2.17: Variation in Glide formation after C-clusters

a) Glide spelled out		*b) Glide omitted*	
ṣñi + āñcäm	→ *ṣñy āñcäm*	*ṣñi + āñcäm*	→ *ṣñ= āñcäm*
tñi + āriñc	→ *tñy āriñc*	*tñi + āṣānik*	→ *tñ= āṣānik*

Stumpf considers *ṣñ=* and *tñ=* to be mere orthographic variants of *ṣñy* and *tñy* with no phonetic significance. He further observes that the *ñi* 'I' (gen. sg.) never surfaces without the glide in a sandhi context. The corpus exhibits the following distribution in sandhi.

Table 2.18: Distribution of word-final glide in sandhi

	ṣñi	*tñi*	*ñi*
with glide	7	3	11
without glide	11	6	0

While the total number of attestations is too small to be truly conclusive, the distribution in (T:2.18), in combination with that in (T:2.16), does lend some support to the hypothesis that glides are optionally dropped word-finally after a consonant cluster. I refer to this hypothetical rule as *glide deletion*. This generalization further allows us to account for a small number of additional forms that show apparently aberrant behavior in sandhi.

Table 2.19: Deletion of *e* in sandhi

a)	*sne* 'without'	+	*āñu* 'break'	→ *sñ= āñu*
b)	*sne* 'without'	+	*ālak* 'another'	→ *sn= ālak*
c)	*kuc ne* 'what'	+	*ākṣiññā* 'announced'	→ *kuc n= ākṣ(i)ññā*
d)	*ekre* 'empty'	+	*āyäntu* 'bones'	→ *ekr= āyäntu*

As in the previous examples, the sandhi forms in (T:2.19) lack the expected glide after a consonant cluster. Stumpf treats *sñ āñu* as an orthographic variant of *sny āñu* (also

attested), but considers *sn= ālak* a possible scribal error. In c) *ne* is an enclitic forming a prosodic unit with the Wh-word *kuc* 'who/what'. Therefore, c) can also be considered as showing omission of an expected glide after a consonant cluster. What is striking about d) is that, unlike in the other cases, the segment preceding the omitted glide is neither a palatal, nor a nasal, showing that glide deletion extends beyond those segment types.

If glide deletion is indeed the correct explanation for the behavior of forms in word-final *yo*, we need to address the two cases in (T:2.16) that do not seem to fit the pattern. First (T:2.16a) *āsley= āsleyo* is generally[29] taken to represent two consecutive instrumental forms of an otherwise unattested noun **āsle*. The meaning of this hypothetical noun, however, is entirely obscure. Even if we do assume that the final *yo* in the sequence *āsleyāsleyo* is the instrumental morpheme, there is an alternative analysis that does not involve external sandhi. The form could simply be an āmreḍita (i.e. reduplication) compound bearing a single inflection, similar in structure to the forms in (T:2.20).

Table 2.20: Āmreḍita compounds

kälyme kälymeyäṣ abl.	'from every direction'
ype ypeyā per.	'to every country'

In that case, the first glide in *āsleyāsleyo* would simply be due to the same morphophonemic rule that inserts a glide between stems ending in front and round vowels and a suffix, e.g. in *ype-y-ā*. A possible objection to this analysis is that the examples in (T:2.20) can also be analyzed as a form of phrasal inflection (which is well-attested in Tocharian), in which case we would not expect a word-level morphophonological process such as glide insertion to operate at the word boundary. There are, however, a number of cases of āmreḍita compounds unambiguously behaving as single words. The following forms are taken from Bernhard (1958: 118).

[29]See Sieg et al. (1931: 132 fn. 1); Stumpf (1971b: 99); Carling et al. (2009: 58).

Table 2.21: Āmreḍita-compounds with *a* at boundary

kropa-krop	'group by group'
kumpa-kump	'pot by pot' → 'in crowds' (Carling et al. 2009: 151)
praṅk(a)-praṅk	'moment by moment'
loka-lok	'far far'
(mā)ka-māk[30]	'very much'
waṣta-wa(ṣt)[31]	'house by house'

Bernhard argues that the forms in (T:2.21) must be true compounds since the first member ends in *a*. The vowel *a* has a special status within TA phonology in that it is banned from word-final position.[32]

Therefore, an analysis of *āsleyāsleyo* as an āmreḍita compound with a single inflection is at least possible. In any event, the form itself, as well as the context in which it appears, are too obscure to provide any evidence for *o*-deletion in external sandhi.

The second case that shows the absence of an expected glide after a single consonant is *wsāluy= āmpi* 'both with ... clothes' (T:2.16b), which is a genuine counterexample. One noteworthy fact about this example is that it occurs in the same verse as one of the cases that does fit the pattern (T:2.16f).

(2.7) *rätraṃ* *wsāluy=* *āmpi* *;* *wsā* *yokāṃ* *yatsy=* *āmpi* *;* *lānt*
red.PL clothes.INS.PL both gold color.OBL.SG skin.INS both king.SG
sew(ā)ññ *āmpi* *:*
son.NOM.PL both
Both sons of the king had both red clothes and golden skin. (A 144 a2, translation mine, 6/6/5)

[30]Index: */// ka māk*

[31]Index: *waṣta wa ///*

[32]The only exception to this generalization is the preposition *śla* 'with', and this is usually attributed to its status as a proclitic. See e.g. Stumpf (1971b: 98) and chapter 3.

Therefore, it is at least conceivable that the scribe was influenced by the second pair of words when spelling the first.

Finally, we have yet to determine whether glide deletion is a phonological process or, as Stumpf suggested for *ṣñ=* and *tñ=*, merely a form of orthographic variation. Based on the admittedly small amount of data available, I adopt an extended version of Stumpf's view, i.e. that all of the glide-zero alternations discussed above are of an orthographic nature with no linguistic significance. This conclusion is, in part, dependent on the analysis of glide formation adopted above. If the resulting glide is, in fact, part of a rising diphthong, we would not expect the number of consonants in the preceding onset to have any effect on its realization. Even if it did, the syllable structure of the various cases undergoing glide deletion is not identical. While in forms such as *ṣñi* and *tñi* the resulting glide would follow a branching onset, *kuc ne* and *ekre* most likely have to be syllabified as *kuc.ne* and *ek.re*[33] with a single consonant in the onset. Most of the forms undergoing glide deletion do, however, share an orthographic property, i.e. that spelling out the full cluster would require the scribe to use a three-member ligature in the brāhmī alphabet. In the next section we will see additional evidence that the Tocharian scribes (optionally) left out consonant characters at word boundaries to avoid complex ligatures.

To conclude, treating the behavior of word-final *yo* in sandhi as due to orthographic glide deletion is not without problems. The only example where the glide resulting from final *o* is actually spelled out also involves a consonant cluster (*maltw āksisam*), while none of the cases of final *yo* are written as <yw> in sandhi. Under the current analysis this has to be treated as an accident, although the optional status of glide deletion is independently motivated by the coexistence of forms such as *ṣñy* and *ṣñ=*. Furthermore, there is at least one counterexample with no straightforward explanation. On the other hand, the current

[33]These syllabifications are based on forms such as *rätraṃ* and *wäknā*. Since schwa in TA cannot occur at the end of a syllable, the syllable boundary has to be located before the sonorant in each case.

proposal does have the advantage over Stumpf's treatment that it can be supported with independent evidence.

All of the remaining cases of vowel deletion attested in TA, which can be neither treated as cases of coalescence nor as subject to glide deletion, involve the vowel *a*, i.e. they are either of the structure a#V or V#a. I begin by discussing the latter type, every example of which involves V_2-deletion.

Examples of the type V#a fall into two types. The first consists of cases in which the second word is an enclitic, which led Stumpf to the conclusion that combinations of host + clitic generally undergo V_2-deletion. However, there are only four attestations of this kind and all of them involve the same enclitic, i.e. the emphatic particle *aśśi*.[34]

(2.8) *All attestations of V_2-deletion in enclitics*

a) *tā =śśi* 'where'
b) *ke =śśi* 'whose'
c) *tuṅkyo =śśi* 'with love'
d) *ārśo =śśi* 'today'

Additionally, from what we have seen so far, we expect final mid vowels to undergo glide formation in sandhi, contrary to what we observe in b)-d). I will return to this issue momentarily.

The second type of V#a consists of cases in which the first word is the negation *mā*. Again, there are only four attestations in TA.

(2.9) *All attestations of V_2-deletion after negation*

a) *mā =pärmāt* 'not despised'
b) *mā =ryu* (3x) 'not long'

[34]Stumpf treats examples such as *wiyoki* (from *wiyo* + *oki*) as cases of V_2-deletion by analogy to cases with *aśśi*. The current analysis simply treats this as vowel coalescence, no different from other examples involving identical vowels.

Within Stumpf's analysis, V_2-deletion in cases such as (2.9) is due to an idiosyncratic property of the negation. Therefore, the cases of V_2-deletion in (2.8) and (2.9) are entirely unrelated.

I submit an alternative proposal that assigns a uniform analysis to both sets of examples, and which capitalizes on the observation that all attested cases of V_2-deletion in TA involve the vowel *a*. We have already seen that *a* has a phonologically weak status in that it generally cannot occur word-finally. There is also diachronic evidence for the weak status of *a* coming from a sound change within the prehistory of TA called Vowel Balance.[35] Vowel Balance is the collective term for two sound changes that caused a reduction of $\bar{a}$ to *a* in the second syllable of a word if the preceding syllable contained a [-high] vowel. If both the preceding and the following syllable contained a [-high] vowel the resulting *a* was further reduced to *ä*.[36]

(2.10) *Vowel Balance*

$*\bar{a} > a \quad /\#C^*[\text{-high}]C^* \quad \_$
$*a > ä \quad /\#C^*[\text{-high}]C^* \quad \_ \quad C^*[\text{-high}]$

In the synchronic grammar of TA, Vowel Balance has come to be reanalyzed as a morphophonological processes (Kim 2007: 12), but the original sound change indicates that *a* had a phonologically weak status, at least relative to $\bar{a}$.

Based on these observations, I submit the hypothesis that in a context involving deletion within the sequence V#a, it is always V_2 that undergoes deletion in TA, regardless of the morphosyntactic status of either of the two words in question. In terms of OT constraints, we can express this hypothesis by distinguishing between two different MAX constraints, one penalizing the deletion of *a* (MAX(V_{weak})) and one penalizing deletion of any other vowel

[35]For details see e.g. Krause and Thomas (1960: 45-7), Winter (1994), and Kim (2007).

[36]The exact details of this second change are not fully agreed upon by scholars. The version presented here is based on Kim (2007).

(MAX(V_{strong})). Ranking MAX(V_{strong}) above MAX(V_{weak}) ensures that deletion always targets *a* over any other vowel.

(2.11) MAX(V_S) >> MAX(V_W)

/tā aśśi/	ONSET	MAX(V_S)	MAX(V_W)
☞ a. *tāśśi*			*
b. taśśi		*!	
c. tāaśśi	*!		

The same ranking that yields *tā =śśi* will now also produce *mā =ryu*, as well as the cases of V_1-deletion involving the preposition *śla* 'with'. This is because the current constraint system is only sensitive to the quality of the deleted vowel, not its position or the type of morpheme it is part of.[37]

(2.12) *Deletion of* V_1

/śla oko/	ONSET	MAX(V_S)	MAX(V_W)
☞ a. *śloko*			*
b. ślako		*!	
c. ślaoko	*!		

The ranking in (2.11) and (2.12) constitutes a unified account for all the attested cases of vowel deletion in TA. It is true that V_2 deletion in external sandhi, at least between two content words, is quite rare crosslinguistically (see Casali 1997).[38] However, cases of vowel-specific V_2 deletion do exist. In Sanskrit, word-initial *a* is deleted after word-final *e* and *o* (Whitney 1896: 47).

[37]In the following tableau I assume that vowel coalescence between *a* and *o* is ruled out by an undominated faithfulness constraint IDENT(ROUND).

[38]Casali proposes a constraint MAX-WI, which penalizes the deletion of a word-initial vowel and he argues that MAX-WI universally outranks the more general constraint MAX(V).

(2.13) *V_2-deletion in Sanskrit*

té 'they' *abruvan* 3pl. 'said' → *tè 'bruvan*

Similarly, in Ancient Greek a word initial ε [ɛ] (sometimes α [a]) can be deleted following a long vowel. The process is relatively rare and mostly attested in dramatic and comedic texts, as well as in a few inscriptions (cf. Lejeune 2005: 320).

(2.14) *V_2-deletion in Ancient Greek*

ἐπεί 'after' ἐδάκρῡσα 1sg 'wept' → ἐπεὶ δάκρῡσα (example from Lejeune 2005: 320)

Let us now return to the cases in which *a* is deleted following *e* and *o*. The fact that these examples do not undergo glide formation instead requires some explanation.

(2.15) *Deletion of* a *after mid vowels*

a) *ke =śśi* 'whose'
b) *tuṅkyo =śśi* 'with love'
c) *ārśo =śśi* 'today'

Note that Stumpf's hypothesis that *o* in *yo* is [-labial] does predict b) but it does not account for the absence of glide formation in a) and c). I argue that the cases in (2.15) need to be treated as instances of variation. This view receives support from the following examples.

(2.16) *Glide formation of mid vowels before* a

a) *ke* 'whose' + *artār* 'you love' → *ky artār*
b) *ākläṣlye* 'students' + *oki* 'like' → *ākläṣly oki*
c) **kuc kälyme* 'what direction' + *aśśi* ptcl. → *(kuc käly)my aśś(i)*

Example a) shows that *e* does undergo glide formation before *a*. We cannot attribute the difference between *ke =śśi* and *ky artār* to the enclitic status of *aśśi*, as indicated by

b), where *e* does undergo glide formation before another enclitic, namely *oki*.[39] Example c) is highly fragmentary but if the restoration by Ji et al. (1998) is correct, it would form a near-minimal pair with (2.15a), in that *e* undergoes glide formation before the same enclitic. I therefore conclude that in the sequences e#V and o#V both glide formation and vowel deletion are available to avoid a violation of ONSET. We can capture this variation by allowing both relative rankings of the faithfulness constraints IDENT(HI) and MAX(V_W).

(2.17) IDENT(HI) >> MAX(V_W)

/ke aśśi/	ONSET	MAX(V_S)	IDENT(HI)	MAX(V_W)
☞ a. *keśśi*				*
b. kyaśśi			*!	
c. kaśśi		*!		
d. keaśśi	*!			

(2.18) MAX(V_W) >> IDENT(HI)

/ke artār/	ONSET	MAX(V_S)	MAX(V_W)	IDENT(HI)
☞ a. *kyartār*				*
b. kertār			*!	
c. kartār		*!		
d. keartār	*!			

To sum up so far, the current proposal regarding V_2-deletion is necessarily of a tentative nature due to the small amount of available data. It covers the same set of TA data as Stumpf's analysis but does so without having to assume a special prosodic status for the negation *mā*. Unlike Stumpf's proposal, the current analysis predicts that, within a V-deletion context, a sequence V#a will always undergo V_2-deletion. Unfortunately, there is

[39]Note that the glide in *ākläṣly oki* is not actually spelled, presumably due to orthographic glide deletion (see above). The character <y> is part of the digraph <ly> representing the segment [λ].

currently not enough data available to test this prediction[40] and therefore it is currently not possible to fully evaluate which of the two analyses is the correct one.

2.2.2.3 Predictions of the Analysis

At this point it is useful to provide a summary of the predictions the current analysis makes for V#V-sequences that are not attested as undergoing sandhi. The following is an updated version of the table in (T:2.7). The entries in angled brackets are the outcomes we expect for the various unattested types of V#V-sequences based on the constraint ranking employed in the analysis. I have excluded all entries that I treated as mere orthographic variants (such as /o#ā/ → *ā* instead of *wā*). The purpose of this overview is twofold. On the one hand, it makes the consequences of the constraint ranking fully explicit. On the other hand, it conveniently illustrates how to reinforce/falsify the analysis once additional TA data becomes available.

Table 2.22: Predicted sandhi outcomes of V_1#V_2-pairs

V_2→		a	ā	e	o	i	u
V_1↓	a	<a>	ā	<e>	o	<i>	<u>
	ā	ā	ā	?	?	?	?
	e	ya, e	yā	e	yo	<i>[41]	<yu>
	o	<wa>, o	wā	we	o	<wi>	<u>
	i	ya	yā	ye	yo	i	yu
	u	<wa>	wā	<we>	wo	<wi>	u

Note that the current proposal is still incomplete in that it does not make any predictions about the behavior of final *ā* before an unlike vowel other than *a*. CS shows deletion of final

[40]One crucial test case would be sandhi in an ā#a-sequence in which the first word is not the negation and the second word is not an enclitic.

[41]Both /e#i/ → *i* and /o#u/ → *u* are attested sandhi effects in CS. Baković (2007) uses the undominated constraint NoLong to account for the phenomenon. Since I have adopted the same constraint above we are predicting this outcome for TA as well.

low vowels in this context (Baković 2007: 2), and Stumpf shows that in TB the general rule is to delete V_1 over V_2. It is therefore not unlikely that TA *ā* could undergo the same kind of deletion. However, without additional data I leave this issue open for the time being.

2.2.3 C#V-sequences

The crucial point about the various forms of V#V-sandhi discussed so far is that, regardless of the exact analytical details, all sandhi effects can be viewed as a means of avoiding onsetless syllables. In the following section I show that the same generalization applies to sandhi in C#V-sequences.

Sandhi effects involving C#V can be classified into two types. The first and more general type is consonant gemination. A word final consonant is doubled before a word-initial vowel. Gemination is by far the best attested sandhi effect in TA and, unlike any of the other forms of sandhi, it frequently occurs in prose as well as verse. Gemination is attested for almost every possible word-final consonant.

Table 2.23: Segments attested undergoing gemination

Obstruents					
p	*klopp oki*	'like pain'	*ṣ*	*tmäṣṣ aci*	'from that'
t	*mañkätt oki*	'like the moongod'	*s*	*wäss oki*	'like poison'
c	*kucc aśśi*	'who'	*ś*	*ākāśś oki*	'like the sky'
k	*ṣäkk āṅkari*	'six tusks'	*ts*	*ṣakk atsts aśśi*	'indeed'
Sonorants					
n	*weñānn anac*	's/he said to him'	*r*	*aptsarr oki*	'like an Apsaras'
m	*sne nākämm oki*	'like without fault'	*l*	*wäll oki*	'like a king'
ñ	*ṣulañṅ oki*	'like mountains'	*ly*	–	

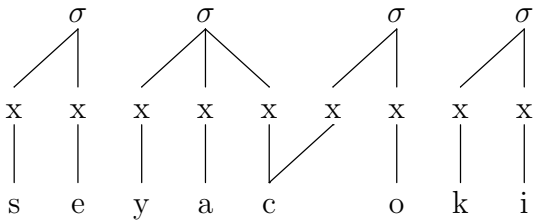
The phenomenon is not discussed by Stumpf (1971b), who only covers sandhi in V#V and R#V sequences, but had already been mentioned by Sieg et al. (1931). Remarks on Tocharian gemination across word boundaries can also be found Krause and Thomas (1960:

74), Klingenschmitt (1994: 349) and Malzahn (2010: 20-2), although none of these scholars attempted a synchronic analysis.

In TA geminates generally do not occur in word-initial position,[42] and word-finally they almost exclusively appear before word-initial vowels.[43]

Furthermore, all the remaining word-final clusters in TA (i.e. clusters that are neither geminates, nor due to glide formation) consist of two consonants with falling sonority. From these observations I conclude that geminates in TA must be heterosyllabic. In the case of gemination sandhi, this means that the geminated consonant is linked to both the final syllable of the first word, as well as the inital syllable of the second word.

(2.19) *Consonant gemination*



Therefore, gemination sandhi is simply another instance of phrase-level resyllabification in order to prevent a syllable from being realized without an onset. Let us now see how we can incorporate sandhi gemination into our current OT-analysis. First, I make the standard assumption that gemination violates a markedness constraint No-Gem (Rose 2000: 102), which assigns a violation to a consonantal segment that is linked to more than one syllabic position. No-Gem has to be ranked below the two Max-constraints in order to prevent vowel deletion from applying instead of gemination. We further have to rule out

[42]The imperative is expressed by a prefix *p*, and in verbs with stem-inital *p* both labials are spelled, e.g. *ppärksār* 'ask!', *ppāṣār* 'protect!'. These forms differ from the kind of gemination under discussion, however, in that the two segments are heteromorphemic.

[43]I could only find 14 exceptions to this rule, and in 9 of these the apparent geminate is spelled with a combination of anusvāra and a second nasal consonant, e.g. <ṃn>, <ṃñ>, etc. Geminates spelled this way have a special status, as I argue in the next section.

the option of exclusively linking the word-final consonant to the following onset by deleting its original association line. I assume that such a candidate is ruled out by the constraint AlignRight(Wd,σ), which requires the right-most segment of a word to be linked to the right edge of a syllable.[44] In the following tableau a word-boundary is indicated by a bracket, while a syllable boundary is indicated by a dot.

(2.20) Max(V_w), AlignRight(Wd,σ) >> No-Gem

/ñās aru/	Onset	Max(V_w)	AlignRight(Wd,σ)	No-Gem
☞ a. ñās.]saru				*
b. ñā.s]aru			*!	
c. ñās.]ru		*!		
d. ñās.]aru	*!			

Therefore, ranking Max(V_w) and AlignRight(Wd,σ) above No-Gem ensures that C#V-sequences can only undergo gemination.

The final sandhi phenomenon to be discussed only affects R#V-sequences. While nasals and liquids are attested undergoing gemination like obstruents, they can alternatively undergo a process that involves the deletion of a preceding schwa. In what follows I will refer to this phenomenon as relinking.

Table 2.24: Examples of relinking involving schwa-deletion

sumnātär 'take away'	+ *oki* 'like'	→	*sumnātr oki*
**kareñcäm* 'they laugh at them'	+ *oki* 'like'	→	*kareñcm oki*
poñcäṃ 'all'	+ *ārkiśoṣi* 'world'	→	*poñcn ārkiśoṣi*

Stumpf (1971b: 108) describes the process as sequences of the shape äR turning into a semivowel, analogous to glide formation. Under the current proposal, the parallel between glide formation and relinking lies in the resyllabification of a word-final segment into the

[44] I am operating with the version of Align argued for by Itô and Mester (1999), which, unlike previous versions, allows the edge-segment to be linked to more than one syllable.

following onset. That resyllabification is involved in (T:2.24) is supported by the fact that the resulting forms exhibit word-final clusters with rising sonority, which are unattested outside of sandhi contexts. The deletion of the preceding schwa is therefore not a defining feature of the sandhi effect in question, but merely a side effect. Once the final sonorant no longer occupies the coda-position of the final syllable, the schwa must be deleted as TA does not tolerate schwas in open syllables. This means that, unlike in Stumpf's analysis, we have no reason to expect that relinking requires the final segment to be preceded by schwa, but it is the alternation between schwa and zero that makes the phenomenon visible to us. There is, in fact, a way to identify cases of relinking in the absence of schwa deletion, which I address in the section on TA orthography.

It is important to note that relinking is simply an alternative to gemination. All the segments that are attested undergoing relinking also exhibit gemination.

Table 2.25: Segments undergoing relinking and gemination

a) Relinking		b) Gemination	
sumnātr oki	'take away'	*wājärr oki*	'like a vajra'
kareñcm oki	'they laugh at them'	*sne nākämm oki*	'like without fault'
poñcn ārkiśoṣi	'the whole world'	*tsratränn oki*	'like separated

The difference between the forms in a) and b) is that the latter violate No-Gem while the former violate AlignRight(Wd,σ). Therefore, the simplest way to capture this variation is to allow both rankings of the two constraints relative to each other. However, this solution does not account for the observation that, while both sonorants and obstruents can be geminated, relinking is only attested with sonorants. For example, a combination of *pälkoräṣ* 'having seen' and the emphatic particle *ats* yields *pälkoräṣṣ ats* in sandhi but never **pälkorṣ ats*.

Considering the relatively low number of attestations[45] showing relinking, we need to ask whether we could be dealing with a mere accidental property of the corpus. Therefore, before looking into a possible explanation let us first establish whether it is actually necessary to account for the observed variation. In order to determine this we can look at the ratio between gemination and relinking involving liquids and nasals and then compare this ratio with the number of cases in which an obstruent is geminated in äC#V sequences.

Table 2.26: Gemination vs. relinking in obstruents and sonorants

	ä[-son]#V	ä[+son+cons]#V
Gemination	138	23
Relinking	0	16

Of the 39 cases of äC#V-sequences involving liquids and nasals in a sandhi context, 16 (41%) undergo relinking as opposed to gemination. Considering the relatively high frequency of sandhi involving äC#V-sequences with obstruents (138), it is likely not a coincidence that none of them shows relinking.

A potential explanation for the distribution in (T:2.26) lies in the observation that geminate sonorants are crosslinguistically more marked than geminate obstruents. Taylor (1985) established the implicational generalization that the existence of geminate sonorants in a language presupposes the existence of geminate obstruents. Kawahara (2007) posits a set of universally ranked markedness constraints that reflect the correspondence between the sonority and markedness of geminates. *GG penalizes geminates involving glides, *LL laterals and *NN nasals.

(2.21) *Kawahara (2007)*

*GG >> *LL (lateral) >> *NN >> *ObsGem

[45] 14 cases in Stumpf's article + 2 from the YQ-manuscript.

Since in TA there is no detectable difference between nasals and liquids with respect to sandhi and therefore no evidence for the more fine-grained distinction in (2.21), the present analysis merely distinguishes between *ObsGem and the general constraint *SonGem, which penalizes any kind of geminate involving sonorants. In order to account for the coexistence of gemination and relinking in TA, as well as the restriction of the latter phenomenon to sonorants, we simple have to allow both relative rankings of *SonGem and Align(Wd, σ). In the following tableaux I am making the additional assumption that the deletion of *ä* is triggered by an undominated markedness constraint *schwa., which bans *ä* from occurring in open syllables. I am further taking schwa-deletion to violate a faithfulness constraint that is ranked lower than both Max_{strong} and Max_{weak}.

(2.22) Max(V_w) >> *SonGem >> AlignL(Wd,σ)

/ciñcr akmal/	Onset	*schwa.	Max(V_w)	*SonGem	AlignL(Wd,σ)
☞ a. ciñc.r akmal					*
b. ciñcärr.]rakmal				*!	
c. ciñcär.]kmal			*!		
d. ciñcä.r]akmal		*!			*
e. ciñcär.]akmal	*!				

(2.23) AlignL(Wd,σ) >> *SonGem

/wājär oki/	Onset	*schwa.	AlignL(Wd,σ)	*SonGem
☞ a. wājär.r oki				*
b. wājär.r oki		*!	*	
c. wājär. oki	*!			

The two tableaux illustrate how each of the two possible rankings of *SonGem and Align(Wd, σ) produce relinking and gemination of sonorants respectively. *ObsGem, on the other hand, is consistently outranked by AlignL(Wd,σ), preventing obstruents from ever undergoing relinking.

2.2.4 Summary

The main argument of this section is that all sandhi phenomena involving X#C-sequences can be interpreted as strategies to satisfy a constraint against onsetless syllables, a constraint that can be independently motivated based on the avoidance of word-internal hiatus in TA. Specifically, the phenomena conspiring to avoid violations of ONSET can be grouped into three different types: a) coalescence of identical vowels, b) vowel deletion, and c) linking the final segment of a word to the initial syllable of the following word. The latter can be achieved by either adding an additional association line to the segment in question (as in gemination) or by disassociating the segment from its original syllable (as in glide formation and relinking). One question that has so far been ignored is the status of initial onsetless syllables in the absence of external sandhi and whether these cases are simply permitted to remain V-initial or whether they are repaired by some other mechanism. The next section addresses this question using evidence from Tocharian orthography.

2.3 Virāma Spelling and Tocharian A Prosody

The main focus of this section is an aspect of the Tocharian corpus that has thus far been neglected in the study of Tocharian prosody, namely the use of the Tocharian brāhmī alphabet. More specifically, I will demonstrate that the distribution of sentence-internal virāma spelling (see section 2.3.2 for an explanation of the term) in Tocharian A manuscripts is at least partially predictable and mainly occurs in two environments: a) where virāma spelling renders unnecessary the use of a complex akṣara and b) in prevocalic position in the absence of external sandhi.

I argue that the distribution of virāma spelling in pre-vocalic position is not a mere orthographic idiosyncrasy, but provides independent evidence for the phonological constraint discussed in connection with external sandhi, i.e. a global ban on onsetless syllables. As in

the previous section, the present study focuses primarily on the situation in TA, leaving a survey of the TB facts for future research.

2.3.1 Main Features of the Northern-Turkestan Brāhmī Alphabet

The term 'Nordturkistanische Brāhmī' was first introduced by Sander (1968). The brāhmī alphabet most Tocharian texts are written in is based on an Indian alphabet, although the character shapes have been modified and several characters were added to the inventory. While all attested TA texts are written in the same standard ductus, a comparatively small number of TB manuscripts are written in earlier forms of the writing system.[46] This indicates that the brāhmī alphabet was first adopted to write TB and its use was only later extended to TA, as suggested by Krause (1952: ix) (cf. also Peyrot 2008: 208).

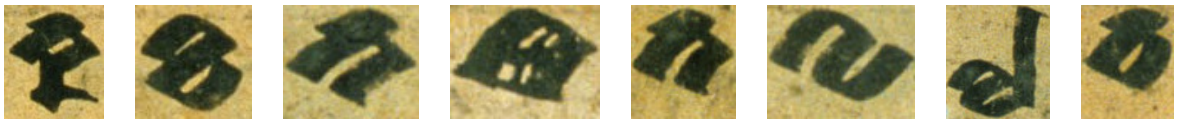
I begin by giving a brief overview of some of the main properties of the alphabet used for most Tocharian manuscripts, focusing solely on the properties relevant to this study.[47] Much of this introduction is based on the much more comprehensive overview in Malzahn (2007a).

The brāhmī alphabet is written left-to-right. Each basic consonant character is associated with a following default vowel, which is either *a* or *ä*. Characters with inherent *ä*, as well as the corresponding vowel diacritic (see below), are innovations of the Tocharian scribes, as the vowel *ä* does not exist in Sanskrit.

[46] See Malzahn (2007b) on the different paleographical layers of TB manuscripts.

[47] All the sample characters are cutouts from photographs of the manuscript leaves THT 698 (= A 56), THT 690 (= A 57) and THT 641 A 8. The manuscript leaves belong to the Depositum der BERLIN-BRANDENBURGISCHEN AKADEMIE DER WISSENSCHAFTEN in der STAATSBIBLIOTHEK ZU BERLIN - Preussischer Kulturbesitz Orientabteilung. The photographs of A 56 and A 57 have been taken from the website of the IDP (http://idp.bbaw.de/), the photograph of A 8 is from TITUS (http://titus.fkidg1.uni-frankfurt.de/framee.htm?/index.htm). The full images of all three leaves can be found in appendix D.

(2.24) *Some basic consonant characters*



<ka> <kä> <na> <nä> <ta> <wa> <la> <ca>

Vowels other than *a* can be rendered by diacritics above or below an akṣara.[48] Consonant characters whose inherent vowel is *ä*, on the other hand, are never marked with vowel diacritics.

(2.25) *Notation of vowels*

<kā> <ke> <ki> <ko> <ku> <ce> <ci> <cä>

The brāhmī alphabet also has independent vowel characters, which are confined to word-initial position, although word-initial vowels can also be written using diacritics (see below).

(2.26) *Some independent vowel characters*

<a> <ā> <e> <o>

Consonant clusters are generally rendered by complex akṣaras, in which the individual consonant characters are arranged in vertical order starting from the top.

[48]The term akṣara refers to any set of characters occupying a single space on the horizontal axis, which includes all combined consonant characters (see below), as well as diacritics. An akṣara can also consist of a single unmodified character as in the examples in (2.24).

(2.27) *Ligatures*

<kta> <tka> <lko> <klo> <cpā> <ppo> <cca> <lntw>

A nasal in coda position can be represented by anusvāra, a diacritic dot above the akṣara containing the preceding vowel. In word-final position the anusvāra dot always represents the consonant /n/, while word-internally it represents a nasal homorganic with the following consonant.

(2.28) *Anusvāra*

<naṃ> <näṃ> <nāṃ> <laṃ>

Following this basic overview of some of the orthographic principles underlying the brāhmī alphabet, let us now take a more detailed look at what is going to be the main focus of this section, i.e. virāma spellings.

2.3.2 Virāma Spelling in Tocharian Orthography

One way to render final consonants in Tocharian brāhmī is to put them in virāma position, which means that the lack of a following vowel within the same word is indicated by a connecting line between the akṣara in question and the preceding akṣara.[49] This connecting line can be accompanied by a diacritic dot above the final or the penultimate akṣara. A peculiar feature of Tocharian brāhmī orthography is that only akṣaras marked for *ä*-vocalism are used in virāma position. Specifically, this means that either the last consonant character

[49]If the two akṣaras are separated by a line break, the virāma stroke is drawn from the second akṣara towards the beginning of the line.

within the akṣara must have inherent *ä* or (usually if no such character is available) the corresponding vowel-diacritic is used, as in the last example.[50]

(2.29) *Virāma spelling*

<ñi-k> <tu-ṅk> <mā-t> <slā-nt> <na-c>

At the end of a sentence or verse, virāma spelling is the only way to render a consonant that is not followed by a vowel. On the other hand, within a sentence or verse there are two possible ways of spelling a word-final consonant relative to the following word. In what follows I will use the term *C#X-pair* to refer to pairs of adjacent words in which the first word ends in a consonant. The Tocharian brāhmī alphabet provides two main strategies to render C#X pairs. The first option, which I will call *vi-disjunction*, is to put the word-final consonant(s) in virāma position, as depicted in (2.29), and rendering the initial segment(s) of the following word with a separate akṣara. The second option is to express the word-final consonant(s) and the following segment(s) using a single akṣara. I will refer to this strategy as *a-linking*.[51]

Table 2.27: Orthographic Rendering of C#X pairs

	... C # (C)V ...
a-linking	... <C (C)V> ...
vi-disjunction	<C> <(C)V> ...

[50]Malzahn (2007b: 283) shows that this idiosyncratic property of Tocharian orthography can be explained through the phonology of the oldest attested variety of TB, where word-final *ä* was still realized. She thereby provides further evidence that the Tocharian brāhmī alphabet was initially used to write TB, as opposed to TA.

[51]A third way of rendering C#X pairs is to spell word-final /n/ or /m/ (the latter only in Sanskrit) using anusvāra and starting the next word with a separate akṣara. While its distribution overlaps in many ways with that of vi-disjunction, I will leave the details for future study.

The following examples illustrate these two alternative spellings for the same phrase in TA.[52]

(2.30) *Variation between a-linking and vi-disjunction in* lokit yeṣ *'came as a guest'*

a) 
<lo ki **tye** - ṣ>

b) 
<lo ki **- t ye** ṣ...>

What makes vi-disjunction particularly important for our purposes is that its use in Tocharian has much wider scope than it has in Sanskrit. In Sanskrit orthography a-linking is the rule and virāma spelling is generally restricted to the end of a verse or sentence, while in Tocharian orthography vi-linking also frequently occurs sentence-internally. The same orthographic difference is also present in TA - Sanskrit bilingual texts, indicating that it reflects some real linguistic difference between the two languages.

The use of virāma spelling in Tocharian raises two immediate questions. The more general question concerns what linguistic property caused the Tocharian scribes to extend the use of virāma in the way described above. Stumpf (1990: 106) and Peyrot (2008: 176) remark in rather broad terms that a-linking is better suited to render Sanskrit than it is for Tocharian. Peyrot mentions the "characteristic sandhi" as the feature that makes a-linking especially well suited to write Sanskrit, without specifying the details of this relationship between phonology and orthography.

The other question we can raise about Tocharian orthography is what factors determined which strategy a scribe used in order to render a particular C#X-pair. Stumpf (1990: 106) showed, based on an earlier study by Winter (1955: 220-1), that in TB the use of vi-disjunction increases over time, and is most frequently attested in the youngest layer of TB

[52]I will romanize brāhmī orthography in the following way: Each group of letters separated by a space or a hyphen represents one akṣara, while the hyphen itself represents the virāma line connecting two akṣaras. I am thereby deviating from the more traditional practice of rendering a consonant character in virāma position using the schema C◟.

manuscripts. However, neither author offers any explanation as to why vi-disjunction is used in individual cases.

In order to gain more insight into both of these questions, I compared the frequency with which each type of C#X-pair is rendered using vi-disjunction. I used an electronic search of the TA corpus in order to cover as much data as possible in a maximally consistent fashion. The data set was selected from the CEToM database using a number of (partially arbitrary) criteria, which are summarized in appendix C.

2.3.3 General Distribution of Virāma Spelling in Tocharian A

Table (T:2.28) shows the distribution of virāma spelling in C#X-pairs where the second member starts in a consonant, henceforth *C#C-pairs*. The leftmost column of the table refers to the number of consonants involved in the cluster spanning the word boundary.[53] The different types of clusters are illustrated with an example in (T:2.29).

[53]In five cases the first word ends in three consonants. In addition to being phonotactically aberrant, all of the attested examples exhibit special behavior in other respects.

a-linking:	*rājavartt yok*, *rājavartt yokās*
vi-disjunction:	*mäṃnt pat*, *winuṃnt dhyāṃ*, *ānaṃnd ṣñi*

The a-linked cases are peculiar in that they show word-final gemination before a consonant. One possible explanation is that *rājavartt yok* 'multicolored' (translation by Ji et al. 1998) is a compound showing a sporadic process of word-internal gemination before a sonorant consonant (on this phenomenon see Klingenschmitt 1994 and Malzahn 2010). The cases involving vi-disjunction are special in that they all involve combinations of anusvāra and another nasal. I will return to this phenomenon below.

Table 2.28: Distribution of vi-disjunction in C#C-pairs

C#C	a-linked		vi-disjoined		total
C#C	5525	61%	3449	38%	8974
C#CC	771	24%	2323	75%	3094
CC#C	176	13%	1113	86%	1289
C#CCC	4	5%	65	94%	69
CCC#C	2	66%	1	33%	3
CC#CC	11	2%	434	97%	445
CC#CCC	0	0%	10	100%	10
CCC#CC	0	0%	2	100%	2
total	6489	46%	7397	53%	13886

Table 2.29: Examples of C#C-pairs

	a-linking		vi-disjunction	
C#C	*täm pälkoräṣ*	<tä **mpä** lko rä-ṣ>	*täm pälkoräṣ*	<tä**-m pä** lko rä-ṣ>
CC#C	*yärk yatsi*	<yä **rkya** tsi>	*yärk yāmurā*	<yä**-rk yā** mu rā>
C#CC	*puk wrasas*	<pu **kwra** sa-s>	*puk wrasas*	<pu **-k wra** sa s...>
CC#CC	*oṅk träṅkäṣ*	<o **ṅkträ** ṅkä-ṣ>	*oṅk träṅkäṣ*	<o-**ṅk trä** ṅkä ṣ...>
C#CCC	*śtwar śtwar*	<śtwa **rśtwa** r...>	*ces śtwar*	<ce**-s śtwa** r...>
CC#CCC			*śmeñc śtwar*	<śme**-ñc śtwa** r...>

The table in (T:2.28) shows that the number of attestations of vi-disjunction vis-à-vis a-linking increases with the number of consonants. We can make this pattern even more explicit by grouping the table entries according to the number of consonants alone, disregarding the position of the word boundary.

Table 2.30: Distribution of vi-disjunction grouped by number of consonants

CC	a-linked		vi-disjoined		total
CC	5525	61%	3449	38%	8974
CCC	947	21%	3436	78%	4383
CCCC	17	3%	500	96%	517
CCCCC	0	0%	12	100%	12
total	6489	46%	7397	53%	13886

In order to ensure that we can interpret the distribution in (T:2.30) as a relationship between virāma spelling and C-clusters, we can use Pearson's chi-squared test. The resulting χ^2 value is 2307.423, yielding a probability of under .00001 that the two properties are unrelated (the performed calculations are provided in appendix B).

This is in line with the observation in Sieg and Siegling (1921: ix) that in those rare cases where virāma spelling occurs word-*internally*, it is used to avoid complex ligatures.

Table 2.31: Word-internal Virāma

a)	*pältskumäñcsā*	<pä-**l tsk**u mä-**ñc s**ā>
b)	*wākmtsune*	<wā-**k mts**u ne>
c)	*kātkmāṃ*	<kā-**t km**āṃ>
d)	*mālkluneyo*	<mā-**l kl**u ne yo>
e)	*säkskiñciṃ*	<sä-**k sk**i ñciṃ>
f)	*puklāyo*	<pu-**k l**ā yo>
g)	*kāśyapnäṣṣ*	<kā śya-**p n**ä ṣṣ...>
h)	*weñṣäm*	<we-**ñ ṣ**äm...>
i)	*pälkoräṣṣ ats*	<pä lko rä-**ṣ ṣ**a-ts>

The table includes all the examples of word-internal virāma in TA that I am aware of. 5 out of 9 cases involve consonant clusters with more than two members. Of the four remaining cases, the treatment of the cluster <kl> in f) <pu-k lā yo> is in line with its distribution across word boundaries, with the cluster attested 8 times with a-linking and 24 times with

vi-disjunction.

Based on the data so far, I agree with Peyrot (2008) that the phonological differences between Tocharian and Sanskrit were an important factor in the extended use of virāma spelling in the former. Vi-disjunction could have been an orthographic reflex of the higher frequency of complex consonant clusters across word boundaries in Tocharian. It provided a means to avoid some of the ligatures that would have otherwise been required to render these clusters. Within every ligature, the initial character is written below a fixed imaginary base line and every additional character is written below the previous one. Large ligatures are therefore in danger of intruding onto the next line, depending on the number and size of the individual combined characters.

While the avoidance of complex ligatures is clearly a factor in the distribution of virāma spelling within C#C-pairs, it is only a partial explanation, especially in the light of the kinds of doublets we have seen in (2.30) where the exact same consonant cluster is rendered in both ways (<lo ki t̲ye-ṣ> vs. <lo ki-t̲ ye ṣ>). Hence, there must have been at least one additional orthographic principle that prevented a scribe from avoiding ligatures at word boundaries altogether. A likely candidate for such a principle is the need to save paper, since a set of characters that has been combined into a complex akṣara only takes up a single horizontal space. The necessity to save space/paper is also reflected in the complete avoidance of line breaks other than at the edge of the manuscript page, regardless of any form of linguistic or metrical boundaries.

Let us now turn to the distribution of *C#V-pairs*, i.e. C#X-pairs in which the second member starts in a vowel. The data in table (T:2.32) is arranged according to the number of consonants preceding the word boundary. Each of the cells in the table is further illustrated with a concrete example in (T:2.33).

Table 2.32: Distribution of vi-disjunction in C#V-pairs

C#V	a-linked		vi-disjoined		total
C#	51	5%	860	94%	911
CC#	651	79%	170	20%	821
CCC#	74	100%	0	0%	74
total	776	42%	1030	57%	1806

Table 2.33: Examples of C#V-pairs

C#V	a-linking		vi-disjunction	
C#	*śtwar īmeyis*	<śtwa **rī** me yi-s>	*śtwar amoktse*	<śtwa**-r a** mo ktse>
CC#	*yukk oki*	<yu **kko** ki>	*tuṅk oki*	<tu**-ṅk o** ki>
CCC#	*kācky arässi*	<kā **ckyā** rä ssi>		

Even though the use of virāma spelling appears again to be dependent on the number of consonants,[54] the relationship appears to be reversed. The number of cases involving a-linking after a single consonant is relatively small at only 5%. After two consonants we have a significant increase to 79% and after three consonants vi-disjunction is entirely unattested. In view of the pattern we saw in C#C-pairs above, the distribution in (T:2.32) is somewhat unexpected. If vi-disjunction is, in fact, an orthographic tool to avoid certain kinds of complex ligatures, we would expect *all* C#V-pairs to be a-linked, independent of the number of consonants involved. The reason for this is the following: An akṣara in virāma position has to directly precede a word boundary, which can be independently established based on the placement of virāma akṣaras in C#C-pairs.[55] In C#V-pairs the word boundary always falls at the *end* of a consonant cluster (by the very definition of the term) and therefore,

[54]Running Pearson's chi-squared test on the table in (T:2.32) yields a value of 1059.458. This places the probability that virāma spelling is independent of C-clusters at less than .00001 (see appendix B for details).

[55]The cases that involve word-internal virāma spelling are rare enough to be considered orthographic anomalies.

ligatures in C#V-pairs cannot be broken up through vi-disjunction.

Not only is vi-disjunction in C#V-pairs useless for avoiding ligatures, it is also graphically more costly than a-linking in requiring an additional akṣara, instead of a simple diacritic on an existing one.

It follows that C#V-pairs involving vi-disjunction have nothing to do with the avoidance of complex ligatures and require a different explanation. In the following section I will show that, unlike in C#C-pairs, the distribution of a-linking and vi-disjunction in C#V-pairs is entirely determined by a single factor, namely external sandhi.

2.3.4 Complementary Distribution of Vi-Disjunction and A-linking in C#V-Pairs

The distribution of a-linking in C#V-pairs can be summed up with the following schema.

Table 2.34: General rule for using vi-disjunction in C#V-pairs

C#V-pairs		1806 cases in the dataset	
+external sandhi	→	a-linking	(56 exceptions)
-external sandhi	→	vi-disjunction	(1 exception)

A C#V-pair is a-linked if the two words in question are affected by external sandhi, and vi-disjoined otherwise. The table in (T:2.35) provides one example for each type of case.

Table 2.35: Spelling examples: sandhi vs. no sandhi

a)	glide formation	*praski äräṣ*	→	*prasky äräṣ*	<pra **skya** räṣ>
b)	relinking	*sumnātär oki*	→	*sumnātr oki*	<su mnā **tro** ki>
c)	vowel deletion	*poryo āriñc*	→	*pory= āriñc*	<po **ryā** ri-ñc>
e)	gemination	*weṣ eṃtsuräṣ*	→	*weṣṣ eṃtsuräṣ*	<we **ṣṣeṃ** tsu rä ṣ>
f)	no sandhi	*ypic oki*	→	*ypic oki*	<ypi**-c o** ki>

Example f) is especially interesting in that it is one of the rare cases in which the en-

clitic *oki* does not trigger gemination on the preceding obstruent. Just as predicted by the generalization in (T:2.35), the lack of sandhi corresponds to the use of vi-disjunction.

Among the 1806 C#V-pairs in the selected data set there were only 57 exceptions to this pattern. 56 involve a-linking in the absence of sandhi, while only one of them involves vi-disjunction in the presence of sandhi.

In the previous section I made the generalization that all forms of external sandhi listed in (T:2.35) are instances of resyllabification. Having established that a-linking in C#V-pairs is dependent on the presence of sandhi, we can now use this generalization to give a more detailed account of the distribution observed in table (T:2.32), repeated here for the reader's convenience as (T:2.36).

Table 2.36: Distribution of vi-disjunction in C#V-pairs (repeated)

C#V	a-linked		vi-disjoined		total
C#	51	5%	860	94%	911
CC#	651	79%	170	20%	821
CCC#	74	100%	0	0%	74
total	776	42%	1030	57%	1806

First, we need to account for the rarity of a-linking in cases where a single consonant precedes the word boundary (column a), especially compared to its high frequency with two consonants. The reason for this is that glide formation, relinking (accompanied by schwa deletion), and gemination exclusively produce C#V-pairs with at least two consonants preceding the word-boundary. In the case of glide formation, this is because Tocharian does not allow word-internal hiatus and therefore all word-final high-vowels must be preceded by at least one consonant. Furthermore, gemination and glide formation are the most commonly attested forms of sandhi, with gemination accounting for 511 a-linked C#V-pairs and glide formation for 135 within the data set. Of the 51 C#V-pairs that do undergo a-linking, 13 involve vowel coalescence.

The second observation from table (T:2.36) that requires some explanation is that C#V-pairs involving 3 consonants preceding the word boundary are never written using vi-disjunction. This simply follows from the fact that word-final tri-consonantal clusters can only arise through phrasal resyllabification, i.e. external sandhi. The only way a Tocharian A word can end in three consonants is by relinking the rightmost one into the following syllable onset.

We can now rephrase our initial generalization about the rendering of C#V-pairs in Tocharian A thusly: A C#V-pair is spelled using vi-disjunction whenever there is no evidence that resyllabification has taken place. Before providing a phonological evaluation of the correspondence between sandhi and a-linking, I will briefly discuss the cases that do not seem to fit this pattern.

2.3.5 Apparent Exceptions

In this section I address the 57 exceptions to the observed correspondence between sandhi and a-linking, and I show that a large number of these cases are consistent with the proposed generalization after all.

2.3.5.1 Exceptional vi-disjunction

There is a single attested example in which two words are written using vi-disjunction despite the apparent presence of gemination sandhi.

(2.31) ñäkcyāṃñ aptsaräntu 'divine Apsarases'

What is striking about this example is that the geminate is written using a combination of a nasal consonant character and anusvāra (henceforth <ṃN>) instead of a ligature of two identical consonant characters (<NN>). Cases of <ṃN> are ambiguous within Tocharian orthography. On the one hand, they can express a genuine geminate, which is indicated by words attested both with <ṃN> and <NN> spellings.

Table 2.37: Anusvāra in geminates

	<ṃN>	<NN>	
a)	*wṣeṃñesac*	*wṣeññe*	'(to the) camp'
b)	*taṃne*	*tanne*	'so'
c)	*kapśiṃño*	*kapśiññ̃o*	'with the body'

On the other hand, there is also a scribal practice that I will call *redundant anusvāra*. These are cases in which an anusvāra is written in a position where a geminate is extremely unlikely. The phenomenon is relatively well attested, and I am only listing some of the most obvious examples.

Table 2.38: Redundant Anusvāra

	redundant anusvāra	regular spelling	
a)	*mäṃnt*	*mäṃt/mänt*	'how'
b)	*ākṣiṃññunt*	*ākṣiññunt*	'announced'
c)	*keṃnn ats*	*kenn ats*	'wrong' + ptcl.

Example a) seems to contain a geminate preceding a stop, which is unattested outside of <ṃN> spellings in TA. Furthermore, TA words usually do not end in more than two consonants, unless resyllabification has taken place. This is not the case here, since the following word starts in a consonant (*mäṃnt pat*). In the examples b)-c) the anusvāra seems to be hypermarking a preexisting geminate. All of the examples above are also attested without anusvāra and there is no independent evidence that TA distinguished between three levels of consonant length. Therefore, I will assume that the use of anusvāra in cases such as (T:2.38) is entirely redundant and does not correspond to an additional phonological segment.

We can conclude from the use of vi-disjunction that *ñäkcyāṃñ aptsaräntu* does not involve gemination sandhi at all, but rather an instance of redundant virāma.

2.3.5.2 Exceptional a-linking

The cases in which we get a-linking in the absence of sandhi can be grouped into a number of subgroups. The first group consists of examples in which the first word ends in two consonants.

Table 2.39: A-linking after C-cluster

1	*kälpānt*	*äts*[56]	11	*wālant*	*oki*
2	*tuṅk*	*oki*	12	*wālant*	*oki*
3	*lānt*	*oki*	13	*tskont*	*oki*
4	*– nant*	*oki*	14	*śmeñc*	*ats*
5	*wāwrunt*	*oki*	15	*mänt*	*aśśi*
6	*ānt*	*oki*	16	*pont*	*oko*
7	*kāksont*	*oki*	17	*pält*	*o –*
8	*kuront*	*oki*	18	*skeneñc*	*o(ki) ///*
9	*tärkrunt*	*oki*	19	*tsälpṣant*	*i ///*
10	*märsneñc*	*oki*	20	*kātkṣant*	*i- ///*

Sieg and Siegling (1921: 226 fn. 1) suggest treating 1 as a scribal error for *kälpā*{*t*}*t* {*a*}*ts*, so we can dismiss this example. In the remaining cases we would expected gemination of the word-final obstruent, especially since gemination within word-final consonant clusters in TA was clearly possible. The following is a complete list of the attestations I could find.

Table 2.40: Gemination in final C-cluster

sarkk	*oki*	*mäṃtt*	*aśśi*	*kakältsts*	*oki*
kayurṣṣ	*oki*			*waṣtt*	*oki*
papärss	*ats*			*yokeyutts*	*oki*
sarkk	*oki*				

In light of the examples in (T:2.40), I will take the use of a-linking in (T:2.39) as an

[56]Index: *kälpā*{*t*}*t* {*a*}*ts*

indication that gemination sandhi has, in fact, taken place and that the lack of its orthographic representation is due to the aforementioned tendency to avoid complex ligatures at the word boundary. That the superficial lack of geminates in (T:2.39) is orthographically conditioned is further supported by the observation that 5 out of the 8 cases of geminates in (T:2.40) follow either <r> or anusvāra. The cluster-initial <r> is rendered with a diacritic above the akṣara, similar to anusvāra, and hence does count as part of a ligature.[57] The final cluster in *sarkk* is a special case in that *kk* is not rendered by a ligature in the brāhmī alphabet but rather by the character for *k* with an additional stroke in the middle.[58] Hence the spelling of *sarkk* does not involve a ligature at all. Note that this means that we have to find an alternative explanation for the single *k* in *tuṅk oki* (example 2), where spelling out the geminate would not have resulted in a more complex ligature. It is possible that we are simply dealing with a scribal error due to the very similar nature of the characters for *k* and *kk*. Below we will see additional cases of exceptional a-linking that may be explained in this way.

I conclude that the forms in (T:2.39) involve the same kind of phenomenon we observed with glide deletion, i.e. a final segment that is not spelled in order to avoid ligatures with more than two members.

[57]Note how the character <r> is located at the same hight as the vowel diacritic on the preceding akṣara in a), as well as the virāma dot on the following akṣara in b).

a) b)

<wä rpnā> <rku-nt>

[58]

<ka> <kka>

The next subgroup of apparent exceptions involves pairs in which the first member ends in a sonorant.

Table 2.41: A-linking with final sonorants

21	*tkan=*	*ākāśā*	30	*ṣul*	*äntā ne*
22	*yantran*	*oki*	31	*cmol*	*e(ntsāt)*
23	*ṣom*	*atsaṃ*	32	*pärskāl*	*anäṣ*
24	*ṣāñ*	*upāy*	33	/// *-ñcä(ṃ)n*	*oki*
25	*ytār*	*o – –*	34	*(kā)swon=*	*ākālyo*
26	*ytār*	*o ///*	35	*ṣtām*	*ā – ///*
27	*kuryar*	*ukā –*	36	*cami ṣtām*	*i – ///*
28	*śtwar*	*īmeyis*	37	/// *r*	*oki*
29	*ppāṣār*	*āñcäm*	38	/// *·y*	*arthis*

Rather than contradicting the correspondence between a-linking and sandhi, examples 21-32 lend support to the analysis of relinking proposed in the previous section. I argued that schwa-deletion was merely a side effect of shifting a final sonorant into the onset of the following word. A-linking provides us with a diagnostic to identify instances of relinking in the absence of schwa-deletion, such as in (T:2.41). Examples 21-22 are especially informative in that they belong to the rare cases in which a single word-final *n* is spelled using a full consonant character instead of anusvāra, which is restricted to syllable codas. Example 33, on the other hand, must involve a missing anusvāra (see Sieg and Siegling 1921: 174), otherwise the schwa would be located in an open syllable. If 34 is restored correctly, the first word is a sandhi form of *kāswone* 'goodness', although I have no explanation for the lack of a glide at the end of the form.[59] The very fragmentary examples 35-36 are most likely wrongly segmented in the transcription. The form *ṣtāmā* could be the perlative of *ṣtām* 'tree', while in 36 we are probably dealing with the genitive of the same word, i.e. *ṣtāmis*. This is further supported by the preceding genitive demonstrative *cami*. 37 and 38 are even

[59]As mentioned in the previous section, Stumpf considers it likely that the form is a scribal error.

more fragmentary. The lower part of the akṣara containing <r> in 37 is broken off the leaf, and the remaining part is consistent both with <r> and <rr>. 38 is most likely just a case of glide formation.

Finally, the last set of exceptions consists of pairs in which the first word ends in an obstruent.

Table 2.42: A-linking with a single final obstruent

39	*ṣak*	*atsekk ats*	48	*pāk*	*ayi* –
40	*ṣak*	*atsek*	49	*(pu)k*	*āñmaṣ*
41	*ṣak*	*atsek*	50	/// – ·*k*	*oki*
42	*tmäk*	*ats*	51	*träṅkäṣ*	*ā* ///
43	*lok*	*aśśi*	52	*parnoreyis*	*āś·*
44	*sälpiṣṣäk*	*ats*	53	*was*	*a-* ///
45	*āmpuk*	*e* ///	54	*wlyep*	*aṃpl·* – ///
46	*ptsäk*	*upage*	55	/// *s*	*āmpuk*
47	*ñuk*	*ānāsāṃ*	56	/// *(s)s*	*oko*

With obstruents we do not expect relinking to be possible. What is noteworthy here is that 11 out of 18 cases involve word-final *k*, although 48-49, as well as 55-56, are too fragmentary to draw any conclusions about. As mentioned above, the characters <k> and <kk> only differ by a single stroke. The few irregular cases of a-linking involving <k> could therefore simply be scribal errors, as I also suggested for *tuṅk oki*. The form *sälpiṣṣäk ats* stands out due to the unexpected gemination of *ṣ* and was deemed a scribal error for **sälpiṣäkk ats* by Sieg et al. (1931: 303). I have no explanation for 51-52. 53 could be read as *wasa-*, with the *a* being part of a secondary case ending. Finally, part of 54 could be resegmented as *wlyepaṃ*, a feminine plural form of *wlyep* 'soft'.

The final question to be addressed is what property of TA could have caused virāma spellings to be sensitive to external sandhi.

2.3.6 A Phonological Interpretation of the Distribution of Virāma Spelling

I now offer a phonological explanation for the distribution of a-linking in C#V-pairs. As argued above, the prevalence of a-linking in sandhi contexts is unproblematic. If vi-disjunction is used as a means to avoid complex ligatures, there is simply no motivation to use it in those cases. On the other hand, we do have to explain why vi-disjunction is not only allowed but obligatory in the absence of sandhi.

I propose that virāma-spelling before vowel-initial words indicates that the words in question did not, in fact, start in a vowel but instead in an epenthetic consonant, such as [ʔ]. This epenthetic consonant would be the result of a phonological repair strategy that prevents syllables from being realized without an onset. We have already seen that Tocharian A provides evidence for a strict ban on onsetless syllables, both at the word-level and at the phrase-level, and we have further seen that all of the types of external sandhi discussed above conspire to provide a word-initial vowel with a syllable onset. This leaves only a very small number of environments in which vowel-initial syllables seem to be allowed. The following represents an exhaustive list of these environments.

(2.32) *Vowel-Initial Syllables in TA*

a) in absolute initial position
b) word-initially, after a vowel
c) word-initially, after a consonant that has not been resyllabified

As was shown above, environment c) is rendered almost exclusively using vi-disjunction, with the small number of exceptions already discussed.

I am now proposing that the ban on onsetless syllables was, in fact, without exception. In the absence of external sandhi, a vowel-initial word would have to be furnished with an onset by some other means. A typologically common strategy is the insertion of a default consonant, such as glottal stop. I take vi-disjunction to be the orthographic reflex of ʔ-

epenthesis, which, in the absence of external sandhi, provides an underlyingly vowel-initial word with an onset.

We can illustrate the effect of this hypothesized process using an example that combines two of the environments listed in (2.32) and is spelled using vi-disjunction.

(2.33) *el eṣlac* 'to give a gift' <• e-l e ṣla-c>

/el eṣlac/ → [ʔel ʔeṣlac]

The underlying representation in (2.33) contains two instances of word-initial vowels that are targeted by ʔ-epenthesis. According to the current proposal, the second [ʔ] in this example is indirectly reflected in the orthography. There was no special brāhmī character for glottal stop, but the audible presence of an additional consonant between a word-final consonant and a word-initial vowel may well have prevented the scribes from rendering the two sounds with a single akṣara, resulting in the consistent use of vi-disjunction in this particular environment. The first instance of [ʔ] in (2.33) would have been located in absolute initial position (indicated by the punctuation mark •). In this environment, as well as after a word-final vowel, there would have been no orthographic reflex of [ʔ] whatsoever, as virāma spelling is inapplicable in those positions. Nevertheless, the presence of [ʔ] follows as a consequence of adopting the process illustrated in (2.33).

Another language that famously uses [ʔ] to avoid onsetless syllables is German. The examples in (T:2.43) are taken from Alber (2001: 5-6). In German a vowel initial morpheme can be furnished with an onset either by resyllabification or by ʔ-epenthesis, the former mostly in fast speech and the latter mostly in slow speech.[60]

[60] Alber (2001: 5-6), the example numbers and plain orthographic transcriptions were added by me.

Table 2.43: Resyllabification vs. ʔ-epenthesis in German (Alber 2001)

		slow:	fast:	
a)	Ereignis	ʔEr.-ʔéig.nis	ʔE.r-éig.nis	'event'
b)	umarmen	ʔum.-ʔár.men	ʔu.m-ár.men	'to embrace'
c)	erarbeiten	ʔer.-ʔár.bei.ten	ʔe.r-ár.bei.ten	'to obtain by work'
d)	in Ungarn	ʔin.ʔÚn.garn	ʔi.n Ún.garn	'in Hungary'
e)	zum Anfassen	zum. ʔÁn.fas.sen	zu.m Án.fas.sen	'to touch'
f)	beim Essen	beim ʔÉs.sen	bei.m És.sen	'while eating'

Because it provides the exact same strategies to avoid onsetless syllables, German constitutes a close typological parallel for the system I am proposing for TA.[61]

For our OT-analysis this means that the constraint ONSET is undominated, regardless of whether sandhi takes place or not. I take the application of sandhi to be determined by the ranking of the constraint DEP-IO,[62] which assigns a violation to every instance of epenthesis. An undominated ranking of DEP-IO will cause sandhi to apply as described in the previous section.

(2.34) DEP-IO >> UNIF

/śla āñcālyi/	ONSET	DEP-IO	UNIF
☞ a. *ślāñcalyi*			*
b. ślaʔāñcālyi		*!	
c. ślaāñcālyi	*!		

On the other hand, ranking DEP-IO below all the constraints violated by external sandhi will will render ʔ-epenthesis a less costly strategy for avoiding a violation of ONSET.

[61]Insertion of a glottal stop before word-initial vowels also occurs in English, especially at the beginning of larger prosodic constituents (see e.g. Dilley et al. 1996).

[62]Definition of DEP-IO from McCarthy and Prince (1995: 16):

> Every segment of the output has a correspondent in the input.
> (Prohibits phonological epenthesis.)

(2.35) UNIF >> DEP

/śla āñcālyi/	ONSET	UNIF	DEP-IO
☞ a. *śla?āñcālyi*			*
b. ślāñcalyi		*!	
c. ślaāñcālyi	*!		

2.3.7 Summary

In this section, I have argued that the distribution of vi-disjunction in C#C-pairs and in C#V-pairs is governed by entirely distinct factors. Vi-disjunction in the former functions as a means of avoiding certain kinds of ligatures, based on the observation that its frequency increases with the number of consonants in a cluster. In C#V-pairs, on the other hand, vi-disjunction is useless in avoiding ligatures. Instead I argued that virāma spellings were used to indicate that a word-final consonant and an (underlyingly) word-initial vowel were separated by an epenthetic consonant, possibly a glottal stop. This was based on a) the overall tendency of TA to avoid onsetless syllables and b) the observation that the only common environment in which we find vi-disjunction in C#V-pairs is identical with one of the only three environments in which a syllable appears to lack an onset.

Therefore, ʔ-epenthesis correctly accounts for the difference in orthographic treatment between C#C-pairs and C#V-pairs. This can be backed up with cross-linguistic parallels, and is also perfectly in line with the overall phonological behavior of Tocharian A.

Chapter 3

Prosodic Incorporation of Clitics

3.1 Introduction

The present chapter deals with another aspect of Tocharian phonology above the word level, namely the way clitics are prosodically combined with their hosts. The purpose of this section is twofold. First, I review the evidence that allows us to set up a special class of phonological clitics in Tocharian in the first place. Second, I demonstrate how the observed facts can be accounted for within prosodic phonology by taking clitics to form a prosodic unit with their host that is both larger than the prosodic word and smaller than the prosodic phrase, i.e. the Clitic Group (Hayes 1989, Nespor and Vogel 1986 [2007]). Specifically, I will adopt the Clitic Group in combination with the relaxed version of the Strict Layer Hypothesis presented in Vogel (2009) in order to account for a number of seemingly unrelated phenomena associated with clitics, including their behavior with respect to external sandhi and their placement within metrical contexts. The purpose of this section is not to provide an exhaustive discussion of all the elements that can potentially be considered phonological clitics. Rather, I will select a core set of cases based on which the class of elements under discussion can be established. As in the previous chapter, the initial focus will lie on TA. However, after the analysis has been established for TA, I will show how a similar account

can capture the TB facts as well.

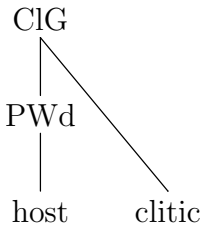
3.2 The Phonology of TA clitics

I begin by discussing a set of TA enclitic particles that is most sharply distinguished from both suffixes and orthotonic words in consistently triggering consonantal sandhi on the preceding word. After laying out the data I show that this can best be captured by operating with a prosodic constituent located between the prosodic word (PWd) and the prosodic phrase (PPhr), i.e. the Clitic Group (ClG), specifically in the version defended in Vogel (2009). I will then show how the same analysis covers additional particles, using evidence from other phonological phenomena as well as metrics.

3.2.1 The Clitic Group within the Prosodic Hierarchy

Within Prosodic Phonology there have basically been two main types of proposals as to how clitics are incorporated into prosodic structure. One type, originating in Hayes (1989) and Nespor and Vogel (1986 [2007]), posits a special prosodic constituent located between the PWd and the PPhr, which directly dominates hosts and their clitics. The most recent version of this proposal is represented by Vogel (2009).

(3.1) *Vogel (2009)*

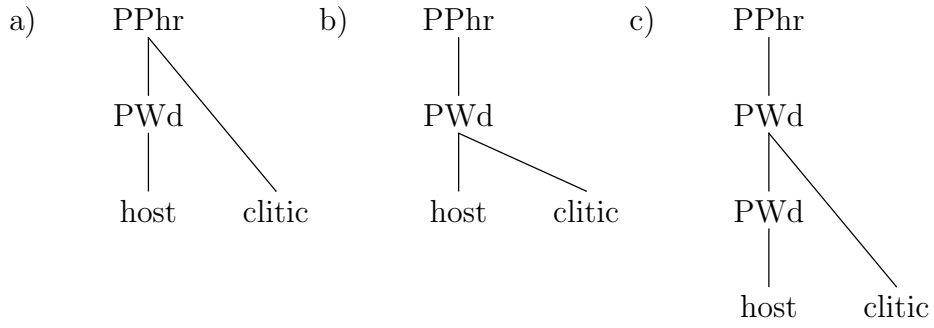


The second type of proposal is one that does not operate with a special prosodic constituent for host+clitic. Instead, clitics are incorporated into independently established prosodic constituents, primarily the PWd or the PPhr. An especially influential version of

this proposal is Selkirk (1996: 188), which posits three possible structures for phonological clitics.

(3.2) *Selkirk 1996*

a) = free clitic, b) = internal clitic, c) = affixal clitic



The main difference between current versions of these two proposals concerns recursivity. A Selkirk-type model can operate with a smaller number of distinct prosodic constituents but has to allow recursive structure within phonology, such as in (3.2c). Vogel's model, on the other hand, can operate without recursion at the price of an additional layer in the prosodic hierarchy. The question of which model is superior is therefore not necessarily one of parsimony but one of whether a) there is independent evidence for recursion in phonology and b) whether there is evidence for a prosodic constituent between the PWd and the PPhr that cannot be captured by a recursive PWd (see Vogel's paper for a detailed criticism of the latter). In what follows I will argue that there is indeed some evidence in TA that suggests that we need such an intermediate constituent.

3.2.2 V-Initial Particles and Gemination Sandhi

Sieg et al. (1931: 285-6, 302-5) observe for a number of vowel-initial particles that they typically trigger gemination on a preceding consonant. The following is a brief overview of the six V-initial particles that support such a generalization, each accompanied by an example sentence.

1. *aci* is a postposition governing the ablative (ending in *-äṣ*). The meaning is roughly '(starting) from'. There are also two attestations of *aci* where it occurs without an ablative, serving an adverbial function. In what follows I will only be referring to postpositional *aci* when referring to the particle.

(3.3) *tmäṣṣ* *aci* *nunak* *kroś* *oksiṣ*
then.ABL.SG starting.from again *krośa*.SG grow.3SG
From there it grows again one krośa, ... (A 4 a6)

2. *aśśi* is a modal particle occurring in questions. In Wh-questions, it usually attaches directly to the Wh-word.

(3.4) *kuss* *aśśi* *sās*
who/what PTCL DEM.NOM.SG
Who may she be? (A 6 a2)

3. *ats* is a particle putting emphasis on the preceding constituent. The particle also occurs as the isofunctional variants *atsaṃ* and *atsek*. In the following discussion I will use *ats* to collectively refer to all of its variants.

(3.5) *śāmāṃ* *śiśkäṣ* , *klisontäṣ* ; *śoll* *ats* *cami*
live.MP lion.ABL.SG sleep.PP.ABL.SG.PRT life.SG emph DEM.GEN.SG
, *wākäm* *ṣeṣ* :
difference.SG was.3SG.IMPF
only life made him different from a living, sleeping lion. (A 12 b5, 4-3/4-3)

4. *oki* is a comparative particle 'like', 'as it were', attaching to the right edge of the constituent it modifies. For a detailed description of the syntax and function of this particle see Thomas (1968).

(3.6) *cokis* *slamm* *oki* *aṣuk* *pärkraṃ* *prārū*
lamp.GEN.SG flame like wide long.PL finger.PL
his fingers slender and elongated like the flame of a lamp; (YQ II.4 b5)

5. *ontaṃ* is a negative polarity item predominantly attaching to the negation or to Wh-words.

(3.7) *marr* *ontaṃ* *nu* ; *puttiśparnäṣ* ; *lotki{ṣ}* *wras(om*
not PTCL CONJ Buddhahood.ABL.SG turn.3SG.OPT being.NOM.SG
:)

Never, either, should a being turn away from the Buddha rank. (A 313 b2, 4/4/4)

6. *an-* is an enclitic pronominal stem attested in the ablative *anäṣ* and the allative *anac*. I will use *an-* to refer to both types of attestations.[1]

(3.8) *pācarr* *oki* ; *späntāll* *anac* , *mäskatär* :
father.SG like trust.GER.SG PRO.ALL become.3SG
One should trust him like a father. (A 61 a4, translation mine, 5/5/4-4/4-3)

The application of consonant gemination indicates that the particles in question are not part of the same prosodic word as their host. Compare the behavior of the particles *oki* and *ats* with that of the locative suffix *-an* when attaching to the noun *wram* 'thing'.

Table 3.1: Gemination: particles vs. locative

	wram 'thing'
particle	*wramm oki wramm ats*
locative	*wram-aṃ*

[1]My translation of the example deviates from that currently given by CEToM, which reads "Like a father he is to be trusted." The pronominal argument in the allative (functioning as a dative) indicates that the gerund has active meaning in this sentence. Such a reading is unproblematic since the Tocharian gerund is voice-indifferent (cf. Pinault 2008: 611).

The examples in (T:3.1) show that V-initial particles trigger consonant gemination while suffixes do not. Vowel-initial suffixes and particles also show different behavior with respect to glide formation. V-initial suffixes can cause a glide to be inserted after a front or round vowel. This behavior differs from external sandhi where a front vowel is turned into a glide instead (cf. Kölver 1965: 4-5).

Table 3.2: Glide formation: particles vs. locative

glide insertion	*cwaṅke* 'lap'	→	*cwaṅkey-aṃ*
glide formation	*ākläṣlye* 'students'	→	*ākläṣly oki*

On the other hand, V-initial particles also differ from V-initial content words. The latter only sporadically trigger gemination on a preceding consonant even if the two words form a closely knit syntactic constituent. Therefore, the close prosodic relationship between V-initial particles and the word they follow indicates that we are dealing with a third category, i.e. enclitics.[2] In what follows I will refer to the six V-initial particles described above as *V-enclitics*.

Before drawing any conclusions from the behavior of V-enclitics with respect to gemination sandhi, however, we need to take a closer look at the attested data. This is because there are a substantial number of cases, most of which have already been noted by Sieg et al. (1931), where gemination before an enclitic does not seem to occur. However, I will argue that, using my findings on TA orthography presented in the previous chapter, the majority of these apparent exceptions have explanations consistent with the application of external sandhi. As a result, the application of consonantal sandhi does not simply differ in relative frequency between host+clitic and word+word. The difference is more clear-cut, in that it is obligatory in the former and optional in the latter, indicating that we are dealing with

[2]Stumpf (1971b) already concluded that V-initial particles in TA are unaccented. He did so based on C-gemination, as well as V_2 deletion in vocalic sandhi, which I argued against in the previous chapter.

different prosodic structures.

The table in (T:3.3) is structured as follows: For each V-enclitic column a) contains the number of attestations in which it triggers gemination on a preceding obstruent, liquid or nasal. Column b) contains the number of cases in which gemination should be possible but is absent.[3]

Table 3.3: Gemination before V-enclitics

	a) Gemination	b) No Gemination
aci	57	0
aśśi	20	5
ats	142	11
oki	166	36
ontaṃ	4	1
an-	13	3
Total:	402	56

The 56 exceptions can be grouped into three major categories based on orthography. The first category consists of 29 cases in which the clitic and its host are spelled using a-linking.

3.2.2.1 A-linking (32 attestations)

7 cases without C-gemination involve relinking sandhi.

Table 3.4: A-linking due to relinking

a)	*kareñcm*	*oki*	e)	*pärskāl*	*anäṣ*
b)	*klawantr*	*oky āñc*	f)	*yantran*	*oki*
c)	*sumnātr*	*oki*	g)	*ṣom*	*atsaṃ*
d)	*ytäṣtr*	*oki*			

[3]In the table, as well in the following discussion, I only considered examples in which the particle is fully preserved, i.e. cases where part (or all) of a particle has been restored by the editors have been disregarded.

Sieg et al. 1931: 305 note that *oki* can trigger schwa-deletion on a preceding syllable, as in a)-d). The relinking analysis of schwa deletion proposed above allows us to subsume e)-g) under the same category. Therefore, even though the cases in (T:3.4) do not show C-gemination, they are consistent with the generalization that V-enclitics obligatorily trigger external sandhi on a preceding consonant.

The remaining 22 appear to involve no external sandhi.

Table 3.5: A-linking, spelling without sandhi

a)	*mänt*	*aśśi*	j)	*kuront*	*oki*	q)	*lok*	*aśśi*
b)	*śmeñc*	*ats*	k)	*lānt*	*oki*	r)	*ṣak*	*atsek*
c)	*śla tuṅk*	*oki*	l)	*kāksont*	*oki*	s)	*tmäk*	*ats*
e)	*tskont*	*oki*	m)	*wālant*	*oki*	t)	*sälpiṣṣäk*	*ats*
f)	*tärkrunt*	*oki*	n)	*(– – – –) nant*	*oki*	u)	*essik*	*ats*
g)	*wāwrunt*	*oki*	o)	*wālant*	*oki*	v)	*ṣak*	*atsek*
h)	*ānt*	*oki*	p)	*märsneñc*	*oki*	w)	*ṣak*	*atsekk*
i)	*lānt*	*oki*						

15 of the 22 examples involve a word-final C-cluster, while 7 involve word-final *k*. In the previous chapter I argued, based on the absence of virāma spelling, that both types of examples likely involved external sandhi after all. According to my proposal the geminates simply did not get spelled in order to avoid complex ligatures at word boundaries (a-p) or due to scribal errors owing to the very similar nature of the brāhmī characters for <k> and <kk>. Therefore, if this proposal is correct, all the examples in (T:3.5) can be dismissed as exceptions to V-enclitics causing external sandhi.

Finally, there are three highly fragmentary cases, which I primarily mention for the sake of completeness. All three of them involve a-linking, although the manuscript fragment including c) is so small that we cannot tell whether we are dealing with a particle at all or merely a fragment of a larger word.

Table 3.6: Fragmentary cases of a-linking

a)	*/// r*	*oki*
b)	*/// -ñcä(ṃ)n*	*oki*
c)	*/// – ·k*	*oki*

3.2.2.2 Vi-disjunction (8 attestations)

There are 8 cases without external sandhi where the enclitic and its host are spelled using vi-disjunction.

Table 3.7: Vi-Disjunction before V-enclitics

a)	*kälk*	*aśśi*	e)	*/// āp*	*oki*
b)	*śla tuṅk*	*oki*	f)	*ypic*	*oki*
c)	*naṣt*	*aśśi*	g)	*kayurṣis*	*oki*
d)	*kus*	*aśśi*	h)	*wäl*	*oki*

Based on the correspondence between a-linking and sandhi established in the previous chapter, these examples look like genuine exceptions. At present, I can offer no explanation for these cases.

3.2.2.3 Anusvāra (16 attestations)

16 of the remaining cases apparently lacking external sandhi involve a word-final /n/ that is spelled with anusvāra.

Table 3.8: Anusvāra before V-enclitics

a)	*ñareyaṃ*	*oki*	f)	*tsräsmāṃ*	*oki*	k)	*spaktānikāṃ*	*ats*
b)	*wraṃ*	*oki*	g)	*ytäs(m)āṃ*	*oki*	l)	*täpreṃ*	*ats*
c)	*tkaṃ*	*oki*	h)	*sälpmāṃ*	*oki*	m)	*täpreṃ*	*atsaṃ*
d)	*yantraṃ*	*oki*	i)	*triwamāṃ*	*oki*	n)	*(lkā)laṃ*	*anäṣ*
e)	*/// lyomaṃ*	*oki*	j)	*ywatr-äṃ*	*oki*	o)	*/// ñāṃ*	*anac*
						p)	*yasäṃ*	*ontaṃ*

Sieg et al. (1931: 148) observe that the locative suffix *an* (spelled <aṃ>) never undergoes C-gemination when preceding *oki*. However, it is highly unlikely that this distribution is anything but an accident of the corpus. First, the locative suffix does undergo gemination with other particles, as already noted by Sieg et. al.

Table 3.9: Gemination of /n/ in the locative

a)	*tampeyaṃn*	*ats*	e)	*pälskaṃn*	*ats*
b)	*tmann*	*ats*	f)	*$p_u$kaṃn*	*atsaṃ*
c)	*tmann*	*ats*	g)	*wraṃn*	*āñc*
d)	*dhyānann*	*ats*			

Furthermore, the total number of examples involving *oki* following a noun in the locative is very low (a-e)[4] and instances of final /n/ not belonging to a locative are also attested without gemination before *oki*. These include the final /n/ of the participle in *mān* (f-i), as well as the pronominal suffix /n/ (j), although both of those suffixes happen to be attested with gemination as well.

Table 3.10: Variation of final /n/ before *oki*

anusvāra		gemination	
sälpmāṃ	*oki*	*mlamānn*	*oki*
ywaträṃ	*oki*	*tsratränn*	*oki*

Finally, we have already seen a case where *oki* triggers another form of external sandhi on the locative suffix, namely *yantran oki*, for which I argued that the spelling of the final nasal as <n> instead of <ṃ> indicates the application of relinking.

Based on these observations, as well as the examples in k)-p), where /n/ fails to undergo gemination before other particles as well, I conclude that the attested behavior of the locative

[4]Sieg et al. (1931: 148) also list *yantran oki* as belonging to this category, on which see below.

suffix *an* before *oki* is part of a more general phenomenon involving word-final anusvāra before V-enclitics. We now need to consider whether the examples in (T:3.8) are indicative of a phonological phenomenon, i.e. the optional absence of sandhi with word-final /n/, or whether we are simply dealing with another instance of orthographic variation obscuring the expected application of sandhi. In what follows I will defend the latter hypothesis. First, anusvāra before a V-enclitic is attested twice as often as all of the cases of vi-disjunction combined (16 vs. 8). It is not at all clear what phonological property would cause /n/ to resist sandhi more frequently than other consonants. Other nasals (/ñ/ and /m/) consistently undergo gemination before a V-enclitic (36 cases combined). On the other hand, /n/ does stand out orthographically in that it is the only consonant within the Tocharian brāhmī alphabet that can be (and most frequently is) written with a diacritic in word-final position instead of a full character. We have already seen that anusvāra is used with a certain degree of inconsistency. It can be used to represent the first member of a nasal geminate but can also hypermark an already existing nasal character, what I referred to as redundant anusvāra. It is conceivable that this orthographic variation also included the use of anusvāra to represent a nasal geminate by itself. If this is correct, there would have been three different ways to spell a word final /nn/, namely <ṃn>, <nn> or <ṃ>. Each of these three spellings is attested with similar frequency before V-enclitics.

Table 3.11: Distribution of gemination and anusvāra before V-enclitics

	<ṃn>	<nn>	<ṃ>
aci	0	0	0
aśśi	1	0	0
oki	6	8	10
ats	4	8	3
an-	3	4	2
ontaṃ	0	0	1
total	14	19	16

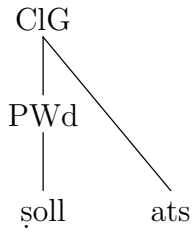
I therefore conclude that the examples (T:3.8) should not be considered as counterexamples to V-enclitics triggering external sandhi.

Summing up, we started out with 56 examples that appear to lack C-gemination before V-enclitics. 7 of those cases showed another type of external sandhi (relinking), while 25 involved evidence from orthography (a-linking) that gemination took place after all. I argued that the 16 cases involving word-final anusvāra spelling should at least be considered consistent with an analysis involving external sandhi. This leaves us with 8 counterexamples with no explanation against 450 that exhibit or are at least consistent with the application of external sandhi. Based on this distribution I now make the following generalization about TA phonology: V-enclitics obligatorily trigger external sandhi on a preceding [+consonantal] segment.

The question is now how to account for this generalization. Treating consonantal sandhi as an idiosyncratic property of the 6 particles in question is unsatisfactory, since both C-gemination and relinking occurs between orthotonic words as well. Instead I argue that the prosodic constituent formed by clitics and their host is the primary domain for consonantal sandhi. I take this constituent to be the ClG. This allows us to account for the fact that combinations of host+clitic exhibit behavior distinct from both words, by undergoing external sandhi, as well as from phrases, by undergoing consonantal sandhi obligatorily. The fact that consonantal sandhi sporadically occurs within larger prosodic constituents as well can then be accounted for by a crosslinguistically common tendency for phonological rules to apply to prosodic constituents beyond their primary bounding domain in fast speech (cf. Hayes 1989: 215).

The following illustrates the prosodic structure I am adopting for host+clitic combinations in TA (in line with the version of the prosodic hierarchy in Vogel 2009).

(3.9) a)



In other words, I am adopting a model in which a host and its clitic(s) form a constituent larger than the PWd and smaller than the PPhr. The following section will further motivate this analysis with additional evidence from the segmental phonology of TA, as well as extend the analysis to proclitics.

3.2.3 Hiatus

For terminological simplicity's sake I will use the term *hiatus* to refer to any sequence of vowels that are orthographically rendered as two adjacent vowel characters, regardless of whether we assume, as I did above, that the two vowels are separated by a glottal stop on the phonetic level. Within the analysis adopted in the previous chapter, a ban on hiatus (in the current usage of the term) means specifically that hiatus cannot be resolved by ʔ-epenthesis. However, nothing about the following discussion depends on this particular analysis and the arguments equally apply if one does not operate with ʔ-epenthesis.

3.2.3.1 Hiatus with Enclitics

I already mentioned in the previous chapter that TA generally does not tolerate word-internal hiatus. On the phrase level, on the other hand, hiatus is at least possible, in that the various vocalic sandhi phenomena discussed above (glide formation, V-deletion, V-coalescence) apply only optionally.

Table 3.12: Hiatus vs. vocalic sandhi with V-enclitics

Hiatus	Sandhi	
wiyo oki	*wiyoki*	(V-coalescence)
tā aśśi	*tā =śśi*	(V-deletion of *a*)
cani ats	*ṣom-cmolṣy ats*	(glide formation of *i*)
kupre ontaṃ	*kācky arṣäntās*	(glide formation of *e*)

The table shows how four[5] of the V-enclitics are attested both in hiatus with a preceding vowel, as well as undergoing vocalic sandhi. The possibility of hiatus is then another property that sets V-enclitics apart from suffixes and confirms that V-enclitics are located outside of the PWd of their host.

Let us now look at the proclitic preposition *śla* 'with', which shares with V-final enclitics the property of having a vowel at the edge that connects to its host, and see what conclusions we can draw about its prosodic constituency based on its behavior with respect to hiatus. The preposition *śla* has long been identified as a proclitic[6] based on the fact that it is the only native TA word ending in *a*.[7] In all other cases *a* (or its ancestor **æ*) underwent apocope during the prehistory of TA (e.g. Proto-Tocharian (PT) **wæræ* > TB *were*, TA *war* 'odor'). If we make the generalization that *a* is banned from word-final position in TA, the occurrence of *a* at the end of the proclitic appears to indicate that *śla*, unlike the clitics we have seen so far, forms a single PWd with its host (e.g. [*śla tuṅk*]$_{\text{PW}}$ 'with love'). However, this cannot be the correct analysis, as it fails to account for the fact that *śla* exhibits exactly the same kind of variation we saw in (T:3.12) with enclitics.

[5]The postposition *aci* only attaches to ablative forms, which end in a consonant. The pronominal stem *an-* is only attested a single time following a vowel, which undergoes glide formation (*/// ṣñy oky anac*).

[6]See e.g. Stumpf (1971b: 98).

[7]In the expression *ślak śkaṃ* 'and likewise', *śla* occurs without a nominal complement. In this case it likely receives accent, as indicated by the fact that it occurs with the emphatic particle *k* and hosts the enclitic conjunction *śkaṃ* 'and'. See Thomas (1979) for details on the use of TA *śla*/TB *śle* with and without nominal complements (the latter are analyzed by Thomas as conjunctions).

Table 3.13: Hiatus vs. vocalic sandhi with *śla*

Hiatus	V-deletion
śla āñcālyi	*śl= āñcālyi*
śla āmāśās	*śl= āmāśās*
śla oko	*śl= oko*

None of the forms in the left column are possible PWds in TA by virtue of exhibiting hiatus. In this respect *śla* behaves exactly like V-enclitics, which I argued are located outside the PWd of their host. Before proposing a solution to this apparently conflicting behavior of *śla* I will turn to another set of data that raises exactly the same kind of problem.

3.2.3.2 Hiatus in Compounds

There is one apparent exception to the ban on word-internal hiatus in TA, namely compounds. Even though TA compounds are notoriously difficult to distinguish from phrases, there is one secure diagnostic for the compound status of two adjacent words, which is a first member ending in *a* (cf. Sieg et al. 1931: 228 with additional discussion on how to identify compounds in TA).

Table 3.14: Final *a* in first compound members

kāsu 'good'	*pälskāñ* 'thoughts'	→	*kāswa pälskāñ*
ṣom 'one'	*kälyme* 'direction'	→	*ṣoma kälyme*

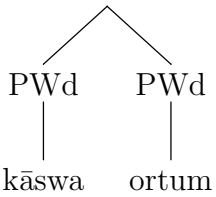
The final *a* at the end of the first member seems to indicate that we are dealing with a single PWd. However, in a small subset of these cases the second member starts in a vowel.

Table 3.15: Hiatus at compound boundaries

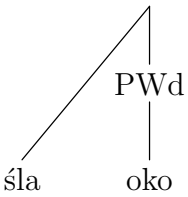
kāsu 'good'	*ortum* 'friendly'	→	*kāswa ortum*
ṣom 'one'	*ārtlune* 'love'	→	*ṣoma ārtlune*

The forms in (T:3.15) exhibit exactly the same kind of apparently conflicting behavior as the examples in (T:3.13) involving *śla*. While, on the one hand, hiatus is only possible at a word boundary, *a* cannot occur at the end of a word. Based on the similarities between the examples in (T:3.13) and the compounds in (T:3.15) it stands to reason they involve similar prosodic structures.

(3.10) a) Compound



b) Proclitic+Host



Since *kāsu* is a content word, it is most likely orthotonic and therefore projects its own PWd. This means that, in order to account for both instances of final *a* in (3.10) we need to make one additional assumption, namely that the domain that does not allow final *a* is not the PWd but the ClG. This is consistent both with the behavior of *śla*, as well as with the general absence of *a* at the end of orthotonic words, since every PWd is dominated by a clitic group (due to the geometry of the prosodic hierarchy argued for in Hayes 1989: 202-20 and Nespor and Vogel 1986 [2007]).

Treating the clitic group as the prosodic constituent formed by compounds, as well as combinations of host+clitic has already been suggested by Vogel (2009: 34-6). Vogel uses the clitic group[8] to account for the fact that accent assignment applies differently to compounds than it does to both PWds and PPhrs in English, as well as in Turkish.

Finally, I should mention one aspect of the analysis I have glossed over so far. I have been operating with a structure in which a clitic is prosodically defective, i.e. does not project its own PWd. This is in line both with the free clitic structure from Selkirk (1996), as well as with the ClG structure from Vogel (2009). However, I am unaware of any TA-internal

[8]Vogel suggests the use of the term Composite Group instead of Clitic Group to capture the larger scope of the constituent. In the current discussion I will continue to use the older term for reasons of consistency.

evidence for clitics being prosodically defective in this way. Apart from the final *a* in *śla*, nothing about the segmental shape of TA clitics sets them apart from orthotonic words.[9] Nevertheless, for the time being I will continue to use the structure in (3.10b) to represent clitics in TA. Not only are clitics frequently prosodically defective crosslinguistically, we will see below that they clearly are in TB. However, I will briefly return to the question of the PWd-status of TA clitics in the next chapter.

3.2.4 Meter

Winter (1959) was the first to note that a number of particles show a restricted distribution in TB verse, showing a preference for specific positions within each colon. Winter concluded based on this observation that TB meter was not merely syllable counting but also involved an "Iktenprinzip" which designated each position within a colon as strong or weak, weak positions being the natural location for clitics. While Winter's Iktenprinzip has not found general acceptance (see e.g. Marggraf 1970 and Malzahn 2012b for arguments against it), there is at least one result of his study that is important for our understanding of both Tocharian meter and prosody, namely that enclitics almost never occur colon-initially. In other words, caesurae that separate an enclitic from its host are strongly disfavored. Bross et al. (2014: 6) make the same observation in their study on TB meter, remarking that "[...] the manner in which the poets realize caesurae suggests that they treated sequences of a lexical word followed by a mono-syllabic clitic as a single word [...]."

In what follows I will show that this generalization holds for TA as well, both for monosyllabic as well as disyllabic clitics, before discussing how it can be captured under the prosodic analysis employed above.

[9] An alternative to the analysis employed above would be to take the vowel *a* to be illegal in PWd-final position, as opposed to ClG-final position, and to attribute the exemption of *śla* from this constraint to its prosodically defective status. In other words, since *śla* does not form a PWd, its *a* is not PWd-final. Such an approach, however, would be unable to account for the behavior of compounds.

3.2.4.1 V-Enclitics

The following tables provide an overview of the distribution of V-enclitics within TA verse. Cola in TA meter are between three and six syllables long. Each row of the table represents one of these colon types, while the columns headed by roman numerals represent the number of attested clitics occupying the respective syllable position within the colon. Disyllabic clitics, which occupy two positions within a colon, are only counted once and only for the first of the two positions. For example, within a 4-syllable colon such as *pācarr oki* the enclitic *oki* is counted as occupying position III. This also means that, according to this schema, a disyllabic clitic can never occupy the last position within a colon, as this would inevitably violate a caesura. The data has been collected by an electronic search of the metrically parsed TA-corpus within the CEToM database. The search I used is highly conservative, in that it only takes into account cola in which each akṣara can be identified, although a colon can appear within a verse that is damaged in other areas. I will address clitics within damaged cola separately where relevant.[10]

Table 3.16: Position of V-enclitics in verse

aci, *aśśi*, *ats*, *oki*, *ontaṃ*, *an-* → 150 total

σ in colon	I	II	III	IV	V	VI
3	1	26	3			
4	2	18	48	8		
5	1	12	10	14	1	
6	0	0	2	1	3	0

Only four of the 150 cola within the dataset have one of the six mentioned V-enclitics within their first syllable. The purpose of providing the more detailed overview in (T:3.16) is to show that position I is the only colon position that rejects enclitics across all different colon

[10]Even in those cases, however, I will only take into account cola in which the clitic is fully preserved, disregarding restorations by the editors.

types. This renders highly unlikely an explanation along the lines of Winter's Iktenprinzip, according to which the first syllable within a colon is only one of several weak positions. It is further worth looking at the four aberrant cases in detail, as only one of them is a genuine counterexample to the generalization that enclitics cannot be separated from their host by a caesura.

One example involves a ceasura preceding the particle *aci*.

(3.11) *mā pe nu cem , aci yeñc ; cam*
not EMPH CONJ DEM.NOM.PL starting.from go.3PL.IMPF DEM.OBL.SG
saṃṅkrāmaṃ :
monastery.LOC.SG
[...] [and] they did not even go close to this monastery. (A 314 a5, 4-3/4-3/4)

Sieg et al. (1931: 297) already argued that in (3.11) *aci* is a preverb instead of a postposition. Even though this would be the only secure case of a preverb *aci*,[11] its use is sharply distinct from the postposition syntactically and phonologically. The preceding nominal form is the subject of the sentence, as opposed to the usual ablative noun governed by the postposition. Furthermore, unlike all postpositional uses of *aci* (see T:3.3), the preceding consonant undergoes neither gemination, nor relinking, indicated by the fact that *cem aci* is spelled vi-disjoined.

The remaining three exceptional cases involve *oki*. Two of these examples are special in that, while *oki* occupies the first syllable within the verse, its host is still located within the same colon.

(3.12) /// *yäṣ , klopasu ; lw oki wärtaṃ :*
go.3SG pain.NOM.SG animal.SG like forest.LOC.SG
he walks around like a miserable animal in the forest ... (A 107 b2, 4-3/4-3/4)

[11]Sieg et al. (1931: 297) only mention two additional potential examples (A 312 a3 and A 328 a2). In Carling et al. (2009: 5) A 328 a2 is considered to involve an adverbial use of *aci* instead.

(3.13) *wtsy* *oky* *akritri* *lap* ; *tseṃ* *ātsäts* *śāku* ;
parasol.SG like round.NOM.SG head.SG blue.SG thick hair.SG
... his head is rounded like a parasol, his hair is thick and blue ... (A 217 b1, 5/5/4-4/4-3)

In *lw oki* 'like an animal' and *wtsy oky* 'like a parasol' the hosts, i.e. *lu* and *wtsi*, have become fully non-syllabic as a result of glide formation and occupy the first syllable of the colon together with the initial vowel of the enclitic. These examples are also noteworthy since they illustrate another phonological property of *oki*. It appears that this enclitic was not necessarily unaccented, but could attract the accent of its host. However, due to the small number of attestations of this kind and our very limited knowledge of TA accent, the exact conditions and scope of this phenomenon have to remain in the dark for the time being.

This leaves us with only one secure example in which an enclitic is separated from its host by a caesura.

(3.14) /// *(—)klisṣiṃ* *sarkk* ; *oki* *tākar* ; *ñäkcyāñ* *tkañi* :
row like be.3PL.PRT divine.NOM.PL earth.NOM.PL
... the heavenly earths were like a flight of step[s]. (A 59 a1, 4/4/4)

The phonological dependence of *oki* on the preceding *sarkk* is further supported by the expected gemination sandhi. If we also take into account damaged padas, there is one additional potential case involving a caesura between *oki* and its host.

(3.15) /// ; *(puk)* *ñareytwäṃ* , *o(ki)* *pat* ; *ñ(areṣiñi)* *(:)*
all hell.LOC.PL like hell
... or like in (all) hells (the inhabitants of hell). (A 314 a5, 4-3/4-3/4)

If the restorations in (3.15) are correct, this makes for a total of two attested cases of V-enclitics following a caesura. Based on the large number of cases that do obey the generalization it is relatively safe to treat (3.14) as a caesura violation, similar to the rare cases in which a colon boundary falls inside a single word.

3.2.4.2 Proclitics

So far we have seen that V-enclitics form a unit with their host that cannot be broken up by a caesura. Above I argued that the same prosodic unit is formed by the proclitic *śla* 'with' and its complement noun. We therefore expect these elements to show the same behavior within verse. The number of attestations of *śla* within verse is unfortunately relatively small but all attested cases are at least consistent with this prediction.[12]

Table 3.17: Position of *śla* in verse

29 attestations

σ in colon	I	II	III	IV	V	VI
3	10	0	0			
4	8	0	0	0		
5	5	0	4	1	0	
6	1	0	0	0	0	0

As expected, the distribution of *śla* mirrors that of the enclitics in that it never occurs in colon-final position, where it would be separated from its host by a caesura.

Tocharian A has another well-attested element with a distribution similar to *śla*, namely the preposition *sne* 'without', which raises the question of whether we are dealing with another instance of a proclitic. Just like in the case of *śla*, the availability of external sandhi, as well as hiatus with the following noun indicates that *sne* is located outside of the PWd formed by its complement noun.

Table 3.18: Hiatus and sandhi with *sne*

a)	*sne kip*	'without shame'	
b)	*sny āk*	'without end'	(glide formation)
c)	*sne eṃts*	'without greed'	(hiatus)

[12]The distribution in the table includes the nine cases, all colon-initial, in which *śla* loses its vowel due to sandhi. In these cases the clitic is categorized as occupying position I within the colon.

This behavior by itself is also compatible with an analysis as an orthotonic word. In the case of *śla* such an analysis could be ruled out due to its phonological shape, i.e. its final *a*. On the other hand, there is nothing about the segmental phonology of *sne* that would set it apart from possible orthotonic words, such as *mñe* 'certainty', *ske* 'effort', or *snum* 'fragrance'. However, its distribution within verse indicates that *sne* is, in fact, a clitic.

Table 3.19: Position of *sne* in verse

71 attestations

σ in colon	I	II	III	IV	V	VI
3	14	0	0			
4	30	1	3	0		
5	16	1	3	1	0	
6	1	0	1	0	0	0

The table in (T:3.19) shows that *sne* has to be located within the same colon as its complement noun.[13] It thereby shows the same kind of distribution as the preposition *śla* and the V-enclitics discussed above. Thus, we can extend the analysis adopted for TA clitics to *sne*.

The metrical distribution of the two proclitic prepositions provides further evidence against the existence of an Iktenprinzip in TA. If the distribution of clitics within verse was determined by whether a metrical position is inherently strong or weak, we would expect proclitics and enclitics to be licensed/avoided in the same set of positions. What we find instead is that the proclitic prepositions occur predominantly in colon-initial position, while enclitics are virtually absent from this position.

[13]I could only find one potential exception among the fragmentary cola.
A 226 a5 *kropñā(nt) – , – – – ; – – lā sne , plā pe – (:)*

3.2.4.3 C-Initial Enclitics

So far I have only discussed clitics which have a vowel on the side facing their host, thereby providing independent evidence from external sandhi and hiatus for their prosodic status as being located outside the prosodic word of their host. There is, however, a class of dependent prosodic elements for which no such evidence is available, i.e. C-initial enclitics. The five best attested examples of this type are the following:

1. The conjunction *śkaṃ* 'and, also'. It typically connects clauses, less frequently phrases, and appears in second position of the second conjunct. The details of its function and syntax are discussed in Thomas (1967).

(3.16) *s$_u$kaṣiṃ* *kärparäṃ* *yā(tse)ñc* *puk* *klopäṃtwaṃ* *śkaṃ* *āñu*
joy.OBL.SG good.SG be.able.3PL all pain.LOC.PL CONJ break.SG
ypeñc
do.3PL
[...] [since] they make possible a joyful dignity and they also cause an end to all sufferings, [...] (A 14 b5)

2. The conjunction *nu* 'and, but'. Its syntactic distribution is very similar to that of *śkaṃ*.

(3.17) *ceṣ* *kenanträ* , *cesäm* *ṣu* ; *cem* *nu*
DEM.NOM.PL call.3PL DEM.OBL.PL PREV DEM.NOM.PL CONJ
cesäs , *kenanträ* :
DEM.OBL.PL call.3PL
These call for those, but those call these. (A 254 b1, 4-3/4-3)

3. The disjunction *pat* 'or'.

(3.18) *(m)äṃ(t* *ne)* *tsekeṣi* *pekeṣi* *pat* *arämpāt* *tāṣ*
how PTCL sculpture.SG painting.SG or form.SG put.3SG
[...] as if it were a fashioned or painted figure. (A 5 a3)

4. The focus particle *pe*. It can occur by itself or, more frequently, in combination with the conjunction *nu*. Even though *penu* is usually listed as a separate lexical entry (e.g. in the glossary of Krause and Thomas 1964), I am deviating from this praxis and simply treat it as a chain of two particles. This decision, however, is not crucial to the following argumentation.

(3.19) *ñuk* *pe* *nu* *ptāñkät* *käṣyāp* *korpac* *kalkam*
I.SG EMPH CONJ Buddha.SG teacher.GEN.SG presence.ALL.SG go.1SG
I too will go to meet the Buddha-god the teacher. (YQ III.5 a4)

5. The relative particle *ne*. It occurs in embedded Wh-clauses, including relative and adverbial clauses.

(3.20) *kus* *ne* *prajñavāṃ* *ṣeṣ* *säm* *tāpärk*
who/what.NOM PTCL PN.SG was.3SG.IMPF DEM.NOM.SG now
śāriputträ *kärsnāl*
PN.NOM.SG know.GER.NOM.SG
He who was Prajñavān, is now to be known as Śāriputra. (A 17 a1)

I will henceforth refer to these five elements as *C-enclitics*. They are subject to the same metrical restriction as V-enclitics, i.e. they have to share a colon with their host and therefore cannot occur colon-initially.

Table 3.20: Position of C-enclitics in verse

ne, *nu*, *śkaṃ*, *pe*, *pat* → 356 attestations

σ in colon	I	II	III	IV	V	VI
3	4	23	37			
4	2	106	38	52		
5	0	32	21	9	18	
6	0	6	1	0	1	6

The number of exceptions (6 out of 356) is again so small that I will list them in their entirety.[14] The first case involves *pe* at the beginning of a tri-syllabic colon.

(3.21) *muso* *säs* ; *ṣom* *nu* *āñc* *dho* , *pe*
rise.PP.NOM.SG.PRT DEM.NOM.SG one.SG CONJ down foot.SG
snālā :

(A 317 b5, 4-3/4-3)

The meaning of this passage is not entirely clear but, according to a note by Wilhelm Siegling in his personal copy of Sieg and Siegling (1921),[15] we are dealing here not with the emphatic particle but with the homophonous word for 'foot'.

Of the five remaining cases, one involves the conjunction *śkaṃ*, two the conjunction *nu*, and two the particle *ne*.

(3.22) *āpāyṣiñik* , *śkaṃ* *wrasañ* ; *naslye* ///
bad.existence.NOM.PL CONJ being.NOM.PL be.GER.PL
And beings of bad existence (were) to be ... (?) (A 116 a4, translation mine, 4-3/4-3)

(3.23) *śāwaṃ* *skeyasyo* ; *puk* *bhūtāñ* *pe* , *nu* *rināmāṃ* ; *tām*
big.PL effort.INS.PL all true.NOM.PL EMPH CONJ seek.MP be.1SG
cindāma,ṇi *(kāmant)* #
PN.SG bring.3PL.PRT
through huge efforts, all the Bhūta [monsters] also desiring [it], (took away) this cintāmaṇi [jewel]. (A 43 and 52 a4, 5/5/4-4/4-3)

(3.24) *ca* — *mi* *pe* ; *nu* *tāṣ* *plāc* ; *ṣotre* *yal* :
EMPH CONJ put.3SG speech.SG sign.SG do.GER.NOM.SG
(A 227-228 a4, 5/4/3/3)

[14] I could not find any additional exception even by taking into account damaged cola.

[15] A scan of the page with Siegling's notes can be found on CEToM under https://www.univie.ac.at/tocharian/images/TS_Siegling/TS_Siegling_Page_173.png

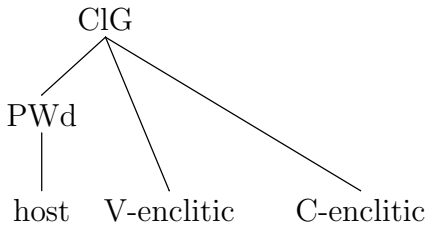
(3.25) /// *nt* ; *mā* *kärsnāṣ* *kus* , *ne* *wrasom* :
not know.3SG who/what.NOM PTCL being.NOM.SG
... which being does not know ... (?) (A 232 b7, translation mine, 4-3/4-3)

(3.26) *mäṃt* *ne* — — , — — — ; *ne* *lyālyoryo,ṣoṣ* *pat* *nu* :
how PTCL PTCL or CONJ
[...] like ... or like ... (A 3 b1, 4-3/4-3)

Again, the small number of attestations suggests that we are simply dealing with caesura violations. This is made all the more likely by the fact that the leaves containing examples (3.23-3.26) all contain other caesura violations as well.

We have seen so far that C-initial clitics have to share a colon with their host, just like V-enclitics and the proclitic prepositions. However, this property by itself merely shows that C-initial clitics behave like suffixes with respect to meter. In order to demonstrate that we are dealing with the same type of clitics discussed above we need to show that they are external to the PWd. We can do this by looking at the relative order of multiple enclitics on the same host. We have already established that V-enclitics are located outside of the PWd of their host. Therefore, if a C-initial clitic can be separated from its host by such an element, it has to be outside of the PWd itself.[16]

(3.27)



Using this diagnostic we can safely establish *nu* as PWd external, as it is attested following four different V-enclitics, *ats*, *oki*, *ontaṃ* and *aśśi*.

[16]The possibility of a morpheme to occur outside of a clitic was already established by Zwicky and Pullum (1983: 504) as a diagnostic to distinguish clitics from suffixes.

(3.28) *tricäṃ* *(dhy)ānan* *yo* *śtärcäṃ* *dhyānann* *ats* *nu*
third.SG meditation.LOC.SG and fourth meditation.LOC.SG emph CONJ
pāk *wewñu*
part.SG say.PP.NOM.SG.PRT
... in the third Dhyāna and in the fourth Dhyāna the part is said. (A 397 b4, translation and segmentation based in Thomas 1967: 268)

(3.29) *śla* *tuṅk* *oki* *nu* ; *räsñā* *poke* , *(paṃ)* *ypamāṃ* :
with love.SG like CONJ stretch.3SG.IMPF arm.SG emph do.MP
[...] but as with love she stretched her arm as she attended [to him]. (A 5 b3, 5/4-3)

(3.30) *marr* *ontaṃ* *nu* ; *puttiśparnäṣ* ; *lotki{ṣ}* *wras(om* :)
not PTCL CONJ Buddhahood.ABL.SG turn.3SG.OPT being.NOM.SG
Never, either, should a being turn away from the Buddha rank. (A 313 b2, 4/4/4)

(3.31) *kucc* *aśśi* *nu* *ptāñkät* *käṣṣi* ///
who/what PTCL CONJ Buddha.SG teacher.SG
How if the Buddha-god the teacher ... (YQ III.10 b5)

Having established that *nu* has the same prosodic status as V-enclitics, we can now use *nu* itself as a diagnostic for other C-enclitics, such as *śkaṃ*.

(3.32) *kuc* *nu* *śkaṃ* *prakṣäl*
who/what.OBL CONJ CONJ beg.GER.NOM.SG
And what should one ask [...] (A 229 a5, translation mine, metrical)

The fact that *śkaṃ* can be separated from its host by *nu* therefore shows that both elements have to be directly dominated by a ClG. Additionally there is one attestation in which *śkaṃ* is preceded by the V-enclitic *ats*.

(3.33) *ṣakk* *ats* *śkaṃ* *ñi* *tmäṣ* *pāk* *plos* *ymā(r*
certainly emph CONJ I.GEN.SG then.ABL.SG part.SG send.2PL.IPV quickly
ṣkārā) ///
back
[...] be sure and send part of it quickly (back) to me! (YQ II.5 b6)

Applying the same reasoning we can now use *śkaṃ* to demonstrate the same prosodic status for the particle *ne*.

(3.34) *kus* *śkaṃ* *ne* *caṣ* *tñi* *waṣtäṣ*
who/what.NOM CONJ PTCL DEM.OBL.SG you.GEN.SG house.ABL.SG
läñclune *pälkānträ*
emerge.ABSTR see.3PL
[...] and those who will observe your going away from the house [...] (YQ II.14 a5)

The same particle can also be preceded by *ats*.

(3.35) *kuss* *ats* *ne* *cam* *peke* *pälkānt*
who/what emph PTCL DEM.OBL.SG paint.1SG.PRT see.3PL.PRT
[...] whoever had seen the picture, [...] (A 9 a6)

Finally, there is a single case in which the emphatic particle *pe* is preceded by another enclitic, namely *aci*.

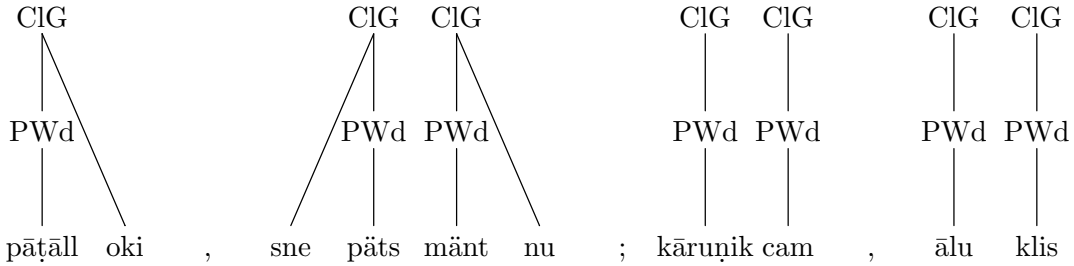
(3.36) *śaci* *rāmbhānäṣṣ* *aci* *pe* *nu*
PN.SG PN.ABL.SG starting.from EMPH CONJ
[...] Śacī, Rambhā, etc., [...] (A 312 b3)

I could not find any attestations in which the disjunction *pat* is separated from its host by another clitic. However the data above should suffice to demonstrate that the pattern observed for V-enclitics (and also proclitics) does apply to TA clitics more generally.

3.2.4.4 A Prosodic Account

We have observed that TA clitics behave as if they formed a single word with their host with respect to meter, even though evidence from external sandhi, hiatus and relative ordering among clitics strongly suggests that they do not. The apparently contradictory nature of this behavior, however, is only due to the traditional definition of *caesura* within Tocharian meter as an obligatory *word* boundary. If we simply take the minimal prosodic unit delimited by a caesura/colon boundary to be the ClG instead, the ban on caesurae between a host and its clitics, as well as between clitics of the same host, directly follows. The example in (3.37) contains a complete verse of the meter 4-4/4-3. The diagram indicates the prosodic structure of the content of each colon up to the ClG.

(3.37) *pāṭāll oki , sne päts mänt nu ; kāruṇik*
underworld like sne ground.SG how CONJ compassionate.NOM.SG
cam , ālu kliṣ :
DEM.OBL.SG other.GEN.PL lead.3SG.OPT
[...] bottomless as the underworld. How could a compassionate one ever bring this [= death] onto others? (A 65 b3, 4-4/4-3)



Since every PWd has to be dominated by a ClG, the generalization equally holds in the absence of clitics, as in the final two cola in (3.37). The ClG, a constituent that has been established for TA based on independent prosodic evidence, straightforwardly accounts for the restricted distribution of clitics in verse.

3.2.5 Summary

This concludes my discussion of TA. We have seen that TA clitics form a prosodic class distinct from both affixes and orthotonic words by patterning with the former in some respects and with the latter in others. The following table gives a summary of these properties.

Table 3.21: Properties shared by clitics and affixes/words

	stem + affix	word + clitic	word + word
a) glide formation of *e* and *o*	-	optional	
b) hiatus at boundary	-	optional	
c) [a] at boundary possible	+		-
d) caesura at boundary possible	-		+
e) external consonant sandhi	-	+	optional

The final property in the table, i.e. the obligatory application of consonant sandhi, sets clitics apart from both suffixes and orthotonic words simultaneously. I have demonstrated that all of these properties can be accounted for by adopting the ClG as a prosodic constituent. The ClG further allowed us to account for the otherwise unexpected behavior of compounds with respect to hiatus.

3.3 The Phonology of TB Clitics

3.3.1 Introduction

In this section I discuss how the analysis proposed for TA clitics above can be extended to a number of particles in TB. Similar to the TA particles, these elements form a prosodic class by showing both affix-like and word-like behavior. This section is organized into three parts. First, I will discuss evidence for TB particles being prosodically dependent in that they a) form a small prosodic constituent with their host and b) are unable to form a PWd by themselves. I then examine data that shows that clitics are not prosodic suffixes, i.e.

they are located outside of the PWd of their host. Finally I discuss how TB compounds, similarly to their TA counterparts, provide evidence for the ClG.

3.3.2 TB Particles are Prosodically Dependent

Instead of attempting to give an exhaustive treatment of all the particle-like elements in TB, I focus on a representative set of core cases of clitic-like elements.

Table 3.22: Selected clitics in TB

Enclitics	
no	'and, but'
ṣpä/ṣp/ṣ/ṣäp	'and, also'
wat	'or'
wa	'therefore'
ka, *tsa*, *nai*, *nta*, *pi*, *ra*	emphatic particles
ksa/kca	indefinite pronoun
ram(t)	'like'
Proclitics (prepositions)	
śle	'with'
snai	'without'

Just as in TA, TB enclitics are located within the same colon as their host, as observed by Bross et al. (2014: 6) in their study on TB meter. This indicates that host+clitic form a relatively small prosodic unit.

In addition to their dependent status with respect to meter, there is evidence that these particles do not form PWds on their own. A subset of the forms in (T:3.22) can be shown not to bear accent, thanks to the accent-governed alternation between the vowels *ä*, *a* and *ā* (on which see the overview on TB accent in chapter 1.1.3). Based on this alternation, the conjunction *ṣpä* 'and' can only be unaccented, as *ä* is the unaccented realization of the underlying segment /ə/. The unaccented status of the conjunction is further confirmed by the variants *ṣ* and *ṣp*, which lack a vowel altogether. The particles containing the vowel *a*, on the

other hand, are ambiguous between an accented /ə/ and an unaccented /a/. Fortunately, for some of these forms we do have historical evidence for the underlying nature of their vowel. In archaic TB, the alternation between *ä*, *a* and *ā* is not systematically realized in the orthography. That is, both accented and unaccented /ə/ can be written as <ä> or <a>, and both accented and unaccented /a/ can be written as <a> or <ā> (for the full details, see Peyrot 2008: 33-41). Therefore, one way to establish the underlying nature of an instance of *a* in Classical TB is to find archaic spellings of the word in question involving either <ä> or <ā>. Krause (1952: 3) observed that the particles *ra* and *wa*, which he already considered enclitics, can be spelled with <ā> in archaic texts. The following table lists additional examples of this type, i.e. where particles that are spelled with <a> in the classical language are rendered with <ā> in an archaic manuscript. Since attestations of this sort are relatively rare due to the small number of available archaic manuscripts, and often ambiguous due to their fragmentary nature, I am giving enough context for each case to show that we are, in fact, dealing with an instance of the clitic in question and not just a fragment of a larger word.

Table 3.23: Archaic spelling of TB particles

classical form	archaic spelling	
ra	# allek poyśī ; ot rā mākā ; sne$_{u}$kī ynāñmä #	^{A}THT 273 b4
wa	# snai keś <u>wā</u> wes ; cī saim yāmoṣ ;	^{A}THT 273 a5
ksa	/// ste nesäṃ <u>ksā</u> :	^{A}THT 139 a4[17]
kca	# mantā <u>kcā</u> nauṣ śärsāre ;	^{A}PK AS 12C b4
ram	; maistsi <u>rām</u> no ; päknāskenträ	^{A}THT 273 a2
wat	# yust-me <u>wāt</u> nai ; tallāñciśkaṃ ;	^{A}THT 273 b3[18]

[17]On *ksā* in connection with unaccented status of of the indefinite see also Hilmarsson (1996: 183-4) and Pinault (1997: 470).

[18]I am deviating from the transcription in Sieg et al. (1953: 166) (also on CEToM), who make the segmentation *wā tnai*, which they then correct to *wā* {*n*}*nai* (see also Peyrot 2013: 368-9). This is probably

The combination of the classical spelling with <a> and the archaic spelling with <ā> lends support to the view that these elements[19] are unaccented, at least within the classical language.

Finally, there is also evidence that the preposition *śle* 'with' (the cognate of TA *śla*) is unaccented. It has an orthotonic counterpart *śale* that is exclusively used without an object, in which case it is usually translated as 'likewise' (e.g. Adams 2013: 680). The form *śale* must have the underlying representation /śəle/ based on the TB accent rules, which exclude a word-final vowel from being accented. If we assume that both forms share the same underlying representation, it follows *śle* has to be unaccented[20] in order to escape accent retraction onto the /ə/.

Additional evidence for the prosodic dependency of at least some particles comes from phonotactics. The particles *ṅke* and *nta* both start in a nasal followed by a homorganic stop, a type of cluster that is otherwise unattested word-initially. This observation is accounted for if we assume that TB[21] has a constraint against PWd-initial clusters of this type and that *ṅke* and *nta* are exempt from it as they do not form PWds.

Another symptom of the prosodically weak status of TB particles is the behavior of the conjunction *no* in combination with *ramt* 'like'. The latter is regularly realized as the

---

because the disjunction *wāt* does not seem to fit the context. However, the combination of *wat* and *nai* is also attested in THT 30 a1, where *wat* also seems to have a purely connective function rather than expressing the meaning 'or'. Therefore, the assumption of a scribal error is unnecessary.

[19]I could not find any secure cases where *ka*, *tsa* or *nta* are spelled with <ā> in archaic manuscripts.

[20]The exact distribution of *śale* and *śle* as actually more complicated than is described here. *śle* can also appear without a nominal complement, in which case it is subject to a rule that deletes accented /ə/, mostly in verse. See Thomas (1979) on the details of *śale* and *śle*, and Winter (1990) and Pronk (2009) on the deletion of accented schwa.

[21]TA appears to have the same constraint, which would explain why an underlying sequence such as /ntā/ is realized as *äntā* with a prothetic schwa.

allomorph *ram* before the conjunction.[22]

(3.38) *; tapāki ramt , ñ(ä)kcy(ai) lak$_u$tsai ;*
mirror.SG like divine.SG shining.SG
like a divine shining mirror (THT 73 a1, translation based on Thomas 1968: 211)

(3.39) *ylaiñäkte ram no ; lare-yok saswe •*
PN.SG like CONJ lovely lord.SG
Like Indra, the lord [is] looking lovely; (PK NS 36 and 20 b2, 5/5)

Thus *ram no* behaves like a sequence of suffixes by showing idiosyncratic morpho-phonological behavior, a phenomenon that is also frequently observed with clitic chains.[23]

3.3.3 Particles are PWd-External

So far we have seen evidence that the particles under discussion do not form PWds by themselves and are in a close prosodic relationship with their host. By itself this would also be compatible with treating them as prosodic suffixes. We therefore still require additional evidence that shows that the elements in question are located outside of the PWd of their host. Within TA this followed from the behavior of V-initial particles and the fact that they either trigger external sandhi or result in hiatus. In TB, on the other hand, there are no well-attested particles that begin with a vowel, and therefore neither X#C-sandhi nor hiatus come into play. The two proclitic prepositions *śle* 'with' and *snai* 'without' are more informative in this respect. Here we do find cases of hiatus between the proclitics and their host.

[22]The allomorph *ram* occasionally occurs in other environments as well, but before *no* it is the only allomorph attested.

[23]A comparable example from another Indo-European language is the Hittite enclitic quotation particle *=wa* which is realized by the allomorph *war* if a following clitic begins in a vowel (Hoffner and Melchert 2008: 354), even though Hittite does not have an intervocalic r-insertion rule outside of its clitic system.

Table 3.24: Hiatus with preposition

śle aklaṣlyeṃ (late form)	'with a student'
snai akālk	'without a wish'

The occurrence of hiatus is just as significant as in TA, since TB has the same ban on word-internal hiatus, which is why the proclitics have to be outside the PWd.[24]

Returning to the enclitic particles, we have seen that the conjunction *ṣpä* 'and' shows suffix-like behavior by virtue of being unaccented, as well as by surfacing as the non-syllabic variants *ṣ* and *ṣp*. However, there is another aspect to the variant *ṣpä*, which distinguishes it from suffixes as well as from orthotonic words. What is crucial about this form is that not only is its only vowel /ə/, but it is located in final position. The fate of word-final /ə/ in TB is rather intricate and has been described in full detail by Malzahn (2012a). In Archaic TB, word-final /ə/ was still realized as *ä*, while in Classical TB prose it is generally realized as Ø, due to apocope. However, in metrical contexts, where the number of syllables within a word is crucial for the well-formedness of a verse, final /ə/ can optionally still be realized. The vowel that surfaces in these cases is either *ä* or, in the majority of cases, *o*. However these facts are to be interpreted linguistically, what is crucial for our purposes is that the *ä* in *ṣpä* shows an entirely different distribution.

Table 3.25: Variants of *ṣpä* as attested in CEToM

	Archaic	Classical	Late
ṣ	2	104	28
ṣp	3	75	0
ṣäp	27	28	2
ṣpä	27	199	9

[24]I should note that I am treating *snai* as a proclitic primarily based on its category membership with *śle*, as well as its TA cognate *sne*, whose prosodic dependence on its complement could be backed up with evidence from the meter. There is nothing about the phonological structure of *snai* itself that would distinguish it from orthotonic words.

The table shows that the variant with final *ä* within the classical language, where final /ə/ in prose is usually realized as Ø, is attested roughly as frequently as all the other variants combined. Furthermore, *ṣpä* is in no way restricted to metrical contexts. The current CEToM corpus contains 70 attestations in manuscripts that are marked as consisting exclusively of prose and 111 in exclusively metrical manuscripts. Finally, despite the robust attestation of the variant in *ä*, there is not a single attestation of a hypothetical **ṣpo*. Therefore, the form *ṣpä* appears to be exempt from the apocope rule that affects PWds in TB prose. We can account for this observation with the assumption that *ṣpä* is not part of any PWd. In other words, it neither forms a PWd of its own, nor is it part of the PWd of its host.

Another property that separates TB clitics from suffixes is their behavior with respect to Marggraf's rule of accent retraction. Suffixes allow an underlying accent to be realized on a preceding stem-final vowel. This is not the case with enclitic particles.

Table 3.26: Particles are located outside the word-accent domain

	a) +enclitic	b) bare form	c) +suffix	
no	*lāre no*	*lāre* 'dear'	*larau-ñe*	'love'
ṣpä	*lakle ṣpä*	*lakle* 'pain'	*läkle-nta*	pl.
nai	*postän nai*	*postäṃ* 'finally'	*posta-ññe*	'later'
ra	*mākā ra no*	*māka* 'much/many'	*makā-ntso*	gen. pl.
wat	*ṣalype wat*	*ṣalype* 'ointment'	*ṣälype-nta*	pl.
ka	*postäṃ ka*	*postäṃ* 'finally'	*posta-ññe*	'later'
ram	*cākkär ram no*	*cākkär* 'wheel'	*cakkar-ntse*	gen.
ṅke	*yākṣe ṅke*	*yākṣe* 'yakṣa' (demon)	*yakṣa-ññe*	'prtng to y.'
ksa	*makte ksa*	*makte* 'self'	*mäkce-pi*	gen.
kca	*totka kca*	*totka* 'little'	*totkā-nts*	gen.pl.
pi	*mākā pi*	*māka* 'much/many'	*makā-ntso*	gen.pl.
nta	*wāki nta*	*wāki* 'difference'	*waki-ssonta*	'different'

Despite their affix-like behavior in other respects, the particles in (T:3.26),[25] unlike suf-

[25]I was not able to find secure examples for *wa* and *tsa*.

fixes, fail to prevent accent retraction on their host. The same kind of independence has long been recognized (cf. Krause and Thomas 1960: 79) for a subset of the so-called secondary case affixes. Based on this observation, Kölver (1965: 5) argued for the non-affixal status of these elements.[26]

Table 3.27: Secondary case morphemes the word-accent domain

		+sec. case	+suffix	
comitative	*lakle* 'pain'	*lakle mpa*	*läkle-ntse*	gen.
allative	*kartse* 'good'	*kartse śc*	*kärtsa-uñe*	'goodness'
locative	*camel* 'birth'	*camel ne*	*cmel-ntse*	gen.
perlative	*lakle* 'pain'	*laklesa*	*läkle-ntse*	gen.

The application of accent retraction in examples such as (T:3.27) led Winter (1967: 739) to the conclusion that the secondary case morphemes in (T:3.27), as well as the ablative morpheme when it does not prevent accent retraction, should be considered postpositions instead of suffixes. Kölver (1965: 6) goes so far as to claim that the secondary case morphemes in TB all have the structure of independent words. However, this is clearly incorrect. Both *mpa* and the perlative morpheme *sa* are not possible PWds by virtue of being unaccented, which is indicated by spellings with <ā> in archaic texts (cf. Stumpf 1974: 408), e.g. in *speltkesā* 'with zeal', *ñäśämpā* 'with me'.[27] Furthermore, the comitative morpheme *mpa*, like the particle *nta*, violates the constraint against word-initial homorganic nasal-stop clusters, while the allative morpheme *śc* is clearly prosodically dependent, as it can surface

[26]The ablative morpheme *meṃ* has a special status in that it shows variation in this respect (for details see Pinault 2006), while the causal morpheme *ñ* always behaves like a suffix. The latter is attested with only a very small number of nouns, specifically nouns denoting some psychological state (cf. Krause and Thomas 1960: 90). Nouns in the so-called causal case may therefore better be treated not as case forms at all but as adverbs reanalyzed from an old genitive formation. On the possible etymological connection between the causal morpheme and the pronominal genitive, see e.g. Pinault (2008: 472-4).

[27]There is also comparative evidence that the underlying vowel of the perlative morpheme *sa* is /a/ as the TA cognate is *ā*. For details on the exact historical relationship between the two forms see Gippert (1987: 26-27).

without a vowel, similar to the conjunction *ṣp*. In fact, both morphemes involve partially parallel sets of variants.

Table 3.28: Variants of *ṣpä* and *ścä*

	Conjunction	Allative
a)	*ṣpä*	*ścä*
b)	*ṣp*	*śc*
c)	*ṣ*	*ś*
d)	*ṣäp*	–
f)	–	*śco*

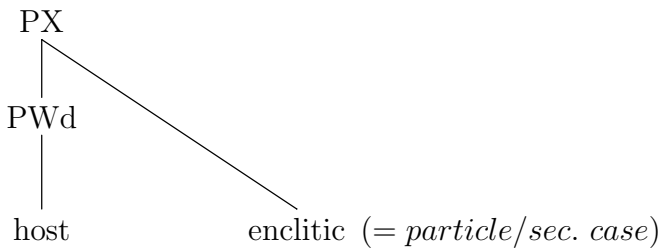
Comparing the variants of the conjunction with those of the allative morpheme, we encounter one potential problem with assigning them the same prosodic structure. I argued above that the lack of a variant in *o* indicates that the conjunction is not subject to the same apocope rule affecting other words in final /ə/ in classical prose. However, the allative morpheme does have such a variant, somewhat unexpectedly if we assume that neither of the two morphemes projects a PWd. There is, however, one piece of evidence that suggests that the final /ə/ in *ścä* does, in fact, have a different status from other instances of final /ə/ that show alternation between *ä*, *o*, and Ø.

Within one non-literary text genre, the so-called caravan passes, the *o*-variant is significantly more common than the variants in *ä* or Ø, while, at the same time, no other instances of word-final /ə/ are realized as *o* within this genre. (cf. Malzahn 2012a: 46-8). The more frequent realization of the final /ə/ of the allative morpheme may indicate that /ścə/, just like /ṣpə/, is PWd-external, although currently no explanation presents itself for why one of the morphemes exhibits an *o*-variant, while the other does not.

We have seen that the secondary case morphemes in (T:3.27) show the same dichotomy between word-like and affix-like behavior as enclitic particles and should therefore be subject to the same analysis. What all of these elements have in common is that they neither form PWds themselves nor are they part of the PWd of their host. We can capture this situation

with the same kind of structure I employed for TA, in which the clitic is directly dominated by a prosodic constituent larger than the PWd. Since I have not yet presented any evidence for the exact identity of the dominating category in TB, I am labelling it PX for the time being.

(3.40)



In the previous section I argued that the category directly dominating clitic particles in TA was the ClG. Now that we have empirically motivated the structure in (3.40), we need to consider a) whether there is evidence for the existence of the ClG in TB at all and b) if so, whether we can identify it with PX in (3.40). In the next section I will review the behavior of TB compounds and show that it points to positive answers to both of these questions.

3.3.4 Compounds

Similar to TA, compounds in TB are often difficult to distinguish from phrasal units, although there is one criterion that enables us to safely identify at least a subset of potential compounds. Bernhard (1958: 21) observes that TB compounds bear a single accent, which is located at the end of the first member. While Bernhard allows for numerous exceptions to this observation, Marggraf reanalyzes it in terms of his accent retraction rule and considers it a necessary property of compound-hood. According to him, the primary compound accent is located on the syllable bearing the underlying accent of the first member (Marggraf 1970: 67).

Table 3.29: Compound accent

Member 1		Member 2		Compound
/maká/	→ *māka* 'much/many'	/yəknə́/	→ *yakne* 'kind, way'	*makā-yäkne*
/totká/	→ *totka* 'little'	/yərmə́/	→ *yarm* 'measure'	*totkā-yärm*

Therefore, the first member of each compound in (T:3.29), unlike their simplex counterpart, escapes accent retraction, while the accent of the second member is deleted entirely. The single word accent combined with the absence of accent retraction gives the impression that compounds in TB are single PWds and, in fact, this is basically the analysis Marggraf proposes.

(3.41) *Derivation of compounds with accent deletion* (Marggraf 1970: 77)

|x x́| + |x x́| + # → |x x́| + |x́ x| → |x x́| + |x x|

Marggraf is operating with a single word boundary #, which correctly accounts for the first member being able to retain its underlying accent. There are, however, two pieces of evidence that indicate that this is not correct. First, TB allow hiatus between the two members.

Table 3.30: Hiatus in compounds

Member 1		Member 2		Compound
/kəntə́/	→ *kante* 'hundred'	/oktə́/	→ *okt* 'eight'	*känte-okt*
/ostə́/	→ *ost* 'house'	/ostə́/	→ *ost* 'house'	*osta-ost* 'house by house'

Since TB does not allow two adjacent vowels within single PWds, the presence of hiatus in the examples in (T:3.30) indicates that compounds form a prosodic unit larger than the PWd.

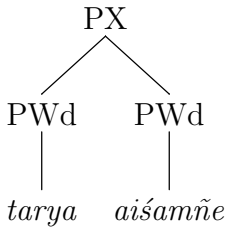
The second piece of evidence comes from Marggraf's own observation that in some cases TB compounds do exhibit two accents, namely when the accented syllable of the second member is not adjacent to the accented syllable of the first.

Table 3.31: Secondary accent in compounds

Member 1		Member 2		Compound
/aścə́/	→ *āścä* 'head'	/sanápalle/	→ **sanāpalle* 'salve'	*aśca-sanāpalle*[28]
/läklé/	→ *lakle* 'pain'	/lyakáñə/	→ **lyakāñ* 'seeing'	*läkle-lyakāñ*
/təryá/	→ *tarya* 'three'	/aiśə́mñe/	→ *aiśamñe* 'wisdom'	*täryā-aiśamñe*

Marggraf accounts for this behavior thusly: The accent of the second member is deleted if it would result in a clash with the accent of a first member and reduced to a secondary accent otherwise. There is only a small number of attestations for this phenomenon, but if Marggraf is correct, the realization of two word accents provides evidence that TB compounds consist of more than one PWd. (I am using PX for the as yet undetermined category dominating the PWds).

(3.42)



We have already seen that PX has to be larger than a PWd, which rules out a recursive structure since TB does not tolerate hiatus within PWds. We can also rule out the prosodic phrase, or there would be no way of distinguishing the accentual properties of compounds from those of PWds forming a small phrase, where both components are accented and subject to accent retraction. In other words, we require a prosodic constituent intermediate between the PWd and the prosodic phrase, and can therefore equate PX with the ClG.

Having established, *pace* Marggraf, that each member of a compound forms its own PWd, we are now forced to reexamine the exact domain of Marggraf's accent retraction rule. The

[28]This form is attested in W-18 a4, which is part of a manuscript edited by Filliozat (1948), not yet included in CEToM.

fact that the first compound member can be accented on the final vowel shows that it cannot be the PWd. I propose that the ban on final accented vowels should be taken as applying to the ClG instead. This correctly exempts the first member of compounds from accent retraction, since its final vowel is no longer final within the relevant domain. Furthermore, it is also not the case that compounds are simply exempt from accent retraction altogether, as observed by Marggraf (see the derivation in 3.41). If this was the case we would expect to find forms such as **ostá-ostá* with a secondary accent realized on the final /ə/ of the second member. Instead, the attested from is *osta-ost* 'house by house', where the final /ə/ is deleted in unaccented position. At the same time, accent retraction within uncompounded nouns (without enclitics) is still predicted to be obligatory, since in this case the right edge of the (only) PWd coincides with the right edge of a ClG.

Now that we have seen that there are indeed empirical reasons to include the ClG as a constituent within TB prosody, we should consider it the prime candidate for the category directly dominating clitics. However, treating the ClG as the prosodic unit containing both compounds and combinations of host+clitic appears to be at odds with the accentual behavior of the latter. If only syllables at the right edge of ClGs were prohibited from bearing accent, we would not expect a host to undergo accent retraction preceding an enclitic, contrary to fact.

Table 3.32: Accent in compounds vs. host+clitic

a) Composition: [totkā́ yärm $_{\text{ClG}}$]
b) Cliticization: [tótka kca $_{\text{ClG}}$]

One potential solution to this problem would be to simply conclude that the prosodic constituent dominating clitics has to be larger than the ClG, such as the prosodic phrase, i.e. that TB clitics are free clitics in the terms of Selkirk (1996). This hypothesis is unattractive from a theoretical viewpoint as it would require a prosodic model that is less restrictive than

both Selkirk (1996), in operating with an additional prosodic constituent, i.e. the ClG, as well as Vogel (2009), in allowing the prosodic phrase to directly dominate a clitic. If we assume instead that, like in TA, compounds and host+clitic in TB form the same prosodic constituent, we are facing an instance of opacity. That is, from a rule-based perspective, (T:3.32b) appears to undergo a rule whose structural description is not met on the surface. The accent on /totká/ gets retracted even though it is not in domain-final position. From a constraint-based view, it seems that a repair strategy has been applied even though the corresponding markedness constraint has not been violated. Such a situation is commonly referred to as counterbleeding.

A similar case of the overapplication of a rule in the context of clitics is discussed by Booij (1995) and Booij (1997) for Dutch. Dutch has a rule that devoices obstruents in a syllable coda. Furthermore, a word-final obstruent gets resyllabified into the onset of a following vowel-initial enclitic. However, resyllabification fails to bleed final devoicing, resulting in surface forms with a devoiced obstruent in an onset position. The following data are taken from Booij (1995: 238).

Table 3.33: Dutch final devoicing before enclitics (from Booij 1995)

vond-ik	/vɔnd ək/	'found I'	$(\text{vɔn})_\sigma(\text{tək})_\sigma$
vond-ie	/vɔnd-i/	'found he'	$(\text{vɔn})_\sigma(\text{ti})_\sigma$
vond et	/vɔnd ət/	'found it'	$(\text{vɔn})_\sigma(\text{tət})_\sigma$

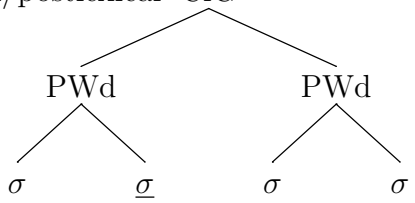
Booij accounts for the data in (T:3.33) in the following way: Coda devoicing takes place within lexical phonology, i.e. at a stage when clitics are not yet present in the phonological representation. Once the clitic is added at the postlexical level the obstruent that is now resyllabified is already voiceless.

I propose to account for the accentual difference between compounds and host+clitic in a similar fashion by taking compounds to be formed within lexical phonology. Since the second member of the compound is already present at this level, the accent at the end of the

first member is not domain-final and does not undergo retraction.

(3.43) *Compounds*

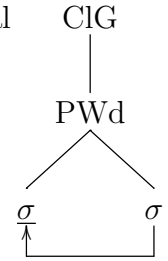
a) lexical/postlexical ClG



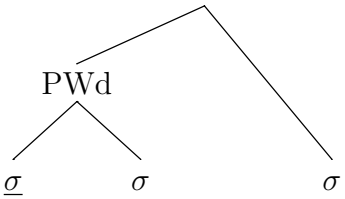
Clitics, on the other hand, are not yet present at the lexical level. Therefore, the final accented vowel of the host is subject to retraction. Once the clitic is added at the postlexical level, there is no reason for the now initial accent to return to its original position. Note that the present analysis presupposes that the ClG is already formed within lexical phonology. This assumption is independently necessary if compounds are formed within the lexicon and I am simply extending it here to uncompounded words.

(3.44) *Host+clitic*

a) lexical ClG



b) postlexical ClG



Note that the above analysis of accent retraction would also be applicable if host+clitic formed a single PWd in TB (as they do in Dutch according to Booij 1995), i.e. the fact that clitics do not prevent accent retraction on their host has nothing to do with the prosodic constituent they form with it. This is in stark contrast with past studies, such as Kölver

(1965: 4-5), which take the accentual behavior of secondary cases as evidence against their status as suffixes. There are, however, still empirical reasons to consider TB clitics PWd-external, such as the possibility of hiatus between V-final proclitics and the initial vowel of their host, as well as the behavior of the final /ə/ in the conjunction *ṣpä* and (possibly) the allative morpheme *-ścä*. In the absence of evidence to the contrary, I adopt the ClG analysis for all the particles and secondary case morphemes discussed above.

3.3.5 Summary

To summarize, I have argued that TB clitics are best analyzed as being dominated by a prosodic constituent intermediate between the PWd and the prosodic phrase, i.e. the ClG. We have therefore arrived at the same analysis as employed for TA, despite the differences in the available evidence. Just as in the previous section, I am providing a list of the phonological properties that set clitics apart from both suffixes and words.

Table 3.34: Properties shared by clitics and affixes/words

	stem + affix	host + clitic	word + word
a) caesura at boundary	-		optional
b) hiatus at boundary	-	optional	
c) final schwa in prose	-	+	-
d) maximal number of accented σ	1	1	2

Chapter 4

On the Internal Structure of Tocharian Wh-Words

4.1 The Accent of TB Wh-words

4.1.1 Introduction

In this section I demonstrate how the analysis of TB clitics argued for in the previous section can be used to solve a persistent problem in the synchronic and diachronic phonology of TB, namely the seemingly irregular accentuation of several Wh-words. After giving a brief overview of the problem, I will argue that a number of TB Wh-words either still contain a phonological clitic or did so in PT.

Most of the diachronic literature on Tocharian Wh-words[1] traces these elements back to combinations of an interrogative element and a (pro)nominal element. The standard etymology of the TB paradigm *kᵤse* 'who/what' involves the PIE interrogative **kʷi/kʷo* and the demonstrative pronoun **so/to*, although the exact details vary somewhat from scholar

[1]cf. Van Windekens (1976: 246), Hilmarsson (1987, 1993, 1996), Ringe (1996: 66), Hackstein (2004, 2005), Adams (2013), Pinault (1989, 1997, 2008).

to scholar. The following etymology is from (Hilmarsson, 1993: 181).

Table 4.1: Etymology of $k_u se$ (Hilmarsson 1993)

TB		PT		PIE
$k_u se$	$<$	$*k^w äsæ$	$<$	$*k^w is\text{-}so$

In accordance with the rule established by Marggraf (1970), a PT form $*k^w äsæ$ should have resulted in TB $*kwase$, as the accent could not have been realized on the final vowel. The problem is not simply a diachronic one, as is indicated by the Wh-words *mäksu* 'which' and *mäkte* 'how'. Both of these forms[2] appear to pose synchronic violations to the accent retraction rule, based on which we expect the first vowel in each case to be realized as *a*. The unstressed nature of the first syllable is further confirmed by the (rare) variant *mkte*, in which the first vowel is unrealized (cf. Winter 1970: 95).

The accentual behavior of *mäkte* led Winter (1970: 95) to consider it a clitic without providing any additional evidence for this claim. There is, in fact, evidence that both *mäkte* and *mäksu* were capable of being accented. Both could occur at the beginning of a sentence hosting an enclitic.

(4.1) *mäksu* *no* *ṣamāne* • *pañäktentse* *wästsintse* *yarmtsa*
which.SG CONJ monk.SG buddha.GEN.SG clothes.GEN.SG measure.PER.SG
wastsi *yamīträ*
clothes.SG do.3SG.OPT
Which monk would make a garment according to the measure of the garment of the Buddha [...] (IOL Toch 247 b4, translation from Peyrot 2013: 335)

[2]I will use *mäksu* to represent the whole paradigm of this pronoun, all forms of which exhibit the same accentual properties.

(4.2) *mäkte* *no* *ñiś* *ceṃ* *tārkau* *saṃsārṣana* *läklentaṃt(s)*
how CONJ I.SG this.OBL.SG dismiss.1SG prtng.to.the.saṃsāra pain.GEN.PL
///

And how can I let go of this ... of the pains pertaining to the saṃsāra? (IOL Toch 156 b5, translation mine)

These examples show that *mäksu* and *mäkte* could occupy a position that is reserved for orthotonic elements.[3,4] Furthermore, the presence of the vowel *ā* in the attested feminine singular forms *mäktā*$_u$/*mäktāw* indicates that the forms of *mäksu* must have been capable of bearing accent.[5] While the second vowel in *mäkte* does not tell us anything about its accentuation, the grammatically similar nature of *mäksu* and *mäkte* (both interrogative forms) makes it unlikely that they would have a radically different prosodic status.

Summing up, I argued that *mäksu* and *mäkte* are not clitics, which is why the exceptional accentual properties of these elements require some other explanation. Instead, I follow Pinault (2008) in assuming that *mäksu* and *mäkte* bear accent on the second syllable. Note further that we cannot attribute the exceptional behavior of the two forms to their status as grammatical function words. Other function words, such as the reflexive pronoun *āñme* and the reflexive adjective *makte* do exhibit accent retraction in the expected environments.

[3]The proclitic preposition *śle* is attested hosting clitics, but only if it occurs without a nominal complement. As with TA *ślak*, most of these cases involve the emphatic particle *k*, indicating that the preposition was accented.

[4]If it were the case that *mäkte* and the following clitic are both unstressed, they could only prosodically attach to the following orthotonic element, functioning as proclitics. This is exactly the analysis proposed by Kloekhorst (2014: 600ff.) for clitics in Hittite. Kloekhorst argues that second-position clitics are *en*clitic when preceded by orthotonic words but *pro*clitic when preceded by clause initial conjunctions, which he considers to be unaccented themselves. Such an analysis is problematic, as it is unclear why the elements in question should be prevented from undergoing prosodic inversion (in terms of Halpern 1995) only if they are preceded by a (proclitic) conjunction. While a detailed discussion of the Hittite facts is beyond the scope of this work, it seems to me that most of Kloekhorst's data are at least compatible with an alternative analysis, in which clause-initial conjunctions are inherently unaccented but can receive accent as a last-resort strategy when required to host enclitics. A somewhat similar situation occurs in Slavic where certain clitics can be stressed when placed in clause-initial position (cf. Embick and Izvorski 1995: 9).

[5]According to Ringe (1996: xxiii) the sequence *ā*$_u$ is the realization of /-áwə/ in word-final position (similarly Pinault 2008: 417).

Table 4.2: The accent of *mäksu* in comparison with other function words

	'which'	'self'	'self'
nom.sg	*mäksu*	*makte*	*āñme*
obl.sg	*mäkce*$_u$	*makce*	*āñm*
gen.sg	*mäkcwi*	*mäkcepi*	*añmantse*

The partial paradigms in (T:4.2) show that both of the reflexives undergo accent retraction whenever the second vowel is located in absolute word-final position, while the initial vowel in *mäksu* is unaccented throughout the paradigm.

4.1.2 Accounting for the Final Accent through Internal Structure

4.1.3 *mäkte/mäksu*

Pinault (2010) provides a detailed diachronic study of the development of the reflexive *makte* compared to the interrogatives *mäksu* and *mäkte*. He operates with the following PT etymologies:[6]

Table 4.3: Pinault (2010) on *mäkte*

	TB		PT
a)	*mäkte*	<	**mæ-k^{w}ä-tæ*
b)	*mäksu*	<	**mæ-k^{w}ä-su*
c)	*makte*	<	**mæ-kät-tæ*

Pinault argues that both the interrogatives and the reflexive continue structures with an initial connective particle **mæ* and a demonstrative pronoun in final position. The presence of the demonstrative is still synchronically reflected in the interrogative forms, each of which

[6]Pinault also discusses the PIE background of the proposed forms. Since I am primarily interested in the synchronic grammar of TB and its development from PT, I am omitting the PIE forms.

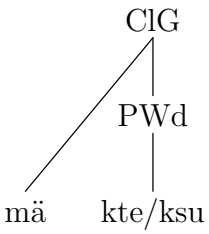
transparently contains an independently attested inflected form of the demonstrative *su*.[7] The only difference between the PT interrogatives and the reflexive is the nature of the second particle. The former contain the same interrogative particle we saw in PT *k^wäsæ 'who/what', [8] while the latter contains a pronominal element *kät. Pinault attributes the lack of initial accent, as well as the special development of the vowel *æ to the status of *mæ as a proclitic. Specifically, he posits a sound change *CæCC > CäCC, where the vowel is either pretonic or located within a clitic (Pinault 2008: 545, Pinault 2010: 360).

It now remains to determine whether *mä* should still be considered a proclitic in the synchronic grammar of TB or whether the accentuation of the interrogatives should be considered a morphophonological idiosyncrasy resulting from its *former* clitic status. Pinault is not fully explicit on the synchronic status of TB *mä* but he does compare its development with that of *päst* 'away', which he considers a synchronic clitic. He further explains the initial accent of the reflexive *makte* through univerbation of the formerly distinct elements in PT *mæ-kät-su, accompanied by the reflexive acquiring adjectival inflection, neither of which took place in the interrogatives. I will adopt the simplest possible hypothesis, namely that *mä* in the interrogatives underwent no further changes and still functions as a proclitic, just like the prepositions *śle* and *snai*. This is also in line with the synchronically productive status of the accent retraction rule (as argued for by Winter 1990: 382). By treating *mä* as a proclitic we arrive at the following (preliminary) prosodic representation:

[7]The internal structure of the interrogative had already been observed by Krause and Thomas (1960). According to their description, the interrogative could also contain forms of the demonstrative *seṃ*, which they based on the plural forms *mäkcenas* and *mäktoynas*. However, Stumpf (1974) demonstrated that the plural forms of the demonstratives *su*, *se* and *seṃ* are identical. Therefore, all forms of *mäksu* can be considered to contain forms of the demonstrative *su*.

[8]On the etymological connection between the *k* in *mäksu* and *mäkte* with the PIE interrogative element in *k^w- see already Van Windekens (1976: 285-286).

(4.3) *Preliminary prosodic structure for* mäkte/mäksu



While this analysis correctly captures the observation that the initial syllable is unaccented in *mäkte* and *mäksu*, one could object that this hypothetical proclitic *mä* would have a highly restricted distribution in TB, occurring exclusively in the two interrogatives in question. This raises the question of what prevented the particle from being reanalyzed in a similar fashion as in the reflexive. Additionally, even if the structure in (4.3) is correct, it does not tell us anything about how the Wh-word k_use could develop out of PT $*k^wäsæ$. In the following section I show that a solution to the latter question can provide some insight into the former.

4.1.4 k_use

(Krause and Thomas, 1960: 50) attribute the development of k_use from $*k^wäsæ$ to weak sentential accent ("unter schwachem Satzaktzent").[9] While they themselves do not provide additional evidence for the special prosodic status of k_use, Pinault (1999) adopts the same hypothesis and connects it to his theory of the diachronic syntax of relative pronouns in Tocharian. The gist of his proposal is contained in the following statement:

> tokh. com. $*kwäsæ$ était faiblement accentué parce qu'il était placé à l'intérieur de la phrase, et non au début. (Pinault 1997: 465)

According to Pinault, this weak accent, resulting from $*k^wäsæ$ being obligatorily placed

[9] A similar account is given by Krause (1952: 12) for *mäkte*, as well as for the personal pronoun *ñäś* 'I' and the adverb *päst* 'away'.

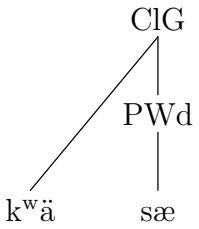
in non-initial position in relative clauses, led to a "contre-accent" (Pinault 1997: 465),[10] a phrasal accent on the second syllable. In short, Pinault's account involves two crucial claims about PT: a) there was a class of weakly accented disyllabic words, which were accented on the second syllable instead of the first when occurring in non-initial position and b) Wh-words in relative clauses were obligatorily placed in non-initial position of the clause. Pinault does not provide any independent evidence for claim a). The only (underlyingly) disyllabic words that surface with an accent on the second syllable in TB are the Wh-words this hypothesis is meant to account for.[11] Furthermore, according to Pinault himself, within the attested stage of Tocharian these elements most frequently occur clause-*initially*, which he considers an innovation. Regarding claim b) Pinault convincingly argues that non-initial instances of Wh-words in relative clauses are a real linguistic property of Tocharian. That is, they are not merely due to stylistic or metrical reasons, nor are they the result of imitating Sanskrit syntax. However, he does not provide any arguments for why Wh-initial relative clauses have to be an innovation and not simply a property inherited into both Tocharian languages from PT.

[10]On p. 472, Pinault explicitly draws a distinction for PT between accented words, unaccented words (i.e. clitics), and weakly accented words, the latter of which receive a "contre-accent" within phrasal phonology.

[11]Hackstein (2000: 101) adopts Pinault's account and provides three additional examples, which he treats as exhibiting accent shift onto the second syllable in enclitics. The first is the postposition *spe* 'close to', whose orthotonic counterpart occurs in the adverb *ysape* 'close'. However, it seems more likely that *spe* is simply unaccented (so Adams 2013: 788), just like the prepositions *sne* 'with' and *snai* 'without'. His second example involves the reduced forms of the copula. In isolation the 3.sg. form is *ste*, which he derives from **skətáe* (with reference to Hackstein 1995: 275-6). However, according to Winter (1970: 101), monosyllabic verb forms, including *ste*, exhibit the same distribution as unaccented particles (see also Bross et al. 2014: 6). Winter (1982: 6) argues that this is also the case in prose. The two forms of the copula that do show accent through their vocalism are 3.sg. *stár-* and the 3.pl. *skentár-*, both of which exclusively occur with suffixed pronouns (e.g. *starne* and *skentarne*, both with a 3.sg. object pronoun). While these forms do exhibit exceptional accentuation, there is an important difference between them and the Wh-words in question. Their accented vowel is never located in word-final position and therefore these forms were always consistent with Marggraf's accent retraction rule, unlike *mäkte* and the PT ancestor of k_use. Finally, Hackstein mentions the Wh-word *kwri* 'if', another case where the PT interrogative element *$k^wä$ failed to receive the word-accent through accent retraction. This, however, cannot be considered independent evidence as this form has exactly the same structure as k_use and therefore belongs to the very set of elements Pinault's phrasal accent rule was designed to account for.

As we have seen, the apparent problem with the accentuation of PT *k^wäsæ is entirely parallel to that of TB *mäksu* and *mäkte*, i.e. the accent fails to be retracted from a word-final vowel. What makes the two cases even more similar is that in both cases the irregular accent falls on a pronominal element. In light of these parallels it seems preferable to attempt a unified account for both of these cases. I propose to extend Pinault's account of the particle *mä* to *k^wä in PT *k^wäsæ.

(4.4) *Prosodic structure for PT* *kʷäsæ



This analysis has the immediate advantage of putting *k^wä into a class of elements whose existence can be independently motivated within the synchronic grammars of TA and TB, i.e. the class of proclitics. PT *k^wä is unaccented for exactly the same reason as TB *śle*, *snai* and *mä*, i.e. it is located outside of the domain within which accent is assigned. It is only to be expected that a PIE phrase such as *$k^w$is so would not be directly reanalyzed as a single word but would first go through a phase in which one component is cliticized to the other. A similar view is also expressed by Kim, who discusses PT *k^wä-sä́ as involving a "suffixed demon. pron." (Kim 1999: 152), although this does not account for the observed accentual facts. Since suffixes are located inside the domain of word-accent, we would not expect *k^wäsä́ to be exempt from accent retraction, if the pronominal element was a suffix.

A clitic analysis for the PIE predecessor of PT *k^wä was already proposed by Hackstein (2004) in the etymology of TB *mäksu*, tracing the latter back to the following PIE clause.

(4.5) **mo(s) -$k^w$is so + u*
which-one who this
which one is this, who this is (Hackstein 2004: 281)

The PIE phrase in (4.5) would have yielded PT *$mä$-$k^wä$-$säw$. Hackstein does not specify his view on the synchronic status of PT *$k^wä$, but if we assume that both it and *$mä$ were still clitics, it follows that neither could have been the target of a retracted accent.

Having argued that treating PT *$k^wä$ as a proclitic can account for the apparently exceptional accentuation of PT *$k^wäsæ$, I will now show that such an analysis can be further motivated by a context-sensitive segmental change that led to the loss of the labial element in *$k^wä$.

4.1.5 The Loss of Rounding in PT *$k^wä$

If the etymologies discussed above are correct and both TB k_use and $mäksu/mäkte$ contain a reflex of the PT element *$k^wä$, we need to determine what caused the loss of the labial element in a subset of the attested forms. The solution in Hackstein (2004: 280) is to analyze the element *-ksu* in *mäksu* as a clitic variant of a hypothetical *k_u-su and thereby to connect it to the development of the indefinite *ksa*, which is historically the clitic counterpart of the Wh-word k_use.[12]

Table 4.4: Hackstein (2004) on loss of rounding

+accent	-accent
k_use	*-ksa*
*k_usu	*-ksu*

One could then extend this analysis to *-kte* in *mäkte*, which appears to contain the neuter form of the demonstrative *se*, the same pronoun that is also contained in k_use (cf. Van Windekens 1976: 286). However, while reducing the loss of rounding in *mäksu* and in

[12]The etymological identity of k_use and the indefinite *ksa* was first pointed out by Hilmarsson (1993: 179) (cf. also Hilmarsson 1996: 184 and Pinault 1997: 471-2). According to Hilmarsson, both forms go back to PIE *k^wis-so, which in unaccented position underwent syncope in Pre-PT. The resulting form was then subject to a sound change that turned the ancestor of PT *$æ$ into *a at the end of monosyllabic words.

the indefinite *ksa* to a single sound change is a desirable result, we have already seen that neither *-ksu* nor *-kte* are clitics. There is, to my knowledge, no evidence that *mäksu* and *mäkte* were any less able to bear accent than $k_u se$, which consistently preserves the rounding of its dorsal.[13]

Pinault (2010: 359) offers a different account of the development of the dorsal in *mäksu/mäkte*. His solution involves a sound change by which a word-internal k^w lost its rounding when it was followed by *s*. In what follows I will refer to this sound change as Unrounding.

(4.6) *Unrounding*

$*k^w > *k$ /X _ s

Additionally, Pinault considers labial dissimilation in $*mä\text{-}k^w su$ as being due to the following *u*. Note that the sound change in (4.6) cannot by itself account for the non-round *k* in *mäksu/mäkte* since we do find internal instances of rounded dorsals preceding consonants, including sibilants.

Table 4.5: Variation between rounded and plain velars

	meaning	/k/[+round]	/k/[-round]
a)	'love'.PER	*täṅkwsa*	*taṅksa*
b)	'brandy'	*śak$_u$se*	*śakse*
c)	'radiance'	*läk$_u$tsauwñaisa*	*läktsauñaisa*
d)	'light'	*läṅk$_u$ci* nom.pl.	*laṅkce* obl.sg.
e)	'trust'.3SG	*pak$_u$trä*	*pakträ*
f)	'years'	*pik$_u$la*	*pikla*

The left examples in a) and b)[14] seem to directly contradict the sound change in (4.6) by containing the sequence /k^{w}s/. The latter sequence is admittedly rare in TB but /k^w/

[13]In later varieties of TB, the dorsal in $k_u se$ and $k_u ce$ can be dropped entirely, which is a separate problem. For details see Peyrot (2008: 69-70).

[14]To what extent *śak$_u$se* 'brandy' can be considered evidence for or against Unrounding depends on the etymology one adopts. Adams (2013: 675) takes it to be a Middle Chinese loan word (see also Tomezzoli and

is also attested before other types of sibilants, as shown in c) and d). These cases can be taken as further counterexamples to Unrounding, since there is no phonetic rationale for why a dorsal should be unrounded before one type of sibilant but not the other, especially considering that /s/ and /ts/ are identical in place of articulation. Rows e) and f) provide further examples of /k^{w}/ preceding coronal consonants.

The right column of the table shows that all of the lexemes in question are also attested with an unrounded dorsal. Ringe (1990: 404) observes that rounded dorsals preceding (voiceless) consonants most commonly occur in western texts, which belong to the archaic layer according to Stumpf (1990). Similarly, Fellner (2006: 60) argues that the forms in the left column belong to a more archaic speech register. What is important for our current purpose is that the change in (T:4.5) happened during the attested history of TB. Specifically, I argue that we are dealing with a phonological neutralization rule that affected rounded dorsals in syllable codas.[15,16] What prevents us from simply treating Unrounding in *mäkte/mäksu* as a subcase of this neutralization rule is the complete absence of the kind of variation we observed in (T:4.5). Despite the very robust attestation of both interrogatives,[17] not a single one involves the spelling <kw> or <k$_{u}$>. I therefore conclude that the loss of rounding in

Kreutz 2011: 72 and Lubotsky and Starostin 2003: 262). Fellner (2006: 60) adopts the same etymology and treats the labiovelar in *śak$_u$se* as a hypercorrection. Pinault (2008: 423) proposes an IE etymology tracing *śakse* back to *deg^{wh}-s-o-*, a derivative of the PIE root *$d^{h}eg^{wh}$ 'burn'. While this development would be in line with Unrounding, it does not account for the existence of the variant *śak$_u$se*.

[15]This generalization includes *pikla*, since forms such as *läklenta* 'pains' and *päklyauṣ* 'hear!' indicate that a word-internal obstruent preceding a sonorant can be syllabified in a coda position. Otherwise, the preceding *ä* would have been subject to deletion in open syllables. However, there appears to be a certain amount of variation with respect to syllabification in TB, as schwa-less forms such as *pklyauṣ* are also attested.

[16]Fellner (2006: 60) provides additional examples for the variation under discussion, including two cases, in which the dorsal appears to be unrounded before a vowel, i.e. *pikala* 'years' (vs. *pikwala*) and *askace* 'darbha-grass' (vs. **askwace*, attested in the gen.sg. *askwacentse*). The former is not conclusive, since the lower part of the akṣara <ka> is destroyed on the manuscript leaf. The latter is somewhat of an anomaly, since before vowels we generally do not find the kind of alternation we saw in (T:4.5).

[17]*mäkte* alone is attested over 200 times within the CEToM corpus.

the interrogatives has to be taken as a separate sound change.

The second part of Pinault's proposal, which involves dissimilation of the dorsal in **mäk^{w}su* by the following *u*, does make sense phonologically, as there is no attestation of a sequence /k^{w}Cu/ or /k^{w}Cw/ in TB. However, one does have to operate with a certain amount of paradigmatic levelling since the only forms of the demonstrative *su* that start with /Cu/ are the nom.sg.masc. and the neuter forms. (Paradigm according to Stumpf 1971a: 4.)

Table 4.6: Paradigm of *su*

	masc.sg.	fem.sg.	neut.	nom.pl.	fem.pl.
nom.	*su*	*sā$_u$*	*tu*	*cey*, *cai*	*toṃ*, *toy*, *to(y)na*
obl.	*ce$_u$*, *cau*	*tā$_u$*	*tu*	*ceṃ*, *ceyna*	*toṃ*, *toy*, *to(y)na*
gen.	*cwi*, *cpi*	*tāy*	*tuntse*	*ceṃts*, *ceynaṃts*	*toṃts*, *to(y)naṃts*[18]

Neither the diphthong *e$_u$*/*au* in the obl.sg.masc., nor the sequence *ā$_u$* in the nom./obl.sg.fem. would likely have led to dissimilation, as is indicated by the form *läk$_u$tsauwñaisa* in (T:4.5). Phonological dissimilation is also conceivable in the gen.sg.masc. and in the fem. pl., where another labial segment follows the initial consonant. However, while levelling *within* the paradigm of *mäksu* is plausible, it is not so clear why the unrounded variant of the dorsal would have been transferred to the adverb *mäkte*, which contains a different demonstrative, i.e. the nom./obl.neut. of *se*. Another drawback of Pinault's proposal is that, unlike Hackstein's account, it draws no connection between the unrounding of the dorsal in *mäksu*/*mäkte* and in the indefinite *ksa*.

If, on the other hand, we analyze the PT interrogative element **k^{w}ä* as a clitic, we now have a means to combine the insights of both Hackstein's and Pinault's accounts for its loss of rounding in *mäksu*/*mäkte*. The PT predecessors of TB *k$_u$se*, *mäksu*, *mäkte* and *ksa* would

[18]I was unable to find an attestation of the variant *toynaṃts*. Therefore, the corresponding entry in the index contains no siglum.

have had the following prosodic structures:

Table 4.7: Prosodic structures of PT Wh-words

a)	$k_u se$	<	[ClG		$*k^w ä$	[PWd $*sæ$]	]
b)	*mäkte*	<	[ClG	$*mä$	$*k^w ä$	[PWd $*te$]	]
c)	*mäksu*	<	[ClG	$*mä$	$*k^w ä$	[PWd $*säw$]	]
d)	*ksa*	<	[ClG	[PWd HOST]	$*k^w sa$		]

The examples in a-c) contain the proclitic $*k^w ä$ attached to a pronominal host, while d) shows the indefinite $*k^w sa$[19] in enclitic position. What all the dorsals, which subsequently lost their rounding in b-d), have in common is that they are 1) non-initial within their ClG and 2) not dominated by a prosodic word by virtue of being part of a clitic. Based on this generalization I propose to modify Pinault's rule for the loss of rounding in *mäksu* thusly:

(4.7) *Unrounding revised*

$k^w > *k$ /[ClG ... X _ s ...]

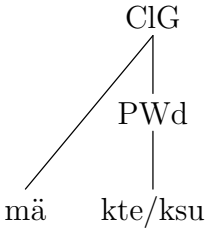
In other words, /k^w/ loses its rounding if it is directly dominated by a ClG and non-initial within that domain. The rule in (4.7) is clearly distinct from the kind of coda-neutralization discussed above, and therefore the fact that it did not lead to variation is unproblematic. It correctly targets *mäkte* and all the forms in the paradigm of *mäksu* without having to resort to any form of analogy. Lastly, it preserves Hackstein's insight that the loss of rounding in the element *-ksu* in *mäksu* can be connected to the development of the indefinite *ksa*, but without having to treat them as morphological or prosodically parallel. The fact that the dorsals in both forms lost their rounding is simply due to the fact that both are part of a clitic.

[19] I am following Pinault (1997: 472) (with the notation $*kwäs\grave{a}$) and Hackstein 2001: 32 in assuming that there was still a labial element in the PT ancestor of TB *ksa*, *pace* Hilmarsson 1993: 179, who posits an unrounded dorsal for the PT form.

4.1.6 The synchronic status of TB $k^{(w)}$

Now that I have argued that PT $*k^{w}ä$ was a proclitic, we may ask about the status of its descendants in TB. Regarding *mäksu/mäkte*, I claimed above that *mä* should be prosodically treated on par with the proclitics *śle* 'with' and *snai* 'without', illustrated by the prosodic representation in (4.3), repeated here for convenience:

(4.8) *Preliminary prosodic structure for* mäkte/mäksu *(repeated)*



This, however cannot represent the full morphological structure of these forms. The fact that every form of *mäksu* contains an independently attested form of the demonstrative *su*[20] indicates that there is morpheme boundary following the element *k*. There are now two possible ways of combining this observation with the treatment of the element *mä* as a clitic. We can either analyze *mä* and *k* as separate morphemes or we can operate with the reanalysis of *mä* and *k* as a single new proclitic *mäk*. In order to evaluate these two options, it is necessary to look at the attested inventory of TB Wh-words more generally. We can first make the generalization that every Wh-word contains one of the interrogative markers *e/in-*, *mäk-*, or $k^{(w)}$-.

[20]The synchronic identification of the pronominal element of *mäksu* and the demonstrative *su* is further supported by Stumpf's observation (Stumpf 1974: 427-8) that the two items share morphological innovations that took place within the attested history of TB.

Table 4.8: Wh-words with *mäk-*

			Pronominal		
masc.		sg.		pl.	
a)	nom.	*mäksu*	/mək-səw/	*mäkcai*	/mək-cai/
b)	obl.	*mäkce*$_u$	/mək-ceu/	–	
c)	gen.	*mäkcpi*	/mək-cpi/	*mäkcenas*	/mək-cenas/
fem.		sg.		pl.	
d)	nom.	*mäksā*$_u$	/mək-saw/	–	
e)	obl.	*mäktā*$_u$	/mək-taw/	–	
f)	gen.	–	–	*mäktoynas*	/mək-toinas/
neut.					
g)	nom./obl.	*mäktu*	/mək-təw/		
h)	gen.	–		–	
			Adverbial		
g)	'how'	*mäkte*	/mək-te/		

Table 4.9: Wh-words with *e/in-*

			Pronominal	
a)	nom.		*intsu*	/in-su/
b)	obl.		*iñcew iñcau*	/in-ceu/, /in-cau/
			Adverbial	
a)	'where, when, if		*ente*, *inte*	/in-te/, /en-te/

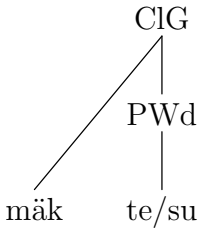
The tables in (T:4.8)-(T:4.9) show that Wh-words with *mäk* and *e/in* are formed in an entirely parallel and morphologically transparent fashion (cf. Adams 2013: 70).

Table 4.10: Combinations of *mäk-* and *i/en-* with *su* and *te*

	su	te
mäk	'which'	'how'
in	'which'	'where, when, if'

Both *mäk-* and *in-* combine with forms of the demonstrative *su* to form a pronominal adjective and with the neuter form of the demonstrative *se* to form an interrogative adverb, although with different meanings. Based on the observation that *mäk-* and *e/in-* compete for the same position within the morphological structure of Wh-words, I conclude that *mäk-* was reanalyzed as a single proclitic yielding the following prosodic structure for *mäksu/mäkte*.

(4.9) *Final prosodic structure for* mäkte/mäksu



According to this analysis the dorsal is still located outside of the PWd in TB. This view receives some support from the phonotactics of the genitive form *mäkcpi*. If the dorsal formed a PWd with the demonstrative this would be the only attested word in the CEToM corpus containing three adjacent obstruents. If we now look at the third type of Wh-words, i.e. those starting with the element k^w, we find a very different picture.

Table 4.11: Pronominal Wh-words ('who, what')

a)	nom.	*$k_u$se*	/k^w-se/
b)	obl.	*$k_u$ce*	/k^w-ce/
c)	gen.	*ket(e)*	/k-et(e)/

Table 4.12: Adverbial Wh-words

a)	'(from) where'	*$k_u$tameṃ/$k_u$temeṃ*	/k^w-tə-meṃ/ /k^w-te-meṃ/
b)	'why'	*kā/kāttsi*	/k-ā/ ?
c)	'if, when'	*kwri/$k_u$ri*	/k^w-ri/ ?
d)	'how much/long'	*kos*	/k-os/ ?

The initial /k^{w}/ from PT *$k^{w}ä$ in $k_{u}se$ and $k_{u}ce$ no longer shows any phonological indication of being a proclitic due to the deletion of schwa in unaccented open syllables. Notice also that in the class of Wh-words shown in (T:4.11) and (T:4.12), only a subset can be decomposed into independently attested morphemes. However, the observation that all the forms start in a dorsal, together with the fact that all the other Wh-words start with a transparently separate interrogative marker, indicates that the forms in (T:4.11) and (T:4.12) had internal structure as well. Therefore, while /k^{w}-/ was likely still a separate morpheme in TB, its status as clitic or affix is unclear.

Finally, the overall compositional nature of Wh-words in TB may also provide some explanation as to why *mäksu/mäkte* never got univerbated (with concomitant accent retraction) as did the reflexive *makte*. The overall pattern of Wh-words always consisting of an interrogative and a (pro)nominal element would have prevented this kind of reanalysis.

4.1.7 *ente/inte* and *intsu*

In the previous sections I argued that all Wh-words in TB that transparently contain a pronominal element involved a proclitic, at least as late as PT. In this section I will address the question of whether we can make a similar case for the element *e/in* in *ente/inte* and *intsu*. Such an analysis seems immediately plausible considering the parallel nature of Wh-words containing *mäk* and *e/in*. Unfortunately, apart from this similarity we have no way of determining the location of the accent in the forms containing *e/in*, since none of the attested forms contains a central vowel. There is, however, some indication that the particle had the status of a proclitic at least in PT.

Before considering the comparative evidence, we have to evaluate the status of the variation between *e* and *i* in *ente/inte* and the lack thereof in *intsu*, since it poses a challenge to the view that the particle in all of these items can be traced back to the same form in

PT. The corpus exhibits the following distribution of forms containing *en-* and *in-*.[21] The distribution of *ente* vs. *inte* is taken from Peyrot (2008: 172), whereas the count of *intsu* is based on the CEToM corpus. The table also contains the equivalent forms in TA for comparison.

Table 4.13: Distribution of TB forms in *in-* and *en-*

	intsu	13	*inte*	15
	**entsu*	0	*ente*	75
TA	*äntsaṃ*		*äntan-ne*	

Pinault (2008: 545-6) projects the variation back to PT, positing an interrogative particle **æn-* that alternated with **än-* in unaccented positions. One problem with this analysis is that there is no evidence for this kind of alternation in TA. Similarly, Adams (2013: 69) posits a PT variation between accented **ænä* and unaccented **änä*, treating the TA variant *antsaṃ* 'which' as a continuation of the former. The attested form *antsaṃ*, however, is likely a scribal error for *äntsaṃ*.[22] Stumpf (1990: 125) and Peyrot (2008: 127) argue that the variation between *ente* and *inte* is an inner-TB development, the latter form being an innovation. Peyrot further supports this view by showing that only *ente* is attested in archaic texts. This raises the question of why we do not find the same kind of alternation for *intsu*. Considering the fact that the initial particle in both forms corresponds to TA *än-* and the small number of attestations of *intsu* in general, I will assume that the lack of alternation is simply an accident. This view is at least consistent with the observation that (within the CEToM corpus) forms of *intsu* are only attested in classical and late texts. Another factor

[21] In the table *intsu* represents all the different attestations of the pronoun, i.e. *intsu*, *iñcew* and *iñcau*, within the current CEToM corpus.

[22] The same manuscript page A 385 also contains an instance of *äntsaṃ*. Furthermore, the only difference between <a> and <ä> is that the latter involves two dots above the akṣara. The scribe made the exact same mistake on the next line, writing *cämpaṣ* instead of *cämpäṣ*, although the latter also occurs on the same line. A similar case is *antā ne*, which Sieg et al. (1931: 182 fn. 2) consider a misspelling for *äntā ne* (see form index under {*ä*}*ntā ne*).

that may have contributed to the absence of **entsu* is another sound change within TB discussed by Stumpf (1990: 125) and Peyrot (2008: 59-60), where an *e* undergoes raising to *i* next to a palatal consonant. This change could have affected all the forms of *intsu* in which the pronominal element started with /c/, such as *iñcew*, possibly affecting the whole paradigm through levelling. Unfortunately, the number of forms that attest this rule is relatively small and Stumpf and Peyrot only give a single example where the vowel *precedes* the palatal.[23,24] Adopting Stumpf's and Peyrot's view, we only need to operate with a single PT ancestor of the particle in question.

(4.10) PT **än-* > TA *än*, TB *en*

I will follow Pinault (2008: 545) in deriving PT **än-* from **æn-*.[25] According to Pinault, the change in question involved the reduction of **æ* in an unaccented closed syllable. However, such a sound change predicts the wrong outcome in the following cases.

Table 4.14: Development of unaccented initial **æ*

(Ringe 1996: 99)

PT		TB	TA
**æṅk^{w}ǽ* 'man'	>	*eṅkwe*	*oṅk*
**ærk^{w}ǽnt-* 'black'	>	**erkent*[26]	**arkant*[27]

[23]The example involves forms of the word for 'iron', attested as TB *iñcwo* and TB *eñcuwo*, as well as the derived adjective TB *iñcuwañ̃eṃ* and TB *eñcuwañeṃ*.

[24]Peyrot (2008: 172) points out that the form *inte* could have arisen in analogy to *intsu* but leaves open the origin of the *i* in the latter.

[25]This step is necessary since the **ä* in **än-* must have postdated the rule that lowered initial schwas to **æ* (see Ringe 1996: 99).

[26]Attested in the fem. pl. *erkenta*

[27]Attested in the masc. obl. pl. *arkañcäs*

Since the outcome of word-initial PT *ä is indistinguishable from that of PT *æ in TB, assuming that both initially yielded *e*, it is the TA form that shows that PT *æ did not get reduced in the examples in (T:4.14). It is also striking that the Wh-words starting in *än-* are the only cases in TA that show a word-initial *ä* (in TB there are none).

If we now take as an additional condition for the reduction of PT *æ that the vowel was part of a clitic and therefore located outside the prosodic word, we can distinguish the development of PT *æ in (T:4.14) from that in the particle PT *æn-. In fact, we have already seen above that Pinault (2008: 545) used exactly such a rule to account for the development *mæ- > *mä- in the forms that yielded TB *mäkte* and *mäksu*, describing the change as *CæCC- > *CäCC- affecting proclitics. Through a slight adjustment we can formulate a sound change that targets both *mæ- and *æn-.

(4.11) *PT Vowel-Reduction*

*æ > *ä / _ C]$_\sigma$ where *æ is directly dominated by a ClG.

Thus there is some evidence that PT *mæ- and *æn- were not only functionally similar (both combining with the same demonstrative), but were also both phonological clitics. While TB *mäk-* still wears its clitic status on its sleeve due to its unaccented schwa, the situation with *e/in-* is not so clear. However, considering the parallel morphological structure of forms with *mäk-* and forms with *e/in-* there is no motivation to assume that the latter but not the former should have undergone a change that incorporated it into the PWd.

4.1.8 Summary

The purpose of this section was to demonstrate that the analysis of clitics motivated in previous sections on a number of TB morphemes can straightforwardly account for the accentual properties of Wh-words in both TB and in PT. On the level of PT, I showed that treating *k^wä as a proclitic, as Pinault (2008) did for the particle *mä, allows us to

give a unified account for the accentuation of both PT *k^{w}äsæ and TB *mäksu/mäkte*. At the same time it connects the unrounding of the dorsal in *mäksu/mäkte* and the indefinites *ksa/kca*, combining the insights from Pinault (2010) and Hackstein (2004). I further argued that a similar analysis of the particle *æn- may explain why its vowel underwent a different development from other instances of initial *æn- in closed syllables. Within TB I argued that the reanalyzed element *mäk-* should be treated as prosodically on par with the proclitic prepositions *śle* and *snai*, which is also in line with the general compositional nature of TB Wh-words.

4.2 Tocharian Wh-Words as Indefinites

4.2.1 Introduction

The present section addresses another aspect of Wh-words in Tocharian, namely their use as indefinites in TA. I will begin by giving what is currently considered to be the indefinite paradigm of TA, followed by a review of Pinault's approach of etymologically connecting the indefinite paradigms of TA and TB. I argue that not only is such a connection untenable but that TA did not, in fact, possess a special indefinite paradigm at all. Instead, Wh-words could simply double as indefinites, a fact that has been obscured by a neutralization rule that rendered a subset of Wh-words phonologically indistinguishable from demonstrative pronouns.

4.2.2 Indefinites in Tocharian A

The TA forms *saṃ*, *caṃ* and *ke* are commonly considered to form a single indefinite paradigm,[28] which has a functional equivalent in TB.

[28]See e.g. Krause and Thomas (1960: 166), Adams (1988: 157), Pinault (1989: 119), and Pinault (2008: 546).

Table 4.15: Indefinites in TA and TB

	TA	TB
Nominative	*saṃ*	*ksa*
Oblique	*caṃ*	*kca*
Genitive	*ke*	*ketra/ketara*

What is peculiar about the TA paradigm is that it seems to overlap with two other pronominal paradigms: The nominative and the oblique are identical with the same case forms of the far-deictic pronoun, while the genitive is identical to that of the Wh-pronoun.

Table 4.16: TA indefinites compared with Wh-pronouns and demonstratives

	Wh-pronoun	Indefinite	Deictic Demonstrative
Nominative	*kus*	*saṃ*	*saṃ*
Oblique	*kuc*	*caṃ*	*caṃ*
Genitive	*ke*	*ke*	*cani*

A diachronic development from demonstratives to indefinites is certainly not unheard of (see e.g. Prince 1981 for the indefinite use of English 'this'). It is not at all clear, however, why only the nominative and the oblique would have undergone this kind of change, with the genitive being supplied by the interrogative paradigm.

Interestingly, contrary to the current standard view, Sieg et al. (1931) posit no special indefinite paradigm for TA. Instead, they equate the *ke* used as an indefinite with the genitive of the Wh-pronoun, observing that Wh-pronouns can, especially in negated contexts, double as indefinites.[29] Sieg et al. (1931: 189) further observe that the oblique of the Wh-pronoun *kus* can function as an indefinite in the expression *kuc yärmaṃ* 'to some extent'. While it is true that *kuc yärmaṃ* is the only attested expression in which *kuc* functions as an indefinite, the number of attestations where the genitive *ke* is used in that function is similarly limited

[29] "In Verbindung mit der Negation dienen die Interrogativa auch als Indefinita" (Sieg et al. 1931: 188).

and only occurs in two examples (listed in Sieg et al. 1931: 188). Based on the indefinite uses of *kuc* and *ke*, we can reasonably hypothesize that the TA Wh-pronoun *kus* 'who/what' could generally double as an indefinite, a typologically extremely common phenomenon. The lack of any nominative attestations in indefinite functions can be viewed as an accident, especially in light of the small number of attested indefinite forms in general. Such an analysis receives further support from another Wh-paradigm that could be used in indefinite function. Sieg et al. (1931: 188) provide one example in which the genitive of *äntsaṃ* 'which' is used as an indefinite.

(4.12) *mā* *äñcanik* *caṃ* *(wraṃ* *klyossi* *yāmäs)* ///
not who/what.GEN.SG DEM.OBL.SG water.LOC.SG hear.INF do.3SG.PRT
He let nobody hear of this issue. (A 294 b2, restoration by and translation based on Sieg et al. 1931: 188)

This example is admittedly highly fragmentary but the interpretation of *äñcanik* as an indefinite is supported by the Uyghur translation of the passage.[30] Additional support for the indefinite use of *äntsaṃ* comes from the particle *ontaṃ*, which Carling et al. (2009) translate as 'somehow, anyhow' and which Sieg et al. (1931: 189) already considered treating as an indefinite. The particle *ontaṃ* looks like the neuter form of *äntsaṃ* preceded by an element *o*, which Van Windekens (1976: 336) considers to be an emphatic prefix (cf. also Adams 2013: 373). While the neuter form **äntaṃ* is not attested by itself, it does occur in interrogative function as part of the Wh-adverb *äntanne* 'where' with the adverbial suffix *ne*.[31] The view

[30]See Geng and et al. (1988: 56-7), who transcribe and translate the relevant passage thusly (brackets indicate restored text).

[kimkä] ymä nang nom nomlayu yrlïqamadï-ï (.)
Noch [keinem] hat er die Lehre zu predigen geruht.

We can render the translation in English as "He has not yet deigned to teach the dharma to anybody."

[31]The adverbial suffix *ne* is most likely etymologically identical with the relative particle *ne*. However, synchronically the two are clearly distinct entities, as can be seen from the fact that the two can cooccur. In fact, all attested cases of *äntanne* are followed by the relative particle, yielding *änta(n)ne ne*.

that *ontaṃ* goes back to an indefinite is further supported by the observation that other Tocharian indefinites are used in particle-like function as well (cf. Malzahn 2012b: 156 on TB *ksa*). The following is an example from TB.

(4.13) *mäkte* *pi* *kca* *tā* *oṅkorñai* *ñiś* *śwātsi* *källālle*
how emph indef this.OBL.SG rice.SG I.SG eat.INF obtain.GER
ṣeym
imperfect.1SG.IMPF
How at all could I get to eat this rice porridge? ($^{\text{L}}$THT 107 a3, late)

In (4.13) the oblique form of the indefinite *ksa* clearly does not function as an argument of the verb and serves as a particle instead. We find a similar situation in TA.

(4.14) *mäñ* *caṃ* *klyom* *wañi* *te* *napeṃsaṃ*
how indef noble.NOM.SG pleasure.SG are human.LOC.PL
What, o noble one, is plaesure among mortals? (YQ II.13 a4, translation based on Geng et al. 2004: 364)

The problem here is that, due to the homophony between indefinites and the forms of the demonstrative *saṃ*, the form *caṃ* in (4.14) is ambiguous. However, due to the parallel nature of this example and the unambiguous (4.13), *caṃ* likely represents the indefinite.

Returning to the indefinite paradigm itself, based on the observation that indefinite use of Wh-words in TA is not restricted to the genitive of *kus*, I argue that *ke* should be removed from the paradigm involving *saṃ*/*caṃ*. This leaves us with the following paradigm of specialized indefinite forms.

Table 4.17: Specialized indefinites

Nominative	*saṃ*
Oblique	*caṃ*
Genitive	?

If this hypothesis is correct and TA Wh-words could generally double as indefinites, the question arises as to the exact status of *saṃ* and *caṃ* within TA grammar. Ideally, the

resulting analysis should also account for the homophony of the two forms with the far-deictic demonstrative. I will first review and argue against the account of Pinault (2008), who treats TA *saṃ* and *caṃ* as related to the TB indefinites *ksa* and *kca*. I will propose a new account, which treats indefinite *saṃ* and *caṃ* as a development internal to TA.

4.2.3 Indefinites in Tocharian B

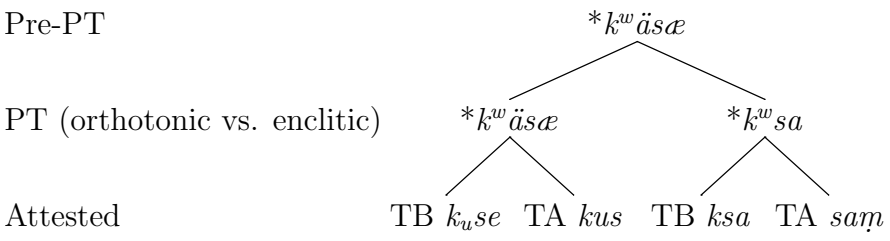
The indefinite paradigm of TB differs considerably from that of TA in that all three forms have a significant affinity with the paradigm of the Wh-pronoun *$k_u$se* 'who/what'. At the same time, none of the forms resembles any of the demonstrative paradigms (in the table I have only listed the paradigm that phonologically corresponds to TA *saṃ*).

Table 4.18: TB indefinites compared with Wh-pronounds and demonstratives

	Wh-pronoun	Indefinites	Deictic Demonstrative
Nominative	*$k_u$se*	*ksa*	*seṃ*
Oblique	*$k_u$ce*	*kca*	*ceṃ*
Genitive	*kete*	*ketra/ketara*	*cwi*

Hilmarsson (1993: 179), Hilmarsson (1996: 184), and Pinault (1997: 471-2) account for the similarity between the paradigms of *$k_u$se* and *ksa* by treating them as the stressed and unstressed counterparts of a common Pre-PT ancestor. Pinault's approach to accounting for the TA indefinites *saṃ* and *caṃ* is to derive them from the same PT forms as their functional equivalents in TB (Pinault 1997: 470 and Pinault 2008: 547-8). The following is a simplified sketch of the development of Pre-PT **k^wäsæ* under this view.

(4.15)



If we assume, as Pinault does, that TA *saṃ* and *caṃ* share a common source with TB *ksa* and *kca*, we would indeed have a motivation to include the genitive *ke* in the TA indefinite paradigm. All three forms would then derive from the same Pre-PT paradigm of $*k^{w}äsæ$. However, since the TA forms cannot be derived from PT $*k^{w}sa$ via regular sound change (unlike the TB forms), Pinault has to posit two additional steps, each of which is problematic.

Pinault attributes the loss of the initial dorsal to cluster reduction, comparing the change to the loss of the dorsal in TB $k_{u}se/k_{u}ce$, yielding *se*/*ce*. This comparison falls short in that it is exactly the orthotonic forms that undergo the change in TB, while the consistently enclitic forms *ksa*/*kca* are not attested without the dorsal. Moreover, we do find initial C-clusters in the TA clitics *śla* 'with', *sne* 'without', and *śkaṃ*, so the reduction rule in question would have to target a very specific type of cluster (e.g. two initial obstruents that are [-continuant]). Pinault (2008: 547) further argues that the initial dorsal of a Pre-TA $*ksā$ could have been lost when following *ālak* 'other', yielding the attested *ālak saṃ* 'some other' (transl. form Carling et al. 2009: 51) from Pre-TA $*ālak\ ksā$ via simplification of the resulting geminate. However, only a subset of the forms of *ālak* end in *k* and could therefore induce such a change.[32] The obl.sg. *ālakäṃ caṃ* could then only be explained by paradigmatic leveling.

Even if we take this hypothesized cluster reduction for granted, the expected outcome of $*k^{w}sa/*k^{w}ca$ would be Pre-TA $*sā/*cā$. The former would have been homophonous with the fem.nom.sg. stem of the demonstrative pronoun.

[32]Paradigm of *ālak* 'other' (from Pinault 2008: 548).

	masculine	feminine
nom.sg.	*ālak*	*ālyāk*
obl.sg.	*āl(y)akäṃ*	*ālyäkyāṃ*
nom.pl.	*ālyek*	*ālkont*
obl.pl.	*ālykes*, *ālyekäs*	*ālkont*
gen.pl.	*ālu*	

Table 4.19: Feminine singular demonstrative stem

Nominative	**sā-**
Oblique	tā-
Genitive	te-

Pinault argues that, to avoid this homophony, the indefinite was partially assimilated to the demonstrative paradigm *saṃ*/*caṃ*. Even if one accepts the existence of language change that works in such a teleological fashion, the advantage of the resulting morphological system is not at all clear. The speakers would have traded one potential homophony (nom.sg.fem. *sā*-) for two (nom.sg.masc. *saṃ* and obl.sg.masc. *caṃ*). Furthermore, TA demonstrative stems never surface in isolation but always feature a deictic suffix (*sās*, *sām*, *sāṃ*). Thus, on the word-level, **sā* would have been unambiguous and it was the morphological assimilation to the demonstrative pronoun that *introduced* homophony.

Thus, I conclude that we should treat TA *saṃ* and *caṃ* as unrelated to TB *ksa* and *kca*. Instead I argue that the TA forms are to be connected with a different Wh-paradigm, namely that of *äntsaṃ*.

Table 4.20: TB indefinites compared with Wh-pronounds and demonstratives (revised)

	Wh-pronoun	Indefinites	Deictic Demonstrative
Nominative	*äntsaṃ*	*saṃ*	*saṃ*
Oblique	*äñcaṃ*	*caṃ*	*caṃ*
Genitive	*äñcani-k*	(**cani*?)	*cani*

Each attested form[33] of the Wh-paradigm *äntsaṃ* transparently contains the corresponding case form of the demonstrative paradigm *saṃ* (see e.g. Krause and Thomas 1960: 166,

[33]The fact that the genitive of *äntsaṃ* is only attested with the emphatic particle -*k* can be considered an accident, since the particle -k can productively attach to pronominal forms, as well as other categories (see below).

Pinault 2008: 545).[34] The interrogative particle *än-* is the cognate of TB *e/in-* discussed above. We have already seen that in TB there is a strong affinity between indefinites and the Wh-pronoun $k_u se$. I argue that, in a similar (albeit not identical) fashion, the TA indefinites *saṃ/caṃ* are related to the paradigm of *äntsaṃ*. The homophony of indefinites with deictic demonstratives is then a side-effect of the relationship between the demonstrative *saṃ* and the Wh-pronoun *äntsaṃ*. A relationship between the indefinite and *äntsaṃ* (as opposed to the demonstrative) is also more plausible from a morpho-syntactic viewpoint, as the use of Wh-words as indefinites (including *äntsaṃ* itself) has already been established by Sieg et al. (1931), and is further attested in numerous other languages (including ancient IE languages such as Greek, Sanskrit, Latin, and Hittite).

The morpho-syntactic relationship between *äntsaṃ* and the indefinite *saṃ* is thus unproblematic, but we still need to establish the exact phonological relationship between the two. TA has a well attested phonological rule that deletes a nasal before a coronal obstruent.[35]

(4.16) *N-deletion*: [+nasal] → Ø/_ [+coronal, -sonorant]

Table 4.21: Examples of N-deletion

	Meaning	a) *Form with nasal*	b) *N-deletion*
a)	good.NOM.PL	*kra<u>ñ</u>ś*	*kraś*
b)	all.FEM.PL	*po<u>n</u>tsāṃ*	*potsāṃ*
c)	queen.GEN.SG	*lā<u>n</u>tse*	*lātse*
d)	split.3PL	—	*kotnec*
e)	good (deed).PL	*kāswoney**<u>ä</u>**<u>n</u>tu*	*kāswoneytu*

[34]The *t* in *äntsaṃ* is due to a completely automatic phonological rule of excrescence between a nasal and a following sibilant.

[35]Krause and Thomas (1960: 68) described the phenomenon as deletion of anusvāra before a sibilant, which does not cover cases such as (T:4.21) d) and e).

Column a) provides a list of examples in which the nasal in question is preserved, while b) shows the same forms having undergone N-deletion. All of the forms in the table are attested,[36] indicating that we are dealing with an optional synchronic phonological rule of TA. Example e) further shows that N-deletion bleeds[37] schwa-epenthesis. This is unsurprising as schwa in TA is banned in open syllables. Therefore, hypothetical forms such as **kāswoneyätu* would be ill-formed. The rule in (4.16) is then exactly what we need to phonologically relate *äntsaṃ* and (indefinite) *saṃ*.

Table 4.22: Deriving *äntsaṃ* and *saṃ* from the same representation

Underlying	Phonological processes	Surface
	(epenthesis + excrescence/assimilation)	*äntsaṃ*, *äñcaṃ*
/nsan/, /ncan/	→	
	(N-deletion)	*saṃ*, *caṃ*

One difference between the forms in (T:4.21) and (T:4.22) is that in the former the deleted nasal is word-internal, while in the latter it is initial (at least underlyingly). However, there is at least one additional case where initial nasal deletion has been posited on entirely independent grounds. TA has the Wh-adverbs *äntā* and *tā* which express the range of meanings 'when, where, if'.[38] Pinault (2008: 546) derives *tā* from **äntā*,[39] without discussing how exactly this split between *tā* and *äntā* came about. Treating the two forms as related is certainly justified based on their similar form and meaning. By attributing this development

[36]The expected counterpart of *kotnec* in (T:4.21) d) is **kotneñc*. While this particular form is unattested, the 3pl.pres./sub. ending occurs predominantly as *-ñc*.

[37]I am using the term 'bleeding' in a pre-theoretical sense here to merely capture the fact that we do not find sequences of the type [ä~~n~~CV] (i.e. with a deleted nasal). Similarly, I am using rule-based notation simply for illustration. The observed facts could be captured equally with ordered rules or with ranked constraints.

[38]I will return to how exactly these meanings apply to each of the two adverbs.

[39]Such a development was already considered by Krause and Thomas (1960: 167).

to N-deletion we thus have a parallel for the deletion of a nasal in (underlying) initial position and nothing prevents us from phonologically deriving indefinite *saṃ* from /nsan/.

4.2.4 The Synchronic Status of Indefinite *saṃ/caṃ*

Taking into consideration the status of N-deletion as a synchronic rule of TA, we may now ask whether we still have to reckon with the existence of a specialized indefinite paradigm within TA morphology. The observation that *äntsaṃ* can serve as an indefinite combined with the optional N-deletion rule appears to render such an assumption unnecessary. Instead we can operate with a single lexical item /nsan/ (Wh/indef), which can be phonetically realized in two different ways (see T:4.22), depending on the application of N-deletion. According to this hypothesis, there is no morphological difference between *äntsaṃ* and (what I previously called 'indefinite') *saṃ* whatsoever. Both are realizations of the same morphological object and as a consequence, TA does not have a specialized indefinite paradigm. We therefore predict that both the surface forms *äntsaṃ* and *saṃ* should be able to serve as indefinites, as well as interrogatives/relatives (int/rel). However, the attested distribution of *äntsaṃ* and *saṃ* appears to contradict such a view.[40]

Table 4.23: Distribution of *saṃ* and *äntsaṃ*

	int/rel	Indefinite
Forms of *äntsaṃ*	~ 18	1
Forms of *saṃ*	0	~ 26

Assuming that *äntsaṃ* and indefinite *saṃ* are derived from the same underlying representation, we expect all four cells in the table to exhibit similar numbers. Thus, the clear-cut divide in distribution between Wh *äntsaṃ* and indefinite *saṃ* requires some additional expla-

[40]The tilde in the table indicates that the counts may vary depending on how one interprets some of the more fragmentary examples.

nation. Specifically, we need to ask a) why indefinite /nsan/ undergoes N-deletion in most cases (with only one exception) and b) why interrogative /nsan/ never does. In the next section, I will argue that N-deletion is obligatory in unaccented clitics, accounting for a). Regarding question b), I will argue that the observed distribution is simply an accident of the attestation, as N-deletion outside of clitics is optional. This view gains plausibility through the variation found in the Wh-adverbs *äntā* and *tā*, which I will argue to be synchronically derived from the same underlying representation /ntā/.

4.2.5 Obligatory N-deletion in Enclitics

A possible explanation for the predominance of n-less forms among indefinites (cf. T:4.23) lies in the prosodic dependency of indefinites. We have already seen evidence that indefinites in TB (*ksa* and *kca*) are unaccented, based on what we know about accent-governed vowel alternations in that language. It is therefore at least plausible that their functional counterparts in TA can be unaccented as well, even though the TA forms themselves are uninformative regarding their accentuation.[41] If we now assume that N-deletion was obligatory within unaccented clitics, the observed distribution is accounted for. There are two pieces of evidence that suggest that this approach is on the right track.

First, the only attested indefinite that surfaces with the nasal particle is the genitive singular masculine form *äñcani-k*. What is noteworthy about this form is that it contains the emphatic particle *k*. This element directly attaches to the word it emphasizes, regardless of what category its host belongs to.[42] While the particle freely attaches to pronouns,

[41]This would also be in line with the claim in Hackstein (2001: 32) that the PT ancestor of the TB indefinite *ksa* was already unaccented. If this is correct, it is not unlikely that the property of indefinites being unaccented has been inherited by both Tocharian languages.

[42]Carling et al. (2009: 96) list the following categories as attested hosts for the particle: nouns, adverbs, verbs, interrogative pronouns, personal pronouns, demonstrative pronouns, indefinite pronouns, and pronominal adverbs.

including the demonstrative *saṃ*, as well as the genitive of *kus* 'who/what' when used as an indefinite, not a single form of indefinite *saṃ* is attested with this particle. If we now make the reasonable assumption that forms containing an emphatic particle are also accented, we have an explanation for why the only attested indefinite that surfaces with the nasal particle also involves the particle *k*. That is, the accent resulting from the emphasis allows the nasal to be preserved. The current proposal requires us to assume that clitics in TA could sometimes be accented. We have already seen evidence for this, based on a case where the host of a clitic surfaces without a vowel, namely *lw oki* 'like an animal'. A similar case is the following.

(4.17) *tmäṣ ṣu tñ= okik , ñy ākāl ṣeṣ ;*
then.ABL.SG PREV you.GEN.SG like I.GEN.SG wish.SG was.3SG.IMPF
... from then on it was my wish just like yours/from that the wish was for me as for you ... (A 268 b4, 4-3/4-3)

The pronominal form *tñi* has become non-syllabic due to glide formation sandhi, indicating that the accent is located on the following clitic instead. What is of additional interest about this example is that the clitic *oki* 'like' is hosting the emphatic particle. Sieg et al. (1931: 306 fn. 1) argue that the particle in this example is nevertheless emphasizing the pronominal *tñi*, despite being separated from it. However, based on the general distribution of *k*, as well as that of *oki*, I consider it more likely that it is, in fact, the latter that receives emphasis in (4.17). The particle *k* is typically very close to the word it emphasizes, including two attestations involving the genitive pronoun *tñi*, yielding *tñik*. Furthermore, there is clearly no rule that requires the particle *k* to be ordered after *oki*, as evidenced by the phrases *nṣäkk oki* 'like me (masc.)' (instead of **näṣṣ okik*) and *ñukk oki* 'like me (fem.)'. If this interpretation of *tñ= okik* is correct, it would constitute a construction parallel to *mā äñcanik* 'not to anyone', where the (indefinite) clitic has been emphasized by the same particle.

Let us now continue to the second piece of evidence in support of the current proposal

regarding N-deletion. There is one more enclitic particle in TA involving a cluster /nt/, namely *ontaṃ*. There are a total of 23 attestations of this particle, none of which involve N-deletion. If my hypothesis regarding N-deletion is correct, *ontaṃ* must be accented in order for it to retain its nasal. There is some indication that this is, in fact, the case. The form *ontaṃ* likely contains the neuter form of *äntsaṃ* used as an indefinite plus an additional element *o*. If Van Windekens is correct in analyzing this *o* as an emphatic prefix, *o-ntaṃ* would be a formation functionally very similar to *äñcani-k*, and if the latter was accented due to its emphatic particle, the former likely was too.[43] The next question is whether we can take this comparison further, i.e. whether there is an unaccented counterpart of *ontaṃ*. We would expect such a form to a) lack the emphatic element *o* and b) have undergone N-deletion. There is at least one attested sentence that appears to contain such a form.

(4.18) *täm* *nu* *mā* *kñasu* *kus* *taṃ* *buddhaṃ*
DEM.SG CONJ not recognize.1SG.PRT who/what.NOM ptcl buddha
mäskaträ
become.3SG
[...], but I did not know what *buddha* referred to. (YQ II.2 a1)

The neuter form *taṃ* cannot be an argument of the verb and the context in which it occurs is consistent with that of *ontaṃ*, which predominantly occurs in the scope of Wh-words[44] and the negation. It is therefore possible to treat *taṃ* as an unaccented variant of *ontaṃ* in this example, exhibiting both N-deletion and the absence of *o*. What poses a complication to this analysis is that one cannot rule out entirely that *taṃ* simply represents the neuter form of the demonstrative *saṃ*, i.e. that it is derived from an underlying representation /tan/ instead of /ntan/. In that case, *taṃ* would simply have adverbial meaning, which is a well-

[43]Even if we assume that *o* had already lost the status as an emphatic prefix within the synchronic grammar of TA, it would still be possible for *ontaṃ* to have inherited its accent from a time when the form was still analyzable.

[44]Of the cases in which *ontaṃ* occurs with a Wh-word, eight involve *kupre* and two *tā*. In light of the small amount of available data, the fact that *ontaṃ* is not attested with *kus* is likely not significant.

attested function of the neuter demonstrative (cf. Stumpf 1971a: 7). However, the Uyghur translation of (4.18) provides some support for treating *taṃ* as a particle in this example. Geng and et al. (1988: 118-9) give the following transcription and German translation.

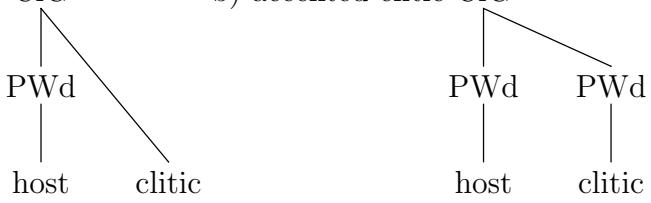
anï adï*ra* uqmaz ärdim.. kim ärki ol buta *yar*lay (.)

Ich kann nicht unterscheiden, wer nun diesen Ausruf 'Oh Buddha!' (tat) (?).[45]

While the Uyghur does not match the Tocharian perfectly, it is conspicuous that the translator used the particle *ärki*, which Erdal (2004: 350) analyses as a question particle with the meaning 'I wonder'. This particle is not obligatory in questions and is, therefore, likely a direct translation of TA *taṃ*, used as a particle, in this example.

Now that we have seen some evidence for the current proposal regarding N-deletion in indefinites, we need to determine how such a view can be incorporated into the prosodic analysis employed for enclitics in the previous chapter. There we saw that none of the TA clitics under discussion showed any direct indications of being prosodically defective, which made it difficult to determine whether they formed PWds or not. Using accent as a criterion, I now adopt an analysis in which clitics form PWds exactly when they are accented.

(4.19) unaccented clitic ClG b) accented clitic ClG



Obligatory N-deletion would then affect nasal-coronal clusters where neither member[46] is dominated by a PWd. In the next section I will provide evidence that Wh-words with

[45] "I cannot discern who may have (made) this exclamation 'o Buddha!'."

[46] So far I have not made any claims about the internal prosodic structure of forms such as *äntsaṃ* themselves. I do not know of any TA-internal evidence that would allow us to establish whether the interrogative particle /n-/ is still a proclitic, as I considered likely for the TB cognate /en-/, or has instead been incorporated into the PWd of the former host. The current formulation of N-deletion, which requires both the nasal and the coronal to be dominated by a PWd, is not dependent on this question.

initial /n/ did optionally undergo N-deletion and that, therefore, the lack of n-less forms of *äntsaṃ*, when used as an interrogative, is likely an accident. The evidence in question comes from *äntā* and *tā*. I will argue that the latter is simply a variant of the former involving optional N-deletion.

4.2.6 *tā* and *äntā*

Sieg et al. (1931) translate *tā* as 'where'[47] and *äntā* as 'when'. They thus treat the two elements as separate lexical items, which is still the standard view. If we now assume that both *äntā* and *tā* are synchronically derived from the same underlying representation /ntā/, it would be entirely unclear how the two forms could have different meanings. However, the later published YQ manuscript fragments contain two examples in which *äntā* seems to have the meaning 'where' (according to the translation by the editors Ji et al. 1998).

(4.20) *///* · *m(ā)gat* *ypeyis* *wampe* *pāṣānak* *ṣul* *äntā* *ne*
PN.SG land.GEN.SG jewel.SG PN.SG mountain.SG where PTCL
säm *tri* *ārkiśo(ṣintwis* *///* *ptāñkä)t* *käṣṣi* • *klyomänt*
DEM.NOM.SG 3 world.GEN.PL Buddha.SG teacher.SG noble.OBL.SG
meträkyāp *waṣtäṣ* *läñclune* *lmāṣtär*
Maitreya.GEN.SG house.ABL.SG emerge.ABSTR sit.3SG
... the Mount Pāṣāṇaka, the jewel of the land of Magadha, where this one, (the teacher) of the three worlds, ... the Buddha-god the teacher will keep waiting for the noble Metraks going away from the house. (YQ II.9 a4)

(4.21) *(ä)ntā* *aśśi* *tāpärk* *säm* *āṣānik* *mäskatär*
where PTCL now DEM.NOM.SG arhat.NOM.SG become.3SG
... Where then is at this moment this venerable one? (A 214 a4)

Due to examples such as these, Carling et al. (2009) translate *äntā* as 'where' and

[47] I am ignoring the difference between interrogative and relative uses of the pronoun.

'when'.[48] These additional cases render the relationship between *äntā* and *tā* reminiscent of what we saw with *äntsaṃ* and *saṃ*, where it appears that the longer forms (i.e. without N-deletion) fulfill a superset of the functions we find with the shorter form. However, there is also one attested case in which *tā* appears to have temporal/conditional meaning.

(4.22) *tā* *ne* *āptā* *yāmäl* *ṣem* *tmäṣ* *kospreṃn*
where PTCL early do.GER.NOM.SG was.1SG.IMPF then.ABL.SG how.much
aśśi *kälpāl* *ṣem*
PTCL obtain.GER.NOM.SG was.1SG.IMPF
... [if] then I had done [it] before, how much then would I have gained from that? (Carling et al. 2009: 38, CEToM)
(Wenn) ich dabei anders(?) als das gehandelt hätte, wie viel hätte ich wohl erlangt? (Thomas 1952: 44) (A 436 a2)

Both Thomas and Carling et al. translate the embedded clause as a conditional. Yet both put the conditional conjunction in the translation in brackets, presumably under the assumption that *tā* can only mean 'where'. However, if we take *tā* and *äntā* to be derived from the same underlying phonological representation /ntā/, we fully expect both items to exhibit the same range of meanings.

Furthermore, there is at least one attestation where *tā* appears to have specifically temporal meaning. Sieg et al. (1931: 182 fn. 1) mention an indefinite form *tācaṃ* 'sometimes', which is only attested twice and within the same sentence.

(4.23) *pācarä* *lānt* *tācaṃ* *märkampalā* *tācaṃ* *mā* *märkampalā*
father.SG king.SG be.LOC.PL law.PER.SG be.LOC.PL not law.PER.SG
watku *wätkäsmāṃ* *pälkāt*
order.SG separate.MP see.3SG.PRT
[...] he saw his father, the king, giving orders sometimes according to the law, sometimes not according to the law. (A 394 a2, translation based on Thomas 1957: 128)

[48] More specifically, they translate bare *äntā* as 'where' and the *äntā* with the relative particle *ne* as 'when'. I see no reason to assume that a Wh-word would have a fundamentally different meaning in a relative clause than it would have in a question. Besides, *äntā* clearly does have the meaning 'where' in the relative clause in (4.20).

Sieg et al. consider the possibility that *tācaṃ* is related to the Wh-word *tā*, partially based on the fact that the original manuscript leaf glosses the form with TB *inte kca*. If we now take the TB gloss at face value, we can simply consider *tā caṃ* as another case where an indefinite is used as a particle following a Wh-word, similar to (4.14). If this is correct, this makes the sentence (4.23) significant for three different reasons: a) it shows that *tā* can have temporal meaning, b) it constitutes another case of a Wh-word doubling as an indefinite, and c) it shows that Wh-words used in indefinite function do not necessarily behave like clitics but can occur in sentence-initial position hosting a clitic themselves.

An additional case of independent Wh-words functioning as indefinites is found in the following sentence.

(4.24) */// kupre ontaṃ ; tā ontaṃ , udumparṣi ; pyāpy oki ,*
if/when PTCL where PTCL type.of.flower.SG flower.SG like
ptāñkät käṣṣi : pākrä mska(tär) ///
Buddha.SG teacher.SG visible become.3SG
At some time, somewhere, like the Udumbara ower, the Buddha-god the teacher makes his appearance. (YQ II.15 b3, 4-3/4-3/4)

Pinault (1997: 487-90) shows, based on the Uyghur translation, that *kupre ontaṃ tā ontaṃ* is an idiom expressing the rarity with which a Buddha manifests itself in the world. He proceeds to demonstrated that the TA idiom is itself a calque from Sanskrit.

(4.25) *kadā -cit karhi -cid*
kupre ontaṃ tā ontaṃ
en de rares occasions [et] en de rares endroits[49] (Pinault 1997: 489)

The Sanskrit particle *cid*, which turns Wh-words into indefinites, is rendered by the TA particle *ontaṃ*. Sanskrit *kadā* means 'when' and therefore matches TA *kupre* 'when, if'. However, under the assumption that *tā* can only mean 'where', there is a mismatch between *tā* and the Sanskrit *karhi*, as the latter specifically means 'when' (cf. Monier-Williams 1899

[49]"on rare occasions [and] in rare places"

[2014]). Such a mismatch would be particularly surprising as the author of the TA verse could have simply used *äntā* as a second form with the meaning 'when' in order to match the Sanskrit.[50] If, on the other hand, we accept that *tā* can also have temporal meaning the Sanskrit phrase forms a perfect match with TA. If we count the attestations of the two items relative to their meanings, we get the following picture.

Table 4.24: Distribution of meanings among *äntā* and *tā*

	'if/when'	'where'
äntā	13	2
tā	3	6

In the table I disregard a number of instances of *tā* and *äntā* which are either too fragmentary for an interpretation or which are compatible with either meaning and are therefore immaterial for the matter at hand. Even though *äntā* is still far more robustly attested with temporal/conditional meaning and *tā* seems to favor locative meaning, the total number is too small to draw any firm conclusions from this distribution. It is thus preferable to capture all of the attested meanings by deriving both forms from the same underlying representation /ntā/. In the next subsection I provide some further support for this conclusion by showing how a form such as /ntā/ fits into the TA system of Wh-words more generally.

[50] One possible objection to this argument is that *äntā* would not fit in the same colon as *tā* due to its additional syllable. However, if the poet wanted to faithfully render the Sanskrit idiom, he could have simple reversed the order of the two TA Wh-words. The phrase *kupre ontaṃ* could easily fit a three-syllable cola a) through glide formation (**kupry ontaṃ*) or b) by rendering the *u* non-syllabic (**k*$_u$*pre ontaṃ*). The monosyllabic form *k*$_u$*pre* is attested 28 times within the CEToM corpus.

4.2.7 The Internal Structure of Wh-Words in TA

Just like in TB,[51] most TA Wh-words can be synchronically analyzed as consisting of one of two interrogative elements (/k$^{(w)}$/ or /n/) and a (pro)nominal element.

Table 4.25: Wh-words involving /k$^{(w)}$/

		Pronominal	
a)	nom.	*kus*	/k^{w}-s/
b)	obl.	*kuc*	/k^{w}-ca/
c)	gen.	*ke*	/k^{w}-e/ (??)
d)	nom.pl.	*k$_u$ce*	/k^{w}-ce/
		Adverbial	
e)	'if, when'	*kupre*	/k^{w}-pre/
f)	'how much'	*kos*	/k^{w}-os/ (??)
g)	'how much'	*kospreṃ*	/k^{w}-os-pre-ṃ/ (??)
h)	'why'	*k$_u$yal*	/k^{w}-yal/ (??)

Table 4.26: Wh-words involving the interrogative element /n/

		Pronominal	
a)	nom.sg.masc.	*äntsaṃ*	/n-saṃ/
b)	obl.sg.masc.	*äñcaṃ*	/n-caṃ/
c)	gen.sg.masc.	*äñcanik*	/n-cāny-k/
d)	obl.sg.fem.	*äntāṃ*	/n-tāṃ/
e)	gen.pl.masc.	*äñcesni*	/n-cesny/
		Adverbial	
f)	'if/when, where'	*äntā*	/n-tā/
g)	'where'	*äntanne*	/n-tan-ne/

While not all of the above forms can be easily decomposed into two recognizable morphemes, we can make the generalization that all of them contain either /k^{w}/ or /n/. In

[51] See section 4.1 for discussion. The etymological literature cited there also applies to TA.

addition to the examples above, there is also the hapax *āṃ* with the meaning 'who'. Due to its isolated nature I have nothing to say about its internal structure but Krause and Thomas (1960: 166) already considered a connection between its final nasal and the *n* in TA *än* and its counterpart in TB *in*. Thus, *āṃ* seems to be at least compatible with this generalization. There only seem to be two exceptions.

Table 4.27: Wh-adverbs apparently lacking the element /n/

h)	'how'	*mänt*	/??/
i)	'where'	*tā*	/??/

We can illuminate the internal structure of TA *mänt* by comparing it with its functional equivalent in TB, i.e. *mäkte*.

Table 4.28: The internal structure of *mänt*

TA	mä	n	t	→	*mänt*
TB	mä	k	te	→	*mäkte*

This parallelism between TA and TB lends support to the identification of the *n* in *mänt* with the interrogative particle in forms such as *äntā* and *äntsaṃ*, considering that the *k* in *mäkte* clearly continues the PT interrogative element *k*^{w}*ä* (see previous section with literature). This would leave *tā* as the only element within the TA Wh-system that is not marked by either of the two Wh-particles. If on the other hand, we assume that *tā* is derived from the same underlying representation as *äntā*, it would pose no exception to the morphological pattern. Thus the strong tendency to mark Wh-words with a particle in TA makes it less likely that the result of applying N-deletion to /ntā/ came to be disassociated from its original source via reanalysis. Therefore *äntā* and *tā* show that the optionality of N-deletion includes the interrogative particle /n-/ in orthotonic Wh-words, which makes it plausible that the lack of attestations of *saṃ* in interrogative function is simply an accident.

4.2.8 Summary

In this section I have argued that indefinites of the shape *saṃ* and *caṃ* are underlyingly identical to the relative/interrogative pronouns *äntsaṃ* and *äñcaṃ*, but a well-attested (optional) nasal-deletion rule often renders them indistinguishable from the demonstrative pronoun. Both the possibility of using Wh-words as indefinites and the N-deletion rule can be independently motivated. The present proposal further avoids all the problems involved in tracing TA *saṃ*/*caṃ* and TB *ksa*/*kca* back to a common source. The two sets of forms are simply unrelated.

Chapter 5

Conclusion

In this final chapter I briefly sum up the main findings resulting from the present study. In the first half of chapter 2, I provided a fresh look at external sandhi in TA. The primary goal of this section was to show that a number of apparently diverse phonological phenomena are all directed towards the same goal, i.e. to ensure that every syllable has an onset, a requirement that can also be observed on the word level. My account of the segmental properties differed from the study in Stumpf (1971b) in a number of important ways. First, I argued for a distinction between vowel deletion and vowel coalescence, *pace* Stumpf. I further argued against a special form of vowel deletion that specifically targets clitics. Instead I took the observed deletion of the initial vowel in clitics as due to the phonologically weak status of the vowel *a*, for which there is independent evidence. Finally I showed that the two-way behavior of *o*, which either undergoes glide formation or deletion, may not, in fact, reflect a form of phonological variation. Instead, the data may simply have been obscured in this case by an orthographic practice of simplifying certain combinations of characters.

The second half of the chapter provided a study of the interaction between a particular aspect of TA orthography and external sandhi. I argued that the use of virāma spellings in TA is governed by two different factors. In preconsonantal position, virāma spellings tend to be used in order to avoid writing akṣaras that take up a certain amount of vertical space.

However, at present it is not possible to establish a set of rules that fully determines the use of vi-disjunction or a-linking in individual cases, as both strategies are often attested for the same consonant cluster. In prevocalic position, however, the use of virāma spelling is fully predictable. A-linking occurs if the two words in question are connected by external sandhi; otherwise vi-disjunction is used. I argued that most of the apparent exceptions to this rule can be accounted for by spelling conventions obscuring some of the phonetic properties of the examples in question. Finally, I accounted for the distribution of prevocalic vi-disjunction using the same phonological principle that drives the application of external sandhi, i.e. avoidance of onsetless syllables. Specifically, I argued that the use of prevocalic virāma is a reflex of another strategy to provide vowel-initial syllables with an onset, i.e. by the insertion of an epenthetic consonant, possibly a glottal stop.

The first half of chapter 3 established the phonological constituent formed by TA clitics and their host. I showed that clitics exhibit behavior that appears to be intermediate between affixes and words, by sharing some properties with the former and others with the latter. Just like two adjacent words, clitics and their host can undergo external sandhi and allow hiatus at their boundary (in so far as I allowed the term *hiatus* to include sequences such as [V?V]). On the other hand, clitics behave like affixes in that they cannot be separated from their host by a caesura in meter. Furthermore, the proclitic *śla* 'with' ends in the vowel *a*, which is banned from word-final position in TA. I then showed that the observed behavior of host+clitic can be accounted for by taking them to form a phonological constituent intermediate between the PWd and the PPhr, i.e. the ClG. This also accounted for the observation that consonantal sandhi within host+clitic is obligatory, while it is only optional between orthotonic words. Furthermore, the only other linguistic unit that allows a) hiatus between two words and b) final *a* at the end of a word are compounds. The ClG enabled us to capture the behavior of host+clitic and compounds using the same phonological constituent.

The second half of chapter 3 addressed the prosodic status of clitics in TB. As in TA,

clitics show affix-like behavior in some respects and word-like behavior in others. They show a certain degree of phonological autonomy by being located outside of the accent domain of their hosts, as well as by allowing hiatus between themselves and their hosts. On the other hand, TB clitics also show a certain degree of dependency, in that, just as in TA, they cannot be separated from their hosts by a caesura. Unlike TA, however, TB also provides some direct evidence that clitics are not only dependent on a host, but also prosodically defective in a number of ways, suggesting that they cannot form a PWd by themselves. They a) are unaccented, b) violate word-level phonotactic constraints, and c) can be smaller than a minimal word, i.e. by lacking a vowel. I argued that, since a number of so-called secondary case affixes show exactly the same properties they should be attributed the exact same prosodic status as enclitic particles. I then concluded that the constituent dominating instances of host+clitic should be taken to be the ClG, based on the fact that the behavior of compounds independently necessitates a phonological constituent between the PWd and the PPhr. On the one hand, compounds cannot be PWds, since they allow hiatus between their members. On the other hand, treating them as a PPhr does not allow us to account for their special accentual behavior.

The first part of chapter 4 dealt with the internal structure of TB Wh-words. I treated the interrogative element contained in most Wh-words as a clitic along the lines of the analysis of the previous chapter, at least as late as PT. By doing so it became possible not only to capture the aberrant accentuation of forms such as *k_use* 'who/what' and *mäkte* 'how', but also to connect it to segmental sound changes that set the Wh-words under discussion apart from content words. These included the loss of rounding of the segment *k^w in *mäkte*, as well as the reduction of the particle **æn* in TA *äntsaṃ* and related forms.

The second part of chapter 4 explored a different aspect of Tocharian Wh-words, namely their functioning as indefinites in TA. I first concluded, based on observations by Sieg et al. (1931), that the genitive indefinite *ke* should not be placed in the same paradigm with the

nominative and oblique forms *saṃ* and *caṃ*. I then took the latter two forms to be derived via a phonological rule from forms of the Wh-word *äntsaṃ* 'which', whose genitive is, in fact, once attested serving as an indefinite. The use of Wh-words to double as indefinites is not only unproblematic from a typological viewpoint, but also fits nicely with the observation that the neuter form of the same Wh-word came to be used as a particle. One obstacle for the present analysis was the fact that interrogative *äntsaṃ* is never attested without the interrogative particle /n-/ while, with one exception, all indefinites lack the particle on the surface. I argued that this property is due to the particle being obligatorily deleted in unaccented enclitics, backed up by the observation that enclitics featuring nasal-coronal clusters also contain an emphatic particle. In the final part of the chapter, I showed that the Wh-adverbs *äntā* and *tā* are morphologically identical and derived from /ntā/ and, therefore, that the deletion of the interrogative particle /n-/ is optional in orthotonic Wh-words. Thus, I concluded that the absence of n-less forms of *äntsaṃ* in interrogative function is simply an accident.

Appendix A: Form Index

The following index contains all the attested Tocharian forms used in the text, not including the words contained in the glossed example sentences. Each form is followed by the page number(s) on which it is mentioned, as well as up to four sigla of manuscript leaves on which it is attested. For TB sigla, ^{A} and ^{L} indicate that a text is archaic or late, respectively. Most of the index has been generated by a Perl script that searches each Tocharian form cited in the dissertation within the digital transcriptions on CEToM, retrieving the sigla of the manuscript leaves on which the forms are attested. In cases where a form is attested more than four times the script prioritizes attestations that are not located within a fragmentary context (i.e. where the target string is not flanked by – · or ///). Where the script produced incorrect results (e.g. due to homophony), I entered the sigla manually.[1]

Tocharian A

aci (postpos.)	67-68, 71, 75, 77, 82-83, 92	A 4 a4, a6; A 54 a1, a3; A 61 b3;
aci (prev.)	83	A 114 a5; A 312 a3;
anac	69	A 61 a4; A 95 b4; A 108 a3; A 113 a4;
anäṣ	69	A 155 b3; YQ III.2 b8; A 179 a1; A 456.b b3;

[1]The forms in question are the following:
Tocharian A: *ṣokyo =kāl*, *ṣoma ārtlune*, *āṃ*, *saṃ*, *caṃ*, *saṃ (dem.)*, *caṃ (dem.)*, *cani (dem.)*, *aci*, *aci (prev.)*, *lu*, *lymā*, *śu*, *śo*, *tā*, *ñi*, *ke*
Tocharian B: *aśca-sanāpalle*, *inte kca*

sälpmāṃ oki	73-74	A 217 a8;
sumnātär	39, 53	A 51 b1;
sumnātr oki	39-40, 53, 71	A 6 a1;
ske	85	A 13 b4; A 16 a2; A 17 a3; A 67 a1;
skeneñc o(ki)	57	A 58 b1;
sñ= āñu	27	A 405 a1;
sn= ālak	27	A 213 b5; THT 1141 a6;
snum	85	A 21 b5; YQ III.11 b3, b6; THT 3336 a1;
sne	12, 27, 85-86, 98, 135	A 2 a3 (2x), a5 (2x); A 4 a1; A 9 b4; A 11 a4, b2;
sne āk	19	A 137 b2; A 218 b4; A 278 b3; A 291 a3;
sne eṃts	85	A 327 b8; YQ I.3 a6; YQ I.5 b7;
sne kip	85	A 215 b1;
sne nākämm oki	37, 40	A 61 b2;
sny āk	12, 19, 85	A 218 a6;
sny āñu	27	A 17 b3;
spaktānikāṃ ats	73	YQ III.12 b5;
/// (s)s oko	60	THT 1141 a5;
tsälpṣant i	57	A 270 b7;
tskont oki	57, 72	A 101 b6;
tsratränn oki	40, 74	A 146 b4;
tsräsmāṃ oki	73	A 298 b6;

Tocharian B

aiśamñe	104-105	THT 15 a8; THT 17 b1; THT 28 b6, b7; THT 31 a6;
añmantse	112	THT 7 b2; THT 28 a2; THT 290 7; IOL Toch 96 a2;
aśca-sanāpalle	104	W-18 a4;
askace	120	[A]THT 224 b2;
askwacentse	120	THT 308 b3;
āñm	112	THT 81 a2; THT 99 b3; THT 239 a3, b4; [A]THT 285 a1;
āñme	112	THT 9 b6; THT 23 a4; THT 55 a4; THT 78 a6;
āścä	104	THT 322 b1; IOL Toch 803 a3;

kärtsauñe	100	THT 64 b8; [A]THT 281 a3; THT 1419.g a1; THT 2558 b2;
k$_u$ce	118, 125, 134	THT 21 a2; THT 22 a8, b1; THT 24 a5; THT 28 b6;
k$_u$tameṃ	125	THT 33 b2; [L]THT 202 b2; [A]THT 283.a b4; THT 409 b3;
k$_u$temeṃ	125	THT 1120 b3; [A]IOL Toch 1244 b1;
k$_u$ri	125	THT 1190 a2; THT 2196 a1; THT 3151 a2; IOL Toch 872 a2;
k$_u$se	10, 110-111, 115-116, 118, 121, 125, 133-134, 136, 152	THT 2 a2, a5; THT 8 a4; THT 13 b4; THT 14 a7;
ketara	130, 134	[L]THT 346 a5; [L]THT 347 a4; [L]THT 511 a2; THT 596 b1, b2;
kete	125, 134	THT 31 b7; THT 93 a6; THT 145 b1; THT 200 b3;
ketra	130, 134	THT 1111 b2; THT 1283 a7; THT 4001 a9;
kos	125	THT 5 a5; THT 15 a5; THT 17 a6; THT 41 b1, b5 (2x);
kca	94, 96, 100, 129-130, 133-134, 136, 139, 149	THT 24 a3; THT 27 b8 (2x); THT 31 b2; THT 33 b6;
kwri	116, 125	THT 5 b2; THT 11 b4; THT 30 a8; THT 31 a4, a6, a7 (2x);
ksa	94, 96, 100, 118, 121-122, 129-130, 132-134, 136, 139-140, 149	THT 2 a4; THT 4 a7; THT 8 a1; THT 15 b4;
ṅke	97, 100	THT 5 a6; THT 21 a7; THT 27 a4, a6, b1; THT 29 a8;
cai	121	Taj 2 (a1); THT 4 b8; THT 8 a4, b1; THT 10 a2;
cau	121	THT 19 b8; THT 34 a5; THT 42 a2, b6; THT 53 a1;
cakkarntse	100	THT 30 b7;
camel	100	THT 42 b4; [L]THT 113 a3; THT 149 b1; THT 175 a6;
camelne	100	[L]THT 109 b2; [A]THT 119 a6; [A]THT 120 a3; [L]THT 580 a3;

totka kca	100, 106	IOL Toch 136 b1; IOL Toch 109 a1;
totkānts	100	THT 31 a6;
totkā yärm	103, 106	THT 5 a1; THT 16 a7;
tona	121	THT 151 a6; THT 158 a1;
tonaṃts	121	PK AS 16.1 b6;
tonts	121	[A]THT 274 a1;
toy	121	[L]THT 107 a5, a10, b1, b3, b6, b8; THT 182 a2 (2x); [L]THT 188 a4; [L]THT 294 13;
toyna	121	[L]THT 107 a2, b2; THT 201 b1; [L]THT 419 b2; [L]THT 500-502 b12;
toynaṃts	121	
nai	94, 100	THT 12 b6; THT 23 b6; THT 30 a1; THT 46 b2;
nauṣw ecce	19	[L]THT 297.c a1;
no	94, 97, 100	THT 1 b3; THT 2 a5, a8; THT 3 b6; THT 4 a7;
nta	94, 96-97, 100	THT 8 a2; THT 11 b3; THT 16 a1; THT 24 a1;
pak$_u$trä	119	THT 523 a5; PK AS 7L b2;
pakträ	119	THT 574 b5; IOL Toch 201 b6; IOL Toch 240 b3; PK AS 16.1 b2;
päklyauṣ	120	THT 77 a5; [A]THT 123 b5; THT 163 b6; THT 1305 b1;
päst	114-115	THT 11 a5 (2x); THT 19 a4; THT 20 a6; THT 23 b6;
pi	94, 100	THT 53 a2; THT 77 a2; THT 79 6; THT 91 b4;
pikala	120	[L]THT 496 a1;
pik$_u$la	119	THT 3 b1, b8; PK AS 7L b2; THT 1262 b2; THT 1262.a b2;
pikla	119-120	[L]THT 415 a3;
pikwala	120	THT 404 a6, b4; [A]THT 407 b1; THT 1399.d b2; THT 3285 a2;
postañña	100	THT 17 a7; THT 36 b8; THT 249.a a3; THT 522 a2;
postäṃ	100	THT 3 b5; THT 12 a6; THT 29 a8; THT 31 b5;
postäṃ ka	100	THT 577 b2; THT 2648 a3;
postän nai	100	[A]THT 284 a5;
pklyauṣ	120	[L]THT 114 a1; THT 1576.d a1; PK AS 17H a2;
pw īke	19	[A]THT 127 a3;
makāṃts	5	THT 15 a8;

ysape	116	THT 78 a3; THT 1507 b3; IOL Toch 86 a1; PK AS 17B a1;
ra	94-96, 100	THT 2 a5; THT 3 a3; THT 5 a8; THT 9 b4;
ram	94, 96-97, 100	THT 7 b1; THT 11 b6; THT 17 b2; THT 45 b4;
ramt	94, 97	THT 1 b4; THT 3 b4 (2x); THT 8 a5; THT 11 a2;
ram no	97	THT 7 b1; THT 11 b6; THT 17 b2; THT 45 b4;
lakle	100, 104	THT 11 a5; THT 14 b5; THT 30 a2 (2x); THT 44 a8, b4;
laklempa	100	THT 3 b6; SI B 75 b4;
lakle ṣpä	100	THT 30 a2;
laklesa	100	THT 46 b6; THT 231 b2; THT 521 b2; THT 1573.a a5;
laṅkce	119	IOL Toch 2 b4;
larauñe	100	THT 12 b6; THT 23 a6; IOL Toch 240 b4; IOL Toch 307 a1;
lāre	100	THT 25 a8; THT 78 b5; THT 86 b1; [A]THT 128 a5;
lāre no	100	PK AS 7I a3;
läk$_u$tsauwñaisa	119, 121	IOL Toch 178 a3;
läklenta	100, 120	THT 17 a8; THT 19 a1 (2x); THT 25 a3; THT 28 b8;
läklentse	100	[A]THT 139 b6; [A]PK AS 9B a3; THT 155 a2, a3 (2x), b1; THT 497 a6;
läkle-lyakāñ	104	PK AS 7E a6;
läktsauñaisa	119	PK AS 14.2 a5; IOL Toch 800 a3;
läṅk$_u$ci	119	[A]THT 245 a3;
wa	94-96, 100	THT 88 a6; THT 214 b2; THT 231 a1; THT 237 b2;
wakissonta	100	THT 1923.a 2;
wat	94, 96, 100	THT 5 a4; THT 7 a7; THT 13 b4; THT 14 b7;
wāki	100	THT 178 b4; THT 333 b1; THT 367 a2; THT 567 b2;
wāki nta	100	PK AS 4B b3;
were	78	THT 30 b2; THT 33 a4; THT 51 a7; THT 176 b5;
śak$_u$se	119	IOL Toch 12 a3; THT 624 b7;
śakse	119	THT 247 a2; IOL Toch 246 b2;
śale	96	THT 498 b4; THT 1167 a5; THT 1574 a2, b2; PK AS 3A a4;

Appendix B: Virāma Statistics

This appendix lays out the calculations of the χ^2 values I used to show that there is a relationship between the number of consonants at word-boundaries and the use of virāma spelling (chapter 2.3.3). The following tables repeat the distribution of a-linking and vi-disjunction in C#C-pairs and C#V-pairs. Additionally, each cell contains in parentheses the number of attestations we would expect if virāma spelling was independent of the number of consonants, i.e. (column total * row total) / table total. To these observed and expected values we can apply Pearson's chi-squared test.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Based on the resulting χ^2 value and the degrees of freedom ((number of rows - 1) * (number columns -1)), I arrived at the given P-values using the online calculator available on Social Science Statistics.[2]

[2]http://www.socscistatistics.com/pvalues/chidistribution.aspx

C#C-Pairs

Categories:	a-linking	vi-disjunction	total
CC	5525 (4193.60)	3449 (4780.40)	8974
CCC	947 (2048.20)	3436 (2334.80)	4383
CCCC	17 (241.60)	500 (275.40)	517
CCCCC	0 (5.61)	12 (6.39)	12
total	6489	7397	13886

$$\frac{(3449.00-4780.40)^2}{4780.40} + \frac{(5525.00-4193.60)^2}{4193.60} + \frac{(3436.00-2334.80)^2}{2334.80} + \frac{(947.00-2048.20)^2}{2048.20} + \frac{(17.00-241.60)^2}{241.60} + \frac{(500.00-275.40)^2}{275.40} + \frac{(12.00-6.39)^2}{6.39} + \frac{(0.00-5.61)^2}{5.61} = 2307.423$$

Degrees of Freedom = 3

P-value: < 0.00001

C#V-Pairs

Categories:	a-linking	vi-disjunction	total
C	51 (391.44)	860 (519.56)	911
CC	651 (352.77)	170 (468.23)	821
CCC	74 (31.80)	0 (42.20)	74
total	776	1030	1806

$$\frac{(860.00-519.56)^2}{519.56} + \frac{(51.00-391.44)^2}{391.44} + \frac{(651.00-352.77)^2}{352.77} + \frac{(170.00-468.23)^2}{468.23} + \frac{(0.00-42.20)^2}{42.20} + \frac{(74.00-31.80)^2}{31.80} = 1059.458$$

Degrees of Freedom = 2

P-value: < 0.00001

Appendix C: Selection of Data on Virāma spellings

The following is a more detailed description of how the data for the study on TA orthography in chapter 2.3 was selected. I used a Perl script to extract and sort the C#X-pairs from the entirety of TA transliterations within the CEToM corpus. The extracted pairs were then subjected to a filter to ensure that mostly relevant examples are contributing to the count. The filter works as follows:

- I compiled a word list consisting of 8171 Tocharian A word forms. The list contains the entries from the verbal index in Malzahn (2010), most of the glossary in Krause and Thomas (1960), as well as a number of entries from Carling et al. (2009).
- Each member of a C#X-pair detected by the script was compared to the word list. If neither of the two members was contained in the list, the C#X-pair was discarded from the data set. This was done to prevent examples from Sanskrit and Tocharian B (in bilingual texts) to contribute to the statistics, as well as overly fragmentary cases.
- In order to take into account orthographic variation and to ensure the correct recognition of words that have undergone external sandhi, a number of neutralizations were applied when comparing members of C#X-pairs with the word list:
  - The distinction between <i>, <ī>, <$_i$> and <y> as well as the distinction between

<u>, <ū>, <$_u$> and <w> were eliminated.

– All instances of <ä> were deleted.

– Word-internal anusvāra was changed to the homorganic nasal with respect to the following stop.

The following tables give an overview of the number of examples that have been discarded by the filter.

(5.1) Distribution of a-linking and vi-disjunction in C#X-pairs, including the number of discarded cases for each type of cluster

C#C	attestations		a-linked		vi-disjoined	
	included	discarded	included	discarded	included	discarded
C#C	8974	731	5525	363	3449	368
CC#C	1289	67	176	7	1113	60
C#CC	3094	223	771	50	2323	173
CCC#C	3	0	2	0	1	0
CC#CC	445	25	11	1	434	24
C#CCC	69	5	4	0	65	5
CCC#CC	2	0	0	0	2	0
CC#CCC	10	1	0	0	10	1

C#V	attestations		a-linked		vi-disjoined	
	included	discarded	included	discarded	included	discarded
C#	911	168	51	63	860	105
CC#	821	32	651	26	170	6
CCC#	74	5	74	4	0	1

Appendix D: Photographs of Manuscript Leaves

The present appendix contains the photographs from which the brāhmī samples in chapter 2.3 were taken. The purpose of including the complete images is to allow the reader to see the various sample characters in context and to provide a more accurate picture of the dimensions of ligatures relative to simple akṣaras. All photographs are reproduced with the permission of the owner:

Depositum der
BERLIN-BRANDENBURGISCHEN AKADEMIE DER WISSENSCHAFTEN
in der STAATSBIBLIOTHEK ZU BERLIN - Preussischer Kulturbesitz
Orientabteilung

5.1 THT 698 (= A 56)

(taken from the IDP http://idp.bbaw.de/)

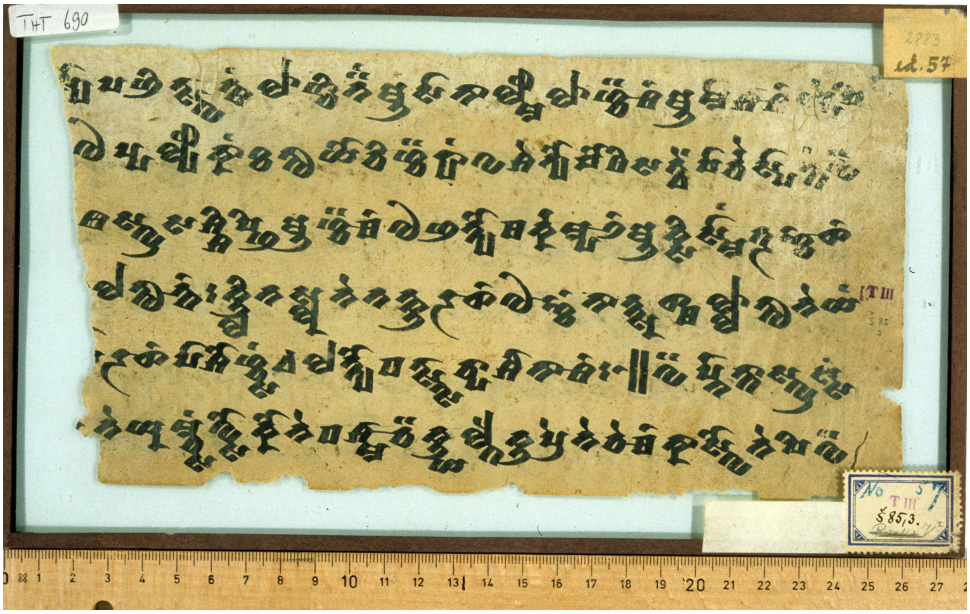
5.1.1 recto

5.1.2 verso

5.2 THT 699 (= A 57)

(taken from IDP http://idp.bbaw.de/)

5.2.1 recto

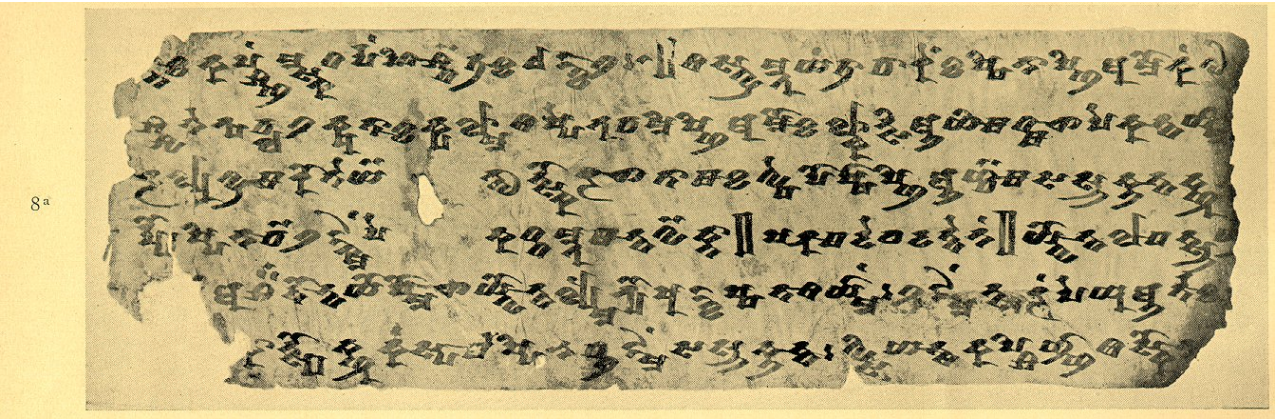


5.2.2 verso

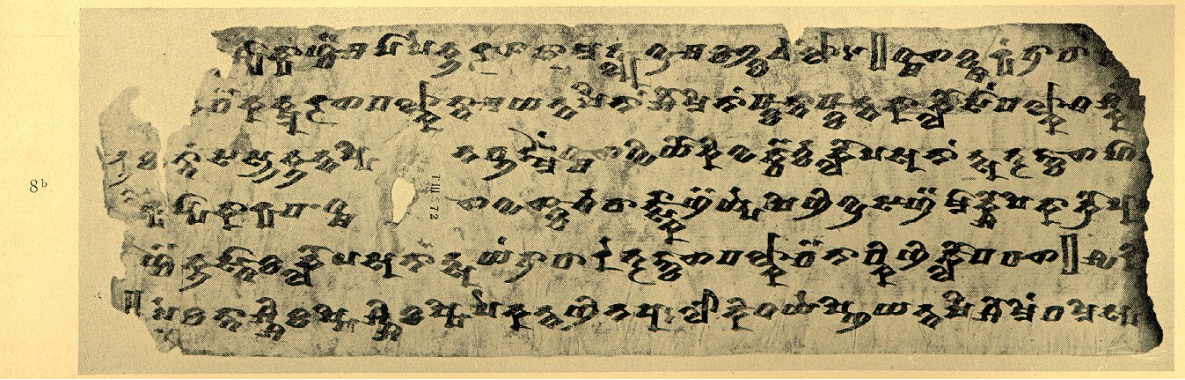
5.3 THT 641 (= A 8)

(taken from TITUS http://titus.fkidg1.uni-frankfurt.de/framee.htm?/index.htm)

5.3.1 recto



5.3.2 verso



Appendix E: Text Editors in CEToM

The following is a list of the manuscript leaves mentioned throughout this dissertation (excluding the form index) with the respective editor(s) currently listed on CEToM.

Bibliography

Adams, Douglas Q. 1988. *Tocharian Historical Phonology and Morphology*, volume 71. New Haven, Connecticut: American Oriental Society.

Adams, Douglas Q. 1999. *A Dictionary of Tocharian B*. Number 10 in Leiden Studies in Indo-European. Amsterdam - Atlanta, GA: Rodopi.

Adams, Douglas Q. 2013. *A Dictionary of Tocharian B. Revised and Greatly Enlarged*. Number 10 in Leiden Studies in Indo-European. Amsterdam - New York: Rodopi.

Alber, Birgit. 2001. Regional variation and edges: Glottal stop epenthesis and dissimilation in standard and southern varieties of German. *Zeitschrift für Sprachwissenschaft* 20:3–41.

Baković, Eric. 2007. Hiatus resolution and incomplete identity. In *Optimality-Theoretic Studies in Spanish Phonology*, ed. Sonia Colina and Fernando Martínez-Gil, 62–73. Amsterdam: John Benjamins.

Bernhard, Franz. 1958. Die Nominalkomposition im Tocharischen. Doctoral Dissertation, Univ. Göttingen.

Booij, Geert. 1995. Cliticization as prosodic integration: The case of Dutch. *The Linguistic Review* 13:219–242.

Booij, Geert. 1997. Non-derivational phonology meets Lexical Phonology. In *Derivations and Constraints in Phonology*, ed. Iggy Roca, 261–288. Oxford: Clarendon Press.

Bross, Christoph, Dieter Gunkel, and Kevin M. Ryan. 2014. Caesurae, Bridges, and the Colometry of Four Tocharian B Meters. *Indo-European Linguistics* 2:1–23.

Carling, Gerd, Georges-Jean Pinault, and Werner Winter. 2009. *Dictionary and Thesaurus of Tocharian A. Volume 1: A-J*. Wiesbaden: Harrassowitz.

Casali, Roderic F. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 493–533.

Dilley, L., S. Shattuck-Hufnagel, and M. Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24:423–444.

Embick, David, and Roumyana Izvorski. 1995. Participle-auxiliary word orders in Slavic. *Annual Workshop on Formal Approaches to Slavic Linguistics: the Cornell Meeting* 4:210–239.

A Comprehensive Edition of Tocharian Manuscripts. http://www.univie.ac.at/tocharian/ (retrieved: May, 21, 2015).

The International Dunhuang Project: The Silk Road Online. http://idp.bl.uk/.

Thesaurus Indogermanischer Text- und Sprachmaterialien. http://titus.fkidg1.uni-frankfurt.de/framee.htm?/index.htm.

Erdal, Marcel. 2004. *A Grammar of Old Turkic*. Leiden: Brill.

Fellner, Hannes A. 2006. On the Development of Labiovelars in Tocharian. In *Proceedings of the Seventeenth Annual UCLA Indo-European Conference, Los Angeles October 27-28, 2005*, ed. Karlene Jones-Bley et al., 51–65. Washington, DC: Institute for the Study of Man (JIES Monogr. 52).

Fellner, Hannes A. 2007. The Expeditions to Tocharistan. In *Instrumenta Tocharica*, ed. Melanie Malzahn, 13–36. Heidelberg: Universitätsverlag Winter.

Filliozat, Jean. 1948. *Fragments de textes koutchéens de médecine et de magie. Texte, parallèles sanskrits et tibétains, traduction et glossaire*. Paris: Librairie d'Amérique et d'Orient Adrien-Maisonneuve.

von Gabain, Annemarie, and Werner Winter. 1958. *Türkische Turfantexte IX. Ein Hymnus an den Vater Mani auf "Tocharisch" B mit alttürkischer Übersetzung*. Abhandl. d. Deutschen Akad. d. Wissenschaften zu Berlin, Kl. f. Sprachen, Literatur und Kunst, 1956, 2. Berlin: Akademie-Verlag.

Geng, Shimin, and Hans-Joachim Klimkeit et al. 1988. *Das Zusammentreffen mit Maitreya. Die ersten fünf Kapitel der Hami-Version der Maitrisimit*. Wiesbaden: Harrassowitz.

Geng, Shimin, Jens Peter Laut, and Georges-Jean Pinault. 2004. Neue Ergebnisse der Maitrisimit-Forschung. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 154:347–369.

Gippert, Jost. 1987. Zu den sekundären Kasusaffixen des Tocharischen. *Tocharian and Indo-European Studies* 1:22–39.

Hackstein, Olav. 1995. *Untersuchungen zu den sigmatischen Präsensstammbildungen des Tocharischen*, volume 38 of *Historische Sprachforschung. Ergänzungsheft*. Göttingen: Vandenhoek & Ruprecht.

Hackstein, Olav. 2000. Review of Hilmarsson (1996). *Kratylos* 45:96–104.

Hackstein, Olav. 2001. Studien zur Grammatikalisierung in älteren indogermanischen Sprachen. *Historische Sprachforschung* 114:15–42.

Hackstein, Olav. 2004. From Discourse to Syntax: The Case of Compound Interrogatives in Indo-European and Beyond. In *Proceedings of the 15th Annual UCLA Indo-European Conference, Los Angeles, November 7-8 2003*, ed. Karlene Jones-Bley et al., 257–298. DC: Institute for the Study of Man (JIES Monogr. 49).

Hackstein, Olav. 2005. Archaismus oder historischer Sprachkontakt. Zur Frage westindogermanisch-tocharischer Konvergenzen. In *Sprachkontakt und Sprachwandel*, ed. Gerhard Meiser and Olav Hackstein, Akten der XI. Fachtagung der Indogermanischen Gesellschaft, 17.23. September 2000, Halle an der Saale, 169–184. Wiesbaden: Reichert.

Halpern, Aaron. 1995. *On the morphology and placement of clitics*. Dissertations in Linguistics. Stanford, CA: CSLI Publications.

Hayes, Bruce. 1989. The Prosodic Hierarchy in Meter. *Phonetics and Phonology* 1:201–260.

Hilmarsson, Jörundur. 1987. Stray notes on the interrogative pronominal stems in Tocharian. *Tocharian and Indo-European Studies* 1:40–48.

Hilmarsson, Jörundur. 1993. Development of labiovelars (and tectals plus *u̯*) in initial position in Tocharian (an overview). *Die Sprache* 35:176–186.

Hilmarsson, Jörundur. 1996. *Materials for a Tocharian Historical and Etymological Dictionary*, volume 5 of *Tocharian and Indo-European Studies Supplementary Series*. Alexander Lubotsky and Guðrún Þórhallsdóttir (eds). Reykjavík: Málvísindastofnun Háskóla Íslands.

Hoffner, Harry A., and H. Craig Melchert. 2008. *A Grammar of the Hittite Language*. Winona Lake, Indiana: Eisenbrauns.

Itô, Junko, and Armin Mester. 1999. Realignment. In *The Prosody-Morphology Interface*, ed. R. Kager, H. v.d. Hulst, and W. Zonneveld, 188–217. Cambridge, U.K.: Cambridge University Press.

Jasanoff, Jay H. 1987. Some Irregular Imperatives in Tocharian. In *Studies in Memory of Warren Cowgill (1929-1985): Papers from the Fourth East Coast Indo-European Conference, Cornell University, June 6-9, 1985*, ed. Calvert Watkins, 92–112. Berlin - New York: Mouton de Gruyter.

Ji, Xianlin, Werner Winter, and Georges-Jean Pinault. 1998. *Fragments of the Tocharian A Maitreyasamiti-Nāṭaka of the Xinjiang Museum, China*, volume 113 of *Trends in Linguistics. Studies and Monographs*. Berlin - New York: Mouton de Gruyter.

Kawahara, Shigeto. 2007. Sonorancy and geminacy. In *Papers in optimality theory iii*, ed. Leah Bateman, Michael O'Keefe, Ehren Reilly, and Adam Werle, volume 32 of *University of Massachusetts Occasional Papers in Linguistics*, 145–186. Amherst: Graduate Linguistics Student Association.

Kaye, Jonathan D. 1992. Do you believe in magic? The story of s+C sequences. *SOAS Working Papers in Linguistics and Phonetics* 2:293–313.

Kim, Ronald I. 1999. The development of labiovelars in Tocharian: a closer look. *Tocharian and Indo-European Studies* 8:139–187.

Kim, Ronald I. 2007. Vowel Weakening in Tocharian A Preterite Participles and Abstract Nouns. *Kyoto University Linguistic Research* 26:1–30.

Kim, Ronald I. 2012. Eine indirekte Fortsetzung urindogermanischer Prosodie? Zweitsilbenakzent bei finiten Verben im Tocharischen. In *Multilingualism and History of Knowledge.* Volume 2: *Linguistic Developments along the Silk Road: Archaism and Innovation in Tocharian*, ed. Olav Hackstein and Ronald I. Kim, Sitzungsberichte der Österreichischen Akademie der Wissenschaften, Philosophisch-historische Klasse 834, 9–32. Wien: Verlag der ÖAW.

Kiparsky, Paul. 2000. Opacity and Cyclicity. *Linguistic Review* 17:351–366.

Kisseberth, Charles W. 1970. On the Functional Unity of Phonological Rules. *Linguistic Inquiry* 1:291–306.

Klingenschmitt, Gert. 1994. Das Tocharische aus indogermanischer Sicht. In *Tocharisch.*

Akten des Kolloquiums der Indogermanischen Gesellschaft, Berlin, September 1990, ed. Bernfried Schlerath, 310–411. Reykjavík: Málvísindastofnun Háskóla Íslands.

Kloekhorst, Alwin. 2014. *Accent in Hittite. A Study in Plene Spelling, Consonant Gradation, Clitics, and Metrics*, volume 56 of *Studien zu den Boğazköy-Texten*. Wiesbaden: Harrasowitz.

Koller, Bernhard. 2008. An Approach to Tocharian Schwa-Zero Alternations in Government Phonology. Master's thesis, Universität Wien.

Kölver, Bernhard. 1965. Der Gebrauch der sekundaären Kasus im Tocharischen. Doctoral Dissertation, Johann Wolfgang Goethe-Universität in Frankfurt am Main.

Krause, Wolfgang. 1952. *Westtocharische Grammatik. Band I: Das Verbum*. Heidelberg: Carl Winter.

Krause, Wolfgang, and Werner Thomas. 1960. *Tocharisches Elementarbuch. Grammatik*, volume 1. Heidelberg: Carl Winter.

Krause, Wolfgang, and Werner Thomas. 1964. *Tocharisches Elementarbuch. Texte und Glossar*, volume 2. Heidelberg: Carl Winter.

Lejeune, Michel. 2005. *Phonétique historique du mycénien et du grec ancien*. Paris: Klincksieck.

Lubotsky, Alexander, and Sergei A. Starostin. 2003. Turkic and Chinese loan words in Tocharian. In *Language in Time and Space. A Festschrift for Werner Winter on the occasion of his 80th Birthday*, ed. Brigitte L. M. Bauer and Georges-Jean Pinault, 157–269. Berlin - New York: Mouton de Gruyter.

Malzahn, Melanie. 2007a. A Tocharian Brahmi Chart. In *Instrumenta Tocharica*, ed. Melanie Malzahn, 223–254. Heidelberg: Universitätsverlag Winter.

Malzahn, Melanie. 2007b. The Most Archaic Manuscripts of Tocharian B and the Varieties of the Tocharian B Language. In *Instrumenta Tocharica*, ed. Melanie Malzahn, 255–297. Heidelberg: Universitätsverlag Winter.

Malzahn, Melanie. 2010. *The Tocharian Verbal System*, volume 3 of *Brill's Studies in Indo-European Languages & Linguistics*. Leiden - Boston: Brill.

Malzahn, Melanie. 2012a. Now you see it, now you don't – Bewegliches *o* in Tocharisch B. In *Multilingualism and History of Knowledge.* Volume 2: *Linguistic Developments along the Silk Road: Archaism and Innovation in Tocharian*, ed. Olav Hackstein and Ronald I. Kim, Sitzungsberichte der Österreichischen Akademie der Wissenschaften, Philosophisch-historische Klasse 834, 33–82. Wien: Verlag der ÖAW.

Malzahn, Melanie. 2012b. Position matters: The placement of clitics in metrical texts of Tocharian B. *Tocharian and Indo-European Studies* 13:153–162.

Marggraf, Wolf-Jürgen. 1970. Untersuchungen zum Akzent in Tocharisch B. Doctoral Dissertation, Christian-Albrechts-Universität zu Kiel.

McCarthy, John J., and Alan Prince. 1995. Faithfulness and reduplicative identity. In *Papers in optimality theory*, ed. Jill N. Beckman, Laura Walsh, and Suzanne Urbanczyk, volume 18 of *University of Massachusetts Occasional Papers in Linguistics*, 249–384. Amherst: Graduate Linguistics Student Association.

Monier-Williams, Monier. 1899 [2014]. *A Sanskrit-English dictionary. Etymologically and philologically arranged with special reference to cognate Indo-European languages*. Retrieved on 06/01/2015 from the digitized version hosted by the Universität zu Köln (http://www.sanskrit-lexicon.uni-koeln.de/monier/).

Nespor, Marina, and Irene Vogel. 1986 [2007]. *Prosodic Phonology*, volume 28 of *Studies in Generative Grammar*. Berlin: Mouton de Gruyter.

Peyrot, Michaël. 2008. *Variation and change in Tocharian B*. Amsterdam - New York: Rodopi.

Peyrot, Michaël. 2013. *The Tocharian Subjunctive. A Study in Syntax and Verbal Stem Formation*, volume 8 of *Brill's Studies in Indo-European Languages & Linguistics*. Leiden/Boston: Brill.

Pinault, Georges-Jean. 1989. *Indroduction au tocharien*, volume 7 of *LALIES. Actes des sessions de linguistique et de littérature*. Paris: Presses de l'École Normale Supérieure.

Pinault, Georges-Jean. 1997. Sur l'assemblage des phrases ('Satzgefüge') en tokharien. In *Berthold Delbrück y la sintaxis indoeuropea hoy. Actas del Coloquio de la Indogermanische Gesellschaft. Madrid, 21-24 de septiembre de 1994*, ed. Emilio Crespo and José Luis García Ramón, 449–500. Madrid: Ediciones de la UAM; Wiesbaden: Dr. Ludwig Reichert Verlag.

Pinault, Georges-Jean. 1999. Tokharien A kapśañi, B kektseñe. In *Compositiones Indogermanicae in memoriam Jochem Schindler*, ed. Heiner Eichner and Hans Christian Luschützky, 457–478. Praha: enigma corporation.

Pinault, Georges-Jean. 2006. Morphologie de l'ablatif tokharien. In $^{GIŠ.ḪUR}$*gul-za-at-tara. Festschrift for Folke Josephson*, ed. Gerd Carling, number 32 in Meijerbergs Arkiv, 248–283. Göteborg: Meijerbergs institut.

Pinault, Georges-Jean. 2008. *Chrestomathie tokharienne*. Leuven - Paris: Peeters.

Pinault, Georges-Jean. 2010. Le pronom d'ipséité en tokharien. In *Typologie et comparatisme. Hommages offerts à Alain Lemaréchal*, ed. Injoo Choi-Jonin, Marc Duval, and Olivier Soutet, 351–365. Leuven: Peeters (Orbis/Supplementa 28).

Prince, Alan, and Paul Smolensky. [1993] 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell Publishing.

Prince, Ellen F. 1981. On the Inferencing of Indefinite-this NPs. In *Elements of Discourse Understanding*, ed. Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag, 231–250. Cambridge: Cambridge University Press.

Pronk, Tijmen. 2009. Reflexes of the deletion and insertion of Proto-Tocharian *ä in Tocharian B. *Tocharian and Indo-European Studies* 11:73–123.

Ringe, Don, Jr. 1996. *On the Chronology of Sound Changes in Tocharian*, volume 80 of *American Oriental Series*. New Haven, Connecticut: American Oriental Society.

Ringe, Donald A. Jr. 1990. Review of Adams (1988). *Language* 66:400–408.

Rose, Sharon. 2000. Rethinking Geminates, Long-Distance Geminates, and the OCP. *Linguistic Inquiry* 31:85–122.

Sander, Lore. 1968. *Palographisches zu den Sanskrithandschriften der Berliner Turfansammlung*. Wiesbaden: Franz Steiner Verlag GMBH.

Selkirk, Elizabeth. 1996. The Phonological Structure of Function Words. In *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*, ed. James L. Mogan and Katherine Demuth, 187–213. Mahwah, NJ: Lawrence Erlabum Associates, Publishers.

Sieg, Emil, and Wilhelm Siegling. 1921. *Tocharische Sprachreste*. Berlin - Leipzig: Walter de Gruyter.

Sieg, Emil, Wilhelm Siegling, and Wilhelm Schulze. 1931. *Tocharische Grammatik*. Göttingen: Vandenhoeck & Ruprecht.

Sieg, Emil, Wilhelm Siegling, and Werner Thomas. 1953. *Tocharische Sprachreste, Sprache B, Heft 2*. Göttingen: Vandenhoeck und Ruprecht.

Stumpf, Peter. 1971a. *Der Gebrauch der Demonstrativ-Pronomina im Tocharischen*. Wiesbaden: Harrassowitz.

Stumpf, Peter. 1971b. Der vokalische Sandhi im Tocharischen. *Zeitschrift für vergleichende Sprachforschung* 85:96–133.

Stumpf, Peter. 1974. Der Plural der westtocharischen Demonstrativ-Pronomina – zugleich ein Beitrag zur Dialekt-Gliederung des Westtocharischen. *Orbis* 23:404–428.

Stumpf, Peter. 1990. *Die Erscheinungsformen des Westtocharischen*, volume 2 of *Tocharian and Indo-European Studies Supplementary Series*. Reykjavík: Skákprent.

Taylor, Martha. 1985. Some patterns of geminate consonants. *The University of Chicago Working Papers in Linguistics* 1:120–129.

Thomas, Werner. 1952. *Die tocharischen Verbaladjective auf -l*, volume 9 of *Deutsche Akademie der Wissenschaften zu Berlin, Institut für Orientforschung*. Berlin: Akademie-Verlag.

Thomas, Werner. 1957. *Der Gebrauch der Vergangenheitstempora im Tocharischen*. Wiesbaden: Harrassowitz.

Thomas, Werner. 1967. Zu wortverbindendem toch. A śkaṃ/B ṣpä. *Zeitschrift für vergleichende Sprachforschung* 81:161–180.

Thomas, Werner. 1968. Zur Verwendung von toch. A oki/B ramt und A mäṃtne/B mäkte in Vergleichen. *Orbis* 17:198–231.

Thomas, Werner. 1979. Zur Verwendung von A śla, B ṣale, ṣle im Tocharischen. *Zeitschrift für vergleichende Sprachforschung* 93:150–173.

Tomezzoli, GT, and Joseph Kreutz. 2011. The linguistic position of the tocharian. In *Proceedings of the 9th International Topical Conference: Origin of Europeans, Ljubljana*, 67–86.

Van Windekens, Albert Joris. 1976. *Le tokharien confronté avec les autres langues indo-européennes. Vol. I. La phonétique et le vocabulaire*. Travaux publiés par le Centre International de Dialectologie Générale de l'Université Catholique Néerlandaise de Louvain XI. Louvain: Centre International de Dialectologie Générale.

Vogel, Irene. 2009. The status of the Clitic Group. In *Phonological Domains: Universals and Deviations*, ed. Janet Grijzenhout and Barış Kabak, volume 16 of *Interface Explorations*, 15–46. Berlin: Mouton de Gruyter.

Whitney, William Dwight. 1896. *Sanskrit Grammar* (3rd edition). Cambrigde, Massachusetts: Harvard University Press, 2003. Print.

Winter, Werner. 1955. A linguistic classification of "Tocharian" B texts. *Journal of the American Oriental Society* 75:216–225 = Winter (2005a: 1–10).

Winter, Werner. 1959. Zur "tocharischen" Metrik. In *Akten des XXIV. Internationalen Orientalistenkongresses München 1957*, ed. Herbert Franke, 520–521. Wiesbaden: Deutsche Morgenländische Gesellschaft. = Winter (2005a: 26-27).

Winter, Werner. 1967. A Note on Cases. In *To Honor Roman Jakobson*, 2250–2253. The Hague - Paris: Mouton. = Winter (2005b: 737-740).

Winter, Werner. 1970. Über tocharische Verbformen in ihrem Verhältnis zum Satzakzent. *Untersuchungen zum Akzent in Tocharisch B.* Wolf-Jürgen Marggraf. Christian-Albrechts-Universität zu Kiel, 95–103. = Winter (2005a: 141–149).

Winter, Werner. 1982. Tocharian and Proto-Indo-European. *Lingua Posnaniensis* 25:1–11. = Winter (2005a: 251–261).

Winter, Werner. 1990. The importance of fine points in spelling: deletion of accented vowels in Tocharian B. In *Historical linguistics and philology*, ed. Jacek Fisiak, 371–391. Berlin - New York: Walter de Gruyter. = Winter (2005a: 393-413).

Winter, Werner. 1994. Syncope in Tocharian A. In *In honorem Holger Pedersen*, ed. Jens Elmegård Rasmussen, 401–415. Wiesbaden: Reichert. = Winter (2005a: 450-464).

Winter, Werner. 2005a. *Kleine Schrifte/Selected Writings. Festgabe aus Anlass des 80. Geburtstags.* Volume 1. Bremen: Hempen Verlag.

Winter, Werner. 2005b. *Kleine Schrifte/Selected Writings. Festgabe aus Anlass des 80. Geburtstags.* Volume 2. Bremen: Hempen Verlag.

Zwicky, Arnold M., and Geoffrey K. Pullum. 1983. Cliticization vs. Inflection: English N'T. *Language* 59:502–513.