BRILL

INDO-EUROPEAN LINGUISTICS 2 (2014) 1–23

brill.com/ieul

Caesurae, Bridges, and the Colometry of Four Tocharian B Meters

*Christoph Bross**
Ludwig-Maximilians-Universität München
christoph.bross@lrz.uni-muenchen.de

Dieter Gunkel
Ludwig-Maximilians-Universität München
dieter.gunkel@lrz.uni-muenchen.de

Kevin M. Ryan
Harvard University
kevinryan@fas.harvard.edu

Abstract

The Tocharians composed verse in hierarchical structures, with the verse dominating major cola, and the major colon in turn dominating one or more minor cola. After providing much-needed descriptive data on Tocharian meter, we assess the evidence for the distinction between major vs. minor caesurae in some of the most popular Tocharian B meters, finding support for the commonly assumed colometries in some but not all cases. Of particular interest is the recurring 4+3-syllable colon, since the violability of its internal (putatively minor) caesura varies significantly across meters. We argue that this varying strictness is indeed a function of the meter as opposed to position in the verse, verse length, idiosyncrasies of certain texts, and so forth. We then use a systematic prose comparison method to test the meters for bridges, finding evidence for monosyllable avoidance in (certain) colon-final positions, despite an overall preference for monosyllables in verse vis-à-vis prose. Finally, we discuss the implications that our study has for the restoration of fragmentary Tocharian texts.

* We wish to thank Olav Hackstein, Ron Kim, Melanie Malzahn, Michaël Peyrot, two anonymous referees, and the audiences at the Forschungskolloquium zur historisch-vergleichenden Sprachwissenschaft (LMU Munich, 2013) and Tocharian Texts in Context (University of Vienna, 2013) for insightful questions, suggestions, and criticism.

 | DOI: 10.1163/22125892-00201005

Keywords

bridges – caesurae – colometry – corpus linguistics – metrics – Tocharian

1 Preliminaries

In Tocharian B, poems are composed of stanzas that can be classified according to how many verses they contain and whether or not those verses are isosyllabic. Most stanzas consist of four verses, and most of those consist of four isosyllabic verses, e.g. 4×12 syllables, 4×14, 4×15, and 4×18. Anisosyllabic four-verse stanzas are not uncommon, e.g. 21/21/18/13, 14/11/11/11, and 20/22/10/15. There is one five-verse stanza, which is anisosyllabic, 13/13/13/13/21.[1] The scribes often mark verse-end with a colon (:) or a raised dot (·), and they number the stanzas.[2]

Caesurae articulate verses into cola. Longer cola are generally assumed to be subdivided into minor cola (e.g. Winter 1959; Malzahn 2012a: 34, 2012b: 154; Adams 2013: 3). For example, the 4×14 verse, which is usually composed of 4+3+4+3 syllables—referred to here as the "basic colometry"—contains two seven-syllable major cola (M), each of which contains two minor cola (m) of four and three syllables (σ). The hierarchical structure implicit in the standard metrical analysis can be represented as a tree diagram or with bracketing, as in Figure 1.

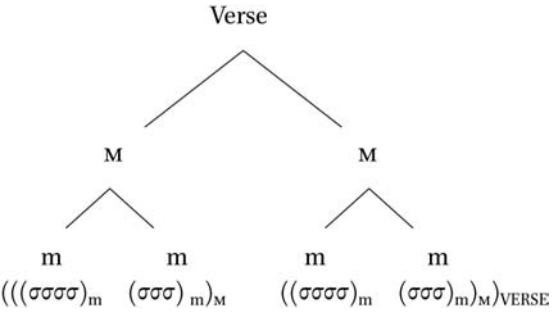

FIGURE 1 *Hierarchical structure posited for 4×14 meter in Tocharian*

1 There is also one fragmentary poem apparently composed in two-verse stanzas, 2×14 (THT 133).

2 The pioneering discussion of Tocharian versification is Sieg and Siegling (1921: x–xi). For an overview of Tocharian versification and an inventory of stanzas and meters, cf. Stumpf (1971a: 71–72), Thomas (1983: 272–276), and Pinault (2008: 399).

We will refer to this structure shorthand as [7||7] or [4|3||4|3], marking caesurae between putative major cola (major caesurae) with "||", caesurae between putative minor cola (minor caesurae) with "|", indicating violations of caesurae with "!", and vowel sandhi with "≠", as in the following 4×14 stanza from the *Udānālaṅkāra* (THT 5 a4–6).

68a wñā-neś (po)yśi | karuntsa || mā tañ ñyātstse | śolantse **:**
68b mā r≠ asānmeṃ | laitalñe || ceṃ sklok ptārka | pälskomeṃ **:**
68c kos tne ñakta | pelaikni || (po) śaiṣṣents≠ ā!naiwacci **:**
68d tary≠ akṣā-ne | pudñäkte || teki ktsaitsñe | srukalñe 68

The omniscient one spoke to him with compassion: 'Your life is not in danger,
nor (will you) fall from the throne. Let this doubt go from your mind,
o master, as unpleasant laws hold here for all the world.'
The Buddha proclaimed three to him: sickness; old age; and death.

In addition to these metrical principles, a number of linguistic phenomena involving various components of the grammar are either confined to, or particularly frequent in, verse (cf. Pinault 2008: 401–405 and 6.2 below).

2 Evidence for a Distinction between Major and Minor Cola

The assumption of a specifically hierarchical distinction between major and minor cola in Tocharian is intuitive for 4×14 due to the rhythmic symmetry of the verse, i.e. the repetition of 4+3-syllable sequences. In verses without this symmetry, e.g. 4×12 (5+**4**+**3**), 4×15 (**4**+**3**+3+5), and 4×18 (**4**+**3**+**4**+**3**+4), the recurrence of 4+3 across the meters, especially verse-peripherally, is suggestive of its metrical coherence. Positing hierarchical structure within the verse permits insightful analyses of various metrical phenomena (cf. Prince 1989; Hayes and MacEachern 1998; Kiparsky 2006; Hayes 2010: 2515–2516). For example, the analysis of the iambic metron of Ancient Greek tragic trimeter given in Figure 2, which involves both hierarchical structure and a strong vs. weak distinction, allows for an intuitive explanation of the asymmetry between the anceps position (x) of the weak foot and the brevis (⏑) of the strong one.

Any asymmetry in the way that the Tocharian B poets composed cola can thus in principle reflect a difference in their metrical status. Whether that difference involves hierarchical structure, binary distinctions such as strong vs. weak, or something else, is a different matter, which we return to in Section 4.

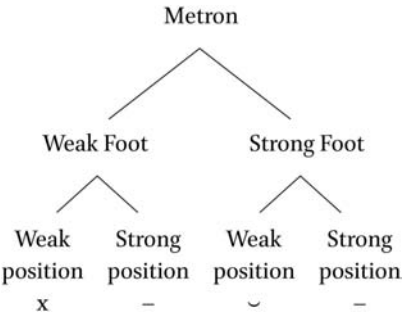

FIGURE 2 *Hierarchical representation of an iambic metron in Ancient Greek*

Crucially, however, it must be shown that such an asymmetry should be attributed to the meter *per se* rather than to some other aspect of the language, e.g. the syntax.

2.1 *Distribution of Clitics*

Winter (1959) identifies such an asymmetry in the distribution of the sentential clitics *ra*, *ka*, *ṣpä*, and *no* in 7-syllable cola in several Tocharian B meters. Positing the sort of structure given for 7-syllable cola in Figure 1, he points out that the clitics occur at the end of the minor colon ($\underline{\sigma}$) ca. twice as often as they do at the end of the major colon (σ).[3]

$$(\,(\sigma\sigma\underline{\sigma})_{\mathrm{m}}\,(\sigma\sigma\sigma)_{\mathrm{m}}\,)_{\mathrm{M}}$$

In a more thorough examination of the phenomenon, Malzahn (2012b) argues convincingly that the distribution should be attributed in the main to an aspect of Tocharian B syntax, specifically to the localization of second position clitics ("Wackernagel's Law"). Of the 205 sentential clitics in her verse corpus, 93% follow the first "orthotonic" word in their syntactic clause, just as they do in prose. Malzahn's study exemplifies the need to rule out potential confounds from syntax and other areas of the grammar when studying meter. Since poetic and prose texts in Tocharian B are roughly contemporary and compatible in genre, prose provides an excellent baseline for comparison. Any systematic

3 Winter considers both positions to be metrically weak; Malzahn (2012a, 2012b) argues that they are strong. On either analysis, the asymmetry—if metrical—would therefore be due to the major vs. minor distinction, not the strong vs. weak one.

differences between verse and prose can then be ascribed to the meter or to other characteristics proper to versification.[4]

2.2 *Caesura Violability*

Von Gabain and Winter (1958: 33–34), the first to propose a distinction between major and minor cola, characterize the minor caesurae as more readily violable than major ones:

> Wir dürften damit berechtigt sein, neben festen Hauptzäsuren auch Nebenzäsuren anzunehmen, d. h. fakultativ aufhebbare Grenzen zwischen Unterabschnitten innerhalb der Kolen.[5]
>
> VON GABAIN AND WINTER 1958: 34

This characterization is repeated in Winter (1959) and later work, but has not yet been systematically studied. Accepting Malzahn's analysis of clitic distribution in verse, caesura violability is to our knowledge the only remaining diagnostic that has been proposed for the distinction between major and minor cola in Tocharian.

3 Violability in Four Meters

In this section, we investigate the violability of the caesurae in 4×12, 4×15, 4×14, and 4×18. We compare the violability of putative major caesurae with putative minor caesurae, and find that the caesurae in 4+3-syllable sequences differ significantly from other caesurae in the same meter. This can be taken as evidence for their minor status and supports the standard colometries found in recent descriptions of Tocharian meter. Against the standard colometries, however, caesura violability provides no evidence for the minor status of the caesura after the 10th syllable in 4×15. The following generalization holds for the meters studied here: from the standpoint of violability, the caesurae in 4+3-syllable sequences are minor caesurae; the others are major caesurae.

4 For the use of prose as a baseline for comparison with verse, cf. Watkins (1999). For the method more generally, cf. Ryan (2011).

5 "In addition to strict major caesurae, this permits us to posit minor caesurae, i.e. optionally violable boundaries between subconstituents within the cola."

3.1 *Caesura Violations in 4×12 (5+4+3)*

The colometry of 4×12 (5+4+3) is standardly given as [5||7], with two major cola.[6] Given the standard analysis of 4+3 sequences ("7ers"), this implies [5||4|3]. If the poets' preference for respecting the major caesura is stronger than their preference for respecting the minor one, we would expect them to violate the caesura after the 9th syllable significantly more frequently than the one after the 5th. Figure 3 plots the incidence of word boundary in 4×12, based on a corpus of 317 verses.[7] Since the manner in which the poets realize caesurae suggests that they treated sequences of a lexical word followed by a monosyllabic clitic as a single word, we did so as well in both the verse and prose corpora.[8]

Note that the numbers along the x-axis of the plot represent verse-internal word boundaries. The peaks at 5 and 9 reflect the caesurae after the 5th and 9th syllables. In support of the major vs. minor distinction, there is a boundary after 5 in 99.3% of the verses, and after 9 in only 87.0%. The error bars give a sense of which differences are significant; the absence of any overlap between the error bars at 5 and 9 in the plot suggests that the difference is significant.

We can confirm the statistical significance of the difference with Fisher's Exact Test of Independence. We see from Table 1 that 38 of 547 caesurae are violated, and that 36 of those violations occur after 9. Assuming the null hypothesis that the poets treat the caesurae equally, Fisher's Exact Test tells us what the probability is that the violations would be at least this unevenly distributed. The probability (p) is less than .00001, meaning that a difference this great would have arisen by chance less than .001% of the time; we take p values less than .05 to be significant. Caesura violability thus supports the [5||4|3] colometry.

---

6 There is another, much less common 4×12 meter with 4+4+4 colometry.

7 The corpus consists of 12-syllable verses from the following texts: PK AS 16.3 (4 verses); PK AS 7B (33); PK AS 7C (47); PK AS 7D (9); PK AS 7K (27); PK AS 7M (41); PK NS 22 (8); THT 123 (2); THT 132 (2); THT 14 (29); THT 21 (36); THT 22 (47); THT 235 (2); THT 244 (8); THT 284 (22). Some of these verses are fragmentary. For such verses, we counted only positions for which the presence or absence of a boundary could be securely determined. It follows that the counts on which the proportions in the figure are based vary from one position to the next (from 267 to 286, in this case); none is based on the full 317 verses. Here and elsewhere we adopt the restorations and emendations supplied in *A Comprehensive Edition of Tocharian Manuscripts* (CEToM).

8 We took the following clitics into account: *ka*, *kca*, *ksa*, *ñke*, *tne*, *nai*, *no*, *nta*, *pi*, *ra*, *ram*(*t*), *wa*, *wat*, *ṣpä*/*ṣäp*, *ṣai*/*ṣey*, *ste*, *tsa*.

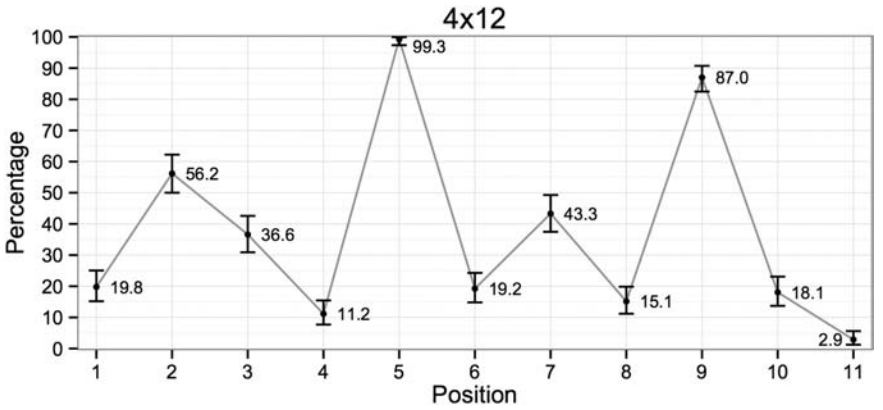

FIGURE 3 *Percentage of verse-internal word boundaries in 4×12 meter. Error bars are 95%-confidence Clopper-Pearson (1934) intervals for proportions.*

TABLE 1 *Violations of caesurae after syllables 5 and 9 in 4×12 meter*

	Violated	**Not violated**	***Row total***
Caesura after 5	2 (.7%)	268	*270*
Caesura after 9	36 (13%)	241	*277*
Column total	*38*	*509*	*547*

The opening stanza of the following passage of the Tocharian B version of the *Karmavibhaṅga* (PK AS 7B a4–5) is fairly representative of the 4×12 corpus as a whole. The caesura after 5 is respected throughout, and the caesura after 9 is violated once, in verse 1b. Note that host-enclitic groups are joined with "+" here and throughout.

1a weñau n$_{ä}$nok yakne(ṃ) ‖ yāmorntats tū | päklyauṣso :
1b krenta yolainaṃts ‖ etrīwaitsāˈnaṃts rano :
1c te keklyauṣormeṃ ‖ epastyañ̈ñe | yänmacer
1d yāmornta yāmtsi ‖ mā+ṣpä triścer | makā-ykne : 1

I will further tell the ways of the deeds—listen to this—
of good and bad ones, of mixed ones also.
Having heard this, you will obtain the skills
to do deeds and you will not err in many ways.

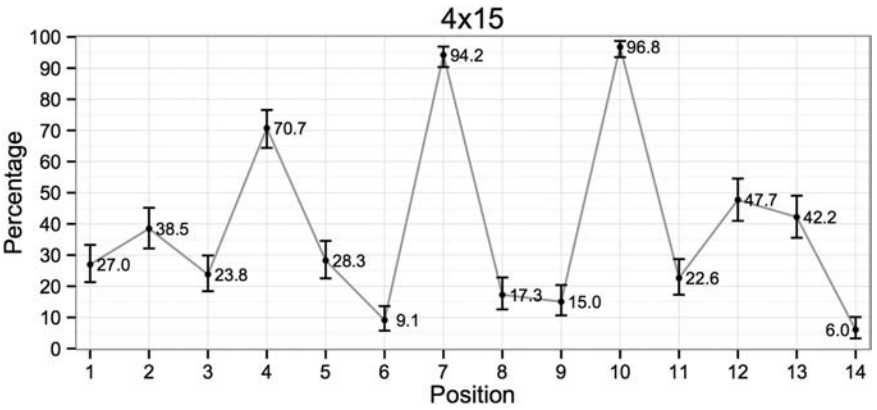

FIGURE 4 *Percentage of verse-internal word boundaries in 4×15 meter*

3.2 *Caesura Violations in 4×15 (4+3+3+5)*

In 4×15, the 7er occurs verse-initially. According to the standard analysis, there are two major cola [7||8]; given the basic colometry 4+3+3+5, this would imply the structure [4|3||3|5].[9] However, the boundary data in Figure 4, based on a corpus of 242 verses,[10] reflect caesurae after syllables 4, 7, and 10. As noted by Stumpf (1971a: 72 with fn. 10), the number of violations decreases numerically towards verse end with the caesura after 10 being the most rigidly enforced.

The poets violate the caesura after 4 significantly more frequently than the caesura after 7 ($p < .00001$) and 10 ($p < .00001$), supporting its minor status. While the putative major caesura after 7 is violated slightly more frequently than the putative minor caesura after 10, the difference can be attributed to chance ($p = .25$), as suggested by the overlapping error bars in Figure 4. From the standpoint of violability, 4×15 could have the structure [4|3||3||5] with three major cola (Bross, Gunkel, and Ryan forthcoming), though independent considerations, such as the symmetry of cola and the bridge data discussed in Section 7, might favor the traditional [4|3||3|5] colometry.

9 There is no support for the supposed alternative colometries 8+7 and 6+4+5 (Bross, Gunkel, and Ryan forthcoming).

10 The corpus consists of 15-syllable verses from the following texts: PK AS 6A (22 verses); PK AS 6B (24); PK AS 6C (26); PK AS 7H (27); PK AS 7I (36); PK AS 7J (14); THT 28 (18); THT 29 (29); THT 291.a (7); THT 30 (39).

TABLE 2 *Violations of caesurae after syllables 4, 7, and 10 in 4×15 meter*

	Violated	**Not violated**	***Row total***
Caesura after 4	67 (29.3%)	162	*229*
Caesura after 7	13 (5.8%)	212	*225*
Caesura after 10	7 (3.2%)	211	*218*
Column total	*87*	*585*	*672*

Stanza 23 of THT 30 (a2–3) is representative of 4×15 meter. The caesura after 4 is violated once in verse 23c, and the other caesurae are respected.

23a cets ce$_u$ silñe | pälskontse || lakle+ṣpä | wīkässi poyśi :
23b toṃ ślokanma | wertsyaine || ākṣa cets | palsko tsārwässiś ·
23c eṃṣketse ā'ratsiśco || yātatsiś | astarñeś ṣeko :
23d serke cmelñe | srukalñents≈ || eṃṣketse | nautalñe yāmtsi 23

To dispel this mental depression and sorrow of theirs, the omniscient one
proclaimed these stanzas in the assembly to comfort their mind:
'For permanent cessation, for continual capacity for purity,
for permanent dispersion of the cycle of birth and death, …'

3.3 *Caesura Violations in 4×14 (4+3+4+3)*

In contrast to 4×12 and 4×15, the putative minor caesurae after syllables 4 and 11 in 4×14 are nearly as strictly enforced as the putative major caesurae. The boundary data in Figure 5 are based on a corpus of 265 verses.[11]

The major caesura after 7 is never violated (in 250 relevant verses), while the minor caesurae after 4 and 11 are violated 9 and 8 times, respectively (in 254 and 252 relevant verses). While numerically small, both differences are significant (p = .004 and p = .007, respectively). Violability thus supports the standard analysis, [4|3||4|3].

11 The corpus consists of 14-syllable verses from the following texts: PK AS 16.2 (4 verses); PK AS 16.3 (3); PK AS 7A (15); THT 127 (8); THT 204 (13); THT 205 (1); THT 228 (12); THT 229 (9); THT 231 (23); THT 240 (3); THT 241 (27); THT 249.a (2); THT 254 (1); THT 255 (37); THT 274 (22); THT 295 (32); THT 296 (7); THT 4 (7); THT 5 (39).

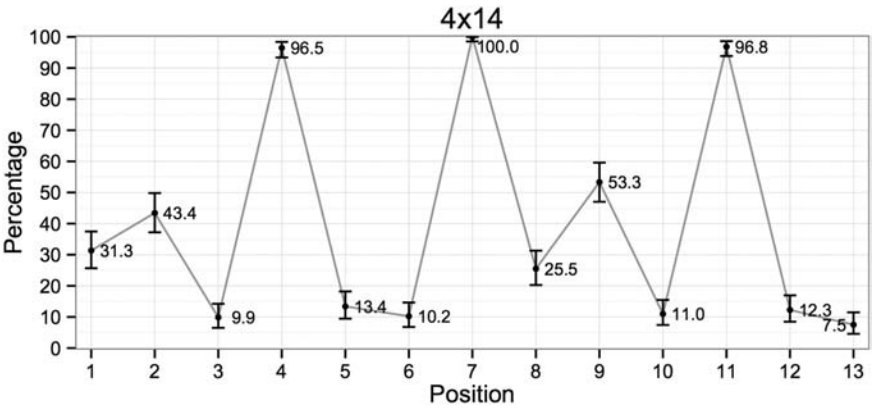

FIGURE 5 *Percentage of verse-internal word boundaries in 4×14 meter*

3.4 *Caesura Violations in 4×18 (4+3+4+3+4)*

The colometry of 4×18 is standardly given as [7||7||4], implying [4|3||4|3||4], as if it were a 4×14 verse plus a 4-syllable colon. The boundary data in Figure 6, based on a corpus of 102 verses,[12] support this colometry to a limited extent. The caesura after 14 is the strongest, being unviolated, and is borderline significantly different from the (putatively minor) caesurae after 4 and 11 (p = .03 in both cases), but not from the (putatively major) caesura after 7 (p = .24).[13] Nevertheless, none of the caesurae after 4, 7, or 11 is significantly different from the others. The numerical trends, for their part, support [4|3||4|3||4], with the caveat that not every "||" is significantly different from every "|" in this scheme. Given the relatively small size of the corpus and the ceiling effect, the p-values are not particularly informative in this case.

The following stanza from PK AS 4A b2–4 (*Udānastotra*) is representative of 4×18 meter:

27a paiykalñesa | drohavārg || akālk kñītär-ñ | serkene || po cmelaṣṣe :
27b mamāntaṣ+ra | yolainne || mā ñi t(ā)koy | māntalyñe || $k_u$se+ṣ krentäṃnne :

12 The corpus consists of 18-syllable verses from the following texts: IOL Toch 5 (22 verses); PK AS 4A (22); THT 10 (2); THT 11 (30); THT 221 (1); THT 266 (3); THT 271 (5); THT 296 (2); THT 8 (15).

13 If these p-values are corrected with a penalty for multiple comparisons (such as the Bonferroni correction), the two .03 values cross over .05. They may therefore be regarded as at best borderline in terms of significance.

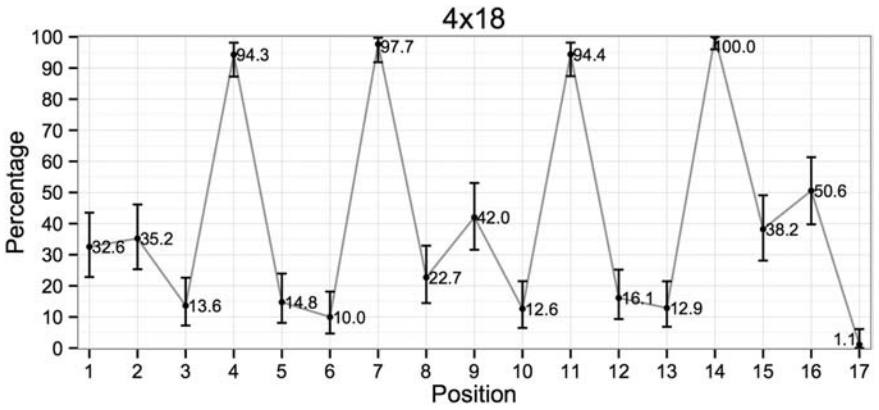

FIGURE 6 *Percentage of verse-internal word boundaries in 4×18 meter*

27c kauṣentai+ra | sanaṃne || mā wer śono | wṣi-ñä+nta || tarkoym trańko :
27d aknātsaimpa | ṣe śmalyñe || mā ñī tāko$_{\text{ī}}$ | śänmīmar || krentäṃmp≠ eṣe 27

May my wish come true in the circle of all births through the writing of the Drohavarga.
May I not bear malice towards an evil person, even if he is malicious, let alone towards those who are good.
May hate and enmity not reside in me at all, even towards a murderous enemy, and may I abandon sin.
May I not meet with an ignorant one, and may I come together with good people.

4 Putative Minor Caesurae in the Four Meters

The poets treat the internal caesurae in 7ers significantly differently in at least three of the four meters. Figure 7 illustrates the variance in violability; the 7ers are arranged from most to least frequently violated and underlined.

This cannot be a function of verse length or of verse-initial vs. verse-final location of the colon, as is clearly demonstrated by the near inviolability of both minor caesurae in 4×14. Nor can it be attributed to the difficulty of metrifying relatively long words in meters with relatively short cola. That would make the opposite prediction. For example, although 5-syllable words can be metrified more comfortably in 4×12 (5+4+3) than in 4×14 (4+3+4+3), the latter exhibits a less violable 7er-internal caesura.

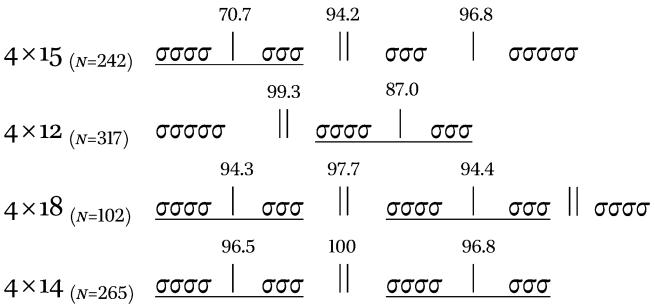

FIGURE 7 *Variance in violability of 7er-internal caesurae*

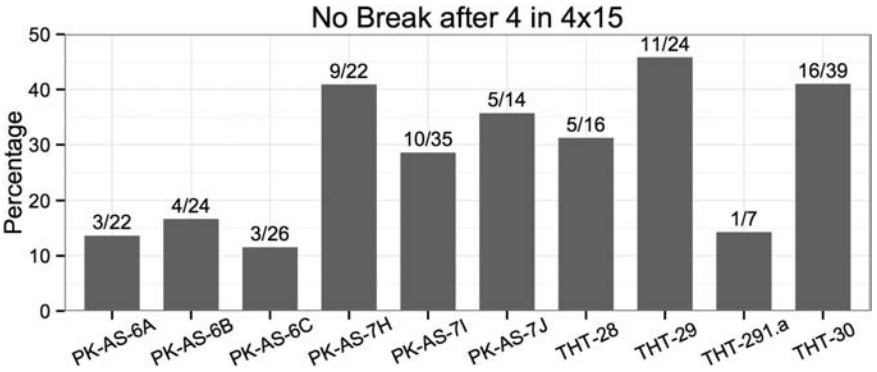

FIGURE 8 *Violations after syllable 4 in 4×15 meter across texts. The counts on which each percentage is based are given above the bar* (*x exceptions/n lines in text*).

We also checked whether certain texts in our corpus were skewing the results. It would be thinkable, for instance, that 7ers in 4×12 are generally quite rigid except in a particular text or group of texts, which could in turn be due to genre, the practice of particular poets, etc. That turns out not to be true: the violations in 4×12 and 4×15 are quite consistent across texts (cf. Figure 8 and Figure 9). For post-4 violations in 4×15, the mean per text is 27.9%, the median 29.9%; for post-9 violations in 4×12, the mean is 9.2%, the median 7.9%. An unpaired t-test on these two vectors is significant ($t(15) = -4.0$, $p = .001$). Thus, taking texts rather than lines as units of analysis, 4×15 is consistently the most violable, followed by 4×12, followed by 4×18 and 4×14, where the minor caesurae in 7ers are nearly inviolable and thus extremely consistent across texts.

The data invite us to consider the hypothesis that the violability of 7ers is a function of the flexibility of the meter as a whole. The most violable 7er-internal

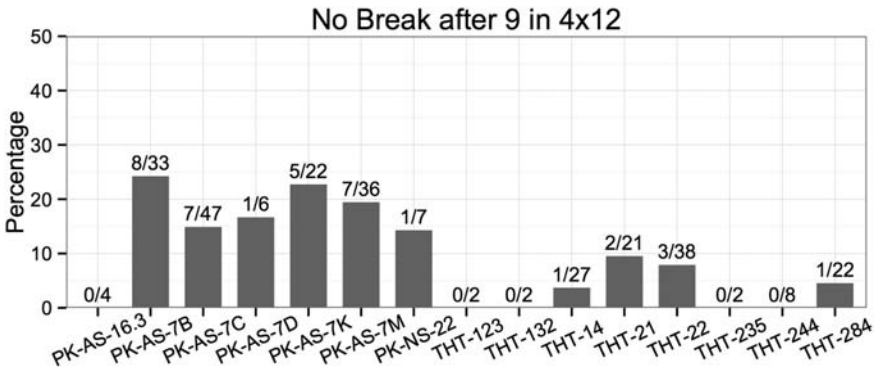

FIGURE 9 *Violations after syllable 9 in 4×12 meter across texts*

caesura is in the meter that also has the most violable major caesura, and the least violable ones are both in 4×14, which also has the strictest major caesura, perhaps encouraged by its symmetry. Furthermore, a sample of three stanzas from each of the meters suggests that enjambment is most frequent in 4×15, and least frequent in 4×18 and 4×14. The correlation is not perfect in either case, however, since 4×12 and 4×18 do not conform to the expected order.

5 Interim Summary

To summarize, in Section 3 we showed that the caesurae in 4+3 sequences ("7ers") are significantly more violable than the other caesurae in the same meter, supporting their minor status. In Section 4, we demonstrated that the violability of the minor caesurae in 7ers is also significantly different across meters. In other words, the 7ers are different from each other. This variance in violability of 7er-internal caesurae cannot be attributed solely to a categorical major vs. minor distinction of the sort sketched in Section 2. It appears to be a function of the meter, and may not be a property that is confined to 7ers, but a property of the meter as a whole.

6 Word Boundary Incidence in Verse and Prose

So far, we have examined the poets' treatment of caesurae using meter-to-meter comparisons. Prose-based comparison is useful for assessing possible bridges in the meters, which, if present, are subtler than caesurae in Tocharian B verse.

TABLE 3 *ICs in prose*[14]

	Constituents	Examples of punctuation
Clause-level	Statements	THT 88 a1; IOL Toch 247 a5
	Commands	IOL Toch 248 b6
	Questions	THT 88 b3
	Correlative clauses	IOL Toch 4 a5
	Conditional protases	IOL Toch 247 b1
	Temporal and causal clauses	IOL Toch 247 a3
	Absolut(iv)e constructions	IOL Toch 178 b8
	Relative protases	IOL Toch 248 a6–b1
	Circumstantial participial clauses	–
	Complement infinitive clauses	–
Phrase-level	Noun phrases in lists	PK AS 16.3 b2–3; THT 108 b4
	Utterance-initial vocatives	THT 107 b1
	Utterance-initial interjections	IOL Toch 247 a5–6

6.1 *Prose Corpus*

For this purpose, we assembled a 2,107-word corpus consisting of the non-metrical, non-fragmentary passages from the following texts: THT 88, 107, 108, 192, 560; IOL Toch 4, 178, 247–248; PK AS 17 A–D, H–K, 16.2–3. The tests require us to identify intonational consituents (ICs) in the prose corpus. We assume that the clause- and phrase-level syntactic constituents in Table 3 were mapped to Intonational Phrases (cf. Nespor and Vogel 2007; Selkirk 2011); scribal punctuation after these constituents was apparently optional, but provides some support for their reality.

The following passage (THT 88 a4–5) illustrates our identification of ICs in the prose corpus.

(*tu lyelyakormeṃ*) (*vṛkṣavāsike ñakte śle māṃtsalyñe śanoś*
this see:ABS tree.dwelling god with sorrow wife:ALL
weṣṣäṃ) (*lariya*) (*pālka+nai mā-ṣekaṃñe wäntarwats*
say:3SG.PRS dear see:SG.IPV-PTC impermanence thing:GEN.PL
sparkālye2 āke)
disappearance end

14 IOL Toch 248 a3 may provide an example of punctuation after a new (i.e. non-given) topic, *uppalavarṇañ aśiyantse* … "To the nun Utpalavarṇā …".

> "Having seen this, the tree-dwelling god says to his wife with sadness: 'Darling, look at the impermanence of things and their ultimate disappearance!'"

We treated host-enclitic units in the prose corpus in the same way that we treated them in the verse corpora.

6.2 *Average Word Length*

An important difference between the prose corpus and the metrical corpus is the average length of words: 2.5 syllables in prose vs. 2.2 in verse. This is likely due in great part to the restrictions that colon size places on verse composition. In 4×14, for example, a word of 5 or more syllables cannot be localized anywhere in the verse without violating a caesura, and 4-syllable words can only be localized spanning positions 1–4 or 8–11. A simple way to quantify average colon size for the different metrical corpora is to divide the number of syllables per verse by the number of cola. This gives a kind of average ideal colon size, since it does not take differences in caesura violability into account. The relationship between colon size and verse length is given in Table 4; all monosyllables are treated as autonomous words.

TABLE 4 *Correlation between average ideal colon size and average word length*

	Average colon size	**Average word length**	**Number of words**
4×14	3.5	2.167	1645
4×18	3.6	2.162	748
4×15	3.75	2.206	1536
4×12	4	2.170	1544
Prose		2.452	2066

The word length data is plotted in Figure 10.[15] It is clear that longer words are avoided in verse. The higher percentage of shorter words in verse is at least in part an artifact of the underrepresentation of longer words, but the higher skew of monosyllables towards verse than di- or trisyllables seems to reflect a favoring of monosyllables in particular, perhaps because they were useful to fill out cola or to fill particular (e.g. weak) metrical positions. We leave this topic for further investigation.

15 No words of 6 or more syllables occur in the verse corpus; the three tokens that occur in the prose corpus, one each of 7, 9, and 10 syllables, are not plotted.

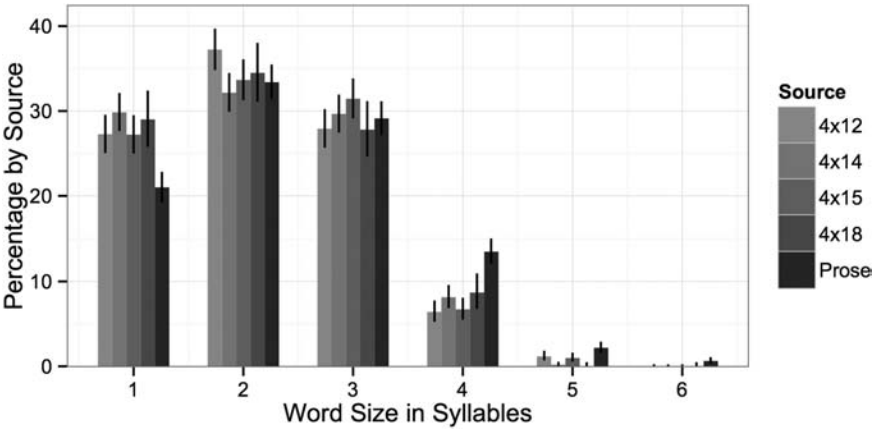

FIGURE 10 *Word length by word type in prose and verse*

The poets employ various tactics to avoid longer words in verse, several of which are exemplified in the stanzas cited above. These include choosing between lexical and morphological alternatives, e.g. the choice between *ṣe* and *eṣe* '(together) with' in 4×18 and the use of nominative plural *pelaikni* in 4×14 for regular *pelaiknenta* 'laws' (cf. Peyrot 2008: 115–116), morphophonological deletion of underlying "weak" vowels, e.g. *wertsyaine* for *wertsiyaine* 'in the assembly' in 4×15 (cf. Winter 1990), and vowel sandhi, e.g. *r≂ asānmeṃ* for *ra asānmeṃ* 'PTC throne:ABL' and *śaiṣṣents≂ ānaiwacci* for *śaiṣṣentse anaiwacci* 'world:GEN unpleasant:NOM.PL' in 4×14 (cf. Stumpf 1971b). This is not to say that the poets simply shorten words wherever they can, of course. In syllable-counting meters, processes that affect syllable count—including augmentation processes such as the retention of an underlying word-final "schwa" as "mobile" *-o* or *-ä*, e.g. *ṣeko* for *ṣek* in 4×15 and *wṣi-ñä* for *wṣi-ñ* in 4×18 (cf. Malzahn 2012a)—are more generally useful. In the aggregate, however, words are shorter in verse than they are in prose.

7 Bridges

Metrical bridges are positions within the verse where poets avoid word boundary. Generally speaking, in order to identify bridges, we want to compare the incidence of colon-internal word boundaries that we observe in verse with what we would expect if the poets were only concerned with respecting caesurae. We can model this expectation by using the syntactic/intonational con-

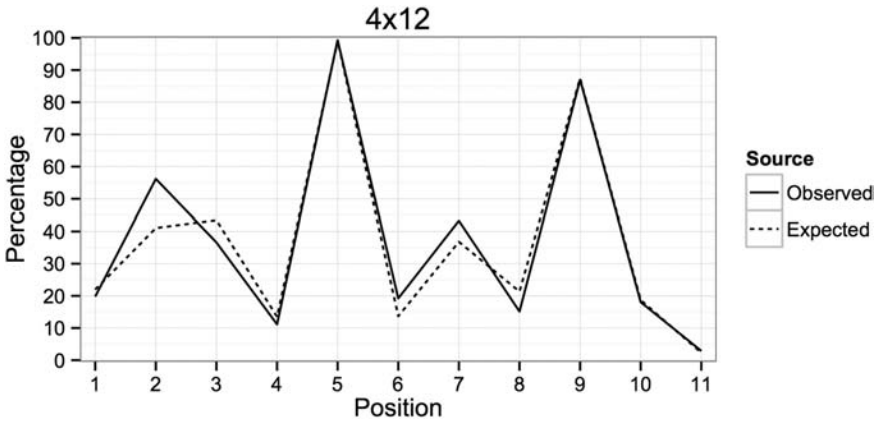

FIGURE 11 *Observed vs. expected boundary incidence in 4×12 meter*

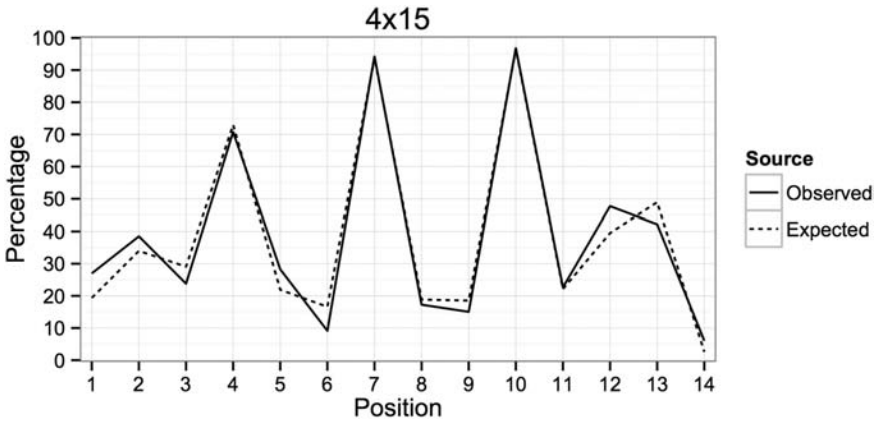

FIGURE 12 *Observed vs. expected boundary incidence in 4×15 meter*

stituents (ICs) from our prose corpus to construct pseudo-verse corpora with caesurae to match the actual verse corpora. In addition to matching caesura position and frequencies, we require the beginnings and endings of constructed verses to align with beginnings and endings of prose ICs, which mimics the poets' avoidance of enjambment. Drawing from prose ICs at random and respecting these constraints, we assembled a 100,000-verse corpus for each of the four meters.

A general pattern that emerges in Figures 11–14 is that there are fewer word boundaries than expected in colon-penultimate position. In the plots, these are the points where the solid line dips below the broken one. This holds

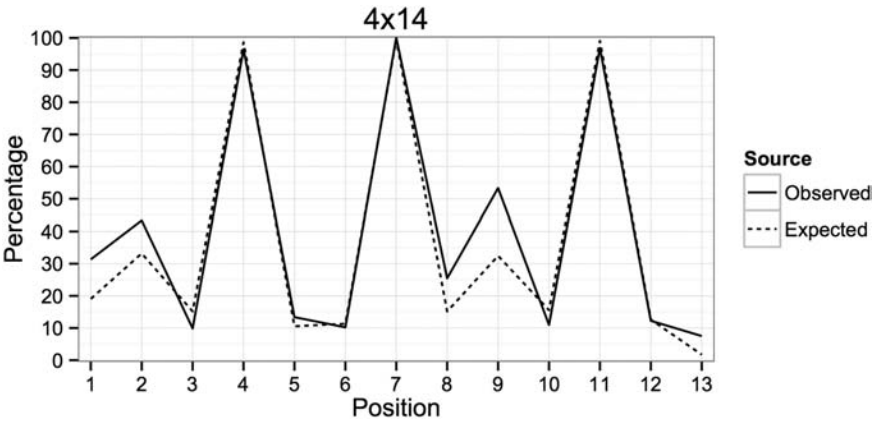

FIGURE 13 *Observed vs. expected boundary incidence in 4×14 meter*

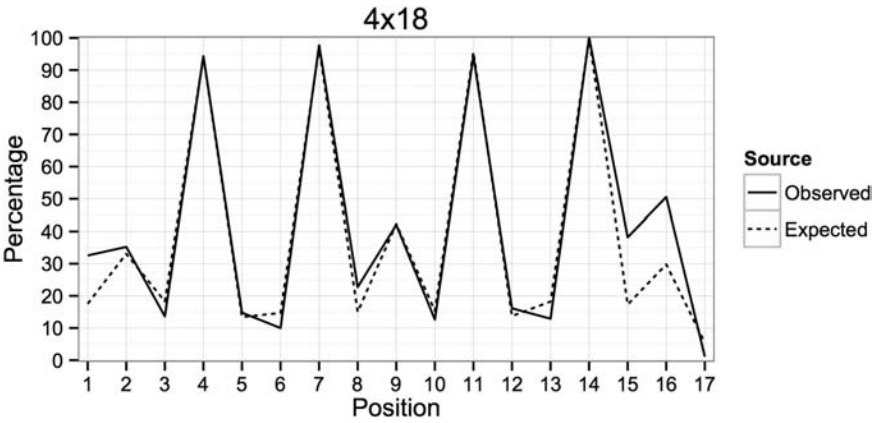

FIGURE 14 *Observed vs. expected boundary incidence in 4×18 meter*

for all colon-penultimate positions in all four meters, but not for all verse-penultimate positions.

To test for the significance of the difference between the observed and expected boundary incidence in these positions, we employ a χ^2 Goodness of Fit Test. Table 5 gives the observed vs. expected boundary incidence after position 6 in 4×15, one of two potential bridge positions where the discrepancy is significant ($p = .0008$), even at a Bonferroni-corrected criterion. This indicates a difference between the poets' treatment of the second and third colon and could be taken to support the traditional colometry [4|3||3|5].

TABLE 5 *Observed vs. expected boundary incidence after syllable 6 in 4×15 meter*

	Boundary	**No boundary**
Observed	21 (9.1%)	209
Expected	17,495 (17.5%)	82,505

However, neither the other significant bridge after 8 in 4×12 (p = .003) nor the borderline significant caesura after 3 in 4×14 (p = .015) is associated with a major caesura.[16] The motivation for the bridges remains unclear.

8 Implications for Textual Restoration

The relatively fragmentary state of the Tocharian corpus regularly requires editors to propose restorations. This can be done with a fair degree of accuracy, especially in cases where parallel texts in other languages exist, which often supply the approximate content of the lacunae. In addition to obvious restrictions such as the physical size of the lacuna in the manuscript, the meter constrains the number of possible restorations. The more exact understanding of the meters that we have arrived at requires that editors review restorations that have already been proposed and check whether they respect the compositional practice of the Tocharian poets. The same holds for new conjectures. We briefly exemplify a number of potentially problematic restorations below.

The restoration of verse 71a at THT 5 b1 proposed by Sieg and Siegling (1949: II, 10 fn. 6) violates the caesura after the 11th syllable in 4×14 and thus has only a 3.2% chance of being correct from a metrical standpoint, to judge from our corpus, where that caesura is violated in 3.2% of the verses.

71a mäkte meski | śeśś(anmoṣ || koklentse śänˈmānmasa) **:**

Just as the joints (of the chariot are connected with straps), …

16 The p-values for the other positions are as follows. 4×12: 4 (p = .10). 4×15: 3 (p = .07); 9 (p = .86). 4×14: 6 (p = .20); 10 (p = .21). 4×18: 3 (p = .45); 6 (p = .97); 10 (p = .51); 13 (p = .25).

Changing the order of the last two words to *śänmānmasa koklentse* would yield a metrically unproblematic restoration.

Restorations that violate the caesura after the 10th syllable in 4×15 likewise have only a 3.2 % chance of being correct according to our data.

PK AS 7J a2 (restoration proposed by Sieg 1938: 44–45)

19c takälñene'+ṣpä wlaiśke ‖ yetse (mäs'keträ cmelane **:**)

And his skin (is) soft to the touch (in the rebirths).

PK AS 6A b1–2 (restoration proposed by the CEToM editors)

12d (kär)ts(au)ñ(e)nta | kraupamar ‖ mā(ka spän'taitse) nervānne 12

I will accumulate virtues in gre(at number having faith) in the nirvāṇa.

Restorations that violate the caesura after 7 in 4×15 have a 5.8 % chance of being correct, e.g. the restoration proposed by Sieg and Siegling (1949: II, 50 fn. 8) for THT 30 b1–2

28d keṣe aiksnar | wä(ntoṣ rup'ne swāñcain'tsa ye)t(se) ysāṣṣe 28

… the golden (skin of his body, covered) completely with fathom-wide (rays).

The metrical abnormality of their restoration was discussed by Thomas (1983: 197), who suggests a metrically regular alternative:

28d keṣe aiksnar | wä(ntoṣo | swañcaintsa | cwi ye)t(se) ysāṣṣe 28

… (his) golden (skin, covered) completely with fathom-wide (rays).

A restoration that violates the minor caesura after the 9th syllable in 4×12 has a 13 % chance of being correct from a metrical standpoint.

PK AS 7B b6 (restoration proposed by Sieg 1938: 10)

8d mā kwipeññenträ ‖ mā onmi(ṃ ya'maske)n(trä) 8

They are not ashamed, they do not show remorse.

There are of course metrically aberrant verses in Tocharian poetry, which means that a metrically abnormal restoration is not impossible. The point of this section is that restorations should roughly follow the metrical practice of the poets as quantified and analyzed in this study.

9 Summary

Caesura violability provides evidence for the following colometries, with the caveat that in the relatively small 4×18 corpus, not every major caesura ("||") is significantly different from every minor caesura ("|"). An overview of this is provided in Table 6.

TABLE 6 *Colometries of four Tocharian B meters*

4×12	5||4|3
4×14	4|3||4|3
4×15	4|3||3||5
4×18	4|3||4|3||4

From the standpoint of violability, all caesurae in 4+3 sequences are minor, and all other caesurae are major. Systematic comparison with prose texts reveals avoidance of colon-final (but not verse-final) monosyllables. It is unclear whether these bridge-like phenomena are metrical in nature or otherwise motivated.

The violability of the internal caesurae in the 4+3 sequences ("7ers") in 4×12 and 4×15 also varies significantly across meters. This cannot be explained by a categorical distinction between major and minor metrical constituents alone. It appears to be a function of the individual 7ers or of the individual meters.

References

Adams, Douglas Q. 2013. More thoughts on Tocharian B prosody. *Tocharian and Indo-European Studies* 14:3–30.

Bross, Christoph, Dieter Gunkel, and Kevin M. Ryan. Forthcoming. The colometry of

Tocharian 4×15-syllable verse. In *Tocharian Texts in Context*, ed. Melanie Malzahn, Michaël Peyrot, Hannes Fellner and Theresa Illès. Bremen: Hempen.

CEToM = *A Comprehensive Edition of Tocharian Manuscripts*. Available online at http://www.univie.ac.at/tocharian/.

Clopper, C.J., and Egon S. Pearson. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26:404–413.

Fabb, Nigel, and Morris Halle. 2008. *Meter in poetry*. Cambridge: Cambridge University Press.

von Gabain, Annemarie, and Werner Winter. 1958. *Türkische Turfantexte IX. Ein Hymnus an den Vater Mani auf "Tocharisch" B mit alttürkischer Übersetzung*. Berlin: Akademie-Verlag.

Hayes, Bruce. 2010. Review of Fabb and Halle (2008). *Lingua* 120:2515–2521.

Hayes, Bruce, and Margaret MacEachern. 1998. Quatrain form in English folk verse. *Language* 64:473–507.

Kiparsky, Paul. 2006. A modular metrics for folk verse. In *Formal approaches to poetry*, ed. D.B. Elan and N. Friedberg, 7–49. Berlin: Mouton de Gruyter.

Malzahn, Melanie. 2012a. Now you see it, now you don't—Bewegliches *-o* in Tocharisch B. In *Linguistic developments along the Silk Road: Archaism and innovation in Tocharian*, ed. Olav Hackstein and Ronald Kim, 33–82. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Malzahn, Melanie. 2012b. Position matters: The placement of clitics in metrical texts of Tocharian B. *Tocharian and Indo-European Studies* 13:153–162.

Nespor, Marina, and Irene Vogel. 2007. *Prosodic Phonology: With a new foreword*. Berlin: de Gruyter.

Peyrot, Michaël. 2008. *Variation and change in Tocharian B*. Amsterdam/New York: Rodopi.

Pinault, Georges-Jean. 2008. *Chrestomathie tokharienne. Textes et grammaire*. Leuven/Paris: Peeters.

Prince, Alan. 1989. Metrical forms. In *Rhythm and Meter*, ed. Paul Kiparsky and Gilbert Youmans, 45–80. San Diego: Academic Press.

Ryan, Kevin. 2011. Gradient weight in phonology. Ph.D. diss., University of California, Los Angeles.

Selkirk, Elisabeth. 2011. The syntax-phonology interface. In *The Handbook of Phonological Theory*. 2nd edition, ed. John Goldsmith, Jason Riggle, and Alan C.L. Yu, 435–484. Malden, MA: Wiley-Blackwell.

Sieg, Emil. 1938. Die Kutschischen Karmavibhaṅga-Texte der Bibliothèque National in Paris. *Zeitschrift für vergleichende Sprachforschung* 65:1–54.

Sieg, Emil, and Wilhelm Siegling. 1921. *Tocharische Sprachreste. I. Band. Die Texte. A. Transcription*. Berlin/Leipzig: de Gruyter.

Sieg, Emil, and Wilhelm Siegling. 1949. *Tocharische Sprachreste. Sprache B. Heft 1. Die*

Udānālaṅkāra-Fragmente. Text, Übersetzung und Glossar. Göttingen: Vandenhoeck & 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.

Thomas, Werner. 1983. *Tocharische Sprachreste. Sprache B. Teil I: Die Texte. Band 1 Fragmente 1–116 der Berliner Sammlung. Herausgegeben von Emil Sieg † und Wilhelm Siegling †, neubearbeitet und mit einem Kommentar versehen von Werner Thomas*. Göttingen: Vandenhoeck & Ruprecht.

Watkins, Calvert. 1999. Questions of syntax and meter in Tocharian. In *Compositiones Indogermanicae in memoriam Jochem Schindler*, ed. Heiner Eichner and Hans Christian Luschützky, 601–614. Prague: Enigma.

Winter, Werner. 1959. Zur "tocharischen" Metrik. In *Akten des XXIV. Internationalen Orientalistenkongresses München 1957*, 520–521 (= Winter 2005: 26–27).

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: de Gruyter (= Winter 2005: 393–413).

Winter, Werner. 2005. *Kleine Schriften, in zwei Bänden. Festgabe aus Anlass des 80. Geburtstags. Ausgewählt und herausgegeben von Olav Hackstein*. Bremen: Hempen.