The colometry of Tocharian 4x15-syllable verse

Christoph Bross, Dieter Gunkel, and Kevin M. Ryan

Abstract

We identify the basic colometry of Tocharian 4x15-syllable verse as 4+3+3+5 (traditionally 7+8), but we find no support for the putative alternative colometries of 4x15 often cited in the literature (viz. 6+4+5 and 8+7). In rare cases in which the medial caesura is violated, a word boundary after position 6 or 8 is highly probable by chance alone, as we confirm through corpus statistics. If the colometry is indeed invariable, one major argument for the influence of Indic on Tocharian metrics is undermined. We further reinforce that the medial caesura is no stronger than the final one, despite their putative statuses as major vs. minor caesurae, respectively. The more complex picture of the metrical practice of the poets demands that we call into question certain proposed restorations.

1 Metrical preliminaries

Tocharian 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. Most of those consist of four isosyllabic verses, e.g. 4x12 syllables, 4x14, 4x15, and 4x18. 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, and one fragmentary poem apparently composed in two-verse stanzas, 2x14 (THT 133). 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, there is general agreement that the 4x14 verse, which is usually composed of 4+3+4+3 syllables — we’ll refer to that 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):

\[(\sigma\sigma\sigma)_{m}(\sigma\sigma)_{m} \) \( \land \)

VERSE

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 4x14 stanza from the Udānālaṅkāra (THT 5 a4–6).

68a wñā-neš (po)yši | karunta | ā | śolantse :
68b mā rā asāņmem | laitaṇe | čem sklokt ā | pālskome :n :
68c kos tne ūakta | pelaiķni | ( po ) šaisšents | ā | naiwacci :
68d tary | aṣā-ne | pudñākte | ā | teki ktsaitsne | srukalne 68

1 Special thanks go to Olav Hackstein and Peter-Arnold Mumm for their comments on this article.
“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, even though, o master, unpleasant laws hold here for all the world.’ The Buddha proclaimed three to him: sickness; old age; and death.”

There is less agreement concerning the colometry of 4x15, despite the fact that it is one of the best-attested verse types in Tocharian. In this contribution, we re-examine its colometry, focusing especially on whether it has built-in colometrical alternatives, and whether there are distinctions between major and minor caesurae in the meter.

2 The corpora

Our corpora of 4x15 are derived from A Comprehensive Edition of Tocharian Manuscripts (CEToM). Verses were required to be exactly 15 syllables long. There were no other selectional criteria. Data from fragmentary verses were taken only from non-fragmentary portions of the line. The corpus of 4x15 consists of 242 verses in Tocharian B, and 48 verses in Tocharian A, taken from the following texts.

<table>
<thead>
<tr>
<th>Tocharian B</th>
<th>Tocharian A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Verses</td>
</tr>
<tr>
<td>PK AS 6A</td>
<td>22</td>
</tr>
<tr>
<td>PK AS 6B</td>
<td>24</td>
</tr>
<tr>
<td>PK AS 6C</td>
<td>26</td>
</tr>
<tr>
<td>PK AS 7H</td>
<td>27</td>
</tr>
<tr>
<td>PK AS 7J</td>
<td>14</td>
</tr>
<tr>
<td>PK AS 7I</td>
<td>36</td>
</tr>
<tr>
<td>THT 28</td>
<td>18</td>
</tr>
<tr>
<td>THT 29</td>
<td>29</td>
</tr>
<tr>
<td>THT 30</td>
<td>39</td>
</tr>
<tr>
<td>THT 291.a</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 1: Verse corpora.

Studies such as Malzahn (2012b) exemplify how important it is to rule out potential confounds from syntax and other areas of the grammar when studying meter. Malzahn argues convincingly that the distribution of sentential enclitics in metrical texts, which Winter (1959) had attributed to a metrical factor, should be ascribed in the main to the syntax of second position clitics in Tocharian B: of the 205 sentential clitics in her verse corpus, 93% follow the first phonological word in their syntactic clause, just as they do in prose. Since poetic and

---

3 There are different types of metrical texts in the Tocharian languages: longer, purely metrical texts (Lehrgedichte, cf. Winter 1999: 74); short isolated poems; and dramatic texts that alternate between prose and short metrical passages (campi genre, cf. Pinault 2008: 407). Although Adams (2003: 9) has suggested otherwise, we assume that there are no significant metrical differences between these types; our findings in section 3 are consistent with this assumption. Likewise, we do not take the “names of tunes” (kene/kam) into account since they apparently don’t refer to the meter, but to some aspect of the performance (Winter 1955: 33). All Tocharian B texts in the corpus are classified by CEToM as Classical except for THT 291.a, which is Archaic.
prose texts in Tocharian B are roughly contemporary and compatible in genre, prose provides an excellent baseline for comparison.3

For the comparisons that we conduct in section 3.2, 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 constituents (ICs) in the prose corpus. We assume that the following clause- and phrase-level syntactic constituents were mapped to Intonational Phrases (cf. Nespor and Vogel 2007; Selkirk 2011); punctuation after these constituents was apparently optional, but provides some support for their reality.

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

Table 2: ICs in prose.

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

(tu lyelyakormeṇ) (vrkṣavāsike ṇakte śle māṃtsalyñe śanoṣ wēssāṃ)
this see:ABS tree.dwelling god with sorrow wife:ALL say:3SG.PRS
(lariya) (pālka nai mā-ṣekamñe wāntarwats sparkāłye āke)
dear see:SG.IPV PTC impermanence thing:GEN.PL disappearance end

“Having seen this, the tree-dwelling god says to his wife with sadness:
‘Darling, look at the impermanence of things and their ultimate disappearance!’”

Since the way that the poets realize caesurae suggests that they treated sequences of a lexical word followed by a monosyllabic enclitic as a single phonological constituent, we did so as well in all three corpora.5

---


5 We took the following clitics into account: (k)ka, kca, ksa, ṇike, tne, nai, no, nta, pi, ra/rā, ram(t), wa, wat, spā/sāp, ñai/ñey, (s)ste, tsa.
3 The colometry of 4x15

3.1 Word boundary distribution in Tocharian B

The distribution of word boundaries in our Tocharian B 4x15 corpus is given in Figure 1. The numbers along the x-axis of the plot represent verse-internal word-boundaries. The peaks in the plot clearly reflect caesurae after the 4th, 7th, and 10th syllables, and a basic colometry of 4+3+3+5. This is at least partly consistent with the standard analysis [7][8]. As noted by Stumpf (1971: 72 with fn. 10), the number of caesura violations decreases numerically towards verse end. There is a break after 4 in 70.7% of verses, after 7 in 94.2%, and after 10 in 96.8%. The fact that the error bars at 7 and 10 overlap suggests that that difference could be attributed to chance; we return to this in section 3.3.

Stanza 23 of THT 30 (a2–3) is representative of 4x15 meter in Tocharian B. The caesura after 4 is violated once in verse 23c, and the other caesurae are respected. Host-enclitic groups are joined with “-”. We discuss the atypical word boundary distribution in stanza 24 (a3–4) in the following section.

23a cets ce_u silñe | pälskontse | lakle-spä | wïkässi poysi :
23b tôṃ ṣlokanna | wertsyaine | ākṣa cets | palsko tsärwässiš •
23c emšketse āratišco | yātatisiš | astarñëš ūeko :
23d serke cmelñe | srukálñents | emšketse | nautalñe yāmtsi 23
24a snai keš cmela | karsatsiš'co-spä-tne | ytárye sā śpälmeṃ :
24b saišše kär(s)au | caisa a'päärtse | yāmusa klyomña :
24c gaṅkne kekmu | makte yai'ku nāki | šesa reššām war •
24d samudrämpra | taiknesa | ytárye sā | oktats | ākṣusa 24

“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, 

and for recognizing the births without number, this superb path 
was made manifest by the one who knows the world as the noble one. 
Just as water, having reached the Ganges, flows flawlessly together 
with the sea, so this eightfold path just proclaimed [leads to immortality].”

3.2 Supposed alternative colometries

The first three verses of stanza 24 exhibit three violations of the caesura after 7. On the basis of rare verses like these — the caesura after 7 is violated in only 13 verses (5.8%) in our corpus — scholars have supposed that poets composing in 4x15 could choose between two or more alternative colometries. Analyses of 4x15 often cite 8+7 and occasionally also 6+4+5 as alternatives to the 4+3+3+5 colometry. Given how rarely the poets violate the caesura after 7, it is prima facie unlikely that either 8+7 or 6+4+5 represented a real compositional alternative for the poets. It is, however, true, that where that caesura is violated in our corpus, there is a word boundary in either the preceding position (as in 24b and c) or the following one (as in 24c) in 12 of the 13 cases (92.3%). The question is how likely it is that a word boundary would occur after 6 or 8 by chance alone in those verses.

We can approximate that probability by assembling 15-syllable units from the intonational constituents (ICs) in our prose corpus and excluding cases where there is a word boundary after the 7th syllable. Since enjambment is rare in Tocharian B poetry, or to be more specific, since the poets usually align the beginnings and ends of verses with the beginnings and ends of the units that we identify as ICs in the prose corpus, we require the prose-based units to do the same. The boundary data plotted in Figure 2 are based on a corpus of 100,000 such units that are randomly assembled from prose ICs. The break incidence after 6 and 8 is 54.5% and 51.3%, and there is a boundary in either the preceding or the following position in 84.7% of the lines. This 84.7% is nowhere near significantly different from the 92.3% observed in poetry. A χ² goodness of fit test against the estimated 84.7% rate gives \( p = .45 \). In other words, a difference at least this great would have arisen by chance 45% of the time. We would consider \( p < .05 \) to be significant.

---


7 This is a Monte Carlo method for gauging probability (Metropolis and Ulam 1949, Robert and Casella 2004, Rubinstein and Kroese 2007).
In order to model the behavior of poets composing 4x15 verses with violations of the caesura after 7 even more closely, we can additionally require the 15-syllable prose-based units to match the word boundary incidence of the verse corpus at the other caesurae, i.e. after the 4th (70.7%) and 10th (96.8%) syllables. These data, plotted in Figure 3, show a word break incidence at 6 and 8 of 57.9% and 51.3%. There is a boundary in either the preceding or the following position in 88.4% of the lines, which again is nowhere near significantly different from the 92.3% observed in poetry: a $\chi^2$ goodness of fit test against the estimated 88.4% rate gives $p = .66$.

In sum, considered in isolation, stanza 24 might seem to be composed in a meter with colometrical alternatives, especially to a scholar familiar with Rigvedic trimeter verse or the
dactylic hexameter. However, upon consideration of a larger corpus of 4x15, there is no support that the meter encoded such alternatives. Consequently, 8+7 and 6+4+5 should no longer be cited as colometrical alternatives for 4x15 in Tocharian B.

### 3.3 Major vs. minor cola

We now return to the question whether the boundary data in Figure 1, which reflect a 4+3|3|5 colometry, are consistent with the standard analysis [7||8], which would imply [4|3||3|5]. As mentioned above, there is a general consensus that there is a distinction between major and minor cola in Tocharian meter. Von Gabain and Winter (1958: 33-34) and Winter (1959) were the first to propose the distinction. According to them, the minor caesurae are more violable than major ones:

Wir dürften damit berechtigt sein, nebem festen Hauptzäsuren auch Nebenzäsuren anzunehmen, d. h. fakultativ aufhebbare Grenzen zwischen Unterabschnitten innerhalb der Kolen. (von Gabain and Winter 1958: 34)

Accepting Malzahn’s (2012b) analysis of clitic distribution in verse, caesura violability is to our knowledge the only remaining diagnostic proposed for the distinction between major and minor cola. Assuming a [4|3||3|5] colometry for 4x15, this predicts that the poets violate the putative minor caesurae after 4 and 10 significantly more frequently than the putative major caesura after 7 (cf. Table 3). From the standpoint of violability, the caesura after 10 is clearly not a minor caesura — it is the most strictly respected — but the caesura after 4 seems to be, to judge from the fact that their error bars in Figure 1 do not overlap.

<table>
<thead>
<tr>
<th>Caesura after 4</th>
<th>Violated</th>
<th>Not violated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caesura after 7</td>
<td>13</td>
<td>212</td>
</tr>
<tr>
<td>Caesura after 10</td>
<td>7</td>
<td>211</td>
</tr>
</tbody>
</table>

Table 3: Violations of caesurae after 4, 7, and 10 in Tocharian B 4x15.

We can test the statistical significance of the difference between the caesura after 4 and that after 7 with Fisher’s Exact Test of Independence. Those two caesurae are violated a total of 80 times in our corpus, and 67 of those violations occur after 4. 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 .0000001, meaning that a difference at least this great would have arisen by chance less than .00001% of the time; the difference is highly significant. The difference between the violability of the caesurae after 7 and 10, however, is not significant. A discrepancy at least as great would have arisen by chance 25% of the time ($p = .25$). In sum, taking caesura violability to be a diagnostic of the major vs. minor distinction, the colometry of 4x15 is [4|3||3|5].

### 3.4 Word boundary distribution in Tocharian A

The distribution of word boundaries in our Tocharian A 4x15 corpus is given in Figure 4. The relatively long error bars are due to the smaller and more fragmentary corpus. Nevertheless,

---

8 15-syllable verses in 20/22/10/15 have the basic colometry 4+4+4+3. This may have contributed to the assumption of an 8+7 alternative for 4x15. However, isosyllabic verses can have different colometries in different meters, e.g. the different 12-syllable verses in the two different 4x12 meters (5+4+3 and 4+4+4) and the different 13- and 21-syllable verses in 13/13/13/13/21 and 21/21/18/13 (Stumpf 1971: 71-72).
the same basic 4+3+3+5 colometry holds in A. The numerical differences between the three caesurae in the two languages can all be attributed to chance ($p = .39, .37, \text{ and } 1$). As in B, the caesura after 4 is violated more frequently than that after 7 and 10, but the differences are not significant ($p = .025 \text{ and } .1$) in the smaller A corpus. Given that there’s a word boundary after 7 100% of the time, there is no support for alternative colometries in Tocharian A either. This is consistent with but does not necessitate Peyrot’s recent proposal that “TA has elaborated the TB metrical tradition, but TB is the source” (Peyrot 2013: 6).

![Figure 4: Percentage of verse-internal word boundaries in Tocharian A 4x15.](image)

4 Implications

4.1 Textual restorations

Our findings have obvious implications for the edition of the texts, especially regarding textual restorations. These must respect the metrical practice of the poets, which was far more constrained than the standard analyses of 4x15 suggest. For example, the restorations bolded below are not impossible, but they are highly unlikely from a metrical standpoint, given that the Tocharian B poets violate the caesura after 7 in 4x15 in only 5.8% of the verses in our corpus, and the caesura after 10 in only 3.2% of them.

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

\[12^c a(\text{saṃ})khyai k(a)l\l(\text{p(an)masa} \| lallyau ň(ā)š | rš(ā)kāñ (\text{em})twecc(ek :)
\]

\[12^d (kā)rts(su)n(ē)n(ta | kraupamar \| mā(ka spān'taitse) nervānne 12
\]

PK AS 7J a2–3 (restoration proposed by Sieg 1938: 44-45)

\[19^a šak pārkāvān'\text{ta wässi} \| aïšënce | kālpāsśām wnlome :
\]

\[19^b takārskān\text{ñe} | eršënce \| māsketā | kārtse lkātsine :
\]

\[19^c takālhënë \| spā nlāške \| yetse (māš'ketā cmelane :)
\]
Possible influence of Sanskrit meter on Tocharian meter

The absence of evidence for colometrical alternatives also bears on the ongoing discussion of the degree of influence that Indic meter may have had on Tocharian meter. From the advent of Tocharian studies until quite recently, the *opinio communis* held that Tocharian meter “est totalement étranger à celui de la métrique du sanskrit” (Pinault 2000: 153). Recently, Widmer (2006) called this into question — quite plausibly — on the general grounds that the majority of Tocharian metrical texts are translations and adaptations of Sanskrit originals and “profondément ancré dans l’état d’esprit du monde indien” (p. 523). Part of Widmer’s specific evidence for Sanskrit influence is based on the following four putative correspondences between Sanskrit and Tocharian meters, which he suggests should not be attributed to chance (pp. 525-526).

<table>
<thead>
<tr>
<th>Stanza structure</th>
<th>Tocharian verse</th>
<th>Sanskrit verse</th>
<th>Sanskrit meter name</th>
</tr>
</thead>
<tbody>
<tr>
<td>4x25</td>
<td>5+5+8+7</td>
<td>5+5+8+7</td>
<td>krauṇicapadā</td>
</tr>
<tr>
<td>4x15</td>
<td>7+8</td>
<td>7+8</td>
<td>candrāvartā</td>
</tr>
<tr>
<td>4x15</td>
<td>†8+7</td>
<td>8+7</td>
<td>maṇiguṇanikarā</td>
</tr>
<tr>
<td>4x15</td>
<td>†8+7</td>
<td>8+7</td>
<td>mālinī</td>
</tr>
</tbody>
</table>

It is important to note that the last two correspondences assume the unsupported 8+7 colometrical alternative for Tocharian 4x15 and may thus be discarded. Furthermore, the fact that syllable weight is regulated in Sanskrit verse but not in Tocharian (which apparently had no binary distinction between heavy and light syllables) is not, as Widmer suggests (p. 526), the only formal difference between the two metrical systems. In the Tocharian system, word boundary distribution is far more strictly regulated, and the number of cola per verse is higher. Thus, 7+8 should be represented as 4+3+3+5 or [4|3||3||5], and 5+5+8+7 as 5+4+4+4+3 (cf. Stumpf 1971: 71). This does not invalidate the general thrust of Widmer’s contribution; the topic deserves further investigation.

### 5 Summary

It emerges from a quantitative corpus-based study of Tocharian 4x15-syllable verse that the basic colometry of the verse is 4+3+3+5. There is no evidence for the alternative colometries 8+7 and 6+4+5 cited in the literature. In Tocharian B, the poets violate the caesura after the 4th syllable significantly more frequently than the caesurae after the 7th and 10th syllables; the numerical difference between the last two can be attributed to chance. Assuming that caesura violability indicates a difference between major (M) and minor (m) cola, the colometry of 4x15 is [4|3][3][5] or

\[
((\sigma\sigma\sigma)_{m}(\sigma\sigma)_{m})(\sigma\sigma)_{M}(\sigma\sigma\sigma\sigma)_{M}\text{VERSE}.
\]

---

Works cited


