1 Introduction

This is a proposal to encode Tocharian and Khotanese scripts in a unified Turkestani script in the Universal Character Set (ISO/IEC 10646). This document outlines the unified system for encoding both writing systems, a tentative code chart and names list, character data, and some specimens. The font used to display the glyphs in this document were designed by the author of the proposal, based on manuscripts available at the International Dunhuang Project websites.

2 Background

The Turkestani script family comprises four Brahmi-derived scripts: Tocharian, Khotanese, Tumshuqese, and so-called Uighur Brahmi, scripts that were all used along the Tarim Basin in the Taklamakan Desert in what is now Xinjiang in western China. As Tumshuqese and Uighur Brahmi are poorly attested, this proposal in its current form seeks only to unify the Tocharian and Khotanese scripts.

Tocharian script was historically used to write the Tocharian languages (ISO 639-3: xto, txb), traditionally referred to as Tocharian A and Tocharian B, which belong to the Tocharian branch of the Indo-European language family. The script was in use primarily during the 8th century CE by the people inhabiting the northern edge of the Tarim Basin.

Tocharian script is attested in over 4,000 extant manuscripts that were discovered in the early 20th century in the Tarim Basin, primarily in Kucha, Karasahr, and Turfan.

Khotanese script was used exclusively to write the Khotanese language (ISO 639-2 kho), one of the two Saka languages alongside Tumshuqese. Khotanese was a Middle-Iranian language spoken from approximately 200 BCE to 1000 CE by people inhabiting the southern rim of the Tarim Basin.

Khotanese script is attested in over 2,300 extant manuscripts found in Dunhuang, among other manuscripts in various other languages. It was spoken in the Kingdom of Khotan, modern-day Hotan.

Though the two languages are only distantly related, their scripts are both closely related both in appearance and functionality, and they were both in use in a similar time and place.
Pros and cons of a unified Turkestani script

Before discussing the issue of unification, it must be reiterated that while the proposal theoretically includes the four scripts listed above (Tocharian, Khotanese, Tumshuqese, and Uighur Brahmi), the current proposal excludes the latter two scripts due to a lack of sufficient original sources to put forth a comprehensive description of either of them. Therefore, the proposal of a unified Turkestani script is essentially only a proposal to unify Tocharian and Khotanese.

The proposal presented herein describes a unified Turkestani script, but it is not necessarily clear if this is the best solution. Below, the pros and cons of unification are discussed.

3.1 Pros

While many would argue that Tocharian and Khotanese are important linguistically due to being the most easterly members of the Indo-European language family, as well as Tocharian representing a heretofore entirely unknown branch of that family, there can also be little doubt that the languages are nevertheless relatively obscure, particularly Khotanese. In addition, Tocharian and Khotanese scripts are closely related, sharing several similarities in glyph shape and function. A script capable of rendering the more complex Tocharian could, with the addition of only three characters and a small amount of adjustment, render Khotanese as well.
With limited public interest and similarly functioning scripts, unification is a valid option.

3.2 Cons

While a comprehensive, character-by-character comparison of Tocharian and Khotanese clearly reveals a close lineage, even a cursory glance at Tocharian and Khotanese manuscripts reveals that the two scripts in practical usage are visually quite different and distinct.

Several letters are significantly different in form. Compare: Tocharian $\text{I}$, $\text{O}$, $\text{KA}$, $\text{CHA}$, $\text{NYA}$, $\text{NNA}$, $\text{THA}$, $\text{BHA}$, $\text{MA}$, $\text{YA}$ vs. Khotanese $\text{I}$, $\text{O}$, $\text{KA}$, $\text{CHA}$, $\text{NYA}$, $\text{NNA}$, $\text{THA}$, $\text{BHA}$, $\text{MA}$, $\text{YA}$.

Scholars well-versed in one script cannot realistically be expected to be able to read the other script without additional training in a language in which they have not specialized. As Tocharian is the more well-known and better-attested language, it is the more likely candidate
for the standard presentation form of Turkestani. Those interested in reading digital Khotanese texts would therefore be forced into two inadequate options: read Khotanese texts in Tocharian script, or obtain a possibly costly Khotanese font in order to display the script as it was originally written. Thus, unification would fail to provide a plain-text solution for recording Khotanese texts digitally.

3.3 Final note on unification

As Tocharian is the more technically complex and the better known of the two scripts, the proposal uses Tocharian glyph shapes by default to demonstrate how the script functions. Nevertheless, cases in which Khotanese differs significantly from Tocharian are illustrated in the native Khotanese script.

4 Structure

4.1 Introduction

Turkestani scripts have typically been referred to as modified forms of Brahmi, indicating that people have traditionally considered these scripts simply to be forms of Brahmi. Although their structure and functionality is indeed clearly within the Brahmic tradition, the Turkestani scripts are nevertheless significantly different in a number of ways from the Aṣokan Brahmi currently encoded both in terms of glyph shape and orthographic conventions.

As is typical with Brahmic scripts, each letter indicates a consonant followed by the inherent vowel $\check{a}$ by default. However, unlike scripts such as Devanagari, there is no visual element that is removed when a letter is used in a conjunct. The vowel is silenced either by a subscript conjunct or the virāma.

The most obviously different aspect of Turkestani is the use in Tocharian of eleven consonant signs, traditionally referred to as Fremdzeichen, which serve the dual function of representing a consonant plus the vowel $\ddot{a}$, and to stand in place of the consonant plus virama (the deciding factor of use being the age of the manuscript; later manuscripts do not use Fremdzeichen alone to indicate consonant + virama).

Turkestani also employs unique compounding and virama usage which will be explained below.

4.2 Representative glyphs

The fonts used in this document were created by the author and are based on the documents preserved in the International Dunhuang Project.

4.3 Character Names

The characters are named in accordance with the UCS convention for Brahmi-based scripts, with the exception of the vowel AE and EI. The rationale for the spelling AE is that the
Freimvokal is traditionally transcribed ä, and ae is the typical replacement for ä in 7-bit ASCII contexts. The rational for EI is that it is the spelling traditionally used by Khotanese scholars when transcribing that vowel.

4.4 Directionality

The script is written from left to right.

4.5 Vowels

There are 14 independent vowel signs:

- Vowel Letter A
- Vowel Letter UU
- Vowel Letter O
- Vowel Letter AA
- Vowel Letter VOCALIC R
- Vowel Letter AU
- Vowel Letter I
- Vowel Letter VOCALIC RR
- Vowel Letter AE
- Vowel Letter II
- Vowel Letter E
- Vowel Letter EI
- Vowel Letter U
- Vowel Letter VOCALIC L
- Vowel Letter VOCALIC LL

The vowels VOCALIC L and VOCALIC LL are not attested in any Turkestani texts, but spaces have been left available in the code block in case of future discovery.

Khotanese allows for diphthongs to be represented by adding vowel diacritics to independent vowel signs:

- ui'
- uai
- uvi

These are likely best represented as character combinations rather than individual characters.

4.6 Vowel Signs

There are 13 dependent vowel signs:

- Vowel Sign AA
- Vowel Sign VOCALIC R
- Vowel Sign AU
- Vowel Sign I
- Vowel Sign VOCALIC RR
- Vowel Sign AE
- Vowel Sign II
- Vowel Sign E
- Vowel Sign EI
VOWEL SIGN U  VOWEL SIGN AI

VOWEL SIGN UU  VOWEL SIGN O

VOWEL SIGN AE is used in both Tocharian and Khotanese. It indicates the vowel /ɨ/ in Tocharian (Krause and Slocum, 2014) and /ə/ in Khotanese (Emmerick and Pulleybank 1993: 45-46). The transcription <ä> is standard.

VOWEL SIGN EI is used only in Khotanese. It indicates the diphthong /aə/ (Emmerick 1998). The transcription <ei> is standard (see Figure 9 c, d).

4.7 Consonants

There are 44 consonant letters:

KA  TTA  PA  SSA  WAE
KHA  TTHA  PHA  SA  SHAE
GA  DDA  BA  HA  SSAE
GHA  DDHA  BHA  BA  SAE
NGA  NNA  MA  TAE
CA  TA  YA  NAE
CHA  THA  RA  PAE
JA  DA  LA  MAE
JHA  DHA  VA  RAE
NYA  NA  SHA  LAE

All letters bear the inherent vowel a. This vowel may be silenced with ₳ VIRAMA or through the use of conjuncts, to be explained below.

4.8 Various signs

There are 5 various signs:
4.9 Numbers

There are 20 numbers:

- ONE  SIX  TWENTY  SEVENTY
- TWO  SEVEN  THIRTY  EIGHTY
- THREE  EIGHT  FORTY  NINETY
- FOUR  NINE  FIFTY  ONE HUNDRED
- FIVE  TEN  SIXTY  ONE THOUSAND

Numbers for various multiples of one hundred also exist, but they are transparent combinations of the digit for one hundred and the digits for multiples of one. It is proposed that the one hundred digit takes virama combined with other digits to form those that are missing.

In numbers 11, 21, etc., the number ONE always stacks vertically, appearing above of the previous number, e.g. 91. As this only occurs with ONE, all other numbers being formed horizontally, e.g. 92, it is best handled at the font level.

4.10 Vowel signs (matras)

Each vowel letter has a corresponding vowel sign. Vowel signs can be found above, below, or to the right of the consonant letter. Vowel signs that appear below the letter often initiate changes in the vowel sign, the consonant letter, or both. The vowel sign ā also takes on several contextual forms, and the consonant letter /l/ takes on irregular forms.

4.10.1 Contextual forms of vowel signs

AA  The vowel sign AA has various contextual forms, outlined below:

1  When combined with open-topped consonants and certain others:
A variation of this occurs in Khotanese, which spans two separate letters:

\[ \text{tāndī} \] (TA, vowel sign AA, NA, DA, vowel sign 1)

This is not mandatory, however, and should be considered a stylistic variant best handled at the font level.

2 A smaller variant occurs with certain round-topped letters:

\[ \text{kā} \] (KHA, vowel sign AA)
\[ \text{gā} \] (GA, vowel sign AA)
\[ \text{dā} \] (DHA, vowel sign AA)
\[ \text{sā} \] (SHA, vowel sign AA)

This does not occur in Khotanese:

\[ \text{kā} \] (KHA, vowel sign AA)
\[ \text{gā} \] (GA, vowel sign AA)
\[ \text{dā} \] (DHA, vowel sign AA)
\[ \text{sā} \] (SHA, vowel sign AA)

3. A tall superscript form also appears with certain letters:

\[ \text{ṅā} \] (NGA, vowel sign AA)
\[ \text{jā} \] (JA, vowel sign AA)
The vowel signs ̃ U and ̈ UU have three contextual variations, outlined below:

1. They both take a distinct form on letters that already have descenders that resemble ̃ U. This form also appears on ̈ DA:

   - ku (KA, vowel sign ̃ U)
   - jhu (JHA, vowel sign ̃ U)
   - du (DDA, vowel sign ̃ U)
   - du (DA, vowel sign ̃ U)
   - ru (RA, vowel sign ̃ U)

   - kū (KA, vowel sign ̈ UU)
   - jhū (JHA, vowel sign ̈ UU)
   - dū (DDA, vowel sign ̈ UU)
   - dū (DA, vowel sign ̈ UU)
   - rū (RA, vowel sign ̈ UU)

2. The second form is similar to the first, and only occurs with subscript ̈ RA:

   - pra (PA, ̈ RA)
   - pru (PA, ̈ RA, vowel sign ̃ U)
   - prū (PA, ̈ RA, vowel sign ̈ UU)

3. Forms superficially resembling the independent vowels ɔ O and ɔ AU appear in combination with certain letters:

   - tu (TA, vowel sign ̃ U)
   - bhu (BHA, vowel sign ̃ U)
   - gu (GA, vowel sign ̃ U)
   - šu (SHA, vowel sign ̃ U)

   - tū (TA, vowel sign ̈ UU)
   - bhū (BHA, vowel sign ̈ UU)
   - gū (GA, vowel sign ̈ UU)
Khotanese script employs an alternate form used for GA and SHA that does not occur in Tocharian. Compare:

\[
\begin{align*}
\text{tu} & \quad \text{TA, vowel sign } \uparrow \\
\text{bhu} & \quad \text{BHA, vowel sign } \uparrow \\
\text{gu} & \quad \text{GA, vowel sign } \uparrow \\
\text{šu} & \quad \text{SHA, vowel sign } \uparrow \\
\text{tū} & \quad \text{TA, vowel sign } \uparrow \uparrow \\
\text{bhū} & \quad \text{BHA, vowel sign } \uparrow \uparrow \\
\text{gū} & \quad \text{GA, vowel sign } \uparrow \uparrow \\
\text{šū} & \quad \text{SHA, vowel sign } \uparrow \uparrow \\
\end{align*}
\]

In Khotanese, \( \uparrow \uparrow \text{U} \) and \( \uparrow \uparrow \text{UU} \) each both have a fourth form that attaches only to subscript \( \text{YA} \):

\[
\begin{align*}
\text{pyu} & \quad \text{PA, YA, vowel sign } \uparrow \uparrow \text{UU} \\
\text{pyū} & \quad \text{PA, YA, vowel sign } \uparrow \uparrow \text{UU} \\
\end{align*}
\]

**Vocalic R and RR** Similar to the vowels \( \text{U} \) and \( \text{UU} \), when these signs attach to a consonant with a descender, it is deleted:

\[
\begin{align*}
\text{kr} & \quad \text{KA, vowel sign } \downarrow \text{VOCALIC R} \\
\text{jhr} & \quad \text{JHA, vowel sign } \downarrow \text{VOCALIC R} \\
\text{dr} & \quad \text{DDA, vowel sign } \downarrow \text{VOCALIC R} \\
\text{kṛ} & \quad \text{KA, vowel sign } \downarrow \text{VOCALIC RR} \\
\text{jhr} & \quad \text{JHA, vowel sign } \downarrow \text{VOCALIC RR} \\
\text{dṛ} & \quad \text{DDA, vowel sign } \downarrow \text{VOCALIC RR} \\
\end{align*}
\]

Note that these do not occur with \( \text{RA} \).

They also cause minor variation in the forms of \( \text{NA} \) and \( \text{BHA} \):
nr (NA, vowel sign VOCALIC R)
bhr (BHA, vowel sign VOCALIC R)
nī (NA, vowel sign VOCALIC RR)
bhī (BHA, vowel sign VOCALIC RR)

LA The consonant letter LA induces a number of irregular vowel sign forms:

li (LA, vowel sign I)
lī (LA, vowel sign II)
le (LA, vowel sign E)
lai (LA, vowel sign AI)
lo (LA, vowel sign O)
lau (LA, vowel sign AU)

I/II/E/AI/O/AU On open topped letters, these vowel signs appear one ascender to the left of the right ascender. In the case of AI, the two elements of the vowel sign appear on different ascenders. Examples:

ghi (GHA, vowel sign I)
pī (PA, vowel sign II)
hī (HA, vowel sign II)
se (SA, vowel sign E)
mai (MA, vowel sign AI)
šo (SSA, vowel sign O)
pau (PHA, vowel sign AU)

Khotanese functions the same for vowel sign I, but the others are different. Vowel sign II spans the open section, except with consonant HA, for which it appears on the right ascender; vowel sign E appears on the right side; vowel sign AI is not separated as in Tocharian; both elements appear on the right ascender:

ghi (GHA, vowel sign I)
pī (PA, vowel sign II)
4.10.2 More than one vowel sign per aksara

Turkestani occasionally allow more than one vowel sign on a single consonant letter or conjunct. This is used for diphthongs and is found only in Khotanese, as Tocharian makes use of subscript vowel letters instead.

\[\text{kuï} \quad (\text{KHA, vowel sign } U, \text{ II})\]

\[\text{ysmuï} \quad (\text{YA, SA, MA, vowel sign } U, \text{ II})\]

4.11 Conjuncts

Turkestani employ subscripts to indicate consonant clusters. Most subscripts are relatively transparent and easily identifiable. There are nevertheless some subscripts that differ to a greater or lesser degree from their base forms.

Turkestani conjuncts typically comprise between 2 and 4 consonant letters, though there is theoretically no limit:

\[\text{jña} \quad (\text{JA, NYA})\]

\[\text{ksûsa} \quad (\text{KA, SSA, TA, SA})\]

3.11.1 Variation in subscript glyph shapes

\[\text{YA and RA form subscripts that are entirely dissimilar to their base forms, while VA is also slightly different:}\]

\[\text{pya} \quad (\text{PA, YA})\]

\[\text{pra} \quad (\text{PA, RA})\]

\[\text{pva} \quad (\text{PA, VA})\]

In Khotanese, \text{VA} takes on a significantly different form when it combines with certain
other letters:

\[ \text{vva} \quad (\text{VA}, \text{VA}) \]

Several other letters gain a supporting bar in subscript form by which they attach to the base letter:

\[ \text{dga} \quad (\text{DA}, \text{GA}) \]
\[ \text{sth} \quad (\text{SA}, \text{TTHA}) \]
\[ \text{ddha} \quad (\text{DA}, \text{DHA}) \]
\[ \text{ntha} \quad (\text{NA}, \text{THA}) \]
\[ \text{wśa} \quad (\text{WAE}, \text{SHA}) \]

All subscripts with a head-like element lose it in subscript form:

\[ \text{lła} \quad (\text{LA}, \text{LA}) \]
\[ \text{csa} \quad (\text{CA}, \text{SA}) \]
\[ \text{kma} \quad (\text{KA}, \text{MA}) \]
\[ \text{lta} \quad (\text{LA}, \text{TA}) \]

In Khotanese, certain subscripts may be additionally reduced in form when they themselves take subscripts, though only with specific letters:

\[ \text{ysma} \quad (\text{YA}, \text{SA}, \text{MA}) \]

Here, \( \text{SA} \) is reduced in form when it combines with \( \text{MA} \), but cf.

\[ \text{ysda} \quad (\text{YA}, \text{SA}, \text{DA}) \]

where \( \text{SA} \) remains in full form.

The position of subscripts in relation to the base consonants to which they attach is entirely dependent on the specific characters involved. Every base and subscript form has an invariable connection point used in the formation of conjuncts. As a result, some subscripts appear directly below the base, while others appear partially or almost fully to the right:

\[ \text{sth} \quad (\text{SA}, \text{TTHA}) \]
\[ \text{kma} \quad (\text{KA}, \text{MA}) \]
\[ \text{sy} \quad (\text{SA}, \text{YA}) \]
This invariable positioning has consequences for subscript  YA, as it typically extends somewhat above the base of the glyph to which it is attached. When the location of the connection point of the base glyph makes this impossible, the height of subscript  YA is truncated. Compare its length and height in the following conjuncts:

\[
\text{csa} \quad (\text{CA, SA})
\]

The conjunct  kka employs an abbreviated form of the subscript:

\[
\text{kka} \quad (\text{KA, KA})
\]

Khotanese in particular employs several ligatures that each represent a single phoneme and that act as single units. Of these,  jsa,  tta, and  rra, would likely best be represented with the akhand feature.

\[
\text{jsa} \quad (\text{JA, SA})
\quad \text{tta} \quad (\text{TA, TA})
\quad \text{rra} \quad (\text{RA, RA})
\]

These conjuncts occur frequently and remain distinct even in subscript form:

\[
\text{stta} \quad (\text{SA, TA, TA})
\quad \text{krpa} \quad (\text{KA, RA, RA})
\]

The frequently occurring conjunct  ysa, while at first a seemingly good candidate to be included as an akhand, is in fact not suitable, as the conjunct for base consonant +  YA  takes precedence over the conjunct for  YA +  SA conjunct:

\[
\text{lysa} \quad (\text{LA, YA, SA})
\]

The basic shape of  ysa has clearly not been preserved in this conjunct, thus invalidating it to be used as an akhand.

Tocharian has one letter,  tsa, which appears frequently and could also be classified as an
akhand. It is noteworthy that \( \text{T} \) has a combining form resembling \( \text{N} \). This likely arises from \( \text{ts} \) appearing frequently, and \( \text{N} \) having the simpler combining form.

### 4.11.2 Variation in base glyph shapes

Conjuncts can also initiate changes in the form of the base consonant. This is most noticeable in the base conjunct forms of consonant letters with descenders. Just as they lose their descenders when combining with subscript vowel signs, so do they lose them in consonant conjuncts, e.g.:

\[
\begin{align*}
\text{kla} & \quad (\text{KA}, \text{LA}) \\
\text{qva} & \quad (\text{DDA}, \text{VA})
\end{align*}
\]

This also occurs with the letter \( \text{R} \), but with an important difference: namely, that it acts as a typical repha. The form of \( \text{RA} \) appears above the writing line and attaches to the full base form of a letter:

\[
\begin{align*}
\text{rha} & \quad (\text{RA}, \text{HA}) \\
\text{r} & \quad (\text{RA}, \text{NNA})
\end{align*}
\]

All vowel signs aside from those that attach to the bottom of consonants must attach to the repha:

\[
\begin{align*}
\text{rgo} & \quad (\text{RA}, \text{GA}, \text{vowel sign } \text{O}) \\
\text{rn} & \quad (\text{RA}, \text{NA}, \text{vowel sign } \text{A})
\end{align*}
\]

The letter \( \text{LA} \) has an irregular form when it combines with repha \( \text{RA} \):

\[
\begin{align*}
\text{rla} & \quad (\text{RA}, \text{LA})
\end{align*}
\]

Repha does not occur with \( \text{YA} \); instead, a regular conjunct is formed:

\[
\begin{align*}
\text{rya} & \quad (\text{RA}, \text{YA})
\end{align*}
\]

The letters \( \text{T} \) and \( \text{NA} \) have unique alterations in shape. The alteration in the base form can differentiate the two letters:

\[
\begin{align*}
\text{twa} & \quad (\text{TA}, \text{WAE})
\end{align*}
\]
This is not always a reliable guide, however, as TA occasionally resembles the combining form of NA:

\( \text{tsa} \) (TA, SA)

### 4.11.3 Aksara conjuncts

Unusually, two consonant signs, each bearing its own a vowel sign, can be combined into a single conjunct. This is most commonly found with the sequence \( \text{ku} \), which represents the Tocharian consonant /kʷ/, but it also occurs for metrical rather than phonological reasons (Hitch 2012: 282) (see Figure 8 j, k). Examples:

- \( \text{ku} \) (KA, vowel sign U)
- \( \text{ce} \) (CA, vowel sign E)
- \( \text{küce} \)
- \( \text{wi} \) (WAE, vowel sign I)
- \( \text{nā} \) (NA, vowel sign AA)
- \( \text{wīnā} \)

As can be seen, the vowel sign from the subscripted aksara is moved to a more convenient location.

### 4.12 Virama

There is 1 virama:

\( \text{Virama} \)

Both Turkestani scripts employ a virama that functions exactly as viramas in other Indic scripts. However, Tocharian employs a second, far more commonly-occurring form of virama that appears visually as a horizontal or diagonal bar that precedes the marked letter or
conjunct and connects it with the preceding, vowel-bearing letter or conjunct. The distinction between the two virama is mostly context-based, as, typically, Fremdzeichen take the bar virama while standard letters take the standard virama, though there are some exceptions (see Figure 8).

Khotanese does not employ the bar virama of Tocharian.

\[\text{šal} \quad (\text{SHA, LAE, VIRAMA})\]

This also occurs with final consonant clusters that include a Fremdzeichen:

\[\text{lkānt} \quad (\text{LA, VIRAMA, KA NA, VIRAMA, TAE, VIRAMA})\]

It is important to note that the bar virama can attach to any portion of the previous aksara, including the base consonant, subscript, or vowel sign.

Occasionally, a consonant may bear a redundant standard virama in addition to the bar virama:

\[\text{mor} \quad (\text{MA, vowel sign O, RAE, VIRAMA, VIRAMA})\]

This is, however, optional; the redundant standard virama is typically absent.

The bar virama should be treated as an alternate form of the standard virama. It appears before the letter it modifies, but this is common in Brahmic scripts (cf. vowel signs in Devanagari, Thai, etc.). Bar virama is the standard form used with Fremdzeichen, while the standard form appears on regular consonants. The exception to this are the letters CA and NYA, which lack Fremdzeichen variants.

\[\text{teñ} \quad (\text{TA, NYA, VIRAMA})\]

Occasionally, the vowel sign AE will appear on a letter carrying a bar virama:

\[\text{picā} \quad (\text{PA, vowel sign I, CA, VIRAMA, vowel sign AE})\]

This is largely restricted to CA and NYA, but does occur on some other letters as well:

\[\text{koyā} \quad (\text{KA, vowel sign O, YA, VIRAMA, vowel sign AE})\]
The proposed implementation is:

- **Fremdzeichen** with virama: virama is realized as bar virama
- **Fremdzeichen** with two viramas: first virama is realized as bar virama, second as standard virama
- standard consonant with virama: virama is realized as standard virama
- standard consonant with virama and vowel sign \( \text{AE} \): virama is realized as bar virama.

### 4.13 Subscript Independent Vowel Letters

In distinct contrast to most Brahmic scripts (but with precedent in e.g. Khmer), Tocharian indicates some diphthongs through the use of subscript independent vowel signs, which are also necessarily marked with virama (see Figure 8 e, f, g, h).

If the base letter has a subscript, the virama is straight and attaches to the subscript. If it does not, the virama angles up to attach to the base consonant. Examples:

- klye-\( \text{u} \) (KA, LA, YA, vowel sign \( \text{E} \) \( \text{U} \))
- lo-\( \text{i} \) (LA, vowel sign \( \text{O} \), \( \text{I} \))

Notice that the subscript \( \text{I} \) takes a different form.

### 4.14 Nasalization

The languages do not have nasalization *per se*, but the script nevertheless employ *anusvāra* both for nasal consonants and for transcription of Sanskrit nasalization. It appears immediately above the base consonant letter.

- gham (GHA, ANUSVARA)
- \( \text{nikmām} \) (NGA, KA, MA, vowel sign \( \text{AA} \))

### 4.15 Aspiration

Turkestani employs three signs for aspiration: the *visarga* sign, which appears to the right of the base consonant sign, and the *jihvāmūliya* and *upadhmānīya*, which respectively indicate velar and labial allophones of \( h \). These differ from visarga in that they act as letters and form conjuncts with the preceding consonant letter.

- khā (KHA, VISARGA)
- \( \text{bka} \) (JIHVAMULIYA, KA)
In Turkestani texts, VISARGA is quite common, but JIHVAMULIYA and UPADHMANIYA are exceedingly rare.

4.16 Punctuation

There are four punctuation marks:

- DANDA
- DOUBLE DANDA
- PUNCTUATION DOT
- PUNCTUATION DOUBLE DOT

5 Summary of differences between Tocharian and Khotanese

5.1 Character inventory

- Tocharian employs eleven characters not needed for Khotanese: \texttt{KAE}, \texttt{TAE}, \texttt{NAE}, \texttt{PAE}, \texttt{MAE}, \texttt{RAE}, \texttt{LAE}, \texttt{WAE}, \texttt{SHAE}, \texttt{SSAE}, \texttt{SAE}, and three characters not currently attested in Khotanese: \texttt{VOWEL LETTER VOCALIC RR}, \texttt{JIHVAMULIYA}, \texttt{UPADHMANIYA}.

- Khotanese employs three characters not needed for Tocharian: \texttt{VOWEL LETTER EI}, \texttt{VOWEL SIGN EI}, \texttt{HOOK}. In addition, Khotanese allows vowel signs and consonant subscripts to be added to independent vowel letters: \texttt{ui’}, \texttt{uai}, \texttt{uvi}.

5.2 Glyph shape

The majority of glyphs are at least slightly different in appearance, while some are different enough to hinder legibility significantly. See Table 1 below for a detailed comparison.

5.3 Variations in vowel signs

- Tocharian employs a variant of vowel sign \texttt{AA} not found in Khotanese. cf. Tocharian \texttt{khā}, Khotanese \texttt{khā}.
Khotanese employs a stylistic variant of sign $\ddot{\alpha}$ not found in Tocharian. cf. Khotanese $\ddot{\alpha} $ or $\ddot{\alpha} $ tändi, Tocharian $\ddot{\alpha} $ tändi.

Khotanese draws a distinction in the forms of $\ddot{u}$ and vowel sign $\ddot{u}$ $U$ not found in Tocharian:

<table>
<thead>
<tr>
<th></th>
<th>Tocharian</th>
<th>Khotanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>$tu$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$bhu$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$gu$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\ddot{s}u$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\ddot{t}u$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$bh\ddot{u}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\ddot{g}u$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\ddot{\ddot{s}}\ddot{u}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

With subscript $\ddot{\dot{\dot{y}}}A$, Khotanese employs variants of $\ddot{u}$ and vowel sign $\ddot{u}$ $U$ not found in Tocharian:

<table>
<thead>
<tr>
<th></th>
<th>Tocharian</th>
<th>Khotanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>$pyu$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$py\ddot{u}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The variant form of $\ddot{\ddot{u}}$ $U$ for $KA$ $\ddot{\ddot{JHA}}$ $\ddot{\ddot{DA}}$ and $\ddot{\ddot{DA}}$ is separated into two parts in Tocharian, but is combined into a single diacritic in Khotanese (Khotanese $\ddot{\ddot{RA}}$ functions the same as Tocharian $\ddot{\ddot{RA}}$):
Varied placement of vowel signs on the consonant letter:

<table>
<thead>
<tr>
<th>Tocharian</th>
<th>Khotanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>pī</td>
<td>p?</td>
</tr>
<tr>
<td>hī</td>
<td>hë</td>
</tr>
<tr>
<td>se</td>
<td>s~</td>
</tr>
<tr>
<td>mai</td>
<td>m}</td>
</tr>
<tr>
<td>so</td>
<td>s</td>
</tr>
<tr>
<td>pau</td>
<td>p</td>
</tr>
</tbody>
</table>

Khotanese forms some diphthongs by placing multiple vowel signs on a single consonant. This does not occur in Tocharian:

\[
\begin{array}{c}
\text{Tocharian} \quad \text{Khotanese} \\
p\text{kū} \quad (\text{KHA, vowel sign U, II}) \\
\text{ysmu} \quad (\text{YA, SA, MA, vowel sign U, II})
\end{array}
\]

5.4 Differences in conjuncts

Like the base letters, conjuncts often have a difference in appearance that can impede legibility. Examples:

<table>
<thead>
<tr>
<th>Tocharian</th>
<th>Khotanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>kya</td>
<td>k</td>
</tr>
<tr>
<td>vva</td>
<td>v</td>
</tr>
<tr>
<td>lya</td>
<td>l</td>
</tr>
<tr>
<td>ysmma</td>
<td>ysmma</td>
</tr>
</tbody>
</table>
• Khotanese and Tocharian have different candidates for akhand ligatures:

<table>
<thead>
<tr>
<th></th>
<th>Tocharian</th>
<th>Khotanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>jsa</td>
<td>—</td>
<td>[ä]</td>
</tr>
<tr>
<td>tta</td>
<td>—</td>
<td>[ä]</td>
</tr>
<tr>
<td>rra</td>
<td>—</td>
<td>[ö]</td>
</tr>
<tr>
<td>tsa</td>
<td>[ä]</td>
<td>[ö]</td>
</tr>
</tbody>
</table>

• Tocharian allows aksara conjuncts, Khotanese does not

<table>
<thead>
<tr>
<th></th>
<th>1st aksara</th>
<th>2nd aksara</th>
<th>combined aksara</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[ä]</td>
<td>[ä]</td>
<td>[ä]</td>
</tr>
<tr>
<td>ku</td>
<td>[ä]</td>
<td>[ä]</td>
<td>[ä]</td>
</tr>
<tr>
<td>ce</td>
<td>[ä]</td>
<td>[ä]</td>
<td>[ä]</td>
</tr>
<tr>
<td>külce</td>
<td>[ä]</td>
<td>[ä]</td>
<td>[ä]</td>
</tr>
<tr>
<td>wi</td>
<td>[ä]</td>
<td>[ä]</td>
<td>[ä]</td>
</tr>
<tr>
<td>nã</td>
<td>[ä]</td>
<td>[ä]</td>
<td>[ä]</td>
</tr>
<tr>
<td>wĩnã</td>
<td>[ä]</td>
<td>[ä]</td>
<td>[ä]</td>
</tr>
</tbody>
</table>

5.5 Virama

• Tocharian uses a pre-character, bar-shaped virama variant that is highly context-dependent, while Khotanese does not:

<table>
<thead>
<tr>
<th></th>
<th>Tocharian</th>
<th>Khotanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>šas</td>
<td>[ä]</td>
<td>[ä]</td>
</tr>
<tr>
<td>mor</td>
<td>[ä]</td>
<td>[ä]</td>
</tr>
</tbody>
</table>

• Tocharian allows VIRAMA and vowel sign AE to appear on the same letter, while Khotanese does not:
5.6 Subscript vowel letters

- Tocharian employs subscript vowel letters, while Khotanese does not:

Tocharian

\textit{klye-u} \hspace{1cm} \textit{lo-i}

6 Character properties

Turkestani character properties are as follows:

\begin{verbatim}
11F50;TURKESTANI LETTER A;Lo;0;L;;;;;N;;;;;
11F51;TURKESTANI LETTER AA;Lo;0;L;;;;;N;;;;;
11F52;TURKESTANI LETTER I;Lo;0;L;;;;;N;;;;;
11F53;TURKESTANI LETTER II;Lo;0;L;;;;;N;;;;;
11F54;TURKESTANI LETTER U;Lo;0;L;;;;;N;;;;;
11F55;TURKESTANI LETTER UU;Lo;0;L;;;;;N;;;;;
11F56;TURKESTANI LETTER VOCALIC R;Lo;0;L;;;;;N;;;;;
11F57;TURKESTANI LETTER VOCALIC RR;Lo;0;L;;;;;N;;;;;
11F58;<RESERVED>
11F59;<RESERVED>
11F5A;TURKESTANI LETTER E;Lo;0;L;;;;;N;;;;;
11F5B;TURKESTANI LETTER EI;Lo;0;L;;;;;N;;;;;
11F5C;TURKESTANI LETTER O;Lo;0;L;;;;;N;;;;;
11F5D;TURKESTANI LETTER AU;Lo;0;L;;;;;N;;;;;
11F5E;TURKESTANI LETTER AE;Lo;0;L;;;;;N;;;;;
11F5F;TURKESTANI LETTER GHA;Lo;0;L;;;;;N;;;;;
11F60;TURKESTANI LETTER GA;Lo;0;L;;;;;N;;;;;
11F61;TURKESTANI LETTER KA;Lo;0;L;;;;;N;;;;;
11F62;TURKESTANI LETTER GHA;Lo;0;L;;;;;N;;;;;
11F63;TURKESTANI LETTER JHA;Lo;0;L;;;;;N;;;;;
11F64;TURKESTANI LETTER NGA;Lo;0;L;;;;;N;;;;;
11F65;TURKESTANI LETTER CA;Lo;0;L;;;;;N;;;;;
11F66;TURKESTANI LETTER CHA;Lo;0;L;;;;;N;;;;;
11F67;TURKESTANI LETTER JA;Lo;0;L;;;;;N;;;;;
11F68;TURKESTANI LETTER JHA;Lo;0;L;;;;;N;;;;;
11F69;TURKESTANI LETTER NYA;Lo;0;L;;;;;N;;;;;
11F6A;TURKESTANI LETTER TTA;Lo;0;L;;;;;N;;;;;
11F6B;TURKESTANI LETTER TTHA;Lo;0;L;;;;;N;;;;;
11F6C;TURKESTANI LETTER DDA;Lo;0;L;;;;;N;;;;;
\end{verbatim}
Preliminary Proposal to Encode the Turkestani Script

Lee Wilson

11F6D; TURKESTANI LETTER DDHA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F6E; TURKESTANI LETTER NNA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F6F; TURKESTANI LETTER TA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F70; TURKESTANI LETTER THA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F71; TURKESTANI LETTER DA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F72; TURKESTANI LETTER DHA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F73; TURKESTANI LETTER NA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F74; TURKESTANI LETTER PA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F75; TURKESTANI LETTER PHA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F76; TURKESTANI LETTER BA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F77; TURKESTANI LETTER BHA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F78; TURKESTANI LETTER MA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F79; TURKESTANI LETTER YA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F7A; TURKESTANI LETTER RA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F7B; TURKESTANI LETTER LA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F7C; TURKESTANI LETTER VA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F7D; TURKESTANI LETTER SHA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F7E; TURKESTANI LETTER SSA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F7F; TURKESTANI LETTER SA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F80; TURKESTANI LETTER HA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F81; TURKESTANI LETTER KAE; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F82; TURKESTANI LETTER TAE; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F83; TURKESTANI LETTER NAE; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F84; TURKESTANI LETTER PAE; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F85; TURKESTANI LETTER MAE; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F86; TURKESTANI LETTER RAE; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F87; TURKESTANI LETTER LAE; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F88; TURKESTANI LETTER WAE; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F89; TURKESTANI LETTER SHAE; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F8A; TURKESTANI LETTER SSAE; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F8B; TURKESTANI LETTER SAE; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F8C; TURKESTANI VOWEL SIGN AA; Mc; 0; L; ; ; ; ; N; ; ; ; ;
11F8D; TURKESTANI VOWEL SIGN I; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F8E; TURKESTANI VOWEL SIGN II; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F8F; TURKESTANI VOWEL SIGN U; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F90; TURKESTANI VOWEL SIGN UU; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F91; TURKESTANI VOWEL SIGN VOCALIC R; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F92; TURKESTANI VOWEL SIGN VOCALIC RR; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F93; <RESERVED>
11F94 <RESERVED>
11F95; TURKESTANI VOWEL SIGN E; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F96; TURKESTANI VOWEL SIGN AI; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F97; TURKESTANI VOWEL SIGN O; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F98; TURKESTANI VOWEL SIGN AU; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F99; TURKESTANI VOWEL SIGN AE; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F9A; TURKESTANI VOWEL SIGN EI; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F9B; TURKESTANI SIGN ANUSVARA; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F9C; TURKESTANI SIGN VISARGA; Mc; 0; L; ; ; ; ; N; ; ; ; ;
11F9D; TURKESTANI SIGN HOOK; Mn; 0; NSM; ; ; ; ; N; ; ; ; ;
11F9E; TURKESTANI SIGN JIHVAMULIYA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
11F9F; TURKESTANI SIGN UPADHMANIYA; Lo; 0; L; ; ; ; ; N; ; ; ; ;
Preliminary Proposal to Encode the Turkestani Script

Lee Wilson

11FA0; TURKESTANI VIRAMA; Mn; 9; L; ; ; ; N; ; ; ; ;
11FA1; TURKESTANI NUMBER ONE; No; 0; L; ; ; ; 1; N; ; ; ; ;
11FA2; TURKESTANI NUMBER TWO; No; 0; L; ; ; ; 2; N; ; ; ; ;
11FA3; TURKESTANI NUMBER THREE; No; 0; L; ; ; ; 3; N; ; ; ; ;
11FA4; TURKESTANI NUMBER FOUR; No; 0; L; ; ; ; 4; N; ; ; ; ;
11FA5; TURKESTANI NUMBER FIVE; No; 0; L; ; ; ; 5; N; ; ; ; ;
11FA6; TURKESTANI NUMBER SIX; No; 0; L; ; ; ; 6; N; ; ; ; ;
11FA7; TURKESTANI NUMBER SEVEN; No; 0; L; ; ; ; 7; N; ; ; ; ;
11FA8; TURKESTANI NUMBER EIGHT; No; 0; L; ; ; ; 8; N; ; ; ; ;
11FA9; TURKESTANI NUMBER NINE; No; 0; L; ; ; ; 9; N; ; ; ; ;
11FAB; TURKESTANI NUMBER TEN; No; 0; L; ; ; ; 10; N; ; ; ; ;
11FAC; TURKESTANI NUMBER TWENTY; No; 0; L; ; ; ; 20; N; ; ; ; ;
11FAD; TURKESTANI NUMBER THIRTY; No; 0; L; ; ; ; 30; N; ; ; ; ;
11FAE; TURKESTANI NUMBER FORTY; No; 0; L; ; ; ; 40; N; ; ; ; ;
11FAF; TURKESTANI NUMBER FIFTY; No; 0; L; ; ; ; 50; N; ; ; ; ;
11FB0; TURKESTANI NUMBER SIXTY; No; 0; L; ; ; ; 60; N; ; ; ; ;
11FB1; TURKESTANI NUMBER SEVENTY; No; 0; L; ; ; ; 70; N; ; ; ; ;
11FB2; TURKESTANI NUMBER EIGHTY; No; 0; L; ; ; ; 80; N; ; ; ; ;
11FB3; TURKESTANI NUMBER NINETY; No; 0; L; ; ; ; 90; N; ; ; ; ;
11FB4; TURKESTANI NUMBER ONE HUNDRED; No; 0; L; ; ; ; 100; N; ; ; ; ;
11FB5; TURKESTANI NUMBER ONE THOUSAND; No; 0; L; ; ; ; 1000; N; ; ; ; ;
11FB6; TURKESTANI DANDA; Po; 0; L; ; ; ; N; ; ; ; ;
11FB7; TURKESTANI DOUBLE DANDA; Po; 0; L; ; ; ; N; ; ; ; ;
11FB8; TURKESTANI PUNCTUATION DOT; Po; 0; L; ; ; ; N; ; ; ; ;
7 Code charts

It is important to note that this proposal in its current form requires one column to be taken from the proposed Satavahana block.

<table>
<thead>
<tr>
<th></th>
<th>11F5</th>
<th>11F6</th>
<th>11F7</th>
<th>11F8</th>
<th>11F9</th>
<th>11FA</th>
<th>11FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>1</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>2</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>3</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>4</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>5</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>6</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>7</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>8</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>9</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>A</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>B</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>C</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>D</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>E</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
<tr>
<td>F</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
<td>❖❖❖</td>
</tr>
</tbody>
</table>
**Independent vowels**

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+1F50</td>
<td>TURKESTANI LETTER A</td>
</tr>
<tr>
<td>U+1F51</td>
<td>TURKESTANI LETTER AA</td>
</tr>
<tr>
<td>U+1F52</td>
<td>TURKESTANI LETTER I</td>
</tr>
<tr>
<td>U+1F53</td>
<td>TURKESTANI LETTER II</td>
</tr>
<tr>
<td>U+1F54</td>
<td>TURKESTANI LETTER U</td>
</tr>
<tr>
<td>U+1F55</td>
<td>TURKESTANI LETTER UU</td>
</tr>
<tr>
<td>U+1F56</td>
<td>TURKESTANI LETTER VOCALIC R</td>
</tr>
<tr>
<td>U+1F57</td>
<td>TURKESTANI LETTER VOCALIC RR</td>
</tr>
<tr>
<td>U+1F58</td>
<td>&lt;RESERVED&gt;</td>
</tr>
<tr>
<td>U+1F59</td>
<td>&lt;RESERVED&gt;</td>
</tr>
<tr>
<td>U+1F5A</td>
<td>TURKESTANI LETTER E</td>
</tr>
<tr>
<td>U+1F5B</td>
<td>TURKESTANI LETTER AI</td>
</tr>
<tr>
<td>U+1F5C</td>
<td>TURKESTANI LETTER O</td>
</tr>
<tr>
<td>U+1F5D</td>
<td>TURKESTANI LETTER AU</td>
</tr>
<tr>
<td>U+1F5E</td>
<td>TURKESTANI LETTER AE</td>
</tr>
<tr>
<td>U+1F5F</td>
<td>TURKESTANI LETTER EI</td>
</tr>
</tbody>
</table>

**Consonants**

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+1F60</td>
<td>TURKESTANI LETTER KA</td>
</tr>
<tr>
<td>U+1F61</td>
<td>TURKESTANI LETTER KHA</td>
</tr>
<tr>
<td>U+1F62</td>
<td>TURKESTANI LETTER GA</td>
</tr>
<tr>
<td>U+1F63</td>
<td>TURKESTANI LETTER GHA</td>
</tr>
<tr>
<td>U+1F64</td>
<td>TURKESTANI LETTER NGA</td>
</tr>
<tr>
<td>U+1F65</td>
<td>TURKESTANI LETTER CA</td>
</tr>
<tr>
<td>U+1F66</td>
<td>TURKESTANI LETTER CHA</td>
</tr>
<tr>
<td>U+1F67</td>
<td>TURKESTANI LETTER JA</td>
</tr>
<tr>
<td>U+1F68</td>
<td>TURKESTANI LETTER JHA</td>
</tr>
<tr>
<td>U+1F69</td>
<td>TURKESTANI LETTER NYA</td>
</tr>
<tr>
<td>U+1F6A</td>
<td>TURKESTANI LETTER TTA</td>
</tr>
<tr>
<td>U+1F6B</td>
<td>TURKESTANI LETTER TTHA</td>
</tr>
<tr>
<td>U+1F6C</td>
<td>TURKESTANI LETTER DDA</td>
</tr>
<tr>
<td>U+1F6D</td>
<td>TURKESTANI LETTER DDHA</td>
</tr>
<tr>
<td>U+1F6E</td>
<td>TURKESTANI LETTER NNA</td>
</tr>
<tr>
<td>U+1F6F</td>
<td>TURKESTANI LETTER TA</td>
</tr>
<tr>
<td>U+1F70</td>
<td>TURKESTANI LETTER THA</td>
</tr>
<tr>
<td>U+1F71</td>
<td>TURKESTANI LETTER DA</td>
</tr>
<tr>
<td>U+1F72</td>
<td>TURKESTANI LETTER DHA</td>
</tr>
<tr>
<td>U+1F73</td>
<td>TURKESTANI LETTER NA</td>
</tr>
<tr>
<td>U+1F74</td>
<td>TURKESTANI LETTER PA</td>
</tr>
<tr>
<td>U+1F75</td>
<td>TURKESTANI LETTER PHA</td>
</tr>
<tr>
<td>U+1F76</td>
<td>TURKESTANI LETTER BA</td>
</tr>
<tr>
<td>U+1F77</td>
<td>TURKESTANI LETTER BHA</td>
</tr>
<tr>
<td>U+1F78</td>
<td>TURKESTANI LETTER MA</td>
</tr>
<tr>
<td>U+1F79</td>
<td>TURKESTANI LETTER YA</td>
</tr>
<tr>
<td>U+1F7A</td>
<td>TURKESTANI LETTER RA</td>
</tr>
<tr>
<td>U+1F7B</td>
<td>TURKESTANI LETTER LA</td>
</tr>
<tr>
<td>U+1F7C</td>
<td>TURKESTANI LETTER VA</td>
</tr>
<tr>
<td>U+1F7D</td>
<td>TURKESTANI LETTER SHA</td>
</tr>
<tr>
<td>U+1F7E</td>
<td>TURKESTANI LETTER SSA</td>
</tr>
<tr>
<td>U+1F7F</td>
<td>TURKESTANI LETTER SA</td>
</tr>
</tbody>
</table>

**Dependent vowel signs**

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+1F80</td>
<td>TURKESTANI LETTER HA</td>
</tr>
<tr>
<td>U+1F81</td>
<td>TURKESTANI LETTER KAE</td>
</tr>
<tr>
<td>U+1F82</td>
<td>TURKESTANI LETTER TAE</td>
</tr>
<tr>
<td>U+1F83</td>
<td>TURKESTANI LETTER TAE</td>
</tr>
<tr>
<td>U+1F84</td>
<td>TURKESTANI LETTER PAE</td>
</tr>
<tr>
<td>U+1F85</td>
<td>TURKESTANI LETTER MAE</td>
</tr>
<tr>
<td>U+1F86</td>
<td>TURKESTANI LETTER RAE</td>
</tr>
<tr>
<td>U+1F87</td>
<td>TURKESTANI LETTER LAE</td>
</tr>
<tr>
<td>U+1F88</td>
<td>TURKESTANI LETTER WAE</td>
</tr>
<tr>
<td>U+1F89</td>
<td>TURKESTANI LETTER SHAE</td>
</tr>
<tr>
<td>U+1F8A</td>
<td>TURKESTANI LETTER SSAE</td>
</tr>
<tr>
<td>U+1F8B</td>
<td>TURKESTANI LETTER SAE</td>
</tr>
</tbody>
</table>

**Various signs**

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+1F9A</td>
<td>TURKESTANI SIGN AA</td>
</tr>
<tr>
<td>U+1F9B</td>
<td>TURKESTANI SIGN I</td>
</tr>
<tr>
<td>U+1F9C</td>
<td>TURKESTANI SIGN II</td>
</tr>
<tr>
<td>U+1F9D</td>
<td>TURKESTANI SIGN U</td>
</tr>
<tr>
<td>U+1F9E</td>
<td>TURKESTANI SIGN UU</td>
</tr>
<tr>
<td>U+1F9F</td>
<td>TURKESTANI SIGN VOCALIC R</td>
</tr>
<tr>
<td>U+1F9G</td>
<td>TURKESTANI SIGN VOCALIC RR</td>
</tr>
<tr>
<td>U+1F9H</td>
<td>&lt;RESERVED&gt;</td>
</tr>
<tr>
<td>U+1F9I</td>
<td>&lt;RESERVED&gt;</td>
</tr>
<tr>
<td>U+1F9J</td>
<td>TURKESTANI SIGN E</td>
</tr>
<tr>
<td>U+1F9K</td>
<td>TURKESTANI SIGN AI</td>
</tr>
<tr>
<td>U+1F9L</td>
<td>TURKESTANI SIGN O</td>
</tr>
<tr>
<td>U+1F9M</td>
<td>TURKESTANI SIGN AU</td>
</tr>
<tr>
<td>U+1F9N</td>
<td>TURKESTANI SIGN AE</td>
</tr>
<tr>
<td>U+1F9O</td>
<td>TURKESTANI SIGN EI</td>
</tr>
</tbody>
</table>

**Virama**

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+1FA0</td>
<td>TURKESTANI VIRAMA</td>
</tr>
</tbody>
</table>

**Numbers**

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+1FA1</td>
<td>TURKESTANI NUMBER ONE</td>
</tr>
<tr>
<td>U+1FA2</td>
<td>TURKESTANI NUMBER TWO</td>
</tr>
<tr>
<td>U+1FA3</td>
<td>TURKESTANI NUMBER THREE</td>
</tr>
<tr>
<td>U+1FA4</td>
<td>TURKESTANI NUMBER FOUR</td>
</tr>
<tr>
<td>U+1FA5</td>
<td>TURKESTANI NUMBER FIVE</td>
</tr>
<tr>
<td>U+1FA6</td>
<td>TURKESTANI NUMBER SIX</td>
</tr>
<tr>
<td>U+1FA7</td>
<td>TURKESTANI NUMBER SEVEN</td>
</tr>
<tr>
<td>U+1FA8</td>
<td>TURKESTANI NUMBER EIGHT</td>
</tr>
<tr>
<td>U+1FA9</td>
<td>TURKESTANI NUMBER NINE</td>
</tr>
<tr>
<td>U+1FAA</td>
<td>TURKESTANI NUMBER TEN</td>
</tr>
<tr>
<td>U+1FAB</td>
<td>TURKESTANI NUMBER TWENTY</td>
</tr>
</tbody>
</table>
8 Samples

<table>
<thead>
<tr>
<th>BRAHMI</th>
<th>TURKESTANI</th>
<th>BRAHMI</th>
<th>TURKESTANI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>H</td>
<td>DDA</td>
<td>RAE</td>
</tr>
<tr>
<td>AA</td>
<td>H</td>
<td>DDHA</td>
<td>LAE</td>
</tr>
<tr>
<td>I</td>
<td>NNA</td>
<td>TA</td>
<td>WAE</td>
</tr>
<tr>
<td>II</td>
<td>TA</td>
<td>THA</td>
<td>SHAE</td>
</tr>
<tr>
<td>U</td>
<td>L</td>
<td>DA</td>
<td>SSAE</td>
</tr>
<tr>
<td>UU</td>
<td>L</td>
<td>DHA</td>
<td>SAE</td>
</tr>
<tr>
<td>VOC.R</td>
<td>X</td>
<td>NA</td>
<td>SIGNAA</td>
</tr>
<tr>
<td>VOC.RR</td>
<td>X</td>
<td>PA</td>
<td>SIGNI</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>PA</td>
<td>SIGNII</td>
</tr>
<tr>
<td>AI</td>
<td>E</td>
<td>PHA</td>
<td>SIGNU</td>
</tr>
<tr>
<td>O</td>
<td>O</td>
<td>BA</td>
<td>SIGNUU</td>
</tr>
<tr>
<td>AU</td>
<td>A</td>
<td>BHA</td>
<td>SIGNVOC.R</td>
</tr>
</tbody>
</table>
### Table 1: Comparison of basic letters and signs of Aṣokan Brahmi and Tocharian Turkestani.

<table>
<thead>
<tr>
<th>Aṣokan Brahmi</th>
<th>Tocharian Turkestani</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE</td>
<td>MA</td>
</tr>
<tr>
<td>EI</td>
<td>YA</td>
</tr>
<tr>
<td>KA</td>
<td>RA</td>
</tr>
<tr>
<td>KHA</td>
<td>LA</td>
</tr>
<tr>
<td>GA</td>
<td>VA</td>
</tr>
<tr>
<td>GHA</td>
<td>SHA</td>
</tr>
<tr>
<td>NGA</td>
<td>SSA</td>
</tr>
<tr>
<td>CA</td>
<td>SA</td>
</tr>
<tr>
<td>CHA</td>
<td>HA</td>
</tr>
<tr>
<td>JA</td>
<td>KAE</td>
</tr>
<tr>
<td>JHA</td>
<td>TAE</td>
</tr>
<tr>
<td>NYA</td>
<td>NAE</td>
</tr>
<tr>
<td>TTA</td>
<td>PAE</td>
</tr>
<tr>
<td>TIHA</td>
<td>MAE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Signs</th>
<th>Aṣokan Brahmi</th>
<th>Tocharian Turkestani</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA</td>
<td>Ꝑ</td>
<td>Ꝑ</td>
</tr>
<tr>
<td>YA</td>
<td>ꝑ</td>
<td>ꝑ</td>
</tr>
<tr>
<td>RA</td>
<td>ꝑ</td>
<td>Ꝓ</td>
</tr>
<tr>
<td>LA</td>
<td>ꝑ</td>
<td>Ꝓ</td>
</tr>
<tr>
<td>VA</td>
<td>ꝑ</td>
<td>Ꝓ</td>
</tr>
<tr>
<td>SHA</td>
<td>Ꝑ</td>
<td>Ꝑ</td>
</tr>
<tr>
<td>SSA</td>
<td>ꝑ</td>
<td>Ꝓ</td>
</tr>
<tr>
<td>SA</td>
<td>ꝑ</td>
<td>Ꝓ</td>
</tr>
<tr>
<td>HA</td>
<td>Ꝑ</td>
<td>Ꝑ</td>
</tr>
<tr>
<td>KAE</td>
<td>Ꝑ</td>
<td>Ꝑ</td>
</tr>
<tr>
<td>TAE</td>
<td>Ꝑ</td>
<td>Ꝑ</td>
</tr>
<tr>
<td>NAE</td>
<td>Ꝑ</td>
<td>Ꝑ</td>
</tr>
<tr>
<td>PAE</td>
<td>Ꝑ</td>
<td>Ꝑ</td>
</tr>
<tr>
<td>MAE</td>
<td>Ꝑ</td>
<td>Ꝑ</td>
</tr>
</tbody>
</table>
Figure 6: A table of the basic letters of Tocharian Turkestani (from Krause and Thomas 1960:41, Malzahn 2007b:227-8).
Figure 7: A table of the basic letters, signs, and digits of Khotanese Turkestani as well as a selection of conjuncts (from Leumann, 1934: 39).
Figure 8: examples of bar virama, subscript independent vowel letters, and stacked aksaras. 
  a. ceṃts,  b. ṅkūl,  c. ttoṣ,  d. cār, e. ssoī, f. loī, g. ksāū, h. cē, i kiuce, j. maīcu, k. winā.

Figure 9: Examples of Khotanese-specific signs and aksaras with double vowel signs: 
  a. e’,  b. vo’,  c. rei,  d. ysei,  e. kuī,  f. ysmuī
Figure 8: Original Tocharian manuscript displaying a list of velar and palatal conjuncts.

Figure 9: Original Tocharian manuscript displaying a list of palatal and retroflex conjuncts.

Figure 10: Original Tocharian manuscript displaying a list of dental and bilabial conjuncts.

Figure 11: Original Tocharian manuscript displaying a list of bilabial, liquid, and fricative conjuncts.
Figure 12: Original Tocharian manuscript displaying a list of bilabial, liquid, and fricative conjuncts.

Figure 13: Original Tocharian manuscript displaying a list of fricative, affricative, and velar conjuncts.
9 References


Preliminary Proposal to Encode the Turkestani Script

### A. Administrative

1. **Title:** Preliminary Proposal to Encode the Turkestani Script

2. **Requester's name:** Lee Wilson

3. **Requester type (Member body/Liaison/Individual contribution):** Individual contribution

4. **Submission date:** 2014-08-02

5. **Choose one of the following:**
   - This is a complete proposal:
   - (or) More information will be provided later:

### B. Technical – General

1. **Choose one of the following:**
   - a. This proposal is for a new script (set of characters):
     - Proposed name of script:
     - Turkestani
   - b. The proposal is for addition of character(s) to an existing block:
     - Name of the existing block:

2. **Number of characters in proposal:** 101

3. **Proposed category (select one from below - see section 2.2 of P&P document):**
   - A-Contemporary
   - B.1-Specialized (small collection)
   - B.2-Specialized (large collection)
   - C-Major extinct
   - D-Attested extinct
   - E-Minor extinct
   - F-Archaic Hieroglyphic or Ideographic
   - G-Obscure or questionable usage symbols

4. **Is a repertoire including character names provided?**
   - a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?
   - b. Are the character shapes attached in a legible form suitable for review?

5. **Fonts related:**
   - a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
   - b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):

6. **References:**
   - a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
   - b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

7. **Special encoding issues:**
   - Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

8. **Additional Information:**
   - Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at [http://www.unicode.org](http://www.unicode.org) for such information on other scripts. Also see Unicode Character Database ([http://www.unicode.org/reports/tr44/](http://www.unicode.org/reports/tr44/)) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

---

## C. Technical - Justification

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Has this proposal for addition of character(s) been submitted before?</td>
<td>No</td>
</tr>
<tr>
<td>2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?</td>
<td>n/a</td>
</tr>
<tr>
<td>If YES, with whom?</td>
<td></td>
</tr>
<tr>
<td>If YES, available relevant documents:</td>
<td></td>
</tr>
<tr>
<td>3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?</td>
<td>extinct</td>
</tr>
<tr>
<td>Reference:</td>
<td></td>
</tr>
<tr>
<td>4. The context of use for the proposed characters (type of use; common or rare)</td>
<td>rare</td>
</tr>
<tr>
<td>Reference:</td>
<td></td>
</tr>
<tr>
<td>5. Are the proposed characters in current use by the user community?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, where? Reference:</td>
<td></td>
</tr>
<tr>
<td>6. After giving due considerations to the principles in the P&amp;P document must the proposed characters be entirely in the BMP?</td>
<td></td>
</tr>
<tr>
<td>If YES, is a rationale provided?</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?</td>
<td>Yes</td>
</tr>
<tr>
<td>8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, is a rationale for its inclusion provided?</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, is a rationale for its inclusion provided?</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, is a rationale for its inclusion provided?</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>11. Does the proposal include use of combining characters and/or use of composite sequences?</td>
<td>Yes</td>
</tr>
<tr>
<td>If YES, a rationale for such use provided?</td>
<td>Yes</td>
</tr>
<tr>
<td>If YES, reference: Combining signs</td>
<td></td>
</tr>
<tr>
<td>Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>12. Does the proposal contain characters with any special properties such as control function or similar semantics?</td>
<td>Yes</td>
</tr>
<tr>
<td>If YES, describe in detail (include attachment if necessary) Virama</td>
<td>Virama</td>
</tr>
<tr>
<td>see proposal for details</td>
<td></td>
</tr>
<tr>
<td>13. Does the proposal contain any Ideographic compatibility characters?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, are the equivalent corresponding unified ideographic characters identified?</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
</tbody>
</table>