Title: Preliminary Proposal to Encode the Turkestani Script
Author: Lee Wilson (ttlwilson@hotmail.com)
Date: 2014-08-02

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.

Figure 1: Map showing the locations of Kucha, Karisahr (Qarahisar), Turfan, Dunhuang, and Hotan in what is now the westernmost region of China. (Mladjov)

3 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.

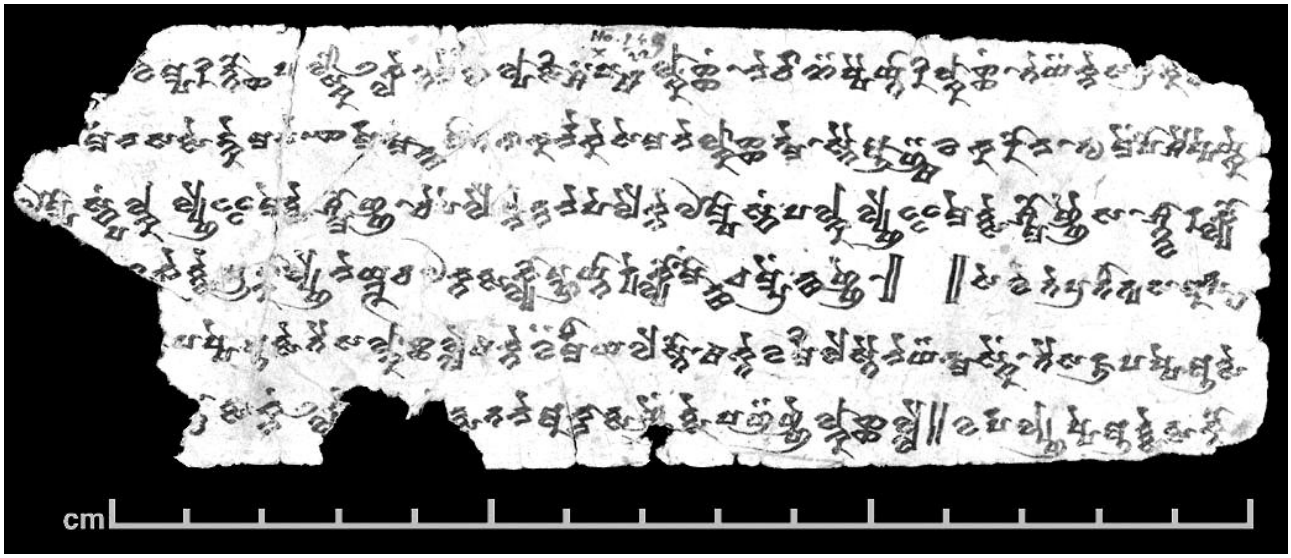


Figure 2: A manuscript written in Tocharian (from International Dunhuang Project).

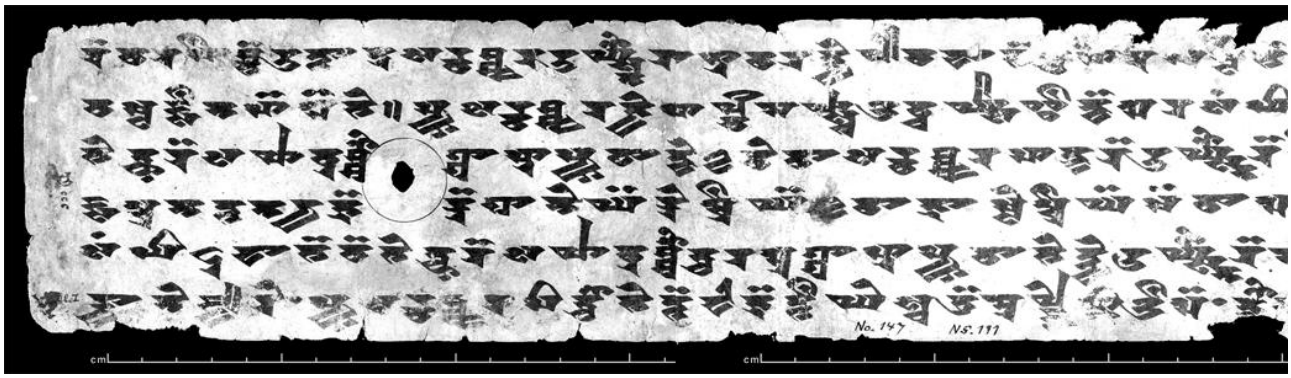


Figure 3: A manuscript written in Khotanese (from International Dunhuang Project).

Several letters are significantly different in form. Compare: Tocharian 𑀇 I, 𑀏 O, 𑀓 KA, 𑀙 CHA, 𑀜 NYA, 𑀡 NNA, 𑀣 THA, 𑀦 BHA, 𑀫 MA, 𑀬 YA vs. Khotanese 𑀆 I, 𑀉 O, 𑀑 KA, 𑀔 CHA, 𑀕 NYA, 𑀖 NNA, 𑀘 THA, 𑀙 BHA, 𑀚 MA, 𑀛 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 Aśokan 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 *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 *ä*, 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

Fremdvokal 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 AI

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:

	ANUSVARA		HOOK		UPADHMANIYA
	VISARGA		JIHVAMULIYA		

HOOK is a sign peculiar to Khotanese, which indicates "the recent loss of an internal sound," usually /ẓ/ (Emmerick, 1979, p. 9) (see Figure 9 a, b).

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:

 *ghā* (GHA, vowel sign AA)
ñā (NYA, vowel sign AA)
pā (NYA, vowel sign AA)
phā (NYA, vowel sign AA)
mā (NYA, vowel sign AA)
yā (YA, vowel sign AA)
ṣā (YA, vowel sign AA)
sā (YA, vowel sign AA)
hā (YA, vowel sign AA)

A variation of this occurs in Khotanese, which spans two separate letters:

 *tāndi* (TA, vowel sign AA, NA, DA, vowel sign I)

The 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:

 *khā* (KHA, vowel sign AA)
gā (GA, vowel sign AA)
dhā (DHA, vowel sign AA)
śā (SHA, vowel sign AA)

This does not occur in Khotanese:

 *khā* (KHA, vowel sign AA)
gā (GA, vowel sign AA)
dhā (DHA, vowel sign AA)
śā (SHA, vowel sign AA)

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

 *ṅā* (NGA, vowel sign AA)
jā (JA, vowel sign AA)

 *ṭā* (TTA, vowel sign AA)
 *ṇā* (NNA, vowel sign AA)

U/UU 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)
 *ḍu* (DDA, vowel sign U)
 *du* (DA, vowel sign U)
 *ru* (RA, vowel sign U)

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

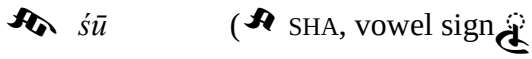
2. The second form is similar to the first, and only occurs with subscript RA:

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

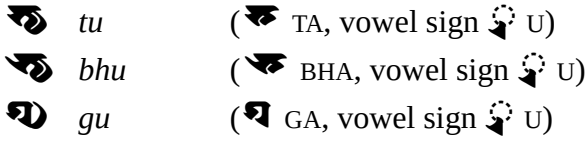
3. Forms superficially resembling the independent vowels O and AU appear in combination with certain letters:

 *tu* (TA, vowel sign U)
 *bhu* (BHA, vowel sign U)
 *gu* (GA, vowel sign U)
 *śu* (SHA, vowel sign U)

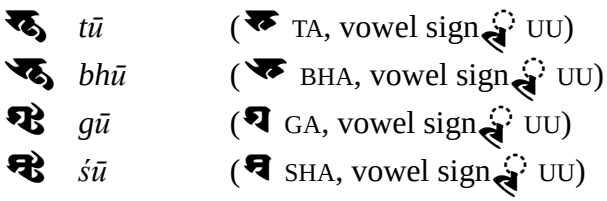
 *tū* (TA, vowel sign UU)
 *bhū* (BHA, vowel sign UU)
 *gū* (GA, vowel sign UU)

 *śū* (SHA, vowel sign UU)

Khotanese script employs an alternate form used for GA and SHA that does not occur in Tocharian. Compare:

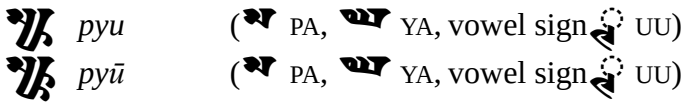


tu (TA, vowel sign U)
bhu (BHA, vowel sign U)
gu (GA, vowel sign U)
śu (SHA, vowel sign U)



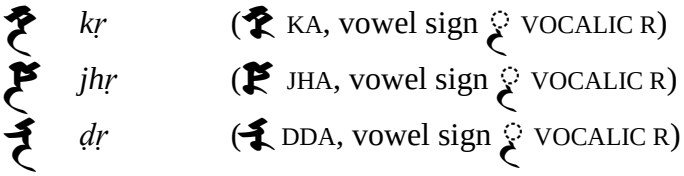
tū (TA, vowel sign UU)
bhū (BHA, vowel sign UU)
gū (GA, vowel sign UU)
śū (SHA, vowel sign UU)

In Khotanese, U and UU each both have a fourth form that attaches only to subscript YA:

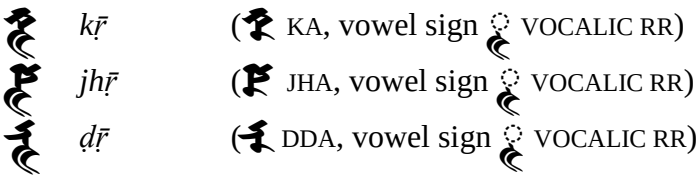


pyu (PA, YA, vowel sign UU)
pyū (PA, YA, vowel sign UU)

VOCALIC R AND RR Similar to the vowels U and UU, when these signs attach to a consonant with a descender, it is deleted:



kṛ (KA, vowel sign VOCALIC R)
jhṛ (JHA, vowel sign VOCALIC R)
ḍṛ (DDA, vowel sign VOCALIC R)



kṝ (KA, vowel sign VOCALIC RR)
jhṝ (JHA, vowel sign VOCALIC RR)
ḍṝ (DDA, vowel sign VOCALIC RR)

Note that these do not occur with RA.

They also cause minor variation in the forms of NA. and BHA:

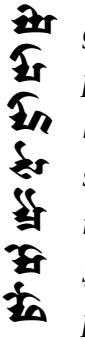
 *nṛ* (NA, vowel sign VOCALIC R)
 bhṛ (BHA, vowel sign VOCALIC R)

 *nṝ* (NA, vowel sign VOCALIC RR)
 bhṝ (BHA, vowel sign VOCALIC RR)

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

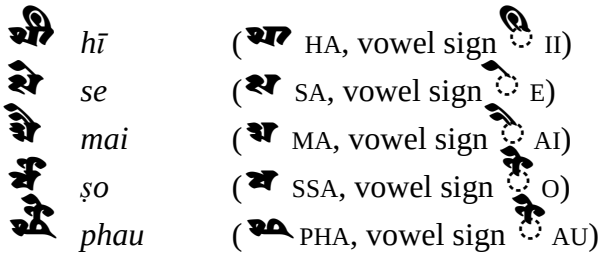
 *li* (LA, vowel sign I)
lī (LA, vowel sign II)
le (LA, vowel sign E)
lai (LA, vowel sign AI)
lo (LA, vowel sign O)
lau (LA, vowel sign AU)

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)
ṣo (SSA, vowel sign O)
pau (PHA, vowel sign AU)

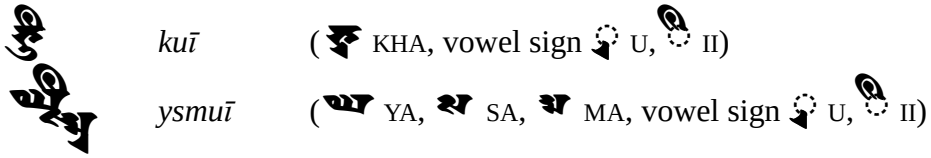
Khotanese functions the same for vowel sign I, but the others are different. Vowel sign II spans the open section, except with consonant HA, for which it appears on the right ascender; vowel sign E appears on the right side; vowel sign AI is not separated as in Tocharian; both elements appear on the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

 *ghi* (GHA, vowel sign I)
pī (PA, vowel sign II)



4.10.2 More than one vowel sign per aksara

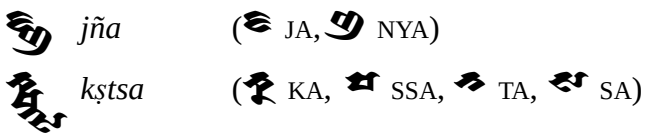
Turkestani occasionally allow more than one vowel sign on a single consonant letter or conjunct. This is used for diphthongs and is found only in Khotanese, as Tocharian makes use of subscript vowel letters instead.



4.11 Conjuncts

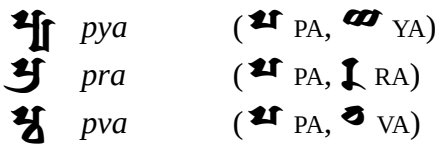
Turkestani employ subscripts to indicate consonant clusters. Most subscripts are relatively transparent and easily identifiable. There are nevertheless some subscripts that differ to a greater or lesser degree from their base forms.

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



3.11.1 Variation in subscript glyph shapes

YA and RA form subscripts that are entirely dissimilar to their base forms, while VA is also slightly different:



In Khotanese, VA takes on a significantly different form when it combines with certain

other letters:

 *vva* (VA, VA)

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

 *dga* (DA, GA)
 *ṣṭha* (SSA, TTHA)
 *ddha* (DA, DHA)
 *ntha* (NA, THA)
 *wśa* (WAE, SHA)

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

 *lla* (LA, LA)
 *csa* (CA, SA)
 *kma* (KA, MA)
 *lta* (LA, TA)

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

 *ysma* (YA, SA, MA)

Here, SA is reduced in form when it combines with MA, but cf.

 *ysda* (YA, SA, DA)

where 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:

 *ṣṭha* (SSA, TTHA)
 *kma* (KA, MA)
 *sya* (SA, YA)

 *csa* (CA, SA)

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:

 *sya* (SA, YA)
 *kya* (KA, YA)

The conjunct *kka* employs an abbreviated form of the subscript:

 *kka* (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.

 *jsa* (JA, SA)
 *tta* (TA, TA)
 *rra* (RA, RA)

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

 *stta* (SA, TA, TA)
 *krra* (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:

 *lysa* (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 TA has a combining form resembling NA. This likely arises from *tsa* appearing frequently, and NA 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.:

kla (KA, LA)
ḍva (DDA, VA)

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

rha (RA, HA)
rṇa (RA, NNA)

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

rgo (RA, GA, vowel sign O)
rnā (RA, NA, vowel sign AA)

The letter LA has an irregular form when it combines with repha RA:

rla (RA, LA)

Repha does not occur with YA; instead, a regular conjunct is formed:

rya (RA, YA)

The letters TA and NA have unique alterations in shape. The alteration in the base form can differentiate the two letters:

twa (TA, WAE)

 *nwa* ( NA,  WAE)

This is not always a reliable guide, however, as TA occasionally resembles the combining form of NA:

 *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 *ku*, which represents the Tocharian consonant /k^{w}/, but it also occurs for metrical rather than phonological reasons (Hitch 2012: 282) (see Figure 8 j, k). Examples:

 *ku* ( KA, vowel sign  U)

 *ce* ( CA, vowel sign  E)

 *küce*

 *wi* ( WAE, vowel sign  I)

 *nā* ( NA, vowel sign  AA)

 *wïnā*

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:

 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.

 *śal* (SHA, LAE, VIRAMA)

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

 *lkānt* (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:

 *mor* (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.

 *teñ* (TA, NYA, VIRAMA)

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

 *picä* (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:

 *koyä* (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 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-u* (KA, LA, YA, vowel sign E U)

 *lo-i* (LA, vowel sign O, I)

Notice that the subscript 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.

 *ghaṃ* (GHA, ANUSVARA)

 *ṅkmāṃ* (NGA, KA, MA, vowel sign 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 *jihvāmūlīya* 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.

 *khaḥ* (KHA, VISARGA)

 *ẖka* (JIHVAMULIYA, KA)

 *ḫpa* (UPADHMANIYA, PA)

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: KAE,  TAE,  NAE,  PAE, MAE,  RAE,  LAE, WAE, SHAE,  SSAE, SAE, and three characters not currently attested in Khotanese:  VOWEL LETTER VOCALIC RR,  JIHVAMULIYA, UPADHMANIYA.

- Khotanese employs three characters not needed for Tocharian:  VOWEL LETTER EI, VOWEL SIGN EI,  HOOK. In addition, Khotanese allows vowel signs and consonant subscripts to be added to independent vowel letters:  *ui’*, *uai*, *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  AA not found in Khotanese. cf. Tocharian  *khā*, Khotanese *khā*.

- Khotanese employs a stylistic variant of sign 𑀆 AA not found in Tocharian. cf. Khotanese 𑀓𑀸𑀦𑁆𑀝𑀺 or 𑀓𑀻𑀦𑁆𑀝𑀺 *tāndi*, Tocharian 𑀓𑀸𑀦𑁆𑀝𑁂 *tāndi*.

- Khotanese draws a distinction in the forms of 𑀉 U and vowel sign 𑁂 UU not found in Tocharian:

	Tocharian	Khotanese
tu	𑀢𑀼	𑀢𑀼
bhu	𑀪𑀼	𑀪𑀼
gu	𑀕𑀼	𑀕𑀼
śu	𑀰𑀼	𑀰𑀼
tū	𑀢𑀽	𑀢𑀽
bhū	𑀪𑀽	𑀪𑀽
gū	𑀕𑀽	𑀕𑀽
śū	𑀰𑀽	𑀰𑀽

- With subscript 𑁆𑀬 YA, Khotanese employs variants of 𑀉 U and vowel sign 𑁂 UU not found in Tocharian:

	Tocharian	Khotanese
pyu	𑀧𑁆𑀬𑀼	𑀧𑁆𑀬𑀼
pyū	𑀧𑁆𑀬𑀽	𑀧𑁆𑀬𑀽

- The variant form of 𑁂 UU for 𑀓 KA 𑀚 JHA 𑀡 DDA and 𑀤 DA is separated into two parts in Tocharian, but is combined into a single diacritic in Khotanese (Khotanese 𑁆𑀭 RA functions the same as Tocharian 𑁆𑀮 RA):

	Tocharian	Khotanese
kū	𑀓𑁆𑀼𑀽	𑀓𑁆𑀽
jhū	𑀚𑁆𑀼𑀽	𑀚𑁆𑀽
ḍū	𑀡𑁆𑀼𑀽	𑀡𑁆𑀽

dū  

- Varied placement of vowel signs on the consonant letter:

	Tocharian	Khotanese
pī		
hī		
se		
mai		
ṣo		
pau		

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

kuī ( KHA, vowel sign U, II)

ysmuī ( YA, SA, MA, vowel sign U, II)

5.4 Differences in conjuncts

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

	Tocharian	Khotanese
kya		
vva		
lya		
ysma		

- Khotanese and Tocharian have different candidates for akhand ligatures:

	Tocharian	Khotanese
jsa	—	
tta	—	
rra	—	
tsa		

- Tocharian allows aksara conjuncts, Khotanese does not

1st aksara	2nd aksara	combined aksara
		
ku	*ce*	*küce*
		
wi	*nā*	*wïnā*

5.5 Virama

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

	Tocharian	Khotanese
śas		
mor		

- Tocharian allows VIRAMA and vowel sign  AE to appear on the same letter, while Khotanese does not:

Tocharian	Khotanese

picä   or 

koyä   or 

5.6 Subscript vowel letters

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

Tocharian

klye-u 

lo-i 

6 Character properties

Turkestani character properties are as follows:

```
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 AI;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 EI;Lo;0;L;;;;;N;;;;;
11F60;TURKESTANI LETTER KA;Lo;0;L;;;;;N;;;;;
11F61;TURKESTANI LETTER KHA;Lo;0;L;;;;;N;;;;;
11F62;TURKESTANI LETTER GA;Lo;0;L;;;;;N;;;;;
11F63;TURKESTANI LETTER GHA;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;;;;;
```

```
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;;;;;
```

```
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;;;;;
11FAA;TURKESTANI NUMBER TEN;No;0;L;;;;10;N;;;;;
11FAB;TURKESTANI NUMBER TWENTY;No;0;L;;;;20;N;;;;;
11FAC;TURKESTANI NUMBER THIRTY;No;0;L;;;;30;N;;;;;
11FAD;TURKESTANI NUMBER FORTY;No;0;L;;;;40;N;;;;;
11FAE;TURKESTANI NUMBER FIFTY;No;0;L;;;;50;N;;;;;
11FAF;TURKESTANI NUMBER SIXTY;No;0;L;;;;60;N;;;;;
11FB0;TURKESTANI NUMBER SEVENTY;No;0;L;;;;70;N;;;;;
11FB1;TURKESTANI NUMBER EIGHTY;No;0;L;;;;80;N;;;;;
11FB2;TURKESTANI NUMBER NINETY;No;0;L;;;;90;N;;;;;
11FB3;TURKESTANI NUMBER ONE HUNDRED;No;0;L;;;;100;N;;;;;
11FB4;TURKESTANI NUMBER ONE THOUSAND;No;0;L;;;;1000;N;;;;;
11FB5;TURKESTANI DANDA;Po;0;L;;;;;N;;;;;
11FB6;TURKESTANI DOUBLE DANDA;Po;0;L;;;;;N;;;;;
11FB7;TURKESTANI PUNCTUATION DOT;Po;0;L;;;;;N;;;;;
11FB8;TURKESTANI PUNCTUATION DOUBLE 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.

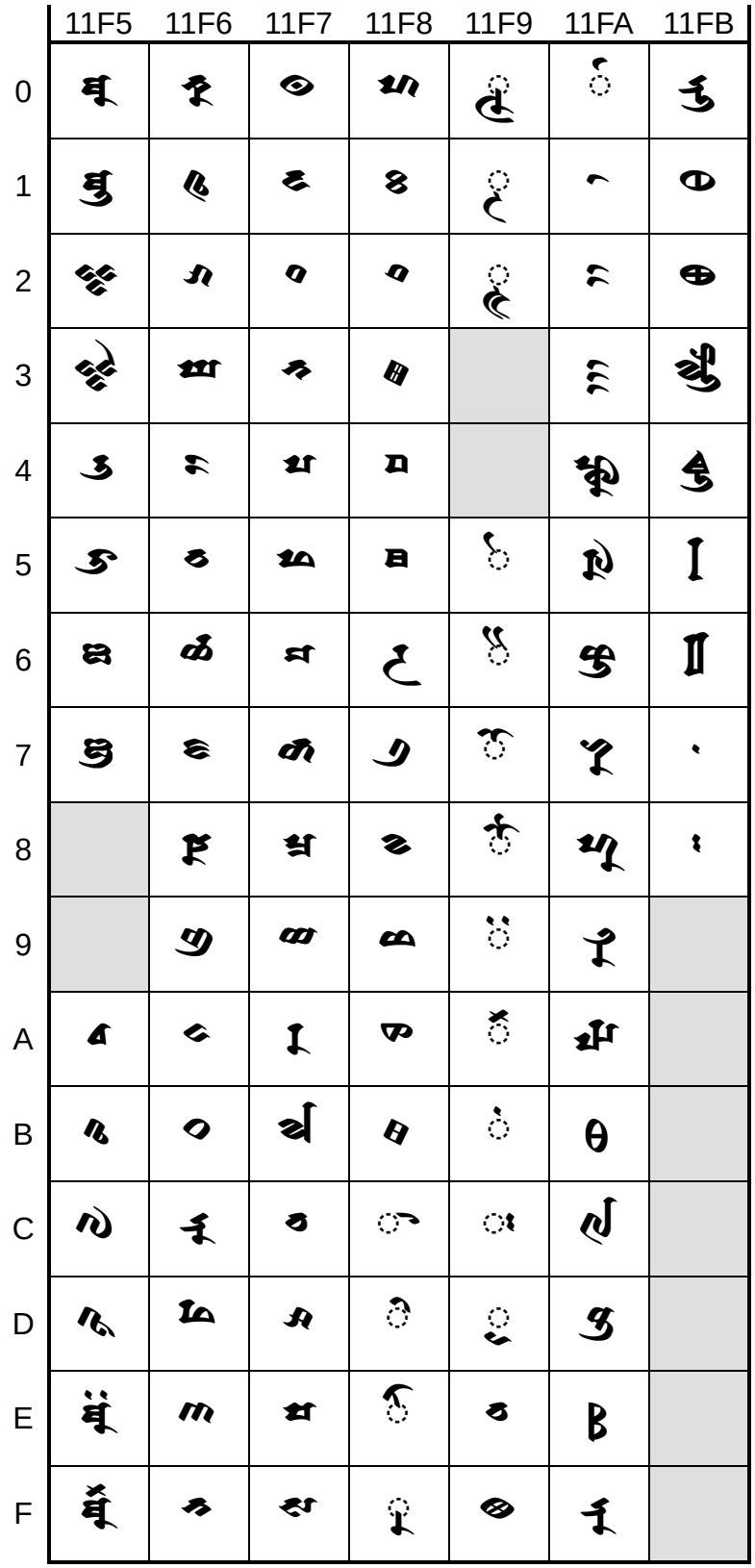


Figure 4: Proposed code chart for Turkestani

Independent vowels

11F50 TURKESTANI LETTER A
11F51 TURKESTANI LETTER AA
11F52 TURKESTANI LETTER I
11F53 TURKESTANI LETTER II
11F54 TURKESTANI LETTER U
11F55 TURKESTANI LETTER UU
11F56 TURKESTANI LETTER VOCALIC R
11F57 TURKESTANI LETTER VOCALIC RR
11F58 <RESERVED>
11F59 <RESERVED>
11F5A TURKESTANI LETTER E
11F5B TURKESTANI LETTER AI
11F5C TURKESTANI LETTER O
11F5D TURKESTANI LETTER AU
11F5E TURKESTANI LETTER AE
11F5F TURKESTANI LETTER EI

Consonants

11F60 TURKESTANI LETTER KA
11F61 TURKESTANI LETTER KHA
11F62 TURKESTANI LETTER GA
11F63 TURKESTANI LETTER GHA
11F64 TURKESTANI LETTER NGA
11F65 TURKESTANI LETTER CA
11F66 TURKESTANI LETTER CHA
11F67 TURKESTANI LETTER JA
11F68 TURKESTANI LETTER JHA
11F69 TURKESTANI LETTER NYA
11F6A TURKESTANI LETTER TTA
11F6B TURKESTANI LETTER TTHA
11F6C TURKESTANI LETTER DDA
11F6D TURKESTANI LETTER DDHA
11F6E TURKESTANI LETTER NNA
11F6F TURKESTANI LETTER TA
11F70 TURKESTANI LETTER THA
11F71 TURKESTANI LETTER DA
11F72 TURKESTANI LETTER DHA
11F73 TURKESTANI LETTER NA
11F74 TURKESTANI LETTER PA
11F75 TURKESTANI LETTER PHA
11F76 TURKESTANI LETTER BA
11F77 TURKESTANI LETTER BHA
11F78 TURKESTANI LETTER MA
11F79 TURKESTANI LETTER YA
11F7A TURKESTANI LETTER RA
11F7B TURKESTANI LETTER LA
11F7C TURKESTANI LETTER VA
11F7D TURKESTANI LETTER SHA
11F7E TURKESTANI LETTER SSA
11F7F TURKESTANI LETTER SA
11F80 TURKESTANI LETTER HA
11F81 TURKESTANI LETTER KAE
11F82 TURKESTANI LETTER TAE
11F83 TURKESTANI LETTER NAE
11F84 TURKESTANI LETTER PAE
11F85 TURKESTANI LETTER MAE
11F86 TURKESTANI LETTER RAE
11F87 TURKESTANI LETTER LAE
11F88 TURKESTANI LETTER WAE
11F89 TURKESTANI LETTER SHAE
11F8A TURKESTANI LETTER SSAE
11F8B TURKESTANI LETTER SAE

Dependent vowel signs

11F8C TURKESTANI SIGN AA
11F8D TURKESTANI SIGN I
11F8E TURKESTANI SIGN II
11F8F TURKESTANI SIGN U
11F90 TURKESTANI SIGN UU
11F91 TURKESTANI SIGN VOCALIC R
11F92 TURKESTANI SIGN VOCALIC RR
11F93 <RESERVED>
11F94 <RESERVED>
11F95 TURKESTANI SIGN E
11F96 TURKESTANI SIGN AI
11F97 TURKESTANI SIGN O
11F98 TURKESTANI SIGN AU
11F99 TURKESTANI SIGN AE
11F9A TURKESTANI SIGN EI

Various signs

11F9B TURKESTANI SIGN ANUSVARA
11F9C TURKESTANI SIGN VISARGA
11F9D TURKESTANI SIGN HOOK
11F9E TURKESTANI SIGN JIHVAMULIYA
11F9F TURKESTANI SIGN UPADHMANIYA

Virama

11FA0 TURKESTANI VIRAMA

Numbers

11FA1 TURKESTANI NUMBER ONE
11FA2 TURKESTANI NUMBER TWO
11FA3 TURKESTANI NUMBER THREE
11FA4 TURKESTANI NUMBER FOUR
11FA5 TURKESTANI NUMBER FIVE
11FA6 TURKESTANI NUMBER SIX
11FA7 TURKESTANI NUMBER SEVEN
11FA8 TURKESTANI NUMBER EIGHT
11FA9 TURKESTANI NUMBER NINE
11FAA TURKESTANI NUMBER TEN
11FAB TURKESTANI NUMBER TWENTY

11FAC TURKESTANI NUMBER THIRTY
11FAD TURKESTANI NUMBER FORTY
11FAE TURKESTANI NUMBER FIFTY
11FAF TURKESTANI NUMBER SIXTY
11FB0 TURKESTANI NUMBER SEVENTY
11FB1 TURKESTANI NUMBER EIGHTY
11FB2 TURKESTANI NUMBER NINETY
11FB3 TURKESTANI NUMBER ONE HUNDRED
11FB7 TURKESTANI PUNCTUATION DOT
11FB8 TURKESTANI PUNCTUATION DOUBLE DOT
11FB4 TURKESTANI NUMBER ONE THOUSAND

Punctuation

11FB5 TURKESTANI DANDA
11FB6 TURKESTANI DOUBLE DANDA

Figure 5: Proposed names list for Turkestani

8 Samples

	BRAHMI	TURKESTANI TOCH	TURKESTANI KHOT
A			
AA			
I			
II			
U			
UU			
VOC.R			
VOC.RR			
E			
AI			
O			
AU			

	BRAHMI	TURKESTANI TOCH	TURKESTANI KHOT
DDA			
DDHA			
NNA			
TA			
THA			
DA			
DHA			
NA			
PA			
PHA			
BA			
BHA			

	BRAHMI	TURKESTANI TOCH	TURKESTANI KHOT
RAE			
LAE			
WAE			
SHAE			
SSAE			
SAE			
SIGNAA			
SIGNI			
SIGNII			
SIGNU			
SIGNUU			
SIGNVOC.R			

Table 1: Comparison of basic letters and signs of Aśokan Brahmi and Tocharian Turkestani.

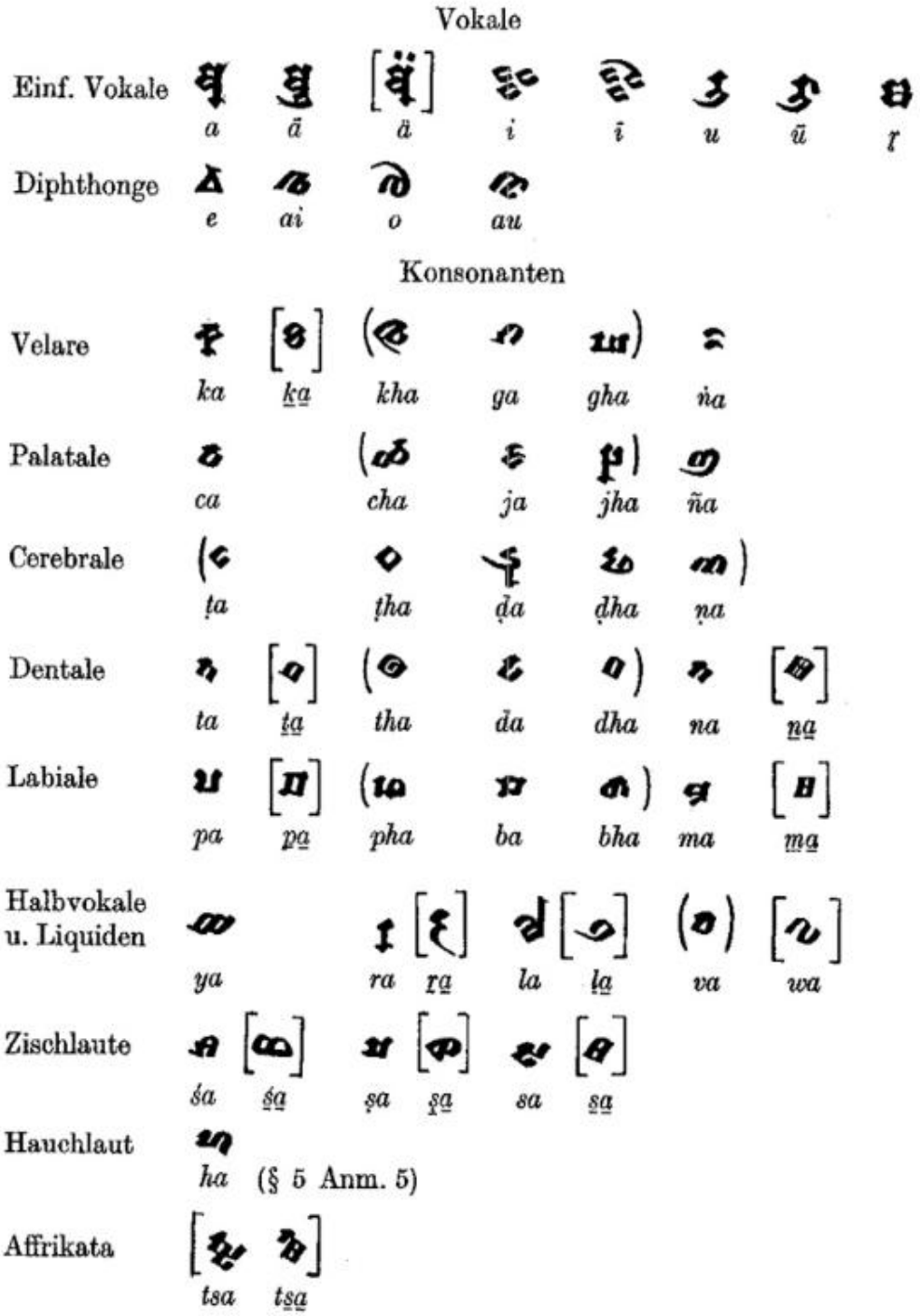


Figure 6: A table of the basic letters of Tocharian Turkestani (from Krause and Thomas 1960:41, Malzahn 2007b:227-8).

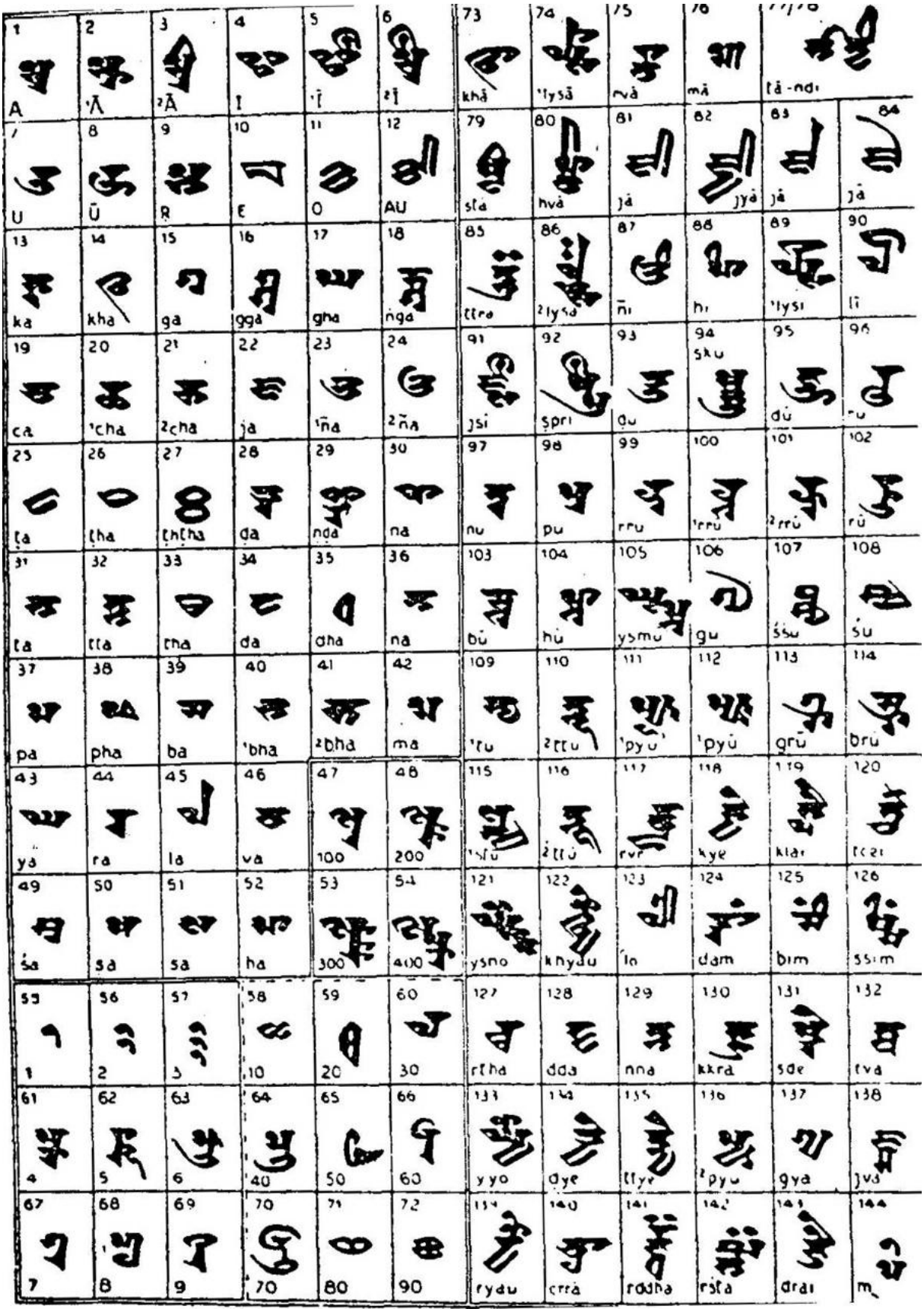


Figure 7: A table of the basic letters, signs, and digits of Khotanese Turkestani as well as a selection of conjuncts (from Leumann, 1934: 39).

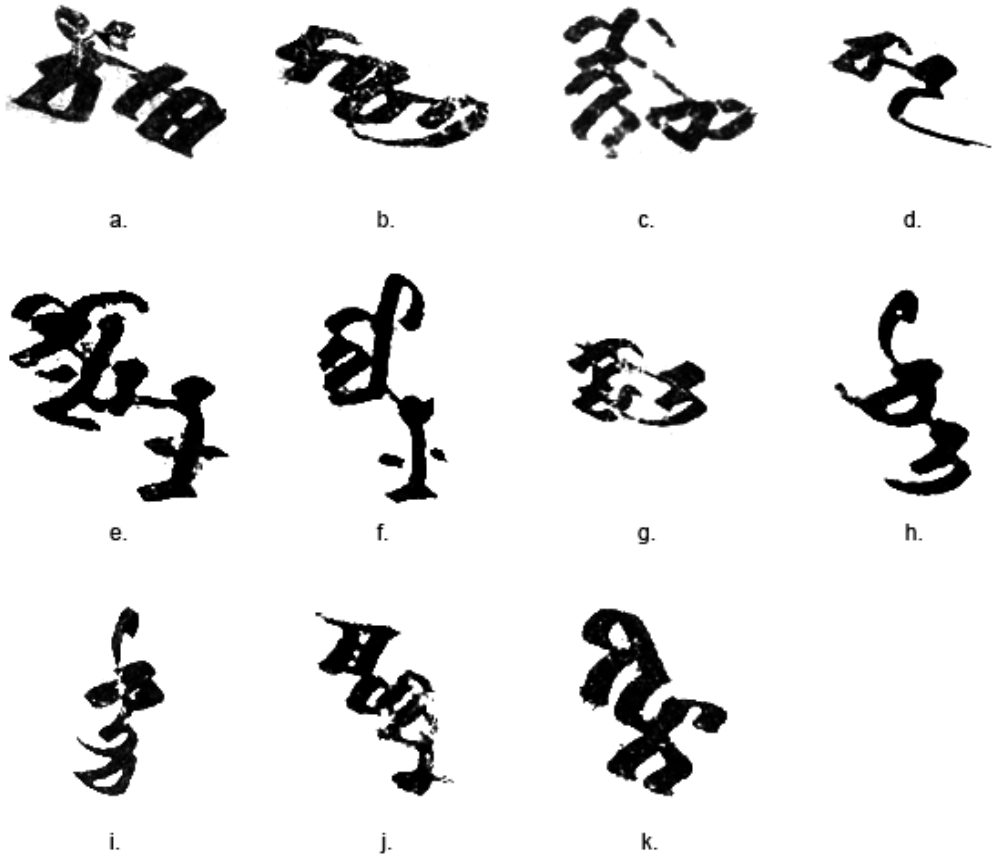


Figure 8: examples of bar virama, subscript independent vowel letters, and stacked aksaras. a. *ceṃts*, b. *ṅk̲ä̲l̲*, c. *ttos̲*, d. *cār̲*, e. *ssoî*, f. *loî*, g. *ksāû*, h. *ceû*, i *küce*, j. *mañcu*, k. *winā*.

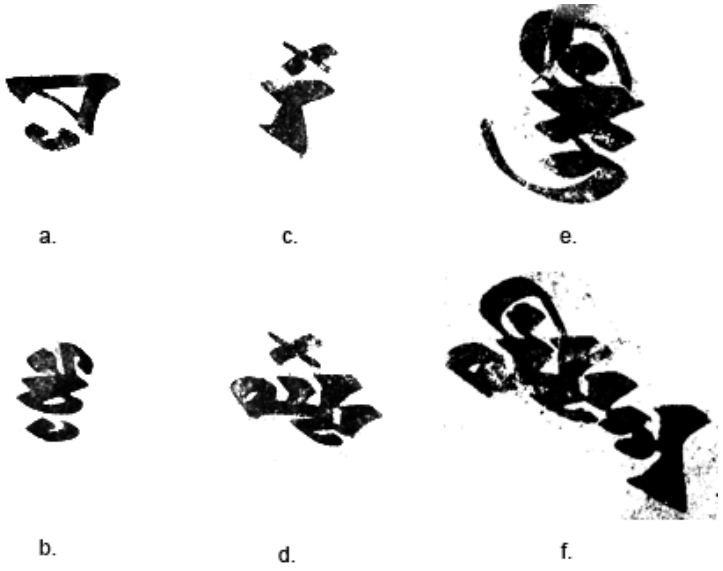


Figure 9: Examples of Khotanese-specific signs and aksaras with double vowel signs: a. *e'*, b. *vo'*, c. *rei*, d. *ysei*. e. *kuī*, f. *ysmuī*

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.

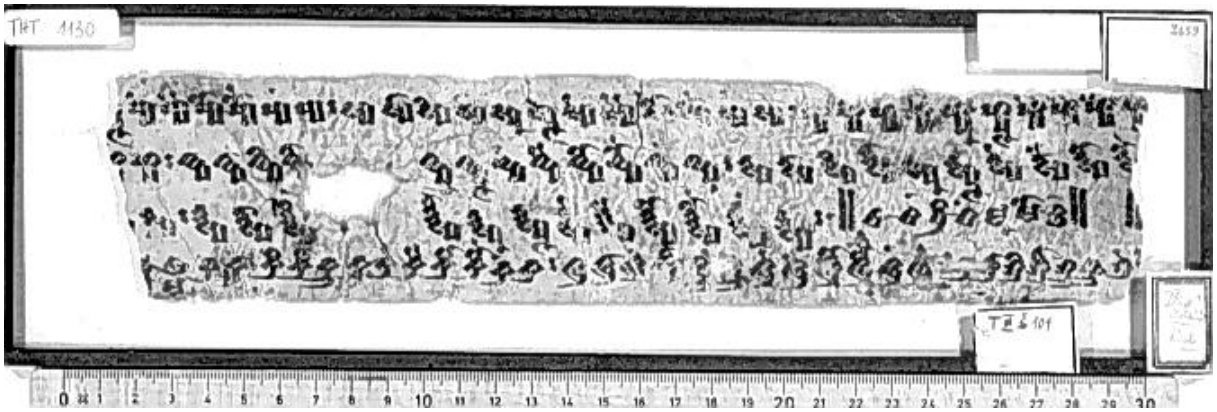


Figure 13: Original Tocharian manuscript displaying a list of fricative, affricative, and velar conjuncts.

9 References

Emmerick, Ronald E. “Khotanese and Tumshuqese” In: *The Iranian Languages*, Windfuhr, Gernot ed. 2013, pp 377-415. Routledge.

Hitch, Doug. “Review: Melanie Malzahn (ed.) *Instrumenta Tocharica*” In *Tocharian and Indo-European Studies*, vol 13. Jens Elmegård Rasmussen, Michaël Peyrot, Thomas Olander, eds. 2012, pp 277-290. Copenhagen: Museum Tusculanum Press.

International Dunhuang Project. `http://idp.bl.uk/`. Accessed August 2014.

Krause, Todd B., Jonathan Slocum. Tocharian Online. May 13, 2014. `http://www.utexas.edu/cola/centers/lrc/eieol/tokol-1-X.html`. Accessed July, 2014.

Krause, Wolfgang, Werner Thomas. *Tocharisches Elementarbuch*.1960. Heidelberg: Carl Winter - Universitätsverlag.

Leumann, M. *Sakische Handschriftproben*. Zürich: Manu Leumann, 1934.

Mladjov, Ian. “Tang China” `http://www.historyandcivilization.com/Maps---Tables---Chinese-History.html`. Accessed July 2014.

Pulleybank, Edwin G., Ronald E. Emmerick. *A Chinese text in Central Asian Brahmi script: New evidence for the pronunciation of Late Middle Chinese and Khotanese*. Rome: Istituto Italiano per il Medio ed Estremo Oriente, 1993.

TITUS. “The Tocharian “Alphabet””. `http://titus.fkidg1.uni-frankfurt.de/didact/idg/toch/tochbr.htm`. Accessed July 2014.

---. “Tocharian Manuscripts”. `http://titus.fkidg1.uni-frankfurt.de/texte/tocharic/tht.htm`. Accessed August 2014.

Wojciech Skalmowski, Alois van Tongerloo, Katholieke Universiteit te Leuven. *Middle Iranian Studies*. Leuven: Peeters Publishers, 1984.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html **.**
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Preliminary Proposal to Encode the Turkestani Script*
2. Requester's name: *Lee Wilson (ttlwilson@hotmail.com)*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *2014-08-02*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *yes*
   (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): *yes*
      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	X	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? *Yes*
   a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? *Yes*
   b. Are the character shapes attached in a legible form suitable for review? *Yes*
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
   Lee Wilson (TrueType or OpenType format)
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   Lee Wilson (ttlwilson@hotmail.com)
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *No*
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)?
   No

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 for such information on other scripts. Also see Unicode Character Database (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.

[1] Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

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