Title: Preliminary Proposal to Encode the Khotanese Script
Author: Lee Wilson (ttlwilson@hotmail.com)
Date: 2015-01-26

1   Introduction

This is a proposal to encode the Khotanese script in the Universal Character Set (ISO/IEC 10646). This document outlines the unified system for encoding Khotanese, 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

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.

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

3   The Issue of Representing Khotanese in Unicode

The primary issue facing any proposal to encode Khotanese in the UCS is that of unification with the Brahmi script. While it is true that many sources on Khotanese refer to its script as a variant of Brahmi or some similar appellation, the Khotanese script nevertheless presents several differences from Brahmi as laid out in the UCS in terms of glyph shapes, character repertoire, and rendering behaviours. With this in mind, there are two main models for representing Khotanese in the UCS:

1. Encoding Khotanese as an independent script
2. Encoding Khotanese as a subset of Brahmi

3.1   Assessment of the Models for Representing Tocharian

Due to the traditional description of Khotanese script as a variant of Brahmi, their similar character repertoire, and the numerous Sanskrit loanwords found in Khotanese, some may argue that Khotanese is simply a regional variation of Brahmi and is accordingly a candidate for being encoded as a subset of Brahmi. In such a case, the distinctive elements of Khotanese would need to managed at the presentation level through fonts and by encoding characters unique to Khotanese as Brahmi extensions. This approach poses problems, which are outlined below:

2

- *Failure to provide a plain text solution*: The Brahmi script as represented in the UCS is based on Aśokan Brahmi from the 3rd century BCE. The first and most obvious issue facing the encoding of Khotanese as a subset of Brahmi is the visual dissimilarity of Khotanese and Brahmi characters. Nearly all characters in Khotanese are considerably different from their Brahmi counterparts, as illustrated in the following selection of letters:

Any reader of Brahmi-encoded Khotanese texts would be required to obtain a Brahmi font with character design based on Khotanese in order to read the texts, and would subsequently be unable to view Aśokan Brahmi texts properly. As a result, considering Khotanese as a subset of Brahmi fails to provide a means for plain text representation of the script.

- *Fundamental differences in structure*: Khotanese, while descended from Brahmi, employs a structural form not used in Brahmi, namely the attachment of multiple vowel signs to a single aksara, which will be outlined below. As such, Khotanese script cannot be accurately rendered with Brahmi encoding.

Considering these problems, model 1, encoding Khotanese as an independent script, appears to be the best option.
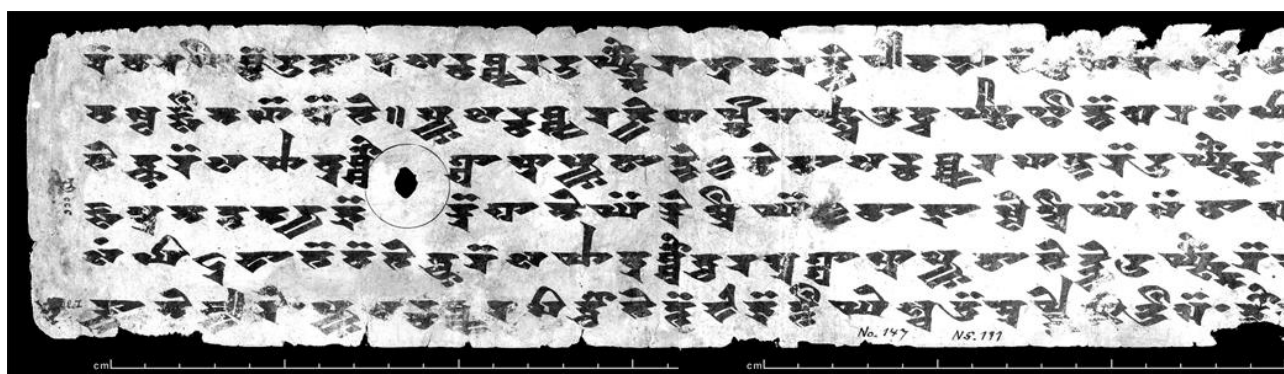

Figure 3: A manuscript written in Khotanese (from International Dunhuang Project).

4   Structure

4.1   Introduction

Khotanese script has typically been referred to as a modified form of Brahmi, indicating that

3

people have traditionally not considered this to be an independent script. While it is true that Khotanese structure and functionality is indeed clearly within the Brahmic tradition, the it is 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*.

Khotanese also employs unique compounding, 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 13 independent vowel signs:

अ	VOWEL LETTER A		VOWEL LETTER UU		VOWEL LETTER AU
	VOWEL LETTER AA		VOWEL LETTER VOCALIC R		VOWEL LETTER AE
	VOWEL LETTER I		VOWEL LETTER E		VOWEL LETTER EI
	VOWEL LETTER II		VOWEL LETTER AI		
	VOWEL LETTER U		VOWEL LETTER O		

The vowel VOCALIC RR is not attested in Khotanese texts, but a space has been left available in the code block in case of future discovery.
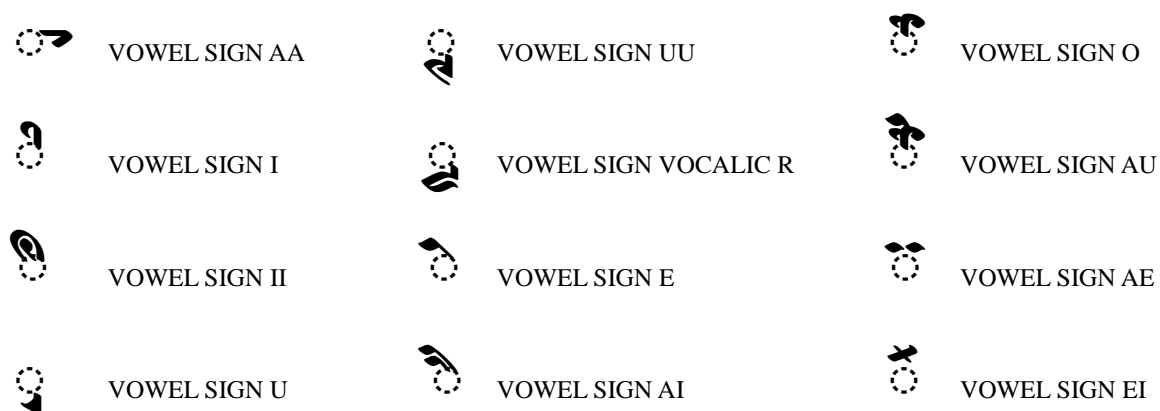
Khotanese allows for diphthongs to be represented by adding vowel diacritics to independent vowel signs:

ui'   *uai*   *uvi*
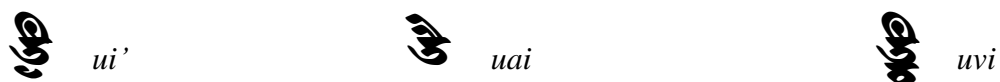
These are likely best represented as character combinations rather than individual characters.

4.6   Vowel Signs

There are 12 dependent vowel signs:

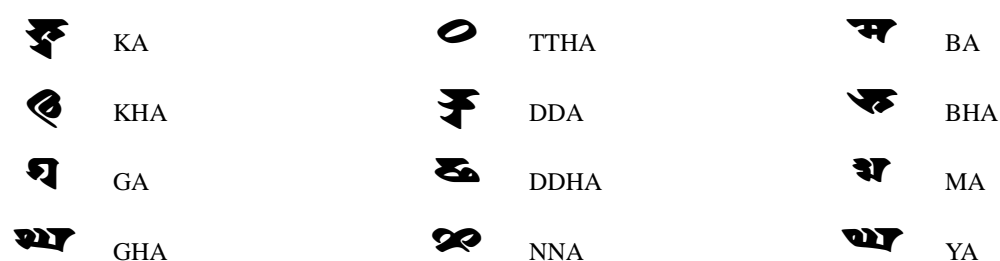|		
VOWEL SIGN AA	VOWEL SIGN UU	VOWEL SIGN O
VOWEL SIGN I	VOWEL SIGN VOCALIC R	VOWEL SIGN AU
VOWEL SIGN II	VOWEL SIGN E	VOWEL SIGN AE
VOWEL SIGN U	VOWEL SIGN AI	VOWEL SIGN EI

VOWEL SIGN AE indicates the vowel /ə/ in Khotanese (Emmerick and Pulleybank 1993: 45-46). The transcription <ä> is standard.

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

4.7   Consonants

There are 33 consonant letters:

KA	TTHA	BA
KHA	DDA	BHA
GA	DDHA	MA
GHA	NNA	YA

5

𑦮	NGA	𑦵	TA	𑦱	RA
𑦡	CA	𑦶	THA	𑦯	LA
𑦢	CHA	𑦷	DA	𑦰	VA
𑦣	JA	𑦸	DHA	𑦳	SHA
𑦤	JHA	𑦹	NA	𑦴	SSA
𑦥	NYA	𑦺	PA	𑦲	SA
𑦦	TTA	𑦻	PHA	𑦽	HA

All letters bear the inherent vowel *a*. This vowel may be silenced with $\overset{?}{\circ}$ VIRAMA or most typically through the use of conjuncts, to be explained below.

Note that many scribes failed to differentiate between the default forms of the letters 𑦵 TA and 𑦹 NA in Khotanese manuscripts. However, their combinations with certain vowels and subscript consonants remain distinct, as well as their forms in Khotanese cursive script, thus requiring the letters to be encoded separately. Their forms remain distinct in the font for convenience.

4.8   Various signs

There are 2 various signs:

$\overset{\bullet}{\circ}$ ANUSVARA	$\overset{\cdot}{\circ}$ HOOK

$\overset{\cdot}{\circ}$ HOOK indicates "the recent loss of an internal sound, usually /ẓ/" (Emmerick, 1979, p. 9) (see Figure 9 a, b).

4.9   Numbers

There are 19 numbers:

𑣠	ONE	𑣥	SIX	𑣩	TWENTY	𑣭	SEVENTY
𑣡	TWO	𑣦	SEVEN	𑣪	THIRTY	𑣮	EIGHTY
𑣢	THREE	𑣧	EIGHT	𑣫	FORTY	𑣯	NINETY
𑣣	FOUR	𑣨	NINE	𑣬	FIFTY	𑣰	ONE HUNDRED

Ⰽ FIVE   **∞** TEN   **Ⰼ** SIXTY

Numbers for various multiples of one hundred also exist (Ⰽ 200, Ⰽ 300, Ⰽ 400), 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. A space has been left open in the code chart pending the discovery of a character for 1000.

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 signs The vowel signs ◌ⶳ AA, ◌ U, and ◌ UU also takes on several contextual forms, and the consonant letter ⰾ LA takes on irregular forms.

4.10.1   Contextual forms of vowel signs

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

1   When combined with open-topped consonants and certain others:

     ghā   (GHA, vowel sign ◌ⶳ AA)
     pā   (PA, vowel sign ◌ⶳ AA)
     phā   (PHA, vowel sign ◌ⶳ AA)
     mā   (MA, vowel sign ◌ⶳ AA)
     yā   (YA, vowel sign ◌ⶳ AA)
     ṣā   (SSA, vowel sign ◌ⶳ AA)
     sā   (SA, vowel sign ◌ⶳ AA)
     hā   (HA, vowel sign ◌ⶳ AA)

A variation which spans two separate letters also occurs:

     tāndi   (TA, vowel sign ◌ⶳ AA, NA, DA, vowel sign ◌ I)

The is not mandatory, however, and should be considered a stylistic variant best handled at

the font level.

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

         ṅā   (NGA, vowel sign AA)

         jā   (JA, vowel sign AA)

In some texts, the rounded form #1 appears identical to form #3, however, form #3 never appears as form #1.

U/UU The vowel signs U and UU have four 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 and RA:

         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 letters TA and BHA have special forms:

         tu   (TA, vowel sign U)

         bhu   (BHA, vowel sign U)

         tū   (TA, vowel sign UU)

         bhū   (BHA, vowel sign UU)

3.   The letters GA and SHA also have special forms:

         gu   (GA, vowel sign U)

　　　𑐩　*śu*　　　　(𑐩 SHA, vowel sign 𑑂 U)

　　　𑐩　*gū*　　　　(𑐩 GA, vowel sign 𑑂 UU)
　　　𑐩　*śū*　　　　(𑐩 SHA, vowel sign 𑑂 UU)

4.　　The letter 𑐢 NA is slightly altered when it takes these vowel signs:

　　　𑐥　*nu*　　　　(𑐢 SHA, vowel sign 𑑂 U)
　　　𑐥　*nū*　　　　(𑐢 GA, vowel sign 𑑂 UU)

5.　　Subscript 𑐬 RA and 𑐰 RRA take a special form:

　　　𑐕　*gra*　　　(𑐨 GA, 𑐬 RA)
　　　𑐕　*gru*　　　(𑐨 GA, 𑐬 RA, vowel sign 𑑂 U)
　　　𑐕　*grū*　　　(𑐨 GA, 𑐬 RA, vowel sign 𑑂 UU)

　　　𑐳　*krra*　　(𑐳 GA, 𑐬 RA, 𑐬 RA)
　　　𑐳　*krru*　　(𑐳 GA, 𑐬 RA, 𑐬 RA, vowel sign 𑑂 U)
　　　𑐳　*krrū*　　(𑐳 GA, 𑐬 RA, 𑐬 RA, vowel sign 𑑂 UU)

6.　　There is also a form similar to #5 that attaches only to subscript 𑐨 YA:

　　　𑐫　*pyu*　　　(𑐥 PA, 𑐨 YA, vowel sign 𑑂 UU)
　　　𑐫　*pyū*　　　(𑐥 PA, 𑐨 YA, vowel sign 𑑂 UU)

VOCALIC R Similar to the vowels U and UU, when this sign attaches to a consonant with a descender, the descender is deleted:

　　　𑐳　*kṛ*　　　(𑐳 KA, vowel sign 𑑂 VOCALIC R)

Note that these do not occur with 𑐬 RA.

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

　　　𑐮　*li*　　　(𑐮 LA, vowel sign 𑑂 I)
　　　𑐮　*lī*　　　(𑐮 LA, vowel sign 𑑂 II)
　　　𑐮　*le*　　　(𑐮 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. Examples:

ghi (⟨GHA, vowel sign I⟩)
pī (⟨PA, vowel sign II⟩)

4.10.2   More than one vowel sign per aksara

Khotanese occasionally allow more than one vowel sign on a single consonant letter or conjunct.

kuī (⟨KHA, vowel sign U, II⟩)

ysmuī (⟨YA, SA, MA, vowel sign U, II⟩)

4.11 Conjuncts

Khotanese employs 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.

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

ṣṭa (⟨SSA, TTA⟩)

stta (⟨SA, TA, TA⟩)

lysda (⟨LA, YA, SA, DA⟩)

4.11.1   Variation in subscript glyph shapes

YA and RA form subscripts that are entirely dissimilar to their base forms:

ꍏ *bya*   (ꍏ BA, ꍏ YA)
ꍏ *pra*   (ꍏ PA, ꍏ RA)

VA takes on a significantly different form when it combines with certain other letters:

ꍏ *vva*   (ꍏ VA, ꍏ VA)
ꍏ *tva*   (ꍏ TA, ꍏ VA)

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

ꍏ *ṣṭha*   (ꍏ SSA, ꍏ TTHA)
ꍏ *tva*   (ꍏ TA, ꍏ VA)

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

ꍏ   *ysma*   (ꍏ YA, ꍏ SA, ꍏ MA)

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

ꍏ   *ysda*   (ꍏ YA, ꍏ SA, ꍏ DA)

where ꍏ 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:

ꍏ *gga*   (ꍏ GA, ꍏ GA)
ꍏ *pya*   (ꍏ PA, ꍏ YA)
ꍏ *jsa*   (ꍏ JA, ꍏ SA)
ꍏ *ṣṣa*   (ꍏ SSA, ꍏ SSA)

Subscript ꍏ YA, varies in exact length. When attaching to letters that end in a vertical descender on the right side, it extends to the top of the writing line, e.g.:

ꍏ *gya*   (ꍏ GA, ꍏ YA)

ꍏ *pya*   (ꍏ PA, ꍏ YA)

11

For letters lacking this descender, ᨀ YA is shortened and typically has an altered top serif that curves back into the base glyph:

ᨁ *kya*   (ᨂ KA, ᨀ YA)

ᨃ *rya*   (ᨄ RA, ᨀ YA)

ᨅ *dya*   (ᨆ DA, ᨀ YA)

ᨇ *ttya*   (ᨈ TA, ᨈ TA, ᨀ YA)

The exact height and end point varies somewhat from manuscript to manuscript, but it generally does not reach the upper writing line.

With the letter ᨉ LA, subscript ᨀ YA has a special form:

ᨊ   *lya*   (ᨉ LA, ᨀ YA)

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

ᨋ *jsa*   (ᨌ JA, ᨍ SA)

ᨎ *tta*   (ᨈ TA, ᨈ TA)

ᨏ *nda*   (ᨐ NA, ᨆ DA)

ᨑ *rra*   (ᨄ RA, ᨄ RA)

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

ᨒ *stta*   (ᨍ SA, ᨈ TA, ᨈ TA)

ᨓ *krra*   (ᨂ KA, ᨄ RA, ᨄ RA)

The frequently occurring conjunct ᨔ *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 + ᨀ YA takes precedence over the conjunct for ᨀ YA + ᨍ SA conjunct. Compare:

ᨊ *lya*   (ᨉ LA, ᨀ YA)

ᨔ *ysa*   (ᨀ YA, ᨍ SA)

ᨕ *lysa*   (ᨉ LA, ᨀ YA, ᨍ SA)

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

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.:

![glyph]	*klā*	(![KA] KA, ![LA] LA)
![glyph]	*ḍva*	(![DDA] DDA, ![VA] VA)

This also occurs with the letter ![RA] RA, but with an important difference: namely, that it acts as a repha. The position of repha varies from manuscript to manuscript; in some, it appears level with the writing line, while in others, it appears above the writing level. As both forms are frequently attested, either one is a viable option, however, the choice of which one to use would affect implementation, as an above-line initial RA would require the repha function, while an aligned RA would act as a normal conjunct. Examples:

![glyph]	*rna*	(![RA] RA, ![NA] NA)
![glyph]	*rtha*	(![RA] RA, ![THA] THA)

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

![glyph]	*rnā*	(![RA] RA, ![NA] NA, vowel sign ![AA] AA)
![glyph]	*rcä*	(![RA] RA, ![NA] NA, vowel sign ![AE] AE)

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

![glyph]	*rya*	(![RA] RA, ![YA] YA)

4.12 Virama

There is 1 virama:

![VIRAMA glyph]   VIRAMA

The Khotanese virama functions identically to the standard Brahmic virama, silencing the inherent vowel. It has no irregular forms and appears above the consonant.

k   (KA,   VIRAMA)

4.13 Nasalization

The Khotanese language does not have nasalization, but the script nevertheless employs *anusvāra* both for nasal consonants and for transcription of Sanskrit nasalization. It appears immediately above the base consonant letter.

maṃ   (MA,   ANUSVARA)

4.14 Punctuation

There are 3 punctuation marks:

DOUBLE DANDA   PUNCTUATION DOT

PUNCTUATION DOUBLE DOT

It is possible that the single danda is also used. A space has been left open in the code chart for this character..

5   Character properties

Khotanese character properties are as follows:

```
11E65;KHOTANESE LETTER A;Lo;0;L;;;;;N;;;;;
11E66;KHOTANESE LETTER AA;Lo;0;L;;;;;N;;;;;
11E67;KHOTANESE LETTER I;Lo;0;L;;;;;N;;;;;
11E68;KHOTANESE LETTER II;Lo;0;L;;;;;N;;;;;
11E69;KHOTANESE LETTER U;Lo;0;L;;;;;N;;;;;
11E6A;KHOTANESE LETTER UU;Lo;0;L;;;;;N;;;;;
11E6B;KHOTANESE LETTER VOCALIC R;Lo;0;L;;;;;N;;;;;
11E6C;<RESERVED>
11E6D;KHOTANESE LETTER E;Lo;0;L;;;;;N;;;;;
11E6E;KHOTANESE LETTER AI;Lo;0;L;;;;;N;;;;;
11E6F;KHOTANESE LETTER O;Lo;0;L;;;;;N;;;;;
11E70;KHOTANESE LETTER AU;Lo;0;L;;;;;N;;;;;
11E71;KHOTANESE LETTER AE;Lo;0;L;;;;;N;;;;;
```

```
11E72;KHOTANESE LETTER EI;Lo;0;L;;;;;N;;;;;
11E73;KHOTANESE LETTER KA;Lo;0;L;;;;;N;;;;;
11E74;KHOTANESE LETTER KHA;Lo;0;L;;;;;N;;;;;
11E75;KHOTANESE LETTER GA;Lo;0;L;;;;;N;;;;;
11E76;KHOTANESE LETTER GHA;Lo;0;L;;;;;N;;;;;
11E77;KHOTANESE LETTER NGA;Lo;0;L;;;;;N;;;;;
11E78;KHOTANESE LETTER CA;Lo;0;L;;;;;N;;;;;
11E79;KHOTANESE LETTER CHA;Lo;0;L;;;;;N;;;;;
11E7A;KHOTANESE LETTER JA;Lo;0;L;;;;;N;;;;;
11E7B;KHOTANESE LETTER JHA;Lo;0;L;;;;;N;;;;;
11E7C;KHOTANESE LETTER NYA;Lo;0;L;;;;;N;;;;;
11E7D;KHOTANESE LETTER TTA;Lo;0;L;;;;;N;;;;;
11E7E;KHOTANESE LETTER TTHA;Lo;0;L;;;;;N;;;;;
11E7F;KHOTANESE LETTER DDA;Lo;0;L;;;;;N;;;;;
11E80;KHOTANESE LETTER DDHA;Lo;0;L;;;;;N;;;;;
11E81;KHOTANESE LETTER NNA;Lo;0;L;;;;;N;;;;;
11E82;KHOTANESE LETTER TA;Lo;0;L;;;;;N;;;;;
11E83;KHOTANESE LETTER THA;Lo;0;L;;;;;N;;;;;
11E84;KHOTANESE LETTER DA;Lo;0;L;;;;;N;;;;;
11E85;KHOTANESE LETTER DHA;Lo;0;L;;;;;N;;;;;
11E86;KHOTANESE LETTER NA;Lo;0;L;;;;;N;;;;;
11E87;KHOTANESE LETTER PA;Lo;0;L;;;;;N;;;;;
11E88;KHOTANESE LETTER PHA;Lo;0;L;;;;;N;;;;;
11E89;KHOTANESE LETTER BA;Lo;0;L;;;;;N;;;;;
11E8A;KHOTANESE LETTER BHA;Lo;0;L;;;;;N;;;;;
11E8B;KHOTANESE LETTER MA;Lo;0;L;;;;;N;;;;;
11E8C;KHOTANESE LETTER YA;Lo;0;L;;;;;N;;;;;
11E8D;KHOTANESE LETTER RA;Lo;0;L;;;;;N;;;;;
11E8E;KHOTANESE LETTER LA;Lo;0;L;;;;;N;;;;;
11E8F;KHOTANESE LETTER VA;Lo;0;L;;;;;N;;;;;
11E90;KHOTANESE LETTER SHA;Lo;0;L;;;;;N;;;;;
11E91;KHOTANESE LETTER SSA;Lo;0;L;;;;;N;;;;;
11E92;KHOTANESE LETTER SA;Lo;0;L;;;;;N;;;;;
11E93;KHOTANESE LETTER HA;Lo;0;L;;;;;N;;;;;
11E94;KHOTANESE VOWEL SIGN AA;Mc;0;L;;;;;N;;;;;
11E95;KHOTANESE VOWEL SIGN I;Mn;0;NSM;;;;;N;;;;;
11E96;KHOTANESE VOWEL SIGN II;Mn;0;NSM;;;;;N;;;;;
11E97;KHOTANESE VOWEL SIGN U;Mn;0;NSM;;;;;N;;;;;
11E98;KHOTANESE VOWEL SIGN UU;Mn;0;NSM;;;;;N;;;;;
11E99;KHOTANESE VOWEL SIGN VOCALIC R;Mn;0;NSM;;;;;N;;;;;
11E9A;<RESERVED>
11E9B;KHOTANESE VOWEL SIGN E;Mn;0;NSM;;;;;N;;;;;
11E9C;KHOTANESE VOWEL SIGN AI;Mn;0;NSM;;;;;N;;;;;
11E9D;KHOTANESE VOWEL SIGN O;Mn;0;NSM;;;;;N;;;;;
11E9E;KHOTANESE VOWEL SIGN AU;Mn;0;NSM;;;;;N;;;;;
11E9F;KHOTANESE VOWEL SIGN AE;Mn;0;NSM;;;;;N;;;;;
11EA0;KHOTANESE VOWEL SIGN EI;Mn;0;NSM;;;;;N;;;;;
11EA1;KHOTANESE SIGN HOOK;Mn;0;NSM;;;;;N;;;;;
11EA2;KHOTANESE SIGN ANUSVARA;Mn;0;NSM;;;;;N;;;;;
11EA3;KHOTANESE VIRAMA;Mn;9;L;;;;;N;;;;;
11EA4;KHOTANESE NUMBER ONE;No;0;L;;;;1;N;;;;;
```

```
11EA5;KHOTANESE NUMBER TWO;No;0;L;;;;2;N;;;;;
11EA6;KHOTANESE NUMBER THREE;No;0;L;;;;3;N;;;;;
11EA7;KHOTANESE NUMBER FOUR;No;0;L;;;;4;N;;;;;
11EA8;KHOTANESE NUMBER FIVE;No;0;L;;;;5;N;;;;;
11EA9;KHOTANESE NUMBER SIX;No;0;L;;;;6;N;;;;;
11EAA;KHOTANESE NUMBER SEVEN;No;0;L;;;;7;N;;;;;
11EAB;KHOTANESE NUMBER EIGHT;No;0;L;;;;8;N;;;;;
11EAC;KHOTANESE NUMBER NINE;No;0;L;;;;9;N;;;;;
11EAD;KHOTANESE NUMBER TEN;No;0;L;;;;10;N;;;;;
11EAE;KHOTANESE NUMBER TWENTY;No;0;L;;;;20;N;;;;;
11EAF;KHOTANESE NUMBER THIRTY;No;0;L;;;;30;N;;;;;
11EB0;KHOTANESE NUMBER FORTY;No;0;L;;;;40;N;;;;;
11EB1;KHOTANESE NUMBER FIFTY;No;0;L;;;;50;N;;;;;
11EB2;KHOTANESE NUMBER SIXTY;No;0;L;;;;60;N;;;;;
11EB3;KHOTANESE NUMBER SEVENTY;No;0;L;;;;70;N;;;;;
11EB4;KHOTANESE NUMBER EIGHTY;No;0;L;;;;80;N;;;;;
11EB5;KHOTANESE NUMBER NINETY;No;0;L;;;;90;N;;;;;
11EB6;KHOTANESE NUMBER ONE HUNDRED;No;0;L;;;;100;N;;;;;
11EB7;<RESERVED>
11EB8;<RESERVED>
11EB9;KHOTANESE DOUBLE DANDA;Po;0;L;;;;;N;;;;;
11EBA;KHOTANESE PUNCTUATION DOT;Po;0;L;;;;;N;;;;;
11EBB;KHOTANESE PUNCTUATION DOUBLE DOT;Po;0;L;;;;;N;;;;;
```

6   Code charts

	11E6	11E7	11E8	11E9	11EA	11EB
0		𑋀	𑋐	𑋠	𑋰	𑋀
1		𑋁	𑋑	𑋡	𑋱	𑋁
2		𑋂	𑋒	𑋢	𑋲	𑋂
3		𑋃	𑋓	𑋣	𑋳	𑋃
4		𑋄	𑋔	𑋤	𑋴	𑋄
5	𑋅	𑋅	𑋕	𑋥	𑋵	𑋅
6	𑋆	𑋆	𑋖	𑋦	𑋶	𑋆
7	𑋇	𑋇	𑋗	𑋧	𑋷	
8	𑋈	𑋈	𑋘	𑋨	𑋸	
9	𑋉	𑋉	𑋙	𑋩	𑋹	𑋉
A	𑋊	𑋊	𑋚		𑋺	𑋊
B	𑋋	𑋋	𑋛	𑋫	𑋻	𑋋
C		𑋌	𑋜	𑋬	𑋼	
D	𑋍	𑋍	𑋝	𑋭	𑋽	
E	𑋎	𑋎	𑋞	𑋮	𑋾	
F	𑋏	𑋏	𑋟	𑋯	𑋿	

Figure 4: Proposed code chart for Khotanese

Independent vowels

11E65	ℨ	KHOTANESE LETTER A
11E66	℥	KHOTANESE LETTER AA
11E67	℧	KHOTANESE LETTER I
11E68	℧	KHOTANESE LETTER II
11E69	℧	KHOTANESE LETTER U
11E6A	℧	KHOTANESE LETTER UU
11E6B	℧	KHOTANESE LETTER VOCALIC R
11E6C	▨	<RESERVED>
11E6D	℧	KHOTANESE LETTER E
11E6E	℧	KHOTANESE LETTER AI
11E6F	℧	KHOTANESE LETTER O
11E70	℧	KHOTANESE LETTER AU
11E71	℧	KHOTANESE LETTER AE
11E72	℧	KHOTANESE LETTER EI

Consonants

11E73	℥	KHOTANESE LETTER KA
11E74	℧	KHOTANESE LETTER KHA
11E75	℧	KHOTANESE LETTER GA
11E76	℧	KHOTANESE LETTER GHA
11E77	℧	KHOTANESE LETTER NGA
11E78	℧	KHOTANESE LETTER CA
11E79	℧	KHOTANESE LETTER CHA
11E7A	℧	KHOTANESE LETTER JA
11E7B	℧	KHOTANESE LETTER JHA
11E7C	℧	KHOTANESE LETTER NYA
11E7D	℧	KHOTANESE LETTER TTA
11E7E	℧	KHOTANESE LETTER TTHA
11E7F	℧	KHOTANESE LETTER DDA
11E80	℧	KHOTANESE LETTER DDHA
11E81	℧	KHOTANESE LETTER NNA
11E82	℧	KHOTANESE LETTER TA
11E83	℧	KHOTANESE LETTER THA
11E84	℧	KHOTANESE LETTER DA
11E85	℧	KHOTANESE LETTER DHA
11E86	℧	KHOTANESE LETTER NA
11E87	℧	KHOTANESE LETTER PA
11E88	℧	KHOTANESE LETTER PHA
11E89	℧	KHOTANESE LETTER BA
11E8A	℧	KHOTANESE LETTER BHA
11E8B	℧	KHOTANESE LETTER MA
11E8C	℧	KHOTANESE LETTER YA
11E8D	℧	KHOTANESE LETTER RA
11E8E	℧	KHOTANESE LETTER LA
11E8F	℧	KHOTANESE LETTER VA
11E90	℧	KHOTANESE LETTER SHA
11E91	℧	KHOTANESE LETTER SSA
11E92	℧	KHOTANESE LETTER SA
11E93	℧	KHOTANESE LETTER HA

Dependent vowel signs

11E94	℧	KHOTANESE SIGN AA
11E95	℧	KHOTANESE SIGN I
11E96	℧	KHOTANESE SIGN II
11E97	℧	KHOTANESE SIGN U
11E98	℧	KHOTANESE SIGN UU
11E99	℧	KHOTANESE SIGN VOCALIC R
11E9A	▨	<RESERVED>
11E9B	℧	KHOTANESE SIGN E
11E9C	℧	KHOTANESE SIGN AI
11E9D	℧	KHOTANESE SIGN O
11E9E	℧	KHOTANESE SIGN AU
11E9F	℧	KHOTANESE SIGN AE
11EA0	℧	KHOTANESE SIGN EI

Various signs

11EA1	℧	KHOTANESE SIGN HOOK
11EA2	℧	KHOTANESE SIGN ANUSVARA

Virama

11EA3	℧	KHOTANESE VIRAMA

Numbers

11EA4	℧	KHOTANESE NUMBER ONE
11EA5	℧	KHOTANESE NUMBER TWO
11EA6	℧	KHOTANESE NUMBER THREE
11EA7	℧	KHOTANESE NUMBER FOUR
11EA8	℧	KHOTANESE NUMBER FIVE
11EA9	℧	KHOTANESE NUMBER SIX
11EAA	℧	KHOTANESE NUMBER SEVEN
11EAB	℧	KHOTANESE NUMBER EIGHT
11EAC	℧	KHOTANESE NUMBER NINE
11EAD	℧	KHOTANESE NUMBER TEN
11EAE	℧	KHOTANESE NUMBER TWENTY
11EAF	℧	KHOTANESE NUMBER THIRTY
11EB0	℧	KHOTANESE NUMBER FORTY
11FB1	℧	KHOTANESE NUMBER FIFTY
11FB2	℧	KHOTANESE NUMBER SIXTY
11FB3	℧	KHOTANESE NUMBER SEVENTY
11EB4	℧	KHOTANESE NUMBER EIGHTY
11EB5	℧	KHOTANESE NUMBER NINETY
11EB6	℧	KHOTANESE NUMBER ONE HUNDRED
11EB7	▨	<RESERVED>

Punctuation

11EB8	▨	<RESERVED>
11EB9	℧	KHOTANESE DOUBLE DANDA
11EBA	·	KHOTANESE PUNCTUATION DOT
11EBB	℈	KHOTANESE PUNCTUATION DOUBLE DOT

Figure 5: Proposed names list for Khotanese

7　Samples

No.							No.					
1 A	2 'Ā	3 ²Ā	4 I	5 'Ī	6 ²Ī		73 khā	74 'lysā	75 rvā	76 mā	77/78 tā-ndi	
7 U	8 Ū	9 R	10 E	11 O	12 AU		79 sīa	80 hvā	81 jā	82 jyā	83 jā	84 jā
13 ka	14 kha	15 ga	16 gga	17 gha	18 ṅga		85 ttra	86 ²lysā	87 ñi	88 hi	89 'lysi	90 li
19 ca	20 'cha	21 ²cha	22 ja	23 'ña	24 ²ña		91 jsi	92 spri	93 gu	94 sku	95 du	96 ru
25 ta	26 tha	27 ththa	28 da	29 nda	30 na		97 nu	98 pu	99 rru	100 'rru	101 ²rru	102 ru
31 ta	32 tta	33 tha	34 da	35 dha	36 na		103 bu	104 hu	105 ysmu	106 gu	107 ssu	108 su
37 pa	38 pha	39 ba	40 'bha	41 ²bha	42 ma		109 'tu	110 ²ttu	111 'pyu	112 'pyu	113 gru	114 bru
43 ya	44 ra	45 la	46 va	47 100	48 200		115 'stu	116 ²ttu	117 rvr	118 kye	119 klai	120 tcai
49 śa	50 sa	51 sa	52 ha	53 300	54 400		121 ysno	122 khydu	123 lo	124 dam	125 bim	126 ssim
55 1	56 2	57 3	58 10	59 20	60 30		127 rtha	128 dda	129 nna	130 kkra	131 sde	132 tva
61 4	62 5	63 6	64 40	65 50	66 60		133 yyo	134 dye	135 llye	136 'pyu	137 gya	138 jva
67 7	68 8	69 9	70 70	71 80	72 90		134 rydu	140 crra	141 radha	142 rsta	143 drai	144 m

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

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*

9   References

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

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

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.

<table>
<tr><td colspan="2">

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*.**

</td></tr>
</table>

A. Administrative

1. **Title:** *Preliminary Proposal to Encode the Khotanese Script*
2. Requester's name: *Lee Wilson (ttlwilson@hotmail.com)*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *2015-01-26*
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: *Khotanese*
    b. The proposal is for addition of character(s) to an existing block:
        Name of the existing block:
2. Number of characters in proposal: *83*
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.

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: