

Final Proposal to Encode the Khojki Script in ISO/IEC 10646

Anshuman Pandey
Department of History
University of Michigan
Ann Arbor, Michigan, U.S.A.
pandey@umich.edu

January 28, 2011

1 Introduction

This is a proposal to encode the Khojki script in the Universal Character Set (ISO/IEC 10646). It replaces the following documents:

- L2/08-201: “Towards an Encoding for the Khojki Script in ISO/IEC 10646”
- N3596 L2/09-101 “Proposal to Encode the Khojki Script in ISO/IEC 10646”
- N3883 L2/10-326 “Revised Proposal to Encode the Khojki Script in ISO/IEC 10646”

This document is a revision of N3883 L2/10-326. Major changes from the previous proposal include the removal of some vowel and consonant letters, the reassignment of glyphs for some consonant letters, revision of the code chart, and the inclusion of new specimens. The ISO proposal summary form is enclosed.

2 Overview

Khojki is a writing system used by the Nizari Ismaili community of South Asia for recording religious literature. It was developed in Sindh, now in Pakistan, for representing the Sindhi language. The script spread to surrounding regions and was used for writing Gujarati, Punjabi, and Siraiki, as well as several languages allied with Hindi. It was also used for writing Arabic and Persian. The name ‘Khojki’ (ਖੋજکي *khojakī*, Devanagari खोजकी *khojakī*) is derived from the Persian honorific title *خواجه* *khwājah* “master, lord”. It is a translation of the Sanskrit ठाकुर *thākura*, which was used as a title of address by the Lohāṇā community, who were among the early Hindu converts to the Nizari Ismaili tradition in South Asia.¹

Khojki is a Brahmi-based script of the Sindhi branch of the Landa family, which is a class of mercantile scripts related to Sharada.² It is considered to be a refined version of the Lohāṇākī script that was developed as a liturgical script for recording Ismaili literature. Popular Nizari Ismaili tradition states that Khojki was invented and propagated by Pir Sadruddin (Ṣadr al-Dīn), an Ismaili missionary actively working with the Lohāṇā community. Khojki is one of two Landa scripts that were developed into formal liturgical scripts for

¹ Daftary 2008: 26–27.

² See Pandey 2010a and 2011.

use by religious communities; the other is Gurmukhi, which was developed for writing the sacred literature of the Sikh tradition (see tables 2–4 for a comparison). Khojki is also called ‘Sindhi’ and ‘Khwajah Sindhi’.

Khojki was in use by the 16th century CE, as attested by manuscript evidence.³ It was used primarily for recording literature known as *ginān* (𑂔𑂗𑂢𑂰 < Sanskrit ज्ञान *jñāna* ‘knowledge’). These compositions, which were originally transmitted orally, are attributed to Pir Sadruddin, Pir Shamsuddin, and others, and expound upon several mystical, mythological, didactic, cosmological and eschatological themes, and some contain hagiographic accounts of Ismaili saints.⁴ The majority of *ginān* literature exists in manuscript form, but beginning in the 19th century, hundreds of lithographs and books containing these compositions were produced. The printing of Khojki books flourished after Laljibhai Devraj produced metal types for Khojki in Germany for use at his Khoja Sindhi Printing Press in Mumbai.⁵

Khojki materials are archived at several institutions throughout the world. The first collection of Khojki manuscripts belongs to the Ismaili Tariqa and Religious Education Board (formerly the Ismailia Association).⁶ A large collection of manuscripts and books are housed at the Institute of Ismaili Studies in London. Smaller collections are held at Harvard University and the National Archives of Pakistan. Sizeable private collections are also maintained by members of the Ismaili community.

While usage of Khojki has declined over the past century, it is used wherever Nizari Ismaili Muslims of South Asian origin reside. The largest communities are found in Pakistan, India, Canada, United States, United Kingdom, Kenya, Tanzania, and Uganda. Khojki primers continue to be published in Pakistan for teaching the script. Khojki manuscripts and books are used in Ismaili ceremonies not only in South Asia, but in east and south Africa, where large diaspora communities had formed by the 19th century. The script was also used by communities related to the Nizari Ismailis, such as the Imamshahis of Gujarat. Moreover, there are specialists around the globe who study the script, as well as Ismaili history, literature, and culture. An encoding for Khojki in the UCS will provide a means for the Ismaili community and specialists to preserve literature in the script and to advance the study of Ismaili and South Asian religious and literary culture.

3 Proposed Characters

The proposed character repertoire and names list for Khojki is given in Figure 1 and Figure 2. The script is currently allocated to the Supplementary Multilingual Plane (SMP) at the range U+11200..U+1124F.

3.1 Establishing the Character Repertoire

The proposal aims to provide a character set that may be used to encode any Khojki document. This task is somewhat difficult because of the absence of a standard for Khojki and the differences in the historical and modern usage and meaning of several characters. The character repertoire and glyph assignments proposed here are based upon an analysis of Khojki characters presented in primary sources, such as Tajddin (1989), and secondary sources, such as Allana (1969), Asani (1992), Grierson (1905), Shackle and Moir (1992), and Stack (1849). Also used are script charts from various other sources.

The presence of characters with which readers of Khojki may not be familiar should not distract, but rather should be understood as an element of the script that at some historical period and within a particular region was used in a Khojki document, and therefore, as belonging to the broader Khojki repertoire.

³ Daftary 2008: 24.









⁴ Daftary 2008: 23.

⁵ Tajddin 2003.

⁶ Asani 2002: 130.



3.2 Issues With Character-Glyph Assignments

There are many issues related to the assignment of glyphs for certain consonant letters for Khojki.⁷ One is the use of a single glyph for representing multiple phonemes. A reason for such ambiguity is the use of Khojki outside of Sindh by scribes unfamiliar with the Sindhi language. When non-Sindhi speakers wrote Khojki, certain characters that represented sounds specific to Sindhi, such as the implosive consonant letters (GGA, JJA, DDDA, BBA) lost their original phonetic values. As a result of this neutralization, the original glyphs for these letters were reassigned to other consonants within the same articulation class.

The phonetic neutralization and shifts in glyph assignments is evidenced in charts for Khojki published over the years. For example, as shown in Figure 8, the glyph  was used originally for /b/,  for /b̥/, and  for /b̄/. However, as illustrated in the Khojki primer by Tajddin (1989), given in Figure 9, the Sindhi implosive /b̥/ ceased to be representing using a unique letter and the original glyph  assigned to it became conflated with /b/. In turn,  was dissociated from its original value /b/ and was used for writing /b̄/, which was already written as . Glyphs assignments for other groups of consonants letters, such as  DA and  DDDA also shifted in a similar way. These changes were part of a reform of Khojki that took place in the 1930s.⁸

The difficulty of developing a standard for Khojki in the UCS lies in determining if the character-glyph assignments should be made according to historical or modern usage. Historical usage offers a means for resolving the ambiguous assignments of glyphs to characters by reverting to the original Khojki repertoire. On the other hand, the modern assignments may be more familiar to current readers of Khojki and are used in primers that are currently in circulation, as exemplified by the specimen in Figure 9.

The historical Khojki repertoire offers the most practical means for representing the broadest range of Khojki records. Of particular importance is the unique representation of characters used for the Sindhi implosive consonants, which as noted above, are not accommodated with uniformity in several modern Khojki materials. The glyphs and names for these characters are assigned based upon the phonetic values originally used in Sindh, since it was in areas outside of Sindh where these letters were neutralized. A list of the different attested uses of particular glyphs is documented in Section 4.5.

Given the multiple ways in which certain Khojki letters have been used historically, the proposed glyph assignments and character names should be viewed simply as a practical means to standardize the technical representation of Khojki. That the glyph  is assigned the named BBA instead of BA, or that  is named JJA instead of ZA, does not restrict their usage to any particular phoneme. This approach to the Khojki repertoire permits each attested Khojki character, whether historical or modern, to be uniquely identified in the UCS.

3.3 Source for Character Glyphs

The font used here was originally developed by Pyarali Jiwa and is titled ‘Khojki Jiwa’. Several glyphs in the font have been revised and new ones added by the present author.

The glyphs for the digitized Khojki font are based upon the metal fonts developed by Laljibhai Devraj in 1903. Devraj produced the first printed editions of *ginān*-s and other Ismaili literature at his Khoja Sindhi Printing Press in Bombay. Although there were several Khojki printers in Sindh, Gujarat, and Bombay, the majority of books were printed by Devraj, and his typefaces is universally recognized by readers of Khojki.

Glyphs for characters that were not originally part of Devraj’s fonts have been designed with the spirit of Devraj’s style. It is hoped this faithfulness to Devraj’s work will mirror the ‘standard’ appearance of Khojki to which the Nizari Ismaili community is accustomed.

⁷ See Asani (1987: 447–448) for details.

⁸ Shafique Virani, personal correspondence, February 2009.

4 Implementation Details

4.1 Structure

The general structure (phonetic order, *mātrā* reordering, use of *virāma*, etc.) of Khojki is similar to that of other Indic scripts based upon the Brahmi model. It is written from left-to-right.

4.2 Virāma

The Khojki ◌/ VIRAMA is identical in function to corresponding characters in other Indic scripts. It is written to the right of a consonant letter. The form ◌̣ is attested in a few printed books, eg. *Sindhī pahelī copadī* (Bombay: Khoja Sindhi Printing Press, 1910), but it is an attempt to mimic the form of the sign in Devanagari. In some texts, VIRAMA may be represented as ◌̣̣, but this practice was very limited.⁹

4.3 Vowels

Khojki has a smaller repertoire of independent vowel letters than other Brahmi-based scripts. The letters 𑂔 I and 𑂕 U are used for writing both short and long forms of *i* and *u*, respectively.

The letters 𑂔̣ AI and 𑂕̣ AU represent diphthongs. Although they are attested in manuscripts and books, Khojki originally did not have unique letters for these vowels. In early Khojki records, diphthongs are generally represented as diagraphs, eg. <𑂔 A + 𑂕 E> → 𑂔𑂕 AI; <𑂔 A + 𑂕> U → 𑂔𑂕 AU.

Several variant forms of vowel letters are attested: 𑂔̣ = 𑂕 E; 𑂔̣̣ and 𑂕̣̣ = 𑂕 O. Although several charts show unique letters for *II and *UU, the actual use of such forms is rarer and the characters are not proposed for encoding (see Section 5 for more details).

Some independent vowel letters may be represented using a combination of a base vowel letter and a dependent vowel sign. This practice is not recommended. Independent characters should be used instead:

RECOMMENDED	NOT RECOMMENDED
𑂔 VOWEL LETTER AA	𑂔 VOWEL LETTER A + ◌ VOWEL SIGN AA
𑂔̣ VOWEL LETTER AI	𑂔 VOWEL LETTER A + ◌̣ VOWEL SIGN AI
𑂔̣̣ VOWEL LETTER AU	𑂔 VOWEL LETTER A + ◌̣̣ VOWEL SIGN AU

If a variant form of a vowel letter must be represented, it is advised that it be managed at the font level through the creation of a new glyph for the letter, rather than producing by using encoded forms as character primitives. For example, the 𑂔̣̣ variant of 𑂕 O should not be represented as <𑂔 A + ◌̣̣ VOWEL SIGN O>.

4.4 Vowel Signs

The repertoire of dependent vowel signs is larger than that of independent vowel letters. There are separate signs for I and II, but no form for *UU, the single sign ◌̣ being used for both short and long forms. The vowel signs I, II, and U have contextual variant forms (see Section 4.6 for details).

The ◌̣ VOWEL SIGN O is often written by placing the ◌̣ element above the consonant letter, eg. 𑂔̣̣.

⁹ Asani 1987: 449.

4.5 Consonant Letters

As discussed in Section 3.2, there is considerable variation in character-glyph assignments for consonant letters. The table below shows the proposed assignment of glyphs to characters, along with other phonemes the glyph is attested as representing:

GLYPH	ARABIC EQUIVALENT	ORIGINAL PHONEME	CHARACTER NAME	OTHER USES
𑌒	گ	/g/	GA	/g/, /g ^h /
𑌓	گب	/g/	GGA	/g/, /gr/
𑌔	گه	/g ^h /	GHA	/g/, /g/
𑌕	ج	/dʒ/	JA	/f/, /dʒ ^h /
𑌖	جج	/f/	JJA	/dʒ/, /dʒ ^h /, /z/
𑌗	د	/d/	DDA	/n/, /t/
𑌘	ڌ	/n/	NNA	/d/, /t/
𑌙	د	/d/	DA	/d ^h /
𑌚	ڌڌ	/d/	DDDA	/d/, /d ^h /
𑌛	ڌڌڌ	/d ^h /	DHA	/d/, /dʒ ^h /
𑌜	ب	/b/	BA	/b/, /b ^h /
𑌝	بب	/b/	BBA	/b/, /b ^h /
𑌞	بي	/b ^h /	BHA	/b/
𑌟	ڌڌڌڌ	/t/	LLA	/l/

Several consonant letters have attested variant forms. These forms are shown in figures 16–19.

4.6 Consonant-Vowel Combinations

Several combinations of consonant letters and vowel signs are represented as ligatures. Some vowel signs take contextual forms when combined with certain consonant letters. These behaviors are observed with the following vowel signs:

- ◌f VOWEL SIGN I** This is written to the right of the consonant. It has the contextual form ◌[̣]. The regular sign ◌f is written with characters that have right descenders: 𑌒 CA → 𑌒[̣] ci, 𑌖 VA → 𑌖[̣] vi, etc.; and those that possess a final right stroke that terminates at the baseline: 𑌔 KA → 𑌔[̣] ki, 𑌙 DA → 𑌙[̣] di, 𑌛 HA → 𑌛[̣] hi. The variant form ◌[̣] is used with characters that have rounded right edges: 𑌗 CHA → 𑌗[̣] chi, 𑌚 NGA → 𑌚[̣] ni, 𑌞 TTA → 𑌞[̣] ti, 𑌘 DDDA → 𑌘[̣] di, 𑌜 RA → 𑌜[̣] ri.
- ◌l VOWEL SIGN II** This sign is written to the right of the consonant letter. It has the contextual form ◌[̣]l. The regular sign ◌l is used with characters that have rounded right edges: 𑌗 CHA → 𑌗[̣]l chī. The variant form ◌[̣]l is written with characters possessing right descenders: 𑌒 CA → 𑌒[̣]l cī, 𑌖 NA → 𑌖[̣]l nī, etc., and with characters with final strokes that terminate at the baseline: 𑌔 KA → 𑌔[̣]l kī, 𑌛 PHA → 𑌛[̣]l phī, 𑌛 HA → 𑌛[̣]l hī.
- ◌u VOWEL SIGN U** This sign is generally written as an extension of the basic final stroke of a character: 𑌔 KA → 𑌔[̣]u ku, 𑌗 CHA → 𑌗[̣]u chu, etc. With characters that possess a right descender, it is written at

the terminus of the descender: **𑂔** CA → **𑂕** *cu*, **𑂖** GHA → **𑂗** *ghu*, **𑂘** YA → **𑂙** *yu*, etc. It attaches to the bottom of the glyph for characters with rounded shapes: **𑂚** TTHA → **𑂛** *thu*.

4.7 Consonant Conjuncts

Consonant clusters are typically limited to two consonants and are represented in three ways.

1. **Atomic Ligatures** There are four consonant conjuncts that are written in Khojki as atomic ligatures. They are represented as $\langle C_1 + \text{VIRAMA} + C_2 \rangle$:

𑂛 *kṣa* \langle **𑂔** KA + **𑂏** VIRAMA + **𑂖** SA + **𑂐** NUKTA \rangle

𑂛 *jña* \langle **𑂔** JA + **𑂏** VIRAMA + **𑂚** NYA \rangle

𑂛 *tra* \langle **𑂔** TA + **𑂏** VIRAMA + **𑂖** RA \rangle

𑂛 *dra* \langle **𑂔** DA + **𑂏** VIRAMA + **𑂖** RA \rangle

In most Indic scripts, the letter **𑂛** represents a consonant conjunct produced with the sequence \langle KA + VIRAMA + SSA \rangle , eg. Devanagari \langle क + ष + ष \rangle → क्ष. However, Khojki does not have a character SSA analogous to ष U+0937 DEVANAGARI CHARACTER SSA. The absence of SSA in Khojki requires an alternate method of producing **𑂛**. Such circumstances might suggest that **𑂛** be encoded as an atomic character KSSA. However, it is recommended that the glyph be treated as a ligature. It is to be represented as shown above. The sequence **𑂖** SA + **𑂐** NUKTA is used for representing sibilant sounds such as /ʃ/ *śa* (SHA), and may theoretically be used for representing /ʒ/ *ṣa* (SSA).

2. **Full-Form of Consonant + Visible Virama** This is the most common method for writing consonant conjuncts, eg. **𑂔/𑂏** *kya*. Such conjuncts are represented as $\langle C_1 + \text{VIRAMA} + \text{𑂏} \text{ZWNJ} + C_2 \rangle$.
3. **Full-Form of Consonant + Reduced Form of Final Consonant** This type of conjunct is represented as $\langle C_1 + \text{𑂏} \text{ZWJ} + \text{VIRAMA} + C_2 \rangle$. It is typically used for writing clusters involving **𑂖** RA as the second element:

Ms Ism K 15

[190-?]

Local paper; 44 folios; 65; 17 x 14 (13 x 10); 11 lines; multiple sections bound with two sets of sewing: side-sewn and stab-sewn with cords pierced through and laced into laminated layers forming thin paper covers; spine rounded and backed; black ink.

𑂛 = ḍ, ṛ, ṇ; **𑂛** = dh; special form of conjuncts: **𑂛** = br;
𑂛 = kr; **𑂛** = sr; **𑂛** = mr; **𑂛** = ny; **𑂛** = ky; **𑂛** = ly; **𑂛** = ty;
𑂛 = jy; **𑂛** = sv; **𑂛** = gn.

The conjunct /gr/ is represented as **𑂛** (GA + VIRAMA + RA); this is the same glyph for GGA.

In some cases, **𑂘** YA also appears as the second element in a conjunct using the reduced form **𑂛**, eg. **𑂔/𑂛** *kya*.

In Khojki, NUKTA is used primarily to represent Arabic letters and sounds. This is done by writing NUKTA above a Khojki character whose phonetic value most closely matches that of the foreign sound being accommodated. The list of attested NUKTA letters in Khojki is:

	KHOJKI		ARABIC
‘	ا̇	VOWEL SIGN A + NUKTA	ع U+0639 ARABIC LETTER AIN
‘	آ̇	LETTER A + NUKTA	ع U+0639 ARABIC LETTER AIN
‘	آ̇	LETTER E + NUKTA	ع U+0639 ARABIC LETTER AIN
qa	ق̇	KA + NUKTA	ق U+0642 ARABIC LETTER QAF
<i>kha</i>	خ̇	KHA + NUKTA	خ U+062E ARABIC LETTER KHAH
ga	غ̇	GA + NUKTA	غ U+063A ARABIC LETTER GHAIN
za	ذ̇	JA + NUKTA	ذ U+0630 ARABIC LETTER THAL
za	ز̇	JA + NUKTA	ز U+0632 ARABIC LETTER ZAIN
za	ض̇	JA + NUKTA	ض U+0636 ARABIC LETTER DAD
za	ظ̇	JA + NUKTA	ظ U+0638 ARABIC LETTER ZAH
ta	ط̇	TA + NUKTA	ط U+0637 ARABIC LETTER TAH
fa	ف̇	PHA + NUKTA	ف U+0641 ARABIC LETTER FEH
sa	ش̇	SA + NUKTA	ش U+0634 ARABIC LETTER SHEEN
sa	ص̇	SA + NUKTA	ص U+0635 ARABIC LETTER SAD
ha	ح̇	HA + NUKTA	ح U+062D ARABIC LETTER HAH

In some cases, the NUKTA is used for marking abbreviations for common Arabic benedictions:

1. ا̇ = عليه السلام ‘alai-hi as-salām ‘peace be upon him’.
2. ﷺ̇ = صلى الله عليه و سلم و sallā allāhu ‘alai-hi wa sallam ‘God bless him and grant him salvation’. It corresponds to ﷺ̇ U+FDFA ARABIC LIGATURE SALLALLAHOU ALAYHE WASALLAM.

Ms Ism K 13

[190-?]

Foreign paper, watermark: made in England for Shaik Ahmad Shaik Dawood of Bombay; 190 folios; 376; 21 x 17 (16 x 13); 12 - 13 lines; multiple sections, stab-sewn; binding in poor condition; covers missing; black ink.

Same scribe as Ms Ism K 6, Ms Ism K 8, Ms Ism K 11 and Ms Ism K 12; ا̇ = ū, o; آ̇ = z; ز̇ = j; ّ = shaddah; ا̇ = abbreviation for Arabic benediction ‘alai-hi as-salam “peace be upon him”; ﷺ̇ = abbreviation for Arabic benediction sallā allāhu ‘alai-hi wa sallam “God bless him and grant him salvation.”

When NUKTA occurs with other combining signs, the positioning of the sign changes. If it occurs with َ VOWEL SIGN E and ِ VOWEL SIGN AI, the NUKTA is written to the right of the vowel sign. If NUKTA occurs with SHADDA, the two signs are stacked vertically with NUKTA above SHADDA.

4.12 Abbreviations

The ◦ ABBREVIATION SIGN is used for marking abbreviations. In the example below, the abbreviations indicate verse repetition:

જાલ જોખલ જાણેા જાવજો,
 ઢવે જાજો ડજેણું મજાપ... ...જાજે૫૦ (૧૦)
 પજા જાલ જુગજાં જબજાં ડપે ઢું જાણ,
 જાજા જોખ બ જાપ;
 જાજા ડજો જાણેા જાવજા,
 જાજે૫ જાણેા જાહુણ... ...જાજે૫૦ (૧૧)

4.13 Verse Numbering

Verse numbers are often marked in Khojki texts using an overline. The overline spans all of the digits in a number and indicates that they belong to a set. This overline should be managed at the presentation level. The use of ◌ U+0305 COMBINING OVERLINE is not recommended because while it visually represents a ‘verse marker’, it does not possess any semantic properties of such a marker.

4.14 Digits

Khojki digits generally resemble Gujarati forms — ૦ ૧ ૨ ૩ ૪ ૫ ૬ ૭ ૮ ૯ — and may be unified with Gujarati digits (U+0AE6..U+0AEF). See Figure 6 for an example of Gujarati digits in printed Khojki text.

4.15 Number Forms and Unit Marks

Number forms and unit marks are found in Khojki documents (see Figure 3). The most common characters are signs for writing fractions: eg. ૮ U+A833 NORTH INDIC FRACTION ONE SIXTEENTH, ॥ U+A831 NORTH INDIC FRACTION ONE HALF, etc.; placeholder marks: eg. ◦ U+A836 NORTH INDIC QUARTER MARK; and currency marks: eg. ે U+A838 NORTH INDIC RUPEE MARK.

In the specimen below, the parenthesis-like character represents ે U+A838 NORTH INDIC RUPEE MARK. The number form represents the currency value of ‘8 ānā’.



These characters may be represented using those already encoded in the ‘Common Indic Number Forms’ block (U+A830..U+A83F).¹⁰

4.16 Collation

The collating order for Khojki is as follows:

¹⁰ See Pandey (2007) for more information on these characters.

𑂀 A > 𑂁 AA > 𑂂 I > 𑂃 U > 𑂄 E > 𑂅 AI > 𑂆 O > 𑂇 AU > 𑂈 KA > 𑂉 KHA > 𑂊 GA > 𑂋 GGA >
 𑂌 GHA > 𑂍 NGA > 𑂎 CA > 𑂏 CHA > 𑂐 JA > 𑂑 JJA > 𑂒 NYA > 𑂓 TTA > 𑂔 TTHA > 𑂕 DDA >
 𑂖 DDHA > 𑂗 NNA > 𑂘 TA > 𑂙 THA > 𑂚 DA > 𑂛 DDDA > 𑂜 DHA > 𑂝 NA > 𑂞 PA > 𑂟 PHA >
 𑂠 BA > 𑂡 BBA > 𑂢 BHA > 𑂣 MA > 𑂤 YA > 𑂥 RA > 𑂦 LA > 𑂧 VA > 𑂨 SA > 𑂩 HA > 𑂪 LLA >
 𑂫 SIGN AA > 𑂬 SIGN I > 𑂭 SIGN II > 𑂮 SIGN U > 𑂯 SIGN E > 𑂰 SIGN AI > 𑂱 SIGN O >
 𑂲 SIGN AU > 𑂳 ANUSVARA > 𑂴 VIRAMA

Combinations of letters with 𑂳 NUKTA and 𑂴 SHADDA are sorted with the base letter.

4.17 Linebreaking

Khojki has the following linebreaking requirements:

- All vowel letters, vowel signs, and consonant letters behave as in Devanagari.
- The VIRAMA, ANUSVARA, and NUKTA behave as in Devanagari.
- The SHADDA behaves like U+0A71 GURMUKHI SIGN ADDAK.
- The DANDA and DOUBLE DANDA behave like the corresponding characters in Devanagari.
- The SECTION MARK and DOUBLE SECTION MARK behave like the corresponding characters in Kaithi.
- The WORD SEPARATOR may appear at the beginning and end of a line.
- The ABBREVIATION SIGN may not appear at the beginning of a line.

4.18 Character Properties

The properties for Khojki characters in the Unicode Character Database format are:

```

11200;KHOJKI LETTER A;Lo;0;L;;;;;N;;;;;
11201;KHOJKI LETTER AA;Lo;0;L;;;;;N;;;;;
11202;KHOJKI LETTER I;Lo;0;L;;;;;N;;;;;
11203;KHOJKI LETTER U;Lo;0;L;;;;;N;;;;;
11204;KHOJKI LETTER E;Lo;0;L;;;;;N;;;;;
11205;KHOJKI LETTER AI;Lo;0;L;;;;;N;;;;;
11206;KHOJKI LETTER O;Lo;0;L;;;;;N;;;;;
11207;KHOJKI LETTER AU;Lo;0;L;;;;;N;;;;;
11208;KHOJKI LETTER KA;Lo;0;L;;;;;N;;;;;
11209;KHOJKI LETTER KHA;Lo;0;L;;;;;N;;;;;
1120A;KHOJKI LETTER GA;Lo;0;L;;;;;N;;;;;
1120B;KHOJKI LETTER GGA;Lo;0;L;;;;;N;;;;;
1120C;KHOJKI LETTER GHA;Lo;0;L;;;;;N;;;;;
1120D;KHOJKI LETTER NGA;Lo;0;L;;;;;N;;;;;
1120E;KHOJKI LETTER CA;Lo;0;L;;;;;N;;;;;
1120F;KHOJKI LETTER CHA;Lo;0;L;;;;;N;;;;;
11210;KHOJKI LETTER JA;Lo;0;L;;;;;N;;;;;
11211;KHOJKI LETTER JJA;Lo;0;L;;;;;N;;;;;
11212;<reserved>
11213;KHOJKI LETTER NYA;Lo;0;L;;;;;N;;;;;
11214;KHOJKI LETTER TTA;Lo;0;L;;;;;N;;;;;
11215;KHOJKI LETTER TTHA;Lo;0;L;;;;;N;;;;;
11216;KHOJKI LETTER DDA;Lo;0;L;;;;;N;;;;;
11217;KHOJKI LETTER DDHA;Lo;0;L;;;;;N;;;;;
11218;KHOJKI LETTER NNA;Lo;0;L;;;;;N;;;;;
11219;KHOJKI LETTER TA;Lo;0;L;;;;;N;;;;;
1121A;KHOJKI LETTER THA;Lo;0;L;;;;;N;;;;;
1121B;KHOJKI LETTER DA;Lo;0;L;;;;;N;;;;;
1121C;KHOJKI LETTER DDDA;Lo;0;L;;;;;N;;;;;
1121D;KHOJKI LETTER DHA;Lo;0;L;;;;;N;;;;;

```

```

1121E;KHOJKI LETTER NA;Lo;0;L;;;;;N;;;;;
1122F;KHOJKI LETTER PA;Lo;0;L;;;;;N;;;;;
11220;KHOJKI LETTER PHA;Lo;0;L;;;;;N;;;;;
11221;KHOJKI LETTER BA;Lo;0;L;;;;;N;;;;;
11222;KHOJKI LETTER BBA;Lo;0;L;;;;;N;;;;;
11223;KHOJKI LETTER BHA;Lo;0;L;;;;;N;;;;;
11224;KHOJKI LETTER MA;Lo;0;L;;;;;N;;;;;
11225;KHOJKI LETTER YA;Lo;0;L;;;;;N;;;;;
11226;KHOJKI LETTER RA;Lo;0;L;;;;;N;;;;;
11227;KHOJKI LETTER LA;Lo;0;L;;;;;N;;;;;
11228;KHOJKI LETTER VA;Lo;0;L;;;;;N;;;;;
11229;KHOJKI LETTER SA;Lo;0;L;;;;;N;;;;;
1122A;KHOJKI LETTER HA;Lo;0;L;;;;;N;;;;;
1122B;KHOJKI LETTER LLA;Lo;0;L;;;;;N;;;;;
1122C;KHOJKI VOWEL SIGN AA;Mn;0;NSM;;;;;N;;;;;
1122D;KHOJKI VOWEL SIGN I;Mc;0;L;;;;;N;;;;;
1122E;KHOJKI VOWEL SIGN II;Mc;0;L;;;;;N;;;;;
1122F;KHOJKI VOWEL SIGN U;Mn;0;NSM;;;;;N;;;;;
11230;KHOJKI VOWEL SIGN E;Mn;0;NSM;;;;;N;;;;;
11231;KHOJKI VOWEL SIGN AI;Mn;0;NSM;;;;;N;;;;;
11232;KHOJKI VOWEL SIGN O;Mn;0;L;;;;;N;;;;;
11233;KHOJKI VOWEL SIGN AU;Mn;0;L;;;;;N;;;;;
11234;KHOJKI SIGN ANUSVARA;Mn;0;NSM;;;;;N;;;;;
11235;KHOJKI SIGN VIRAMA;Mc;9;NSM;;;;;N;;;;;
11236;KHOJKI SIGN NUKTA;Mn;7;NSM;;;;;N;;;;;
11237;KHOJKI SIGN SHADDA;Mn;0;NSM;;;;;N;;;;;
11238;KHOJKI DANDA;Po;0;L;;;;;N;;;;;
11239;KHOJKI DOUBLE DANDA;Po;0;L;;;;;N;;;;;
1123A;KHOJKI WORD SEPARATOR;Po;0;L;;;;;N;;;;;
1123B;KHOJKI SECTION MARK;Po;0;L;;;;;N;;;;;
1123C;KHOJKI DOUBLE SECTION MARK;Po;0;L;;;;;N;;;;;
1122D;KHOJKI ABBREVIATION SIGN;Po;0;L;;;;;N;;;;;

```

4.19 ‘Confusable’ Characters

Below is a list of Khojki characters that are ‘confusable’ with those in other script blocks in the UCS:

```

1120E KHOJKI LETTER CA ; 091C DEVANAGARI LETTER JA
11223 KHOJKI LETTER BHA ; 091E DEVANAGARI LETTER NYA
1123F KHOJKI WORD SEPARATOR ; 0903 DEVANAGARI SIGN VISARGA

```

5 Characters Not Proposed

The proposed repertoire does not include some characters that are shown in some of the above sources, such as *II, *UU, *JHA. The decision not to encode these characters is based upon the lack of evidence that attests their use. There is sufficient space in the Khojki block for including these characters should additional research suggest the need to encode them.

Although the character set proposed for Khojki is an idealized inventory intended to enable the complete encoding of Khojki documents, there are some characters that appear in Khojki script charts, but are not proposed for encoding at present for one or more of the following reasons: (a) insufficient information regarding the characters and their properties; (b) the possibility of representing a character with another of similar or equal function; (c) unattested use outside of a script chart.

1. **II** Independent forms of I and II are generally not distinguished and both are represented by 𑀓 I. However, Figure 14 shows the glyph 𑀔 for II. This glyph may be a variant of 𑀓 I or 𑀕 YA.
2. **UU** Independent forms of U and UU are generally not distinguished and both are represented by 𑀖 U. However, Figure 14 shows the glyph 𑀗 used for UU. This glyph may be a variant form of 𑀖 U or 𑀘 O.
3. **JHA** Several sources show /dʒʰ/ represented using the seemingly unique letter 𑀙, cf. Allana (1969), Asani (1992). This letter may simply be a form of 𑀚 JA. In fact, JA and 𑀛 JJA are attested for writing /dʒʰ/. Space has been reserved for JHA should additional research prove that a unique letter exists.
4. **SHA** The glyph 𑀜 is used for writing SHA, but it is also used for representing any non-native Sindhi sibilant, eg. 𑀝 U+0635 ARABIC LETTER SAD. It may be composed using the sequence 𑀞 SA + 𑀟 NUKTA.
5. **QA** Some sources show the character 𑀠 used for representing /q/ (ق U+0642 ARABIC LETTER QAF):

P Ism K 14

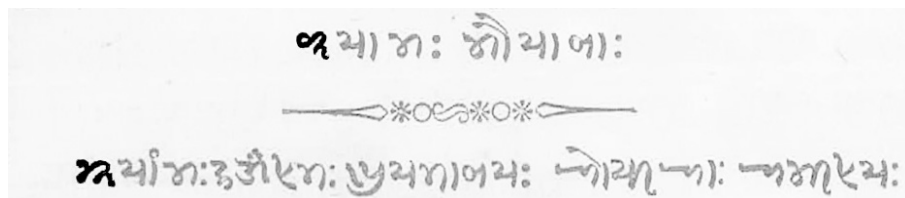
Bombay 1962 s./1905

Elam sār saṅgrah rāg mālā

Publisher: Lālājībhaī Devrāj, Khojā Siṅdhī Čhāpākhānū; 147 pp. in three paginations; 23 x 16; hardboard covers, binding in poor condition; pages in fair condition; frontispiece: full portrait of Aga Khan III in official regalia; preface on the necessity of publishing this book (pp. 1.1-3); tavsīlo (index): pp. 2.135-138; list of publications of the Khojā Siṅdhī Printing Press pp. 3.1-6.

𑀠 = gāḥ; 𑀡 = g, gr; 𑀢 = ain; 𑀣 = 1?; 𑀤 = vowel-less consonant.

The following specimen from a printed book shows KA represented with both 𑀥 and 𑀦. Both glyphs are used to write the word *kalāma* (𑀥𑀦𑀧 / 𑀨𑀦𑀧) and are semantically identical:



Below, the glyph 𑀦 is shown as a variant form of KA:

P Ism K 11

Bombay 1959 s./1902

Ārtī:dhūā:vakhat:trejī:sāṅjījā:śogaḍiā:5

Publisher: Khojā Lālājībhaī Devrāj, Khojā Siṅdhī Čhāpākhānū for the Imāmī Ismāīlī school (nisāl). 64 pp.; 16 x 12; covers missing; binding in poor condition; tafsīlo (index): pp. 63-64; reverse side of title page includes list of sale publications of the Khojā Siṅdhī Printing Press and an announcement of the publication of other works.

𑀦 = z; 𑀧 = j; 𑀨 = k; 𑀩 = ain; 𑀪 = g; 𑀫 = gain; shaddah for doubled consonant or emphasis.

6. **Characters for kṣa, jña, tra, dra** Khojki charts traditionally show the characters 𑀬 *kṣa*, 𑀭 *jña*, 𑀮 *tra*, and 𑀯 *dra* as part of the alphabetic repertoire. The inclusion of these four characters in traditional character charts for Khojki is analogous to the inclusion of the conjuncts क्ष *kṣa*, त्र *tra*, and ज्ञ *jña* in

- Khakee, Gulshan. 1981. “The Dasa Avatara of Pir Shams as Linguistic and Literary Evidence of the Early Development of Ismailism in Sind.” In *Sind Through the Centuries*. Proceedings of an International Seminar held in Karachi in Spring 1975 by the Department of Culture, Government of Sind. Pages 143–155. Edited by Hamida Khuhro. Karachi: Oxford University Press.
- Pandey, Anshuman. 2007. “Proposal to Encode North Indic Number Forms in ISO/IEC 10646” ISO/IEC JTC1/SC2/WG2 N3367 L2/07-354. October 7, 2007. <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3367.pdf>
- . 2009. “Proposal to Encode the Khojki Script in ISO/IEC 10646” N3596 L2/09-101. March 25, 2009. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3596.pdf>
- . 2010a. “A Roadmap for Scripts of the Landa Family” N3766 L2/10-011R. February 9, 2010. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3766.pdf>
- . 2010b. “Revised Proposal to Encode the Khojki Script in ISO/IEC 10646” N3883 L2/10-326. September 12, 2010. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3883.pdf>
- Shackle, Christopher and Zawahir Moir. 1992. *Ismaili Hymns from South Asia: An Introduction to the Ginans*. London: School of Oriental and African Studies, University of London.
- Tajddin Sadik Ali, Mumtaz Ali. 1989. *Khojki Self-Instructor*. Karachi
- . 2003. *101 Ismaili Heroes*. Vol. 1. Karachi: Islamic Book Publisher. <http://www.ismaili.net/Source/mumtaz/Heroes1/hero069.html>
- Virani, Shafique N. 2008. “Idealized Khojki Transliteration Table”. Unpublished.

7 Acknowledgments

I am indebted to Shafique Virani (University of Toronto) for providing Khojki specimens and for offering detailed comments on the forms of Khojki characters and on drafts of this proposal.

I am grateful to Irfan Gowani (Microsoft) for taking an interest in this project and for contacting specialists to provide comments on the proposal.

I would like to thank Amin Khetani for providing me with a copy of Tajddin’s *Khojki Self-Instructor*.

This project was made possible in part by a grant from the United States National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

<i>ka</i>	𑂔	KHOJKI LETTER KA	ڪ	U+06AA ARABIC LETTER SWASH KAF
<i>kha</i>	𑂕	KHOJKI LETTER KHA	ڪ	U+06A9 ARABIC LETTER KAF
<i>ga</i>	𑂖	KHOJKI LETTER GA	گ	U+06AF ARABIC LETTER GAF
<i>gā</i>	𑂗	KHOJKI LETTER GGA	گ	U+06B3 ARABIC LETTER GUEH
<i>gha</i>	𑂘	KHOJKI LETTER GHA	گھ	U+06AF ARABIC LETTER GAF + U+06BE HEH DO CHASHMEE
<i>na</i>	𑂙	KHOJKI LETTER NGA	نگ	U+06B1 ARABIC LETTER NGOEH
<i>ca</i>	𑂚	KHOJKI LETTER CA	چ	U+0686 ARABIC LETTER TCHEH
<i>cha</i>	𑂛	KHOJKI LETTER CHA	چھ	U+0687 ARABIC LETTER TCHEHEH
<i>ja</i>	𑂜	KHOJKI LETTER JA	ج	U+062C ARABIC LETTER JEEM
<i>jā</i>	𑂝	KHOJKI LETTER JJA	جھ	U+0684 ARABIC LETTER DYEHE
<i>ñā</i>	𑂞	KHOJKI LETTER NYA	جھ	U+0683 ARABIC LETTER NYEH
<i>ṭa</i>	𑂟	KHOJKI LETTER TTA	ٺ	U+067D ARABIC LETTER TEH WITH THREE DOTS ABOVE DOWNWARDS
<i>ṭha</i>	𑂠	KHOJKI LETTER TTHA	ٺھ	U+067A ARABIC LETTER TTEHEH
<i>ḍa</i>	𑂡	KHOJKI LETTER DDA	ڊ	U+068A ARABIC LETTER DAL WITH DOT BELOW
<i>ḍha</i>	𑂢	KHOJKI LETTER DDHA	ڊھ	U+068D ARABIC LETTER DDAHAL
<i>ṇa</i>	𑂣	KHOJKI LETTER NNA	ڻ	U+06BB ARABIC LETTER RNOON
<i>ta</i>	𑂤	KHOJKI LETTER TA	ت	U+062A ARABIC LETTER TEH
<i>tha</i>	𑂥	KHOJKI LETTER THA	تھ	U+067F ARABIC LETTER TEHEH
<i>da</i>	𑂦	KHOJKI LETTER DA	د	U+062F ARABIC LETTER DAL
<i>ḍa</i>	𑂧	KHOJKI LETTER DDA	ڏ	U+068F ARABIC LETTER DAL WITH THREE DOTS ABOVE DOWNWARDS
<i>dha</i>	𑂨	KHOJKI LETTER DHA	ڏھ	U+068C ARABIC LETTER DAHAL
<i>na</i>	𑂩	KHOJKI LETTER NA	ن	U+0646 ARABIC LETTER NOON
<i>pa</i>	𑂪	KHOJKI LETTER PA	پ	U+067E ARABIC LETTER PEH
<i>pha</i>	𑂫	KHOJKI LETTER PHA	پھ	U+06A6 ARABIC LETTER PEHEH
<i>ba</i>	𑂬	KHOJKI LETTER BA	ب	U+0628 ARABIC LETTER BEH
<i>ḥa</i>	𑂭	KHOJKI LETTER BBA	بھ	U+067B ARABIC LETTER BEEH
<i>bha</i>	𑂮	KHOJKI LETTER BHA	پھ	U+0680 ARABIC LETTER BEHEH
<i>ma</i>	𑂯	KHOJKI LETTER MA	م	U+0645 ARABIC LETTER MEEM
<i>ya</i>	𑂰	KHOJKI LETTER YA	ي	U+064A ARABIC LETTER YEH
<i>ra</i>	𑂱	KHOJKI LETTER RA	ر	U+0631 ARABIC LETTER REH
<i>la</i>	𑂲	KHOJKI LETTER LA	ل	U+0644 ARABIC LETTER LAM
<i>ṛa</i>	𑂳	KHOJKI LETTER LLA	ڙ	U+0699 ARABIC LETTER REH WITH FOUR DOTS ABOVE
<i>va</i>	𑂴	KHOJKI LETTER VA	و	U+0648 ARABIC LETTER WAW
<i>sa</i>	𑂵	KHOJKI LETTER SSA	س	U+0633 ARABIC LETTER SEEN
<i>ha</i>	𑂶	KHOJKI LETTER HA	ه	U+06C1 ARABIC LETTER HEH GOAL

Table 1: Correspondences between Khojki and Sindhi-Arabic letters. See Section 4.10 for correspondences produced using NUKTA.

	KHOJKI	KHUDAWADI	GURMUKHI	DEVANAGARI		KHOJKI	KHUDAWADI	GURMUKHI	DEVANAGARI
KA	ਕ	ਖ	ਕ	क	DA	ਦ	ड	ਦ	द
KHA	ਖ	ਘ	ਖ	ख	DDDA	ਛ	ਙ	—	डु
GA	ਗ	ਗ	ਗ	ग	DHA	ਘ	ਘ	ਧ	ध
GGA	ਗ	ਘ	—	गु	NA	ਙ	ਙ	ਠ	न
GHA	ਘ	ਙ	ਘ	घ	PA	ਪ	ਪ	ਪ	प
NGA	ਙ	ਙ	ਙ	ङ	PHA	ਘ	ਘ	ਠ	फ
CA	ਚ	ਚ	ਚ	च	BA	ਬ	ਬ	ਬ	ब
CHA	ਚ	ਛ	ਛ	छ	BBA	ਬ	ਘ	—	बु
JA	ਜ	ਜ	ਜ	ज	BHA	ਘ	ਘ	ਭ	भ
JJA	ਘ	ਘ	—	जु	MA	ਮ	ਮ	ਮ	म
JHA	—	ਙ	ਝ	झ	YA	ਯ	ਯ	ਯ	य
NYA	ਙ	ਙ	ਞ	ञ	RA	ਰ	ਰ	ਰ	र
TTA	ਟ	ਟ	ਟ	ट	LA	ਲ	ਲ	ਲ	ल
TTHA	ਠ	ਠ	ਠ	ठ	VA	ਵ	ਠ	ਵ	व
DDA	ਡ	ਡ	ਡ	ड	SHA	—	ਸ਼	ਸ਼	श
DDHA	ਢ	ਢ	ਢ	ढ	SA	ਸ	ਸ	ਸ	स
NNA	ਨ	ਨ	ਨ	ण	HA	ਹ	ਹ	ਹ	ह
TA	ਤ	ਤ	ਤ	त	LLA	ਲ	—	ਲ	ळ
THA	ਥ	ਥ	ਥ	थ	RRA	—	ਤ	ਤ	—

Table 2: Consonant letters of Khojki, Khudawadi, Gurmukhi, and Devanagari.

	KHOJKI	KHUDAWADI	GURMUKHI	DEVANAGARI		KHOJKI	KHUDAWADI	GURMUKHI	DEVANAGARI
A	𑂔	𑂕	ਅ	अ	-A	—	—	—	—
AA	𑂔𑂕	𑂕𑂕	ਆ	आ	-AA	𑂔	𑂕	𑂕	𑂕
I	𑂔𑂕	𑂕	ਇ	इ	-I	𑂔	𑂕	𑂕	𑂕
II	—	𑂕	ਈ	ई	-II	𑂔	𑂕	𑂕	𑂕
U	𑂔	𑂕	ਉ	उ	-U	𑂔	𑂕	𑂕	𑂕
UU	—	𑂕	ਊ	ऊ	-UU	—	𑂕	𑂕	𑂕
E	𑂔	𑂕	ਏ	ए	-E	𑂔	𑂕	𑂕	𑂕
AI	𑂔	𑂕	ਐ	ऐ	-AI	𑂔	𑂕	𑂕	𑂕
O	𑂔	𑂕	ਓ	ओ	-O	𑂔	𑂕	𑂕	𑂕
AU	𑂔	𑂕	ਔ	औ	-AU	𑂔	𑂕	𑂕	𑂕

Table 3: Vowel letters and signs of Khojki, Khudawadi, Gurmukhi, and Devanagari.

	KHOJKI	KHUDAWADI	GURMUKHI	DEVANAGARI		KHOJKI	KHUDAWADI	GURMUKHI	DEVANAGARI
0	—	०	੦	०	5	—	૫	੫	५
1	—	१	੧	१	6	—	૬	੬	६
2	—	२	੨	२	7	—	૭	੭	७
3	—	३	੩	३	8	—	૮	੮	८
4	—	४	੪	४	9	—	૯	੯	९

Table 4: Digits of Khojki, Khudawadi, Gurmukhi, and Devanagari. Khojki uses Gujarati digits.

	1120	1121	1122	1123	1124
0	𑂏 11200	𑂐 11210	𑂑 11220	𑂒 11230	
1	𑂓 11201	𑂔 11211	𑂕 11221	𑂖 11231	
2	𑂗 11202		𑂘 11222	𑂙 11232	
3	𑂚 11203	𑂛 11213	𑂜 11223	𑂝 11233	
4	𑂞 11204	𑂟 11214	𑂠 11224	𑂡 11234	
5	𑂢 11205	𑂣 11215	𑂤 11225	𑂥 11235	
6	𑂦 11206	𑂧 11216	𑂨 11226	𑂩 11236	
7	𑂪 11207	𑂫 11217	𑂬 11227	𑂭 11237	
8	𑂮 11208	𑂯 11218	𑂰 11228	𑂱 11238	
9	𑂲 11209	𑂳 11219	𑂴 11229	𑂵 11239	
A	𑂶 1120A	𑂷 1121A	𑂸 1122A	𑂹 1123A	
B	𑂺 1120B	𑂻 1121B	𑂼 1122B	𑂽 1123B	
C	𑂾 1120C	𑂿 1121C	𑃀 1122C	𑃁 1123C	
D	𑃂 1120D	𑃃 1121D	𑃄 1122D	𑃅 1123D	
E	𑃆 1120E	𑃇 1121E	𑃈 1122E		
F	𑃉 1120F	𑃊 1121F	𑃋 1122F		

Figure 1: Proposed code chart for Khojki.

Independent vowels

11200	𑌀	KHOJKI LETTER A
11201	𑌁	KHOJKI LETTER AA
11202	𑌂	KHOJKI LETTER I
11203	𑌃	KHOJKI LETTER U
11204	𑌄	KHOJKI LETTER E
11205	𑌅	KHOJKI LETTER AI
11206	𑌆	KHOJKI LETTER O
11207	𑌇	KHOJKI LETTER AU

1123C = KHOJKI DOUBLE SECTION MARK
 1123D ◦ KHOJKI ABBREVIATION SIGN

Consonants

11208	𑌈	KHOJKI LETTER KA
11209	𑌉	KHOJKI LETTER KHA
1120A	𑌊	KHOJKI LETTER GA
1120B	𑌋	KHOJKI LETTER GGA
1120C	𑌌	KHOJKI LETTER GHA
1120D	𑌍	KHOJKI LETTER NGA
1120E	𑌎	KHOJKI LETTER CA
1120F	𑌏	KHOJKI LETTER CHA
11210	𑌐	KHOJKI LETTER JA
11211	𑌑	KHOJKI LETTER JJA
11212	𑌒	<reserved>
11213	𑌓	KHOJKI LETTER NYA
11214	𑌔	KHOJKI LETTER TTA
11215	𑌕	KHOJKI LETTER TTHA
11216	𑌖	KHOJKI LETTER DDA
11217	𑌗	KHOJKI LETTER DDHA
11218	𑌘	KHOJKI LETTER NNA
11219	𑌙	KHOJKI LETTER TA
1121A	𑌚	KHOJKI LETTER THA
1121B	𑌛	KHOJKI LETTER DA
1121C	𑌜	KHOJKI LETTER DDDA
1121D	𑌝	KHOJKI LETTER DHA
1121E	𑌞	KHOJKI LETTER NA
1121F	𑌟	KHOJKI LETTER PA
11220	𑌠	KHOJKI LETTER PHA
11221	𑌡	KHOJKI LETTER BA
11222	𑌢	KHOJKI LETTER BBA
11223	𑌣	KHOJKI LETTER BHA
11224	𑌤	KHOJKI LETTER MA
11225	𑌥	KHOJKI LETTER YA
11226	𑌦	KHOJKI LETTER RA
11227	𑌧	KHOJKI LETTER LA
11228	𑌨	KHOJKI LETTER VA
11229	𑌩	KHOJKI LETTER SA
1122A	𑌪	KHOJKI LETTER HA
1122B	𑌫	KHOJKI LETTER LLA

Dependent vowel signs

1122C	𑌬	KHOJKI VOWEL SIGN AA
1122D	𑌭	KHOJKI VOWEL SIGN I
1122E	𑌮	KHOJKI VOWEL SIGN II
1122F	𑌯	KHOJKI VOWEL SIGN U
11230	𑌰	KHOJKI VOWEL SIGN E
11231	𑌱	KHOJKI VOWEL SIGN AI
11232	𑌲	KHOJKI VOWEL SIGN O
11233	𑌳	KHOJKI VOWEL SIGN AU

Various signs

11234	𑌴	KHOJKI SIGN ANUSVARA
11235	𑌵	KHOJKI SIGN VIRAMA
11236	𑌶	KHOJKI SIGN NUKTA
11237	𑌷	KHOJKI SIGN SHADDA

Punctuation

11238	𑌸	KHOJKI DANDA
11239	𑌹	KHOJKI DOUBLE DANDA
1123A	𑌺	KHOJKI WORD SEPARATOR
1123B	𑌻	KHOJKI SECTION MARK

Figure 2: Proposed names list for Khojki.

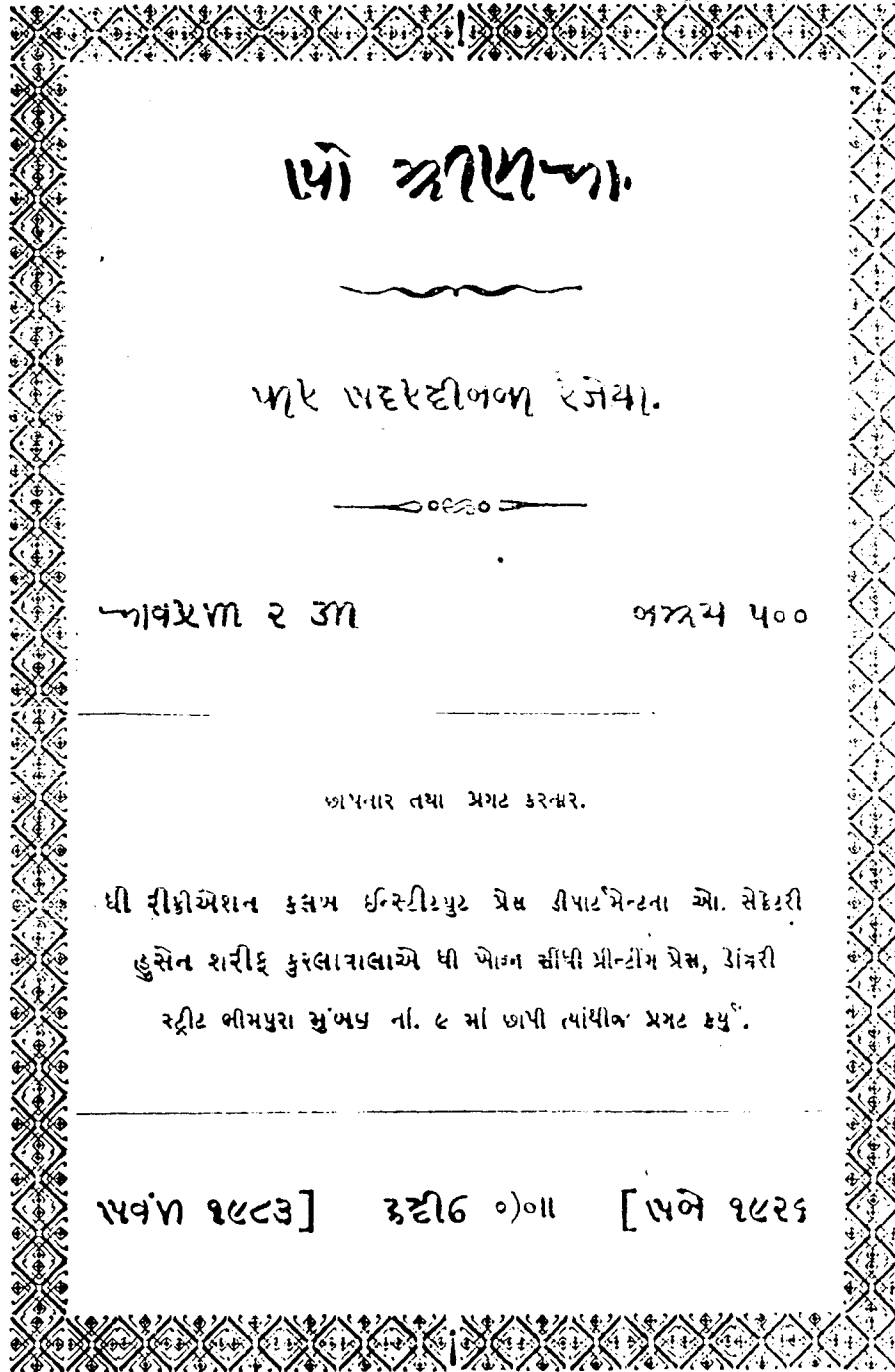


Figure 3: Title page of the *સો કીરયા* (*So Kirā*) printed in Khojki and Gujarati (from Shackle and Moir 1992: 33). North Indic number forms and the U+A838 NORTH INDIC RUPEE MARK are used to indicate the price of the item.

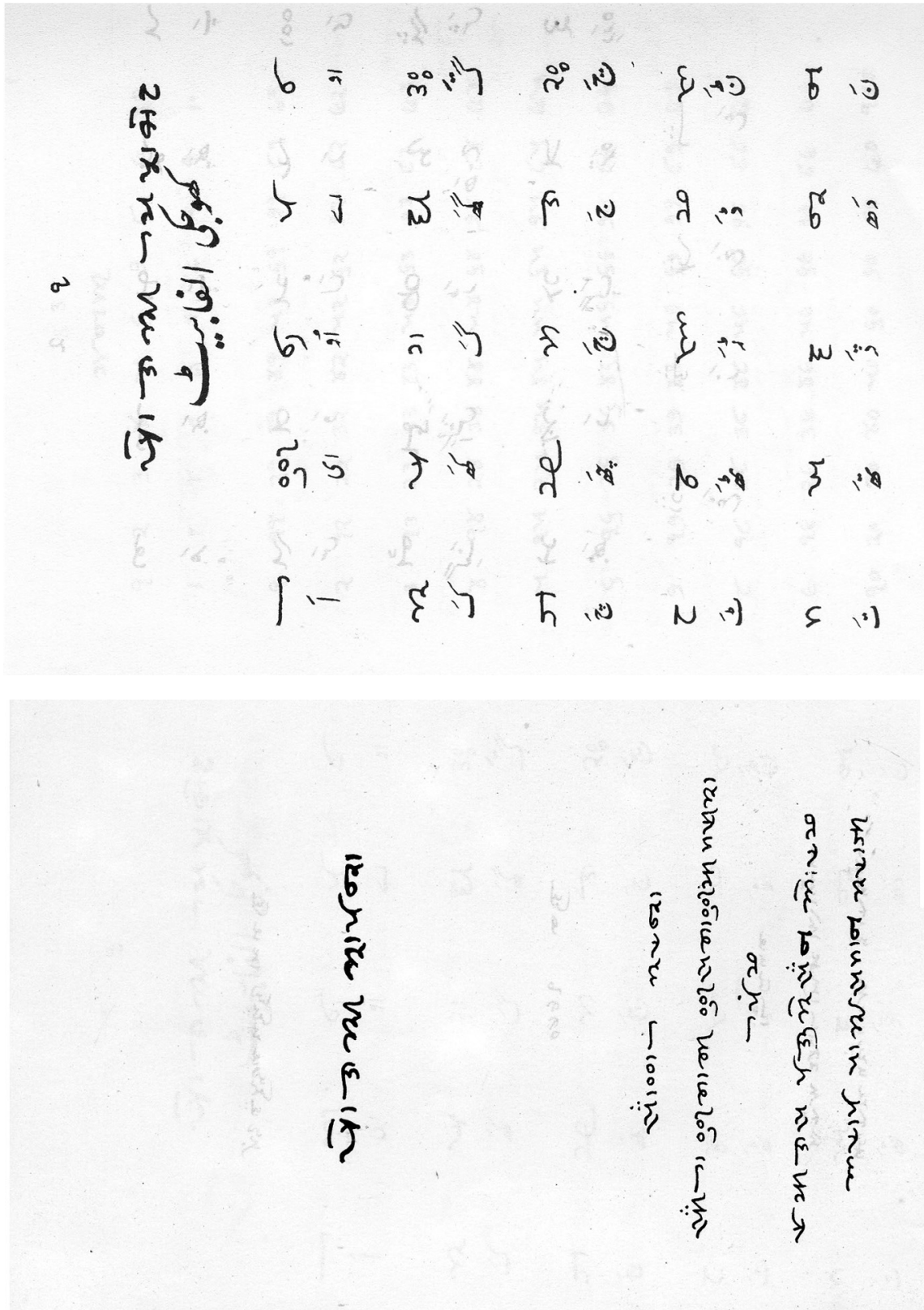
Ms Ism K 1	[Bhadneshwar] [192-?]]	11-13 (same rāg)	māfhe:varāsete:abharat:nūrake ਮਫ਼ਹੇ:ਵਰਾਸੇਟੇ:ਅਬਹਾਰਤ:ਨੂਰਾਕੇ
Foreign wove paper; watermark: E.R. & Co., Made in Austria, chariot; 96 folios; 184 x 16 (17 x 13); 12 - 16 lines; stab/sewn binding laced on hard board with half-cloth covering with decorative paper; black ink.		13-15 (same rāg)	18 verses, G, Pīr Shams tame:ramājo:dhanane:rātake ਮਠੇ:ਰਾਮਾਜੋ:ਦੁਖਾਨਾਨੇ:ਰਾਟਾਕੇ
Scribe: Khojā bhāī Ladhū Devjī (65; 179-80); does not use short 'i' vowel; ਯ = ਚ; ਞ = ਜ; ਝ = h; ਮ = m; ਨ = o; ਞ = pr.		15-18 (same rāg)	19 verses, G, Pīr Shams tame:japājo:dhanane:rātake ਮਠੇ:ਜਾਪਾਜੋ:ਦੁਖਾਨਾਨੇ:ਰਾਟਾਕੇ
Pp. 41-43 torn, no other significant damage; tafsīlo : p. 182 [f. 91v]; collected from Bhadneshwar Jamat, 12th July 1970; sequence of works from pp. 70-102 corresponds to that in P Ism K 27, pp. 19-30.		18-20 (rāg cāitī)	22 verses, G, Pīr Shams bhūlo:bhūlo:te:bhamarāno ਭੁਲੋ:ਭੁਲੋ:ਤੇ:ਭਮਾਰਾਨੋ
[Garbīs. Ref: D 61-64 I 177 N 9,44,81,100 P 300]		20-22 (rāg ṭhāl/ḡhāl?)	22 verses, G, Pīr Shams sat:mārg:samas:pīrake ਸਾਟ:ਮਾਰਗ:ਸਾਮਾਸ:ਪੀਰਾਕੇ
1-2 (a.t.)	ejī:ād:gūr:samas:mūnivar:jā:n ਏਜੀ:ਆਦ:ਗੁਰ:ਸਾਮਾਸ:ਮੁਨੀਵਾਰ:ਜਾ:ਨ	22-25 (same rāg)	18 verses, G, Pīr Shams sate:pāvaṛ:pīo:dhanane:rātake ਸਾਟੇ:ਪਾਵਾਰ:ਪੀਓ:ਦੁਖਾਨਾਨੇ:ਰਾਟਾਕੇ
2-4 (rāg cāl)	āj:ānānā:pāmā:manke:sat:gūr ਆਜ:ਆਨਾਨਾ:ਪਾਮਾ:ਮਾਨਕੇ:ਸਾਟ:ਗੁਰ	25-26 (rāg cāitī)	21 verses, G, Pīr Shams ḡeto:ḡeto:te:ḡafḡal:ḡetīarelo1 ਗੁਟੋ:ਗੁਟੋ:ਤੇ:ਗੁਫਗਲ:ਗੁਟੀਅਰੇਲੋ1
4-6 (same rāg as above)	gūr:āveā:bījī:rātake:āvī ਗੁਰ:ਆਵੇ:ਬੀਜੀ:ਰਾਟਾਕੇ:ਆਵੀ	26-29 (rāg ṭhāl/ḡhāl?)	18 verses, G, Pīr Shams gūr:āveā:sātīmī:rātake ਗੁਰ:ਆਵੇ:ਸਾਟੀਮੀ:ਰਾਟਾਕੇ
6-8 (same rāg as above)	gūr:tāvātā:sarve:rātake:ganān ਗੁਰ:ਟਾਵਾਟਾ:ਸਾਰਵੇ:ਰਾਟਾਕੇ:ਗਾਨਾਨ	29-31 (rāg dohrā)	18 verses, G, Pīr Shams satgūr:samas:kaere:gāfalo ਸਾਟਗੁਰ:ਸਾਮਾਸ:ਕਾਏਰੇ:ਗਾਫਾਲੋ
8-11 (same rāg as above)	ke:tame:amīras:pījo:dhanane:rātake ਕੇ:ਟਾਮੇ:ਅਮੀਰਾਸ:ਪੀਜੋ:ਦੁਖਾਨਾਨੇ:ਰਾਟਾਕੇ	31-32 (rāg cāitī)	18 verses, G, Pīr Shams gūr:kahere:gāfalo:sābhafo ਗੁਰ:ਕਾਹੇਰੇ:ਗਾਫਾਲੋ:ਸਾਬਹਾਫੋ
		33-35 (rāg ṭhāl/ḡhāl?)	18 verses, G, Pīr Shams tathī:āṭhamī:āvā:gāmanā:lokake ਟਾਠੀ:ਆਟਹਾਮੀ:ਆਵਾ:ਗਾਮਾਨਾ:ਲੋਕਾਕੇ
			19 verses, G, Pīr Shams

Figure 7: Khojki text written by Ali S. Asani for his *The Harvard Collection of Ismaili Literature in Indic Languages* (1992: 56-57).

ALPHABET

𑂀	𑂁	𑂂	𑂃	𑂄
a	i	oo	e	o
𑂅	𑂆	𑂇	𑂈	𑂉
ka	kha	ga	gha	anga
𑂊	𑂋	𑂌	𑂍	𑂎
cha	chha	za	ja	nya
𑂏	𑂐	𑂑	𑂒	𑂓
ta	tha	da	dha	nna
𑂔	𑂕	𑂖	𑂗	𑂘
ta	tha	da	dha	na
𑂙	𑂚	𑂛	𑂜	𑂝
pa	fa	ba	bha	ma
𑂞	𑂟	𑂠	𑂡	𑂢
ya	ra	la	va	sa
𑂣	𑂤	𑂥	𑂦	𑂧
sha	ha	la	ksa	gna
𑂨	𑂩	𑂪	𑂫	𑂬
dhr	tr	ga	dha	bha

Figure 9: Chart of Khojki printed in a modern primer (from Tajddin 1989: 10).



Roman characters.	Devanāgarī.	Khudāwādī.*	Shikārpurī.	Sakkar.	THATTĀI.		Lārī.	Wangāl.	Rajjāl.	Khawjā.	MAIMONS.		Sewhāpī Bhabhīrā.	Southern Lahndā.
					Luhandā.	Bhasīā.					Thattā.	Haiderābād.		
a	अ	ṁ	ṁ	ṁ	ʼʼ^	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ
ā	आ	ṁ	ṁ	ṁ	ʼʼ~	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ
i	इ	†ṁ	†ṁ	†ṁ	ṁ.ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ
ī	ई	†ṁ	†ṁ	†ṁ	ṁ.ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ
u	उ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ
ū	ऊ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ
e	ए	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ
ai	ऐ	ṁ	ṁ	ṁ	ʼʼ~	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ
o	ओ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ
au	औ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ

* The Khudāwādī is the character used at Haiderābād and by most of the respectable merchants throughout the country. That of the Shikārpur and Sakkar merchants differs but little from it.

† These are, however, never written in commencing a word. In lieu of them III is used.

Figure 11: Inventory of Khojki ('Khawjās') vowel letters from Grierson (1905: 15).

Roman characters.	Devanagari.	Khudawadi.	Shikarpuri.	Sakkar.	THATTALI.			Wangal.	Rajjat.	Khawaj.	MALMONS.		Sewhapi Bhabhira.	Southern Lahnda.
					Lahanda.	Bhatia.	Lapli.				Thatta.	Haidarabad.		
ka	क	ڪ	ڪ	ڪ	ڪ	ڪ	ڪ	ڪ	ڪ	ڪ	ڪ	ڪ	ڪ	ڪ
kha	ख	ڪھ	ڪھ	ڪھ	ڪھ	ڪھ	ڪھ	ڪھ	ڪھ	ڪھ	ڪھ	ڪھ	ڪھ	ڪھ
ga	ग	گ	گ	گ	گ	گ	گ	گ	گ	گ	گ	گ	گ	گ
gga	ग़	گھ	گھ	گھ	گھ	گھ	گھ	گھ	گھ	گھ	گھ	گھ	گھ	گھ
gha	घ	گھ	گھ	گھ	گھ	گھ	گھ	گھ	گھ	گھ	گھ	گھ	گھ	گھ
na	ङ	ن	ن	ن	ن	ن	ن	ن	ن	ن	ن	ن	ن	ن
cha	च	چ	چ	چ	چ	چ	چ	چ	چ	چ	چ	چ	چ	چ
chha	छ	چھ	چھ	چھ	چھ	چھ	چھ	چھ	چھ	چھ	چھ	چھ	چھ	چھ
ja	ज	ج	ج	ج	ج	ج	ج	ج	ج	ج	ج	ج	ج	ج
jja	झ	جھ	جھ	جھ	جھ	جھ	جھ	جھ	جھ	جھ	جھ	جھ	جھ	جھ
jha	झ	جھ	جھ	جھ	جھ	جھ	جھ	جھ	جھ	جھ	جھ	جھ	جھ	جھ
ña	ञ	ن	ن	ن	ن	ن	ن	ن	ن	ن	ن	ن	ن	ن
ta	ट	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت
tra	ट्र	تھ	تھ	تھ	تھ	تھ	تھ	تھ	تھ	تھ	تھ	تھ	تھ	تھ
tha	ठ	تھ	تھ	تھ	تھ	تھ	تھ	تھ	تھ	تھ	تھ	تھ	تھ	تھ
ra	ड	ر	ر	ر	ر	ر	ر	ر	ر	ر	ر	ر	ر	ر
ḍa	ड़	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ
dra	ड्र	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ
dha	ढ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ	رھ
ra	ण	ڻ	ڻ	ڻ	ڻ	ڻ	ڻ	ڻ	ڻ	ڻ	ڻ	ڻ	ڻ	ڻ

Figure 12: Inventory of Khojki ('Khawājās') consonant letters from Grierson (1905: 16).

Roman character.s.	Dēvanāgarī.	Khudawāqī.	Shikārpurī.	Sakkār.	THAṬṬĀI.			Wangāl.	Rājjar.	Khwājās.	MAIMONB.		Sewhānī Bhabhīras.	Southern Lahnda.
					Juhānās.	Bhātīās.	Larāl.				Thattā.	Haiderābād.		
ta	त	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
tha	थ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
da	द	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
dha	ध	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
na	न	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
pa	प	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
pha	फ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
ba	ब	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
bba	ब	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
bha	भ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
ma	म	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
ya	य	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
ra	र	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
la	ल	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
va	व	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
śa	श	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
sha	ष	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
sa	स	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
ha	ह	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
Numerals				ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ	ٲ
				1	2	3	4	5	6	7	8	9	10	

Figure 13: Inventory of Khojki ('Khwājās') consonant letters from Grierson (1905: 17).

Table showing correspondence of initial vowels in the Roman, Urdu, Sindhi, Gujarati and Khojki scripts

<u>Roman</u>	<u>Urdu</u>	<u>Sindhi</u>	<u>Gujarati</u>	<u>Khojki#</u>
a	ا	ا	अ	𑂀, 𑂁
ā	آ	آ	आ	𑂀, 𑂁 𑂀 [𑂀] (𑂀 𑂀?)
i*	ا	ا	इ	𑂀 [𑂀] (𑂀 𑂀?)
ī*	ای	ای	ई	𑂀 [𑂀] (𑂀 𑂀?)
u*	ا	ا	उ	𑂀 [𑂀]
ū*	او	او	ऊ	𑂀 [𑂀] (𑂀?)
o	او	او	ओ	𑂀 [𑂀, 𑂀] (𑂀?)
e	اے	ای	ऐ	𑂀 [𑂀] (𑂀?)
ai	اے	ای	ऎ	[𑂀, 𑂀]
au	او	او	ऒ	[𑂀, 𑂀, 𑂀]

*On account of ambiguities in the Khojki script, initial and independent short or long "i" vowels in Khojki texts are transliterated as "î". Similarly, short or long "u" vowels are transliterated as "û".

#[] documented variant; (?) variant listed by G. Allana in *Sindhi Śūratkhatī* (Hyderabad, Pakistan: Sindhi Zabān Publications, 1964) or G. Stack in *A Grammar of the Sindhi Language* (Bombay, 1849); () alternate sound employed.

Figure 14: Inventory of Khojki independent vowel letters (from Asani 1992: 678).

Table showing correspondence of non-initial vowels in the Roman, Urdu, Sindhi, Gujarati and Khojki scripts

<u>Roman</u>	<u>Urdu</u>	<u>Sindhi</u>	<u>Gujarati</u>	<u>Khojki#</u>
baḍ	بَڊ	بَڊ	બૈ	૫૬ [૫૬]
bād	باد	باد	બાૈ	૫૬
din	دن	دن	દૈન	૬૫
dīn	دین	دین	દાૈન	૬૫
but*	بُت	بُت	બુત	૫૫
būt*	بوت	بوت	બૂત	૫૫
so	سو	سو	સૌ	૫૦ [૫૦]
be	بے	بی	બૈ	૫૦ [૫૦]
hai	ہائے	ہائی	હૈ	[૫૦, ૫૦-૫૦]
nau	نَو	نَو	નાૈ	[૫૦, ૫૦-૫૦]

*On account of ambiguities in the Khojki script, short and long "u" vowels in Khojki texts are transliterated as "û".

#[] documented variant; (?) variant listed by Allana or Stack;
() alternate sound employed.

Figure 15: Inventory of Khojki dependent vowel signs (from Asani 1992: 679).

Table showing correspondence of
consonants in the Roman, Urdu,
Sindhi, Gujarati and Khojki scripts

<u>Roman</u>	<u>Urdu</u>	<u>Sindhi</u>	<u>Gujarati</u>	<u>Khojki#</u>
b	ب	ب	બ	૧ [મ, ૫]
ḅ		ب̣		૧
bh	بھ	پ	બ	૫, ૫ [મ]
t	ت	ت	ત	૫
th	تھ	تھ	થ	૫ [મ]
ṭ	ط	ت̣	ટ	૮, ૮, (૫?)
ṭh	طھ	ت̣ھ	ઠ	૮
s	س	س	(સ)	(૫)
p	پ	پ	પ	૫ (૫?)
ph	پھ	ف	ફ	૫
j	ج	ج	જ	૩, ૫, ૩ (૫?)
dy		ج̣		૫ (૩?)

#[] documented variant; (?) variant listed by Allana or Stack;
() alternate sound employed.

Figure 16: Inventory of Khojki consonants: *b* to *dy* (from Asani 1992: 680).

<u>Roman</u>	<u>Urdu</u>	<u>Sindhi</u>	<u>Gujarati</u>	<u>Khojki#</u>
jh	جھ	جھ		𑂏
ny		ج	𑂏	𑂏: (𑂏?)
é	ج	ج	𑂏	𑂏, (𑂏?)
éh	جھ	جھ	𑂏	𑂏, (𑂏?)
ḥ	ح	ح	(𑂏)	(𑂏)
kh	خ	خ	(𑂏)	𑂏 (𑂏)
d	د	د	𑂏	𑂏, 𑂏
dh	دھ	دھ	𑂏	𑂏, 𑂏, (𑂏?)
ḍ		دھ		𑂏
ḍ	ڈ	دھ	𑂏	𑂏, 𑂏 (𑂏, 𑂏, 𑂏?)
ḍh	ڈھ	دھ	𑂏	𑂏 (𑂏?)
z	ز	ز	(𑂏)	(𑂏, 𑂏)
r	ر	ر	𑂏	𑂏
ṛ	رھ	رھ	𑂏	𑂏, 𑂏 [𑂏]
ṛh	رھ			
z	ز	ز	𑂏	𑂏 [𑂏]
s	س	س	𑂏	𑂏

Figure 17: Inventory of Khojki consonants: *jh* to *s* (from Asani 1992: 681).

<u>Roman</u>	<u>Urdu</u>	<u>Sindhi</u>	<u>Gujarati</u>	<u>Khojki#</u>
sh	ش	ش	શ	𑂔
ṣ	ص	ص	(સ)	(૫)
z	ض	ض	(ઝ)	(કૌ, ૨૫)
t	ط	ط	(ત, ૨)	(૧, ૨)
ẓ	ظ	ظ	(ઝ)	(કૌ, ૨૫)
ʿ	ع	ع	(અ)	𑂔, 𑂕 (—)
ġ	غ	غ	(ગ)	𑂔 (ગ)
f	ف	ف	ફ	𑂔
q	ق	ق	(ક)	𑂔 (૫)
k	ک	ک	(ક)	𑂔 [𑂔]
kh	کھ	ک	ખ	𑂔, 𑂕 (𑂔?)
g	گ	گ	ગ	𑂔 [𑂔, 𑂕]
ng		گب		𑂔
gh	گھ	گھ	ઘ	𑂔 (𑂔?)
ng		نگ	ડ	3: (2:?)
l	ل	ل	લ	𑂔 [𑂔]

Figure 18: Inventory of Khojki consonants: *sh* to *l* (from Asani 1992: 682).

<u>Roman</u>	<u>Urdu</u>	<u>Sindhi</u>	<u>Gujarati</u>	<u>Khojki#</u>
m	م	م	મ	𑂔
n	ن	ن	ન	𑂕, 𑂖
ṅ		ڻ	ડા	𑂗 (𑂗:?)
w/v	و	و	વ	𑂘 [𑂘]
h	ه	ه	હ	𑂙 [𑂙]
y	ي	ي	ય	𑂚, 𑂛, (𑂜?)
ksh			ક	𑂝, (𑂞?)
gn			જ	𑂟
tr			ટ	𑂠 (𑂡, 𑂢?)
dhr			ઢ	𑂣

Nasalization is indicated by ñ or ṇ̃

Figure 19: Inventory of Khojki consonants: *m* to *dhr* (from Asani 1992: 683).

Table 1
Khojki Script: Vowel System (early 20th century)

(a) Initial and Independent									
a	ā	i	ī	u	ū	o	e	au	ay
𑂀	𑂁	𑂂	𑂃	𑂄	𑂅	𑂆	𑂇	𑂈	𑂉
(b) Medial and Final (with the Consonant 𑂀)									
a	ā	i	ī	u	ū	o	e	au	ay
𑂀	𑂁	𑂂	𑂃	𑂄	𑂅	𑂆	𑂇	𑂈	𑂉

Table 2
Khojki Script: Consonants (early 20th century)

𑂀	k	𑂁	kh	𑂂	g	𑂃, 𑂄, 𑂅	gh	𑂆	ng	𑂇	ng
𑂈	c	𑂉	ch	𑂊	j/jh	𑂋	j/jh	𑂌	ny	𑂍	dy
𑂎	t	𑂏	ṭh	𑂐	ḍ	𑂑, 𑂒, 𑂓	ḍh	𑂔	ṇ	𑂕	ḍ
𑂖	t	𑂗	th	𑂘	d	𑂙, 𑂚	dh	𑂛	ṇ	𑂜	ḍ
𑂝	p	𑂞	ph	𑂟	b	𑂠, 𑂡, 𑂢	bh	𑂣	m	𑂤	b
𑂥	y	𑂦	r	𑂧	l	𑂨	v	𑂩	s	𑂪	sh
𑂫	dr	𑂬	tr	𑂭	h	𑂮, 𑂯	r/l				

Figure 20: (from Asani 1992).

1	2	3	4	5
↵ a	↵) lā	↵) to	↵) hā	↵ r
↵) ā	↵) i	↵) u	↵) e	↵) g ²
↵) k	↵) kh	↵) g ¹	↵) gh	↵) ñ
↵) c	↵) ch	↵) j ¹	↵) j ²	↵) ñ
↵) †	↵) †h	↵) d	↵) dh	↵) ñ
↵) t	↵) th	↵) d ¹	↵) dh	↵) n
↵) p	↵) ph	↵) b ¹	↵) bh	↵) m
↵) y	↵) r	↵) l	↵) v	↵) d ² /dh
↵) s	↵) dr	↵) tr	↵) h	↵) b ² /bh

Figure 21: Traditional arrangement of Khojki characters, beginning with the traditional Sindhi invocation ↵)↵)↵)↵)↵) allā tohār (from Shackle and Moir 1992: 35).

↵) a	↵) ā	↵) i	↵) u	↵) e
↵) o	↵) k	↵) kh	↵)/↵) g	↵)/↵) gh
↵) ñ	↵) c	↵) ch	↵) j/jh	↵) j/jh
↵) ñ	↵) †	↵) †h	↵)/↵) d	↵)/↵) dh
↵) ñ	↵) t	↵) tr	↵) th	↵)/↵) d
↵) dr	↵)/↵) dh	↵) n	↵) p	↵) ph
↵)/↵) b	↵)/↵) bh	↵) m	↵) y	↵) r
↵) l	↵) v	↵) s	↵) h	

Figure 22: Arrangement of Khojki characters based on the traditional Indic phonological pattern (from Shackle and Moir 1992: 36).

Roman	Devana-gari	Gujarati	Sindhi	D MS 1815	Kx MS 1737	Khojaki Primer 1932
k	क	ક	ڪ	क	क	क
kh	ख	ખ	ڪھ	ख	ख	ख
g	ग	ગ	گ	ग	ग ⁵	ग
gh	घ	ઘ	گھ	घ	घ	घ
c	च	ચ	چ	च	च	च
ch	छ	છ	چھ	छ	छ	छ
j	ज	જ	ج	ज	ज ⁶	ज
jh	झ	—	جھ	झ	ज ⁶	—
t	ट	ટ	ٽ	ट	ट ⁷	ट
th	ठ	ઠ	ٺ	ठ	ठ ⁷	ठ
ḍ	ड	ડ	ڊ	ड	ड ⁸	ड

Figure 24: Comparison of Khojki with other scripts: ka to ḍa (from Khaakee 1981).

Roman	Devana-gari	Gujarati	Sindhi	D MS 1815	Kx MS 1737	Khojaki Primer 1932
a	अ	અ	ا	अ	अ	अ
ā	आ	આ	آ	आ	आ	आ
i	इ	ઇ	ي	इ	इ ¹	इ
ī	ई	ઈ	ی	ई	—	—
u	उ	ઉ	و	उ	उ ²	उ
ū	ऊ	ઊ	وو	—	—	—
e	ए	એ	ی	इ	इ ³	इ
o	ओ	ઑ	و	—	उ ²	उ
ai	ऐ	ઐ	ای	ई	—	—
au	औ	ઔ	او	ई	—	—

Figure 23: Comparison of Khojki with other scripts: a to au (from Khaakee 1981).

Roman	Devana-gari	Gujarati	Sindhi	D MS 1815	Kx MS 1737	Khojaki Primer 1932
bh	भ	ભ	پ	𑂔	𑂔 ¹²	𑂔
m	म	મ	م	𑂔	𑂔	𑂔
y	य	ય	ي	𑂔	𑂔 ¹³	𑂔
r	र	ર	ر	𑂔	𑂔	𑂔
ru	—	—	—	𑂔	—	—
l	ल	લ	ل	𑂔	𑂔	𑂔
v	व	વ	و	𑂔	𑂔	𑂔
v s	श	શ	ش	𑂔	𑂔 ¹⁴	𑂔
s	ष	ષ	ش	—	𑂔 ¹⁴	—
s	स	સ	س	𑂔	𑂔	𑂔
h	ह	હ	ه	𑂔	𑂔	𑂔

Figure 26: Comparison of Khojki with other scripts: *bha* to *ha* (from Khakee 1981).

Roman	Devana-gari	Gujarati	Sindhi	D MS 1815	Kx MS 1737	Khojaki Primer 1932
dh	ड	ડ	ڊ	—	𑂔	𑂔
ṛ	ऱ	𑂔	𑂔	—	𑂔 ⁸	𑂔
t	त	ત	ت	𑂔	𑂔	𑂔
th	थ	થ	ٿ	𑂔	𑂔	𑂔
d	द	દ	د	𑂔	𑂔	𑂔
dh	ध	𑂔	ڏ	𑂔	𑂔 ⁹	𑂔
n	न	ન	ن	𑂔	𑂔	𑂔
p	प	પ	پ	𑂔	𑂔 ¹⁰	𑂔
f	फ	ફ	ف	𑂔	𑂔 ¹¹	—
ph	—	—	ق	—	—	—
b	ब	બ	ب	𑂔	𑂔	𑂔

Figure 25: Comparison of Khojki with other scripts: *dha* to *ba* (from Khakee 1981).

Roman	Devana- gari	Gujarati	Sindhi	D MS 1815	Kx MS 1737	Khojaki Primer 1932
r	र	ર	ر	ر	م ⁸	ر
ks	क्ष	ક્ષ	—	—	—	ڪ
gn	ग्न	ગ્ન	—	—	—	ڱ
tr	त्र	ત્ર	—	ر	ر	ڙ
dv	—	ડ	—	ڙڙ	—	—
dy	—	—	ج	—	ڙ ¹⁵	—
ny	—	—	ج	—	ڙ ¹⁶	—
z	—	ڙ	ڙ ڙ	—	ڙ ¹⁷	—
dhr	—	—	—	—	ڙ	ڙ
sr	—	ر	—	ر	—	—

Figure 27: Comparison of Khojki with other scripts: (ra to sra (from Khakee 1981).

Idealized Khojki Transliteration Table						
Shafique N. Virani						
Last updated: June 3, 2008						
	Stops					Nasals
	Unvoiced		Voiced			
	Unaspirated	Aspirated	Implosive	Unaspirated	Aspirated	
Velar Khojki	𑂔	𑂕	𑂖 ¹	𑂗	𑂘	𑂙:
Transliteration ²	ka	kha	ḡa ³ ALA-LC g with diaeresis above	ga	gha	ṅa
Sindhi (Arabic)	ڪ	ڪھ	گھ	گ	گھ	گھ
Sindhi (Dev)	क	ख	U + 097b ग़	ग	घ	ङ
Gujarati	ક	ખ		ગ	ઘ	ઙ
Palatal Khojki	𑂚	𑂛	𑂜	𑂝		𑂞:
Transliteration	ca ⁴	cha	? ALA-LC j with diaeresis above, replacing tittle (maybe use U + 0135 or U + 01F0?)	ja		ña
Sindhi (Arabic)	چھ	چھھ	چھ	چ		چھ
Sindhi (Dev)	च	छ	U + 097c	ज		ञ

Figure 28: A transliteration for Khojki proposed by Shafique Virani (from Virani 2008). Continued in Figure 29.


















			જ			
Gujarati	ઢ	ઠ		ઢ		ઠ
Cerebral Khojki						
Transliteration	ta	ṭha		ḍa	ḍha	ṇa
Sindhi (Arabic)	ٽ	ٺ		ڊ	ڊھ	ڻ
Sindhi (Dev)	ट	ठ		ड	ढ	ण
Gujarati	ટ	ઠ		ડ	ઢ	ણ
Dental Khojki						
Transliteration	Ta	tha	ḍa ALA-LC d with diaeresis below	da	dha	na
Sindhi (Arabic)	ٽ	ٺ	ڙ، ڙھ	ڍ	ڍھ	ڻ
Sindhi (Dev)	त	थ	ड़ (ALA-LC) ड़ (U + 097e) (Lekhwni's CIIL publication - used for Unicode)	द	ध	न
Gujarati	ત	થ		દ	ધ	ન
Labial Khojki						
Transliteration	pa	pha	ḅa ALA-LC b diaeresis below	ba	bha	ma

Figure 29: A transliteration for Khojki proposed by Shafique Virani (from Virani 2008). Continued in Figure 30.

Sindhi (Arabic)	پ	ڦا	ب	ب	پ	م
Sindhi (Dev)	प	फ	U+097F ब	ब	भ	म
Gujarati	પ	ફ		બ	ભ	મ
Semivowels Khojki	ۛۛ	ۛ	ۛ	ۛ	ۛ	
Transliteration	ya	ra	la	la	va	
Sindhi (Arabic)	ي	ر	ل		و	
Sindhi (Dev)	य	र	ल		व	
Gujarati	ય	ર	લ	ળ	વ	
Sibilants Khojki	ۛ					
Transliteration	sa					
Sindhi (Arabic)	س					
Sindhi (Dev)	स					
Gujarati	સ					
Aspirant Khojki	ۛ					
Transliteration	ha					
Sindhi (Arabic)	ه					
Sindhi (Dev)	ह					
Gujarati	હ					

Figure 30: A transliteration for Khojki proposed by Shafique Virani (from Virani 2008). Continued in Figure 31.

Conjuncts Khojki	ꣳ	ꣴ	ꣵ	ꣶ
Transliteration	dra ⁵	j ña ⁶	kṣa ⁷	tra ⁸
Sindhi (Arabic)				
Sindhi (Dev)				
Gujarati				
Diacritic Letters Khojki	ꣷ			
Transliteration	sha			
Sindhi (Arabic)				
Sindhi (Dev)				
Gujarati				

¹ Characters in red are incorrectly represented in this chart.

² The International Phonetic Alphabet (IPA) has never been used, to the best of my knowledge, for transcribing Khojki, and it doesn't make sense to use it when most scholars working with the script have been using systems closer to the ALA-LC standards.

³ Asani transliterates (oddly?) as ng - find out if any standard transliteration system for Sindhi uses this, or whether it's his own invention.

⁴ TDIL recommends cha and chha, as opposed to ALA-LC ca and cha. <http://tdil.mit.gov.in/>. The Department of Information Technology initiated the TDIL (Technology Development for Indian Languages) with the objective of developing Information Processing Tools and Techniques to facilitate human-machine interaction without language barrier; creating and accessing multilingual knowledge resources; and integrating them to develop innovative user products and services.

⁵ Asani changes from dhra (thesis) to dra (Ecstasy), Shackle and Moir give dra (pp. 35-36)

⁶ Omitted by Shackle and Moir. Asani has it in thesis and transcribes as gn, but omits in Ecstasy.

⁷ This letter, in Sankrit and the NIA languages, is ꣳ + ꣴ, the latter representing the retroflex sound sha.

Khojki does not have separate character to distinguish this retroflex sibilant from the more common palatal. To differentiate it, however, from a possible ka + halant + sha combination, it may be best to adopt kṣa as the transliteration, and also to have a separate character for this in Unicode.

⁸ This may have originally designated a ꣴa sound.

Figure 31: A transliteration for Khojki proposed by Shafique Virani (from Virani 2008).

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	Final Proposal to Encode the Khojki Script in ISO/IEC 10646
2. Requester's name:	<i>Script Encoding Initiative (SEI) / Anshuman Pandey <pandey@umich.edu></i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Liaison contribution</i>
4. Submission date:	<i>2011-01-28</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/> Yes
(or) More information will be provided later:	<input type="checkbox"/>

B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	<input checked="" type="checkbox"/> Yes
Proposed name of script:	<i>Khojki</i>
b. The proposal is for addition of character(s) to an existing block:	
Name of the existing block:	
2. Number of characters in proposal:	<i>61</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input checked="" type="checkbox"/> B.1-Specialized (small collection) <input type="checkbox"/> B.2-Specialized (large collection) <input type="checkbox"/>	
C-Major extinct <input type="checkbox"/> D-Attested extinct <input type="checkbox"/> E-Minor extinct <input type="checkbox"/>	
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/> G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<input checked="" type="checkbox"/> Yes
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/> Yes
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/> Yes
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Anshuman Pandey</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Anshuman Pandey <pandey@umich.edu></i>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/> Yes
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/> Yes
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)?	<input checked="" type="checkbox"/> Yes,
	<i>see proposal for additional details</i>

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 <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N3702-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)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	No
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.)? If YES, with whom? If YES, available relevant documents:	Yes
	<i>Shafique Virani <shafique.virani@utoronto.ca> Irfan Gowani <irfango@exchange.microsoft.com></i>
	<i>See citations in text of proposal</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	Yes
	<i>See text of proposal.</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Common
	<i>See text of proposal.</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes
	<i>See text of proposal.</i>
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:	No
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? If YES, is a rationale for its inclusion provided? If YES, reference:	No
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	No
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	Yes Yes
	<i>See text of proposal.</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	Yes Yes
	<i>Combining vowel signs; see text of proposal</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	Yes Virama;
	<i>see text of proposal</i>
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:	No