

Revised 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

September 12, 2010

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”, May 2008
- N3596 L2/09-101 “Proposal to Encode the Khojki Script in ISO/IEC 10646”, March 2009

This document is a revision of N3596 L2/09-101. Major changes from the previous proposal include the removal of some vowel and consonant letters, the reassignment of glyphs and characters, and the inclusion of new specimens. This revision also provides additional information that will facilitate the encoding and implementation of Khojki. A proposal summary form is also enclosed.

This is not a final proposal to encode Khojki. Feedback from specialists and potential users is requested regarding character-glyph assignments for certain consonant letters (Section 3.2), the repertoire of vowel letters (Section 4.3), the proposed encoded representation of consonant conjuncts (Section 4.7), and the list of characters not proposed for encoding (Section 5).

2 Overview

Khojki is the name of the writing system used by the Nizari Ismaili community of South Asia for recording religious literature. It was developed and used in Sindh, now in Pakistan, primarily for representing the Sindhi language. The script was also used in Gujarat and Punjab to write Gujarati, Punjabi, and Siraiki, in addition to several languages allied with Hindi, as well as Arabic and Persian.

The name Khojki (ਖੋજڪي *khōjakī*, Devanagari खोजकी *khōjakī*) is derived from the Persian honorific title خواجه *khwājah* “master, lord”. It is a direct translation of the Sanskrit term ठाकुर *thākura*, which was used as a title of address by the Lohāṇā community of Sindh, 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 closely related to the Lohāṇākī (lit. ‘of the Lohāṇā’) variety of Landa, which was reformed for the specific purpose of developing an ecclesiastical script for Ismaili literature. Popular Nizari Ismaili tradition states that Khojki was invented and propagated by Pir Sadruddin, an Ismaili missionary actively working with the Lohana community. Khojki was one of two Landa-based scripts that were developed as formal literary scripts in order to serve the needs of a religious community; the other is Gurmukhi, which was developed for the Sikh tradition. Khojki is also called ‘Sindhi’ and ‘Khwajah Sindhi’.

Khojki was in use by the 16th century CE, as attested by manuscript evidence. The script was used primarily for recording literature known as *ginān* (𑂔𑂗𑂏𑂛 < Sanskrit ज्ञान *jñāna* ‘knowledge’). The majority of *gināns* are manuscripts, but, hundreds of lithographs and books were published in the 19th and 20th centuries. 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 Bombay (Mumbai).³

Khojki records are available at several institutions throughout the world. A large collection of Khojki manuscripts and books are housed at the Institute of Ismaili Studies in London. A smaller collection of materials are held at Harvard University. Sizeable private collections are also maintained by members of the Ismaili community.

The user community for Khojki is small as compared to other Indic scripts and while usage of it is declining, it is still a living script. Khojki manuscripts and books are used in Ismaili religious ceremonies not only in South Asia, but in east and South Africa, where large diaspora communities had formed by the 19th century. The script is still used by specialists who study Ismaili history and literature. An encoding for Khojki in the Universal Character Set 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 Basis for Character Repertoire

The goal of this proposal is to provide a character set that may be used to encode any Khojki document. This task is somewhat difficult to perform because of the absence of a standard for Khojki. The character repertoire that is proposed here is based upon an analysis of Khojki characters presented by George Stack in *A Grammar of the Sindhi Language* (as represented in George Grierson’s *Linguistic Survey of India*, figures 11–13); Ali S. Asani in *The Harvard Collection of Ismaili Literature in Indic Languages* (figures 14–19); and Christopher S. Shackle and Zawahir Moir in *Ismaili Hymns from South Asia* (figures 20–21); as well as other script charts and primary documents, some of which are included as specimens in this document.

The proposed repertoire does not include some characters that are shown in script charts. The decision to not encode characters such as *II, *UU, *JHA, etc. rests upon the fact that it is difficult to confirm attested usage of such characters. It is possible that these characters were represented in some texts, but whether it

¹ Daftary 2008: 26–27.

² See Pandey 2010a for details on Landa and 2010b for Sindhi.

³ Tajddin 2003.

is normatively representative is uncertain. Nonetheless, space has been reserved in the code chart for these characters in the case that additional research suggests the need to encode them.

The presence of characters with which readers of Khojki may not necessarily 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 belong to the Khojki repertoire.

3.2 Issues With Character-Glyph Assignments

While the character repertoire is fairly stable, there are some issues related to the assignment of glyphs for certain consonant letters.⁴ Several sources show different glyphic representations for the same character. One 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 glyphs that represented sounds specific to Sindhi, such as the implosive consonant letters (GGA, JJA, DDDA, BBA) lost their original phonetic values. The original glyphs for these letters were reassigned to other consonants within the same articulation class, resulting in the use of multiple glyphs for a single character.

For example, in the chart in Figure 8, the glyph 𑂞 is assigned to BBA, but in Figure 9 the glyph is assigned to BA. The change appears in modern usage, as attested in a primer written in Pakistan in 1990 (see in Figure 10). Representative glyphs for several other characters have also changed over time. These changes appear to be part of a reform of Khojki that took place in the 1930s.⁵

For the purposes of this proposal, glyphs have been assigned to characters based upon historical graphical typology.

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 used the fonts to produce the first printed editions of Ismaili literature (*ginān*) at his Khoja Sindhi Printing Press in Bombay. Although different there were several Khojki printers in Sindh, Gujarat, and Bombay, the majority of publications were produced by Devraj. Therefore, Devraj’s font 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.

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

⁵ Personal correspondence with Shafique Virani, February 2009.

4.2 Virāma

The function of Khojki 𑂏 VIRAMA is identical to the corresponding character in Devanagari. It is written to the right of the consonant letter with which it combines. The form 𑂏 is attested in a few printed books, eg. *Sindhī pahelī copadī* (Bombay: Khoja Sindhi Printing Press, 1910), but the use is a deliberate attempt to mimic the form of the sign in Devanagari.

4.3 Vowel Letters

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

The diphthongs are generally represented using diagraphs, eg. 𑂏𑂑 A + 𑂑 → 𑂏𑂑𑂑 AI; 𑂏 + 𑂒 → 𑂏𑂒 AU. The independent letters 𑂏𑂑 AI and 𑂏𑂒 AU are attested in manuscripts and books, and are therefore proposed for encoding as atomic characters.

Several variant forms of vowel letters are attested: 𑂏𑂑 = 𑂑 E; 𑂏𑂑 and 𑂏𑂑 = 𑂑 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. The atomic character for the vowel letter should always be used, as shown below:

RECOMMENDED	NOT RECOMMENDED
𑂏𑂑 VOWEL LETTER AA	𑂏 VOWEL LETTER A + 𑂑 VOWEL SIGN AA
𑂏𑂑 VOWEL LETTER AI	𑂏 VOWEL LETTER A + 𑂑 VOWEL SIGN AI
𑂏𑂒 VOWEL LETTER AU	𑂏 VOWEL LETTER A + 𑂒 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 it using encoded forms. For example, the 𑂏𑂑 variant of 𑂑 O should not be represented as 𑂏 A + 𑂑 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 𑂑 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).

4.5 Consonant Letters

The repertoire of consonant letters follows the pattern for Sindhi-based Landa scripts.

As discussed in Section 3.2, there is great variation in character-glyph assignments for consonant letters. The following assignments of glyphs to characters have been made in this proposal:

- 𑂑 Assigned to GA. Also attested for GGA.
- 𑂒 Assigned to GGA. Also attested for GA. Also used for conjunct GA + VIRAMA + RA.

- 𑌒 Assigned to JA. Also attested for JJA and *JHA.
- 𑌓 Assigned to JJA. Also attested for JA and *JHA.
- 𑌔 Assigned to DA. Also attested for DDDA and DHA.
- 𑌕 Assigned to DDDA. Also attested for DA.
- 𑌖 Assigned to DHA. Also attested for DA and *JHA.
- 𑌗 Assigned to NNA. Also attested for 𑌘 DDA.
- 𑌙 Assigned to BA. Also attested for BBA and BHA.
- 𑌚 Assigned to BBA. Also attested for BA and BHA.
- 𑌛 Assigned to BHA. Also attested for BBA.

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:

- **◌I VOWEL SIGN I** This is written to the right of the consonant. It has the contextual form ◌[◌]. The regular sign ◌I is written with characters that have right descenders: 𑌒 CA → 𑌒[◌] ci, 𑌕 VA → 𑌕[◌] vi, etc.; and those that possess a final right stroke that terminates at the baseline: 𑌔 KA → 𑌔[◌] ki, 𑌔 DA → 𑌔[◌] di, 𑌖 HA → 𑌖[◌] hi. The variant ◌[◌] is used with characters that have rounded right edges: 𑌕 CHA → 𑌕[◌] chi, 𑌛 NGA → 𑌛[◌] ni, 𑌚 TTA → 𑌚[◌] ti, 𑌕 DDDA → 𑌕[◌] di, 𑌖 RA → 𑌖[◌] ri.
- **◌II VOWEL SIGN II** This sign is written to the right of the consonant letter. It has the contextual form ◌[◌]. The regular sign ◌II is used with characters that have rounded right edges: 𑌕 CHA → 𑌕[◌] chī. The variant form ◌[◌] is written with characters possessing right descenders: 𑌒 CA → 𑌒[◌] cī, 𑌕 NA → 𑌕[◌] nī, etc., and with characters with final strokes that terminate at the baseline: 𑌔 KA → 𑌔[◌] kī, 𑌙 PHA → 𑌙[◌] phī, 𑌖 HA → 𑌖[◌] hī.
- **◌U VOWEL SIGN U** This sign is generally written as an extension of the basic final stroke of a character: 𑌔 KA → 𑌔[◌] ku, 𑌕 CHA → 𑌕[◌] chu, etc. With characters that possess a right descender, it is written at the terminus of the descender: 𑌒 CA → 𑌒[◌] cu, 𑌙 GHA → 𑌙[◌] ghū, 𑌛 YA → 𑌛[◌] yu, etc. It attaches to the bottom of the glyph for characters with rounded shapes: 𑌚 TTHA → 𑌚[◌] 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 ligatures in Khojki: 𑌛 kṣa (𑌔 KA + VIRAMA + *SSA); 𑌒 jña (𑌒 JA + VIRAMA + 𑌛 NYA); 𑌙 tra (𑌙 TA + VIRAMA + 𑌖 RA); 𑌚 dra (𑌔 DA + VIRAMA + 𑌖 RA). These are represented as $C_1 + \text{VIRAMA} + C_2$.

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

- 2. Full-Form of Consonant + Visible Virama** This is the most common method for writing consonant conjuncts, eg. **ꣳꣳꣳ** *kya*. Such conjuncts are represented as $C_1 + \text{VIRAMA} + \text{ZWNJ} + C_2$.
- 3. Full-Form of Consonant + Reduced Form of Final Consonant** This type of conjunct is typically used for representing clusters involving **ꣳ** 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.

ꣳ = d, ꣳ, ꣳ; ꣳ = dh; special form of conjuncts: ꣳ = br;
 ꣳ = kr; ꣳ = sr; ꣳ = mr; ꣳ = ny; ꣳ = ky; ꣳ = ly; ꣳ = ty;
 ꣳ = jy; ꣳ = sv; ꣳ = gn.

In some cases, **ꣳꣳ** YA also appears as the second element in a conjunct using the reduced form **ꣳ**, eg. **ꣳꣳꣳ** *kya*.

Ms Ism K 15

[190-?]

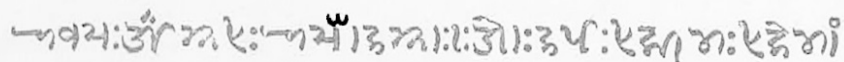
Local paper; 44 folios; 65; 17 x 14 (13 x 10); 11 lines; multiple sections bound with two sets of sewing: side-sewn and stab-sewn with cords pierced through and laced into laminated layers forming thin paper covers; spine rounded and backed; black ink.

ꣳ = d, ꣳ, ꣳ; ꣳ = dh; special form of conjuncts: ꣳ = br;
 ꣳ = kr; ꣳ = sr; ꣳ = mr; ꣳ = ny; ꣳ = ky; ꣳ = ly; ꣳ = ty;
 ꣳ = jy; ꣳ = sv; ꣳ = gn.

This type of conjunct is represented as $C_1 + \text{ZWJ} + \text{VIRAMA} + C_2$.

4.8 Consonant Gemination

Geminated consonants are marked by the sign **ꣳ** SHADDA. The sign is written above the consonant that is doubled, eg. LA + LA is written **ꣳꣳ** (as below), not as a conjunct character. This practice is similar to that in Gurmukhi, where the **ꣳ** U+0A71 GURMUKHI SIGN ADDAK is used to denote doubled consonants, and also in the Arabic script, where the **ꣳ** U+0651 ARABIC SHADDA is used.



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

4.9 Nasalization

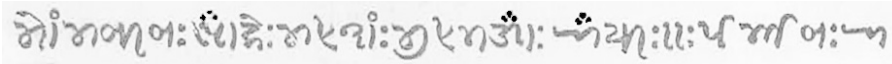
Nasalization is indicated by the sign **ꣳ** ANUSVARA. It is written to the right of the letter or sign with which it combines, as shown below. Also shown is the character **ꣳ**, which is written centered above the combining base. It is regarded as a variant form of ANUSVARA and is attested in several manuscripts and lithographs.



It is possible that it may be represent a contextual variant of ANUSVARA or a character similar to the ॠ CANDRABINDU found in other Indic scripts. In the specimen, the form ॠ appears only in conjunction with ॡ VOWEL SIGN AA, which may affirm its CANDRABINDU-like function of vowel nasalization. Further analysis is required to determine if there is any true semantic distinction between the forms.

4.10 NUKTA

The ॠ NUKTA is used for producing characters to represent sounds not native to Sindhi. The sign may be written with vowel letters, vowel signs, and consonant letters. Unlike in other Indic scripts, the Khojki NUKTA is written above the letter with which it combines:



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
‘	ॠ VOWEL SIGN A + NUKTA	ع U+0639 ARABIC LETTER AIN
‘	𑂀 LETTER A + NUKTA	ع U+0639 ARABIC LETTER AIN
‘	𑂁 LETTER E + NUKTA	ع U+0639 ARABIC LETTER AIN
qa	𑂂 KA + NUKTA	ق U+0642 ARABIC LETTER QAF
kha	𑂃 KHA + NUKTA	خ U+062E ARABIC LETTER KHAH
ga	𑂄 GA + NUKTA	غ U+063A ARABIC LETTER GHAIN
za	𑂅 JA + NUKTA	ذ U+0630 ARABIC LETTER THAL
za	𑂆 JA + NUKTA	ز U+0632 ARABIC LETTER ZAIN
za	𑂇 JA + NUKTA	ض U+0636 ARABIC LETTER DAD
za	𑂈 JA + NUKTA	ظ U+0638 ARABIC LETTER ZAH
ṭa	𑂉 TA + NUKTA	ط U+0637 ARABIC LETTER TAH
fa	𑂊 PHA + NUKTA	ف U+0641 ARABIC LETTER FEH
śa	𑂋 SA + NUKTA	ش U+0634 ARABIC LETTER SHEEN
sa	𑂌 SA + NUKTA	ص U+0635 ARABIC LETTER SAD
ḥa	𑂍 HA + NUKTA	ح U+062D ARABIC LETTER HAH

The NUKTA has a variant, which is the inverted form ॡ.

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

1. 𑂎 = عليه السلام ‘alai-hi as-salām ‘peace be upon him’.
2. 𑂏 = صلى الله عليه و سلم şallā allāhu ‘alai-hi wa sallam ‘God bless him and grant him salvation’.

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; ۛ = ū, o; ۛ = z; ۛ = j; ۛ = shaddah; ۛ = abbreviation for Arabic benediction ʿalai-hi as-salam "peace be upon him"; ۛ = abbreviation for Arabic benediction ḡallā allāhu ʿalai-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.

4.11 Punctuation

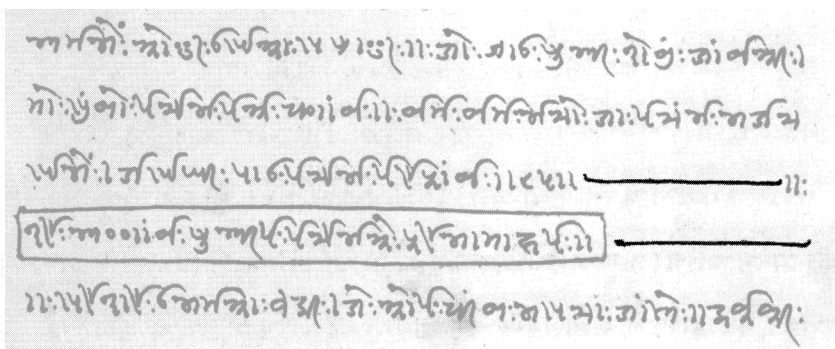
Word Boundaries Khojki separates words using ۛ WORD SEPARATOR:

The WORD SEPARATOR may appear at the beginning and end of a line. It resembles characters such as ۛ U+0903 DEVANGARI SIGN VISARGA; Khojki does not use VISARGA.

Sentence Boundaries The ۛ DANDA and ۛ DOUBLE DANDA are used to mark the end of sentences.

DOUBLE DANDA is also used to mark verse sections. Typically, DOUBLE DANDA is written with WORD SEPARATOR to the left and right of verse numbers:

Section Marks The characters ۛ SECTION MARK and ۛ DOUBLE SECTION MARK appear frequently in Khojki manuscripts as punctuation that delimit the end of a section or another larger block of text. Both section marks are found in manuscripts. The SECTION MARK is generally used to mark the end of a sentence, while DOUBLE SECTION MARK is used to delimit larger blocks of text, such as paragraphs. Both generally extend to the margin of the text-block.



Verse Number Marks Verse numbers are often marked in Khojki texts using an overline. The overline spans all of the digits in a number and marks them as belonging to a set. This feature may be represented using $\overline{\text{U+0305}}$ COMBINING OVERLINE or at the presentation level. The drawback is that these solutions visually represent the verse marker, but they do not capture the semantic properties of the character.

Other Punctuation Printed Khojki employs Latin punctuation marks:

પડા જાન નાન મેબ મ'વા,
 બબા રાખા ઝાપ;
 પડા નજારે નાબા હ' રહી,
 ચડા નજારા કો રાપ... ...શકે ૨૦ (૧૩)

4.12 Abbreviations

The \bullet ABBREVIATION SIGN is used for marking abbreviations:

જાદ જોમલ જાદો નાબાજો,
 કવે નજો રકેહી' ચડાપ... ...શકે ૨૦ (૧૦)
 પડા જાદ ગુગમાં નબા પ્રપે હ' કહા,
 બગરા મોપ બ પાપ;
 બગરા પહો જાદા બાપડા,
 શકે ૨ જરો ડાહુરાપ... ...શકે ૨૦ (૧૧)

4.13 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+0A66..U+0A6F). See Figure 6 for an example of Gujarati digits in printed Khojki text.

4.14 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. $\frac{\text{U+A833}}{\text{U+A831}}$ 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).⁶

4.15 Collation

The collating order for Khojki is as follows:

ॠ A > ॡ AA > ॢ I > ॣ U > । E > ॥ AI > ० O > ॠ AU > ॡ KA > ॢ KHA > ॣ GA > । GGA >
 ॥ GHA > ०: NGA > ॠ CA > ॢ CHA > ॣ JA > । JJA > ०: NYA > ॠ TTA > ॢ TTHA > ॣ DDA >
 । DDHA > ॠ NNA > ॢ TA > ॣ THA > । DA > ॠ DDDA > ॢ DHA > ॣ NA > । PA > ॠ PHA >
 ॢ BA > ॣ BBA > । BHA > ॠ MA > ॠ YA > ॢ RA > ॣ LA > । VA > ॠ SA > ॢ HA > ॣ LLA >
 ॠ SIGN AA > ॢ SIGN I > ॣ SIGN II > । SIGN U > ॠ SIGN E > ॡ SIGN AI > ॢ SIGN O >
 ॠ SIGN AU > ॠ ANUSVARA > ॠ VIRAMA

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

4.16 Linebreaking

Linebreaking for all letters and signs follow the same rules for Devanagari. The SECTION MARK and DOUBLE SECTION MARK function like the corresponding characters in Kaithi. The WORD SEPARATOR may appear at the beginning and end of a line.

4.17 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;<reserved>
11204;KHOJKI LETTER U;Lo;0;L;;;;;N;;;;;
11205;<reserved>
11206;KHOJKI LETTER E;Lo;0;L;;;;;N;;;;;
11207;KHOJKI LETTER AI;Lo;0;L;;;;;N;;;;;
11208;KHOJKI LETTER O;Lo;0;L;;;;;N;;;;;
11209;KHOJKI LETTER AU;Lo;0;L;;;;;N;;;;;
1120A;KHOJKI LETTER KA;Lo;0;L;;;;;N;;;;;
1120B;KHOJKI LETTER KHA;Lo;0;L;;;;;N;;;;;
1120C;KHOJKI LETTER GA;Lo;0;L;;;;;N;;;;;
1120D;KHOJKI LETTER GGA;Lo;0;L;;;;;N;;;;;
1120E;KHOJKI LETTER GHA;Lo;0;L;;;;;N;;;;;
1120F;KHOJKI LETTER NGA;Lo;0;L;;;;;N;;;;;

```

⁶ See Pandey (2007) for more information on these characters.

```

11210;KHOJKI LETTER CA;Lo;0;L;;;;N;;;;;
11211;KHOJKI LETTER CHA;Lo;0;L;;;;N;;;;;
11212;KHOJKI LETTER JA;Lo;0;L;;;;N;;;;;
11213;KHOJKI LETTER JJA;Lo;0;L;;;;N;;;;;
11214;<reserved>
11215;KHOJKI LETTER NYA;Lo;0;L;;;;N;;;;;
11216;KHOJKI LETTER TTA;Lo;0;L;;;;N;;;;;
11217;KHOJKI LETTER TTHA;Lo;0;L;;;;N;;;;;
11218;KHOJKI LETTER DDA;Lo;0;L;;;;N;;;;;
11219;KHOJKI LETTER DDHA;Lo;0;L;;;;N;;;;;
1121A;KHOJKI LETTER NNA;Lo;0;L;;;;N;;;;;
1121B;KHOJKI LETTER TA;Lo;0;L;;;;N;;;;;
1121C;KHOJKI LETTER THA;Lo;0;L;;;;N;;;;;
1121D;KHOJKI LETTER DA;Lo;0;L;;;;N;;;;;
1121E;KHOJKI LETTER DDDA;Lo;0;L;;;;N;;;;;
1121F;KHOJKI LETTER DHA;Lo;0;L;;;;N;;;;;
11220;KHOJKI LETTER NA;Lo;0;L;;;;N;;;;;
11221;KHOJKI LETTER PA;Lo;0;L;;;;N;;;;;
11222;KHOJKI LETTER PHA;Lo;0;L;;;;N;;;;;
11223;KHOJKI LETTER BA;Lo;0;L;;;;N;;;;;
11224;KHOJKI LETTER BBA;Lo;0;L;;;;N;;;;;
11225;KHOJKI LETTER BHA;Lo;0;L;;;;N;;;;;
11226;KHOJKI LETTER MA;Lo;0;L;;;;N;;;;;
11227;KHOJKI LETTER YA;Lo;0;L;;;;N;;;;;
11228;KHOJKI LETTER RA;Lo;0;L;;;;N;;;;;
11229;KHOJKI LETTER LA;Lo;0;L;;;;N;;;;;
1122A;KHOJKI LETTER VA;Lo;0;L;;;;N;;;;;
1122B;KHOJKI LETTER SA;Lo;0;L;;;;N;;;;;
1122C;KHOJKI LETTER HA;Lo;0;L;;;;N;;;;;
1122D;KHOJKI LETTER LLA;Lo;0;L;;;;N;;;;;
1122E;<reserved>
1122F;KHOJKI VOWEL SIGN AA;Mn;0;NSM;;;;N;;;;;
11230;KHOJKI VOWEL SIGN I;Mc;0;L;;;;N;;;;;
11231;KHOJKI VOWEL SIGN II;Mc;0;L;;;;N;;;;;
11232;KHOJKI VOWEL SIGN U;Mn;0;NSM;;;;N;;;;;
11233;<reserved>
11234;KHOJKI VOWEL SIGN E;Mn;0;NSM;;;;N;;;;;
11235;KHOJKI VOWEL SIGN AI;Mn;0;NSM;;;;N;;;;;
11236;KHOJKI VOWEL SIGN O;Mn;0;L;;;;N;;;;;
11237;KHOJKI VOWEL SIGN AU;Mn;0;L;;;;N;;;;;
11238;<reserved>
11239;KHOJKI SIGN ANUSVARA;Mn;0;NSM;;;;N;;;;;
1123A;KHOJKI SIGN VIRAMA;Mc;9;NSM;;;;N;;;;;
1123B;KHOJKI SIGN NUKTA;Mn;7;NSM;;;;N;;;;;
1123C;KHOJKI SIGN SHADDA;Mn;0;NSM;;;;N;;;;;
1123D;KHOJKI DANDA;Po;0;L;;;;N;;;;;
1123E;KHOJKI DOUBLE DANDA;Po;0;L;;;;N;;;;;
1123F;KHOJKI WORD SEPARATOR;Po;0;L;;;;N;;;;;
11240;KHOJKI SECTION MARK;Po;0;L;;;;N;;;;;
11241;KHOJKI DOUBLE SECTION MARK;Po;0;L;;;;N;;;;;
11242;KHOJKI ABBREVIATION SIGN;Po;0;L;;;;N;;;;;

```

4.18 ‘Confusable’ Characters

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

```
1123F KHOJKI WORD SEPARATOR ; 0903 DEVANAGARI SIGN VISARGA
```

Other characters may be added to the list in the future.

5 Characters Not Proposed

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 𑂔 I. However, Figure 14 shows the glyph 𑂕 for II. This glyph may be a variant of 𑂔 I or 𑂕 YA.
2. **UU** Independent forms of U and UU are generally not distinguished and both are represented by 𑂖 U. However, Figure 14 shows the glyph 𑂗 used for UU. This glyph may be a variant form of 𑂖 U or 𑂘 O.
3. **JHA** There does not appear to be a unique glyph for this character. Those assigned to 𑂙 JA and 𑂚 JJA have been used for representing JHA. There are glyphs assigned to the character in Figure 12 and Figure 17, such as 𑂛, but this form is similar in shape to 𑂙.
4. **SHA** The glyph 𑂜 is used for writing SHA, but it is also used for representing any non-native Sindhi sibilant, eg. 𑂜 U+0635 ARABIC LETTER SAD. It may be composed using the sequence 𑂜 SA + 𑂝 NUKTA.
5. **QA** Some sources show the character 𑂞 used for representing /q/ (𑂞 U+0642 ARABIC LETTER QAF):

P Ism K 14

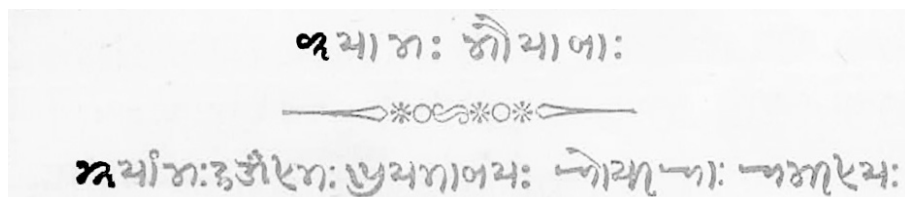
Bombay 1962 s./1905

Elam sār saṅgrah rāg māla

Publisher: Lālājībhaī Devrāj, Khojā Sifdhī Āhāpākhanū; 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); *tavṣīlo* (index): pp. 2.135-138; list of publications of the Khojā Sifdhī Printing Press pp. 3.1-6.

𑂞 = qāf; 𑂟 = g, gr; 𑂠 = ain; 𑂡 = 1?; 𑂢 = vowel-less consonant.

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



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

P I s m K 11

Bombay 1959 s./1902

Ārtī:dhūā:vakhat:trejī:sāñjijā:ćogađiā:5

Publisher: Khojā Lāljiḥhāi Devrāj, Khojā Siñdhī Čhāpākhānū for the Imāmi Ismāiī 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 Siñdhī Printing Press and an announcement of the publication of other works.

𑂔 = z; 𑂕 = j; 𑂖 = k; 𑂗 = ‘ain; 𑂘 = g; 𑂙 = gain; shaddah for doubled consonant or emphasis.

6. **Characters for kṣa, jña, tra, dra** Khojki charts traditionally show the characters 𑂔 kṣa, 𑂕 jña, 𑂖 tra, and 𑂗 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 charts of Devanagari for languages such Hindi. They are enumerated as characters of various scripts on account of their unique shapes, which masks the consonant characters that constitute them. They are not atomic characters, but consonant conjuncts, and they should be handled as such.
7. **VOWEL SIGN UU** A character for a dependent VOWEL SIGN UU that is dissimilar from 𑂚 VOWEL SIGN U has not been identified.
8. **VOWEL SIGN VOCALIC R** Authors of Khojki instructional materials at times attempted to fix certain deficiencies of the script by borrowing orthographic practices from scripts such as Devanagari. One such script primer, *Sindhī pahelī copadī*, introduces 𑂛 VOWEL SIGN VOCALIC R and various diacritics to represent conjunct conjuncts with RA. The specimen below illustrates such uses:

P I s m K 17

Bombay 1966 s./ 1910

Siñdhī pahelī ćopađī. 1st edition.

Publisher: Mūkhī Lāljiḥhāi Devrāj, Manager of the Bombay Shia Imami Ismaili Khoja Jamat Dept of Education, for the Siñdhī Textbook Committee at the Khoja Siñdhī Printing Press; 2000 copies; 92 pp. in various paginations; 17 x 13; covers missing; binding in fair condition; paper in good condition; sañkađiūñ (table of contents): 2.1-3.

According to publisher's note (pp. 1.3-4), since the Khojki script lacks the facility to make "true" conjunct characters, the publication uses Sanskrit ones. Teachers are requested to familiarize themselves with these forms before proceeding with instruction. Special uses: (1) 𑂛 under the first member of a conjunct consonant suppresses the implicit short "a" vowel; (2) if the letter "r" is the second member of a conjunct, the superscript sign 𑂛 is used on the first member i.e. 𑂛𑂔 for "pr"; (3) Subscript 𑂛 with the consonant "r" indicates "r" as the first member of the conjunct i.e. 𑂛𑂔 for "pūrva"; (4) the vowel "ru" is indicated by subscript 𑂛; (5) the following conjuncts introduced: 𑂛 = Gujarati 𑂛; 𑂛 = Gujarati 𑂛;

𑂛 = Gujarati 𑂛; 𑂛 = Gujarati 𑂛;
The script does not use ":" to separate individual words.

6 References

- Asani, Ali S. 1987. "The Khojkī Script: A Legacy of Ismaili Islam in the Indo-Pakistan Subcontinent". *Journal of the American Oriental Society*, vol. 107, no. 3 (Jul. - Sep. 1987), pp. 439-449.
- . 1992. *The Harvard Collection of Ismaili Literature in Indic Languages: A Descriptive Catalog and Finding Aid*. Boston: G. K. Hall & Co.

- . 2002. *Ecstasy and Enlightenment: The Ismaili Devotional Literature of South Asia*. London: I. B. Tauris & Co. in association with The Institute of Islamic Studies.
- Daftary, Farhad. 2008. “Ismaili History and Literary Traditions”. In *An Anthology of Ismaili Literature: A Shi‘i Vision of Islam*. edited by Hermann Landolt, Samira Sheikh, and Kutub Kassam, pp. 1–29. London: I. B. Tauris & Co. in association with The Institute of Islamic Studies.
- Grierson, George A. 1905. *The Linguistic Survey of India*. Vol. VIII. Indo-Aryan family. North-western group: pt. I. Specimens of Sindhī and Lahndā. Calcutta: Office of the Superintendent of Government Printing, India.
- 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.
- Ilana, Ghulam Ali. 1969. *سنڌي سُورَتَخَطِي* [= *Sindhī Sūratkhaṭī*]. 3rd ed. Hyderabad: Sindhi Zaban Publications.
- Institute of Ismaili Studies. 2007. “Khojki Alphabet”. http://iis.ac.uk/view_article.asp?ContentID=108952
- 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. “Proposal to Encode the Sindhi Script in ISO/IEC 10646.” N3871 L2/10-271R. September 10, 2010. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3871.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, Mumtaz Ali Sadik Ali. 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

The proposal author is 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.

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.

	1120	1121	1122	1123	1124
0	𑂔 11200	𑂕 11210	𑂖 11220	𑂗 11230	𑂘 11240
1	𑂙 11201	𑂚 11211	𑂛 11221	𑂜 11231	𑂝 11241
2	𑂞 11202	𑂟 11212	𑂠 11222	𑂡 11232	𑂢 11242
3		𑂣 11213	𑂤 11223		
4	𑂥 11204		𑂦 11224	𑂧 11234	
5		𑂨 11215	𑂩 11225	𑂪 11235	
6	𑂫 11206	𑂬 11216	𑂭 11226	𑂮 11236	
7	𑂯 11207	𑂰 11217	𑂱 11227	𑂲 11237	
8	𑂳 11208	𑂴 11218	𑂵 11228		
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		𑃌 1123E	
F	𑃍 1120F	𑃎 1121F	𑃏 1122F	𑃐 1123F	

Figure 1: Proposed code chart for Khojki.

Independent vowels

11200	𑀀	KHOJKI LETTER A
11201	𑀁	KHOJKI LETTER AA
11202	𑀂	KHOJKI LETTER I
11203	𑀃	<reserved>
11204	𑀄	KHOJKI LETTER U
11205	𑀅	<reserved>
11206	𑀆	KHOJKI LETTER E
11207	𑀇	KHOJKI LETTER AI
11208	𑀈	KHOJKI LETTER O
11209	𑀉	KHOJKI LETTER AU

1123E	𑀒	KHOJKI DOUBLE DANDA
1123F	𑀓	KHOJKI WORD SEPARATOR
11240	𑀔	KHOJKI SECTION MARK
11241	𑀕	KHOJKI DOUBLE SECTION MARK
11242	𑀖	KHOJKI ABBREVIATION SIGN

Consonants

1120A	𑀊	KHOJKI LETTER KA
1120B	𑀋	KHOJKI LETTER KHA
1120C	𑀌	KHOJKI LETTER GA
1120D	𑀍	KHOJKI LETTER GGA
1120E	𑀎	KHOJKI LETTER GHA
1120F	𑀏	KHOJKI LETTER NGA
11210	𑀐	KHOJKI LETTER CA
11211	𑀑	KHOJKI LETTER CHA
11212	𑀒	KHOJKI LETTER JA
11213	𑀓	KHOJKI LETTER JJA
11214	𑀔	<reserved>
11215	𑀕	KHOJKI LETTER NYA
11216	𑀖	KHOJKI LETTER TTA
11217	𑀗	KHOJKI LETTER TTHA
11218	𑀘	KHOJKI LETTER DDA
11219	𑀙	KHOJKI LETTER DDHA
1121A	𑀚	KHOJKI LETTER NNA
1121B	𑀛	KHOJKI LETTER TA
1121C	𑀜	KHOJKI LETTER THA
1121D	𑀝	KHOJKI LETTER DA
1121E	𑀞	KHOJKI LETTER DDDA
1121F	𑀟	KHOJKI LETTER DHA
11220	𑀠	KHOJKI LETTER NA
11221	𑀡	KHOJKI LETTER PA
11222	𑀢	KHOJKI LETTER PHA
11223	𑀣	KHOJKI LETTER BA
11224	𑀤	KHOJKI LETTER BBA
11225	𑀥	KHOJKI LETTER BHA
11226	𑀦	KHOJKI LETTER MA
11227	𑀧	KHOJKI LETTER YA
11228	𑀨	KHOJKI LETTER RA
11229	𑀩	KHOJKI LETTER LA
1122A	𑀪	KHOJKI LETTER VA
1122B	𑀫	KHOJKI LETTER SA
1122C	𑀬	KHOJKI LETTER HA
1122D	𑀭	KHOJKI LETTER LLA

Dependent vowel signs

1122F	𑀮	KHOJKI VOWEL SIGN AA
11230	𑀯	KHOJKI VOWEL SIGN I
11231	𑀰	KHOJKI VOWEL SIGN II
11232	𑀱	KHOJKI VOWEL SIGN U
11233	𑀲	<reserved>
11234	𑀳	KHOJKI VOWEL SIGN E
11235	𑀴	KHOJKI VOWEL SIGN AI
11236	𑀵	KHOJKI VOWEL SIGN O
11237	𑀶	KHOJKI VOWEL SIGN AU

Various signs

11239	𑀷	KHOJKI SIGN ANUSVARA
1123A	𑀸	KHOJKI SIGN VIRAMA
1123B	𑀹	KHOJKI SIGN NUKTA
1123C	𑀺	KHOJKI SIGN SHADDA

Punctuation

1123D	𑀻	KHOJKI DANDA
-------	---	--------------

Figure 2: Proposed names list for Khojki.

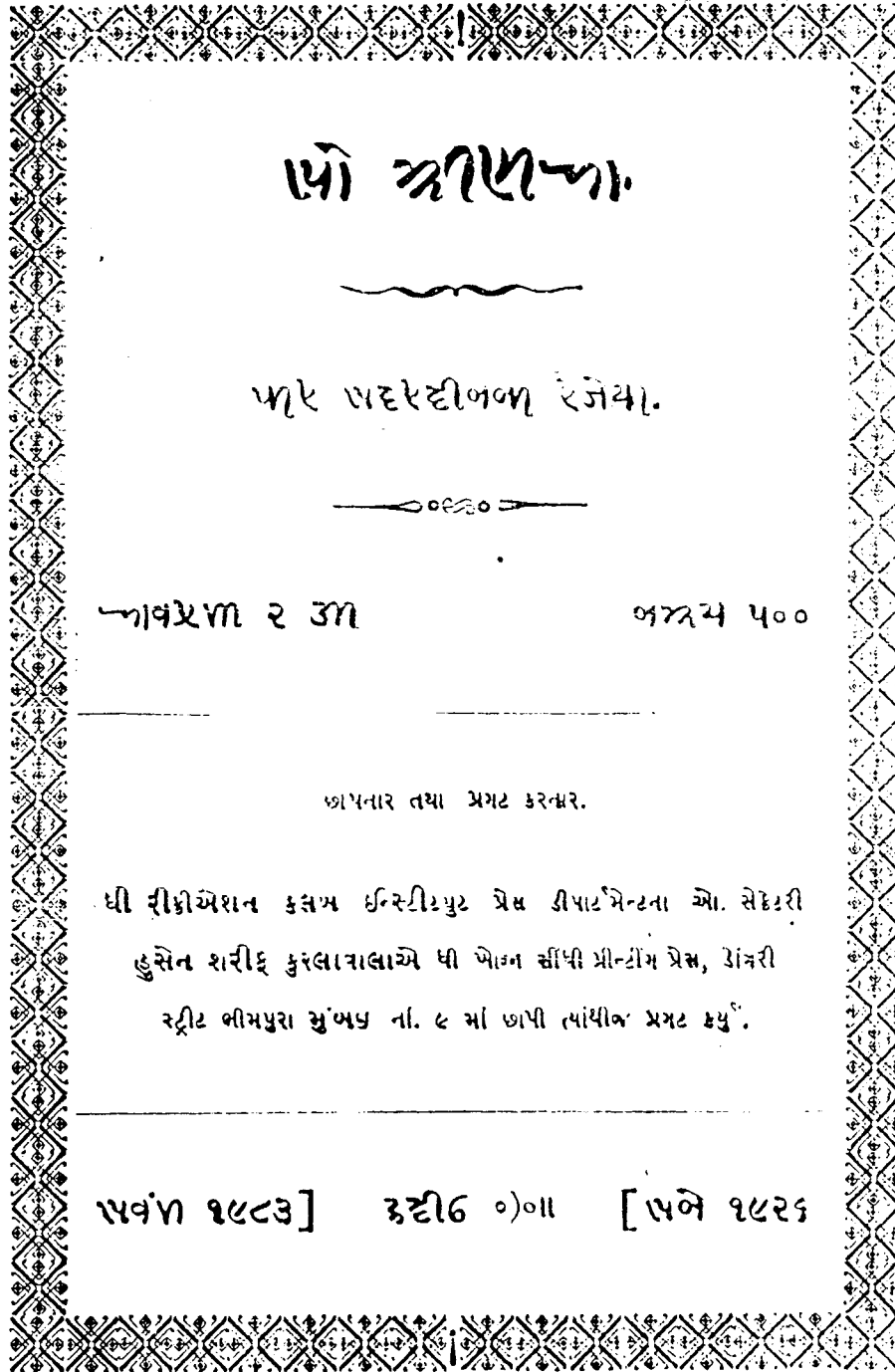


Figure 3: Title page of the *સો કીરણા* (*So Kirā*) printed in Khojki and Gujarati (from Shackle and Moir 1992: 33). North Indic number forms and the U+A838 NORTH INDIC RUPEE MARK are used to indicate the price of the item.

Ms Ism K 1	[Bhadneshwar] [192-?]]	11-13 (same rāg)	māfhe:varāsete:abharat:nūtake ਮਫ਼ਹੇ:ਵਰਾਸੇਟੇ:ਅਬਹਾਰਤ:ਨੂਟਾਕੇ
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:ramājō:dhanane:rātake ਮਠੇ:ਰਮਾਜੋ: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; ਪ = pr.		15-18 (same rāg)	19 verses, G, Pīr Shams tame:japājō:dhanane:rātake ਮਠੇ:ਜਪਾਜੋ: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:ਸਮਾਸ:ਪੀਰਾਕੇ
1-2 (a.t.)	ejī:ād:gūr:samas:mūnivar:jā:n ਏਜੀ:ਆਦ:ਗੁਰ:ਸਮਾਸ: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:ਮਨਕੇ: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:ਸਮਾਸ: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:ਸਾਬਠਾਫੋ
		33-35 (rāg ṭhāl/ḡhāl?)	18 verses, G, Pīr Shams tathī:āṭhamī:āvā:gāmanā:lokake ਟਾਠੀ:ਆਟਠਾਮੀ: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).

𑂀	𑂁	𑂂	𑂃	𑂄
a	i	u	e	o
𑂅	𑂆	𑂇	𑂈	𑂉:
ka	kha	ga	gha	ṅ
𑂊	𑂋	𑂌	𑂍	𑂎:
ca	cha	za	ja	ñ
𑂏	𑂐	𑂑	𑂒	𑂓
ṭa	ṭha	ḍa	ḍha	ṇ
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
śa	ha	ḷa	kṣa	gna
𑂨	𑂩	𑂪	𑂫	𑂬
dhr	tr	ga	dha	bha

Figure 9: Inventory of Khojki letters (from Institute of Ismaili Studies 2007).

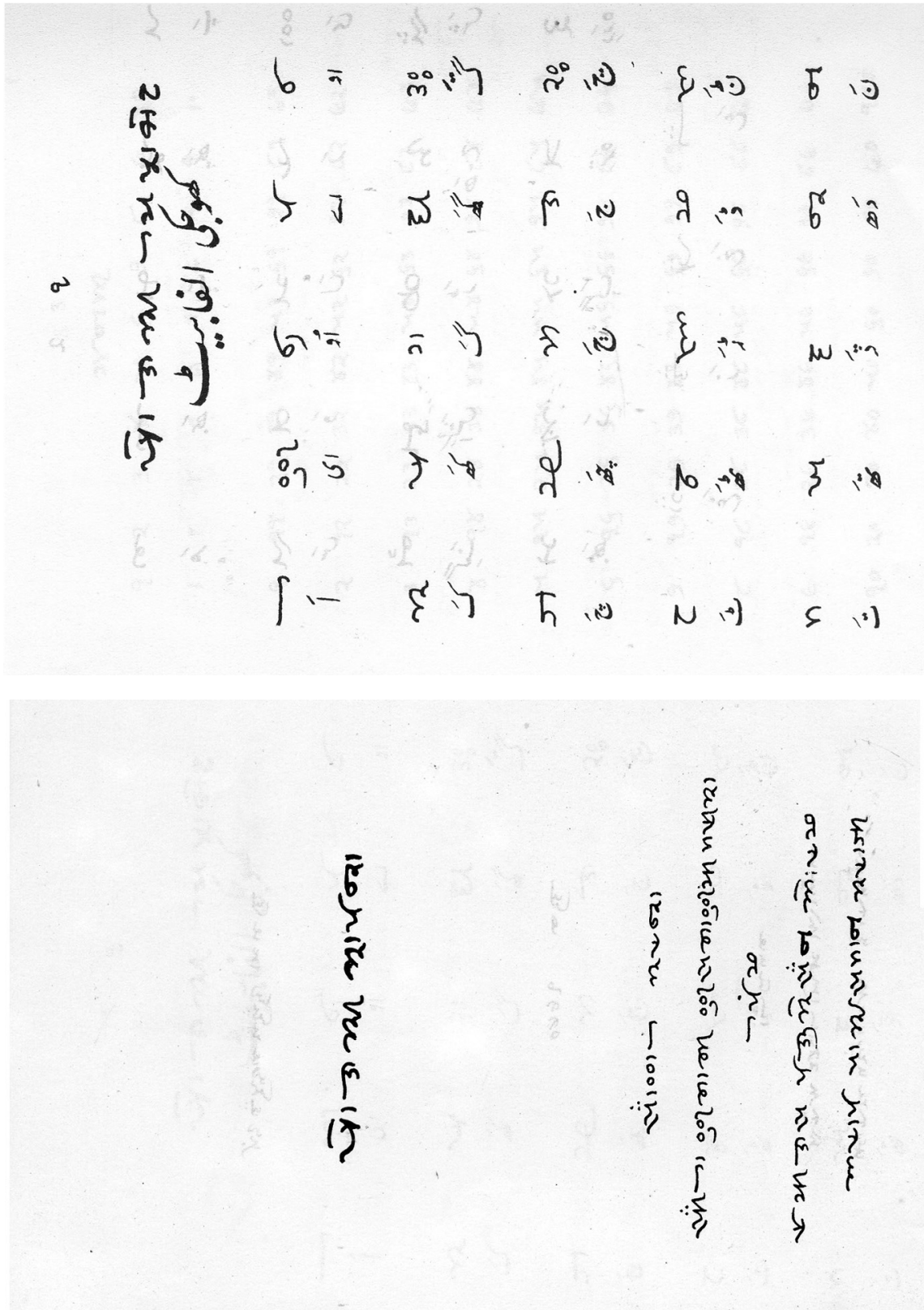


Figure 10: Cover and first page of a Khojki primer from 1990. The chart shows the phonetic values of Khojki letters in present-day Pakistan. Images courtesy of Shafique Virani.

Roman characters.	Devanāgarī.	Khudāwādī.*	Shikārpurī.	Sakkar.	THATTĀI.		Lārī.	Wangāl.	Rājī.	Khawājā.	MĀIMONS.		Sewhāpī Bhabhīrā.	Southern Lohndā.
					Luhāndā.	Bhāsīā.					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 (‘Khawājās’) vowel letters from Grierson (1905: 15).

Roman characters.	Devanagari.	Khudawadi.	Shikarpuri.	Sakkar.	THATTAL.			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.		Larāī.	Wangāī.	Rājjarī.	Khawājās.	MAIMONB.		Sewhānī Bhabhīras.	Southern Lahndā.
					Juhānās.	Bhātīās.					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 ('Khawājās') 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	𑂀, 𑂁 𑂀 [ɔː] (𑂀 𑂀?)
i*	ا	ا	इ	𑂀 [ɔː] (𑂀 𑂀?)
ī*	ای	ای	ई	𑂀 [ɔː] (𑂀 𑂀?)
u*	ا	ا	उ	𑂀 [ɔː]
ū*	اؤ	اؤ	ऊ	𑂀 [ɔː] (𑂀?)
o	او	او	ओ	𑂀 [ɔː, ɔː] (𑂀?)
e	اے	ای	ऐ	𑂀 [ɔː] (𑂀?)
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	બૅ	બૅ [બૅ]
bāḍ	باڊ	باڊ	બાૅ	બાૅ
din	ڊن	ڊن	દિન	દૅન
dīn	ڊين	ڊين	દીન	દાૅન
but*	بُت	بُت	બુત	બુમ
būt*	بوت	بوت	બૂત	બુમ
so	سو	سو	સૅ	સૅ [સૅ]
be	بے	بی	બૅ	બૅ [બૅ]
hai	ہائے	ہائی	હૅ	[હૅ, હૅ-૫]
nau	نَو	نَو	નાૅ	[નાૅ, નૅ-૫]

*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	ભ	૧ [મ, ૫]
b̥		ب̇		૧
bh	بھ	پ	ભ	૫, ૫ [મ]
t	ت	ت	ત	૫
th	تھ	تھ	થ	૫ [મ]
t̥	ط	ت̇	ટ	૮, ૮, (૫?)
th̥	طھ	ت̇ھ	ઠ	૮
s	س	س	(સ)	(૫)
p	پ	پ	પ	૫ (૫?)
ph	پھ	پھ	ફ	૫
j	ج	ج	જ	૩, ૫, ૩ (૫?)
dy		ج̇		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	ر	ر	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	𑂣
ṣ	ص	ص	(સ)	(૫)
z	ض	ض	(ઝ)	(૩, ૨૫)
t	ط	ط	(ત, ૨)	(૧, ૮)
ẓ	ظ	ظ	(ઝ)	(૩, ૨૫)
ʿ	ع	ع	(અ)	𑂣, 𑂣 (—)
ġ	غ	غ	(ગ)	𑂣 (૨)
f	ف	ف	ફ	𑂣
q	ق	ق	(ક)	𑂣 (૪)
k	ک	ک	(ક)	𑂣 [𑂣]
kh	کھ	ک	ખ	𑂣, 𑂣 (—?)
g	گ	گ	ગ	𑂣 [𑂣, 𑂣]
ng		گب		𑂣
gh	گھ	گھ	ઘ	𑂣 (૨?)
ng		نگ	ঙ	𑂣: (૨:?)
l	ل	ل	લ	𑂣 [𑂣]

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			ઘ	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
						𑂊	𑂋	𑂌	𑂍
(b) Medial and Final (with the Consonant 𑂎)									
a	ā	i	ī	u	ū	o	e	au	ay
𑂎	𑂏	𑂐	𑂑	𑂒	𑂓	𑂔	𑂕	𑂖	𑂗

Table 2

Khojki Script: Consonants (early 20th century)

𑂎	k	kh	g	𑂑, 𑂒, 𑂓	gh	𑂔	ng	𑂕	ng
𑂖	c	ch	j/jh	𑂗	j/jh	𑂘	ny	𑂙	dy
𑂚	t	th	𑂛	𑂜, 𑂝, 𑂞	dh	𑂟	𑂠	𑂡	𑂢
𑂣	p	ph	b	𑂤, 𑂥	dh	𑂦	𑂧	𑂨	𑂩
𑂪	y	r	l	𑂫, 𑂬, 𑂭	bh	𑂮	m	𑂯	sh
𑂰	dr	tr	h	𑂱	v	𑂲	s	𑂳	
			r/l	𑂴, 𑂵	r/l				

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 20: 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 21: 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	—	—	झ	ज ⁶	—
t	ट	ટ	ٽ	ट	ट ⁷	ट
th	ठ	ઠ	ٺ	ठ	— ⁷	ठ
ḍ	ड	ડ	ڊ	ड	ड ⁸	ड

Figure 23: 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	ی	इ	इ ¹	इ
ī	ई	ઈ	ی	ई	—	—
u	उ	ઉ	و	उ	उ ²	उ
ū	ऊ	ઊ	و	—	—	—
e	ए	એ	ی	इ	इ ³	इ
o	ओ	ઑ	و	—	उ ²	उ
ai	ऐ	એ	ای	ई	— ⁴	—
au	औ	ઔ	او	ई	— ⁴	—

Figure 22: 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	व्	વ્	ش	𑂔	𑂔 ¹⁴	𑂔
s	ष	શ	ش	—	𑂔 ¹⁴	—
s	स	સ	س	𑂔	𑂔	𑂔
h	ह	હ	ه	𑂔	𑂔	𑂔

Figure 25: 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	ڍ	—	𑂔	𑂔
ṛ	ऋ	ઋ	ڙ	—	𑂔 ⁸	𑂔
t	त	ત	ت	𑂔	𑂔	𑂔
th	थ	થ	ٿ	𑂔	𑂔	𑂔
d	द	દ	ڍ	𑂔	𑂔	𑂔
dh	ध	ધ	ڍ	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 24: 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	ڪ
dv	—	ડ	—	ڙڙ	—	—
dy	—	—	ج	—	ڙ ¹⁵	—
ny	—	—	ج	—	ڙن ¹⁶	—
z	—	ڙ	ڙ ڙ	—	ڙ ¹⁷	—
dhr	—	—	—	—	ڙ	ڙ
sr	—	ر	—	ر	—	—

Figure 26: Comparison of Khojki with other scripts: (ra to sra (from Khakee 1981).

<i>ka</i>	𑂗	KHOJKI LETTER KA	ڪ	U+06AA ARABIC LETTER SWASH KAF
<i>kha</i>	𑂘	KHOJKI LETTER KHA	ڪ	U+06A9 ARABIC LETTER KAF
<i>ga</i>	𑂙	KHOJKI LETTER GA	گ	U+06AF ARABIC LETTER GAF
<i>gā</i>	𑂚	KHOJKI LETTER GGA	گ	U+06B3 ARABIC LETTER GUEH
<i>gha</i>	𑂛	KHOJKI LETTER GHA	گھ	U+06AF ARABIC LETTER GAF + U+06BE HEH DO CHASHMEE
<i>na</i>	𑂜	KHOJKI LETTER NGA	نگ	U+06B1 ARABIC LETTER NGOEH
<i>ca</i>	𑂝	KHOJKI LETTER CA	چ	U+0686 ARABIC LETTER TCHEH
<i>cha</i>	𑂞	KHOJKI LETTER CHA	چ	U+0687 ARABIC LETTER TCHEHEH
<i>ja</i>	𑂟	KHOJKI LETTER JA	ج	U+062C ARABIC LETTER JEEM
<i>ja</i>	𑂠	KHOJKI LETTER JJA	ج	U+0684 ARABIC LETTER DYEH
<i>ña</i>	𑂡	KHOJKI LETTER NYA	ج	U+0683 ARABIC LETTER NYEH
<i>ta</i>	𑂢	KHOJKI LETTER TTA	ت	U+067D ARABIC LETTER TEH WITH THREE DOTS ABOVE DOWNWARDS
<i>tha</i>	𑂣	KHOJKI LETTER TTHA	ت	U+067A ARABIC LETTER TTEHEH
<i>ḍa</i>	𑂤	KHOJKI LETTER DDA	ڊ	U+068A ARABIC LETTER DAL WITH DOT BELOW
<i>ḍha</i>	𑂥	KHOJKI LETTER DDHA	ڊ	U+068D ARABIC LETTER DDAHAL
<i>ṇa</i>	𑂦	KHOJKI LETTER NNA	ڻ	U+06BB ARABIC LETTER RNOON
<i>ta</i>	𑂧	KHOJKI LETTER TA	ت	U+062A ARABIC LETTER TEH
<i>tha</i>	𑂨	KHOJKI LETTER THA	ت	U+067F ARABIC LETTER TEHEH
<i>da</i>	𑂩	KHOJKI LETTER DA	د	U+062F ARABIC LETTER DAL
<i>ḍa</i>	𑂪	KHOJKI LETTER DDA	ڏ	U+068F ARABIC LETTER DAL WITH THREE DOTS ABOVE DOWNWARDS
<i>dha</i>	𑂫	KHOJKI LETTER DHA	ڏ	U+068C ARABIC LETTER DAHAL
<i>na</i>	𑂬	KHOJKI LETTER NA	ن	U+0646 ARABIC LETTER NOON
<i>pa</i>	𑂭	KHOJKI LETTER PA	پ	U+067E ARABIC LETTER PEH
<i>pha</i>	𑂮	KHOJKI LETTER PHA	پھ	U+06A6 ARABIC LETTER PEHEH
<i>ba</i>	𑂯	KHOJKI LETTER BA	ب	U+0628 ARABIC LETTER BEH
<i>ḥa</i>	𑂰	KHOJKI LETTER BBA	ب	U+067B ARABIC LETTER BEEH
<i>bha</i>	𑂱	KHOJKI LETTER BHA	پ	U+0680 ARABIC LETTER BEHEH
<i>ma</i>	𑂲	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>ra</i>	𑂶	KHOJKI LETTER LLA	ڑ	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	SINDHI	GURMUKHI	DEVANAGARI		KHOJKI	SINDHI	GURMUKHI	DEVANAGARI
KA	𑂔	𑂑	ਕ	क	DA	𑂒	𑂓	ਦ	द
KHA	𑂕	𑂒	ਖ	ख	DDDA	𑂆	𑂇	—	डु
GA	𑂖	𑂓	ਗ	ग	DHA	𑂗	𑂔	ਧ	ध
GGA	𑂘	𑂕	—	गु	NA	𑂙	𑂖	ਨ	न
GHA	𑂚	𑂗	ਘ	घ	PA	𑂛	𑂘	ਪ	प
NGA	𑂛	𑂙	ਙ	ङ	PHA	𑂜	𑂚	ਫ	फ
CA	𑂝	𑂚	ਚ	च	BA	𑂞	𑂛	ਬ	ब
CHA	𑂟	𑂜	ਛ	छ	BBA	𑂠	𑂙	—	बु
JA	𑂡	𑂟	ਜ	ज	BHA	𑂢	𑂞	ਭ	भ
JJA	𑂣	𑂡	—	जु	MA	𑂤	𑂟	ਮ	म
JHA	—	𑂢	ਝ	झ	YA	𑂥	𑂡	ਯ	य
NYA	𑂧	𑂣	ਞ	ञ	RA	𑂦	𑂣	ਰ	र
TTA	𑂩	𑂧	ਟ	ट	LA	𑂨	𑂧	ਲ	ल
TTHA	𑂫	𑂩	ਠ	ठ	VA	𑂪	𑂨	ਵ	व
DDA	𑂭	𑂫	ਡ	ड	SHA	—	𑂬	ਸ਼	श
DDHA	𑂯	𑂭	ਢ	ढ	SA	𑂱	𑂬	ਸ	स
NNA	𑂳	𑂱	ਲ	ण	HA	𑂲	𑂳	ਹ	ह
TA	𑂴	𑂴	ਤ	त	LLA	𑂴	—	ਲ਼	ळ
THA	𑂶	𑂶	ਥ	थ	RRA	—	𑂶	ੜ	—

Table 2: A comparison of consonant letters of Khojki, Sindhi, Gurmukhi, and Devanagari.

	KHOJKI	SINDHI	GURMUKHI	DEVANAGARI		KHOJKI	SINDHI	GURMUKHI	DEVANAGARI
A	𑂏	𑂛	ਅ	अ	-A	—	—	—	—
AA	𑂏𑂛	𑂛𑂛	ਆ	आ	-AA	𑂏	𑂏	𑂏	𑂏
I	𑂏𑂛	𑂛	ਇ	इ	-I	𑂏	𑂏	𑂏	𑂏
II	—	𑂛	ਈ	ई	-II	𑂏	𑂏	𑂏	𑂏
U	𑂏	𑂛	ਉ	उ	-U	𑂏	𑂏	𑂏	𑂏
UU	—	𑂛	ਊ	ऊ	-UU	—	𑂏	𑂏	𑂏
E	𑂏	𑂛	ਏ	ए	-E	𑂏	𑂏	𑂏	𑂏
AI	𑂏	𑂛	ਐ	ऐ	-AI	𑂏	𑂏	𑂏	𑂏
O	𑂏	𑂛	ਓ	ओ	-O	𑂏	𑂏	𑂏	𑂏
AU	𑂏	𑂛	ਔ	औ	-AU	𑂏	𑂏	𑂏	𑂏

Table 3: A comparison of vowel letters and signs of Khojki, Sindhi, Gurmukhi, and Devanagari.

	SINDHI	KHOJKI	GURMUKHI	DEVANAGARI		SINDHI	KHOJKI	GURMUKHI	DEVANAGARI
0	—	𑂏	0	०	5	—	𑂏	5	५
1	—	𑂏	1	१	6	—	𑂏	6	६
2	—	𑂏	2	२	7	—	𑂏	7	७
3	—	𑂏	3	३	8	—	𑂏	8	८
4	—	𑂏	4	४	9	—	𑂏	9	९

Table 4: A comparison of digits of Khojki, Sindhi, Gurmukhi, and Devanagari. Khojki uses Sindhi digits.

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	g̃a ³ ALA-LC g with diaeresis above	ga	gha	ṅa
Sindhi (Arabic)	ڪ	ڪھ	گ	گ	گھ	ڱ
Sindhi (Dev)	क	ख	U + 097b ग	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		ṅa
Sindhi (Arabic)	چ	چھ	چ	ج		چ
Sindhi (Dev)	च	छ	U + 097c	ज		ञ

Figure 27: Transliteration of Khojki characters proposed by Shafique Virani (from Virani 2008). Continued in Figure 28.

			𑀓			
Gujarati	ઢ	ઠ		ઢ		ઠ
Cerebral Khojki	𑀓	𑀔		𑀕	𑀖	𑀗
Transliteration	ṭa	ṭha		ḍa	ḍha	ṇa
Sindhi (Arabic)	ٿ	ٺ		ڊ	ڍ	ڻ
Sindhi (Dev)	ट	ठ		ड	ढ	ण
Gujarati	ઢ	ઠ		ડ	ઢ	ણ
Dental Khojki	𑀘	𑀙	𑀚	𑀛	𑀜	𑀝
Transliteration	Ta	tha	ḍa ALA-LC d with diaeresis below	da	dha	na
Sindhi (Arabic)	ت	ٿ	ڙ، ڛ	د	ڏ	ن
Sindhi (Dev)	त	थ	ड़ (ALA-LC) ड़ (U + 097e) (Lekhwni's CIIL publication - used for Unicode)	द	ध	न
Gujarati	ઢ	થ		દ	ધ	ન
Labial Khojki	𑀞	𑀟	𑀠	𑀡	𑀢	𑀣
Transliteration	pa	pha	ḅa ALA-LC b diaeresis below	ba	bha	ma

Figure 28: Transliteration of Khojki characters proposed by Shafique Virani (from Virani 2008). Continued in Figure 29.

Sindhi (Arabic)	پ	ڦا	ب	ب	پ	م
Sindhi (Dev)	प	फ	U+097F ब्र	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 29: Transliteration of Khojki characters proposed by Shafique Virani (from Virani 2008). Continued in Figure 30.

Conjuncts Khojki	𑂗	𑂘	𑂙	𑂚
Transliteration	dra ⁵	j ña ⁶	kṣa ⁷	tra ⁸
Sindhi (Arabic)				
Sindhi (Dev)				
Gujarati				
Diacritic Letters Khojki	𑂛			
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 क + ष, 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 ṭa sound.

Figure 30: Transliteration of Khojki characters 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 10646⁷

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: **Revised Proposal to Encode the Khojki Script in ISO/IEC 10646**
2. Requester's name: **University of California, Berkeley Script Encoding Initiative (Universal Scripts Project); author: Anshuman Pandey (pandey@umich.edu)**
3. Requester type (Member Body/Liaison/Individual contribution): **Liaison contribution**
4. Submission date: **2010-09-12**
5. Requester's reference (if applicable): **N/A**
6. Choose one of the following:
 - (a) This is a complete proposal: **No**
 - (b) or, More information will be provided later: **Awaiting review by specialists**

B. Technical - General

1. Choose one of the following:
 - (a) This proposal is for a new script (set of characters): **Yes**
 - i. Proposed name of script: **Khojki**
 - (b) The proposal is for addition of character(s) to an existing block: **No**
 - i. Name of the existing block: **N/A**
2. Number of characters in proposal: **61**
3. Proposed category: **B.1 - Specialized (small collections of characters)**
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. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?: **Anshuman Pandey; True Type format**
 - (a) If available now, identify source(s) for the font and indicate the tools used: **The characters of the digitized Khojki font are based on normalized forms of printed Khojki characters. The font includes characters from a digitized Khojki font called KhojkiJiwa.ttf (developed by Pyarali Jiwa, provided by Shafique Virani), as well as several new characters not found in KhojkiJiwa and revisions to several glyphs. A new font will be produced by Anshuman Pandey from printed Khojki sources.**
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?: **Yes**
7. Special encoding issues:
 - (a) 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)? **Yes; see proposal for additional details.**
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. **Character properties and numeric information are included.**

⁷ Form number: N3102-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)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?: **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.)? **Yes**
 - (a) If Yes, with whom?:
 - **Shafique Virani (shafique.virani@utoronto.ca), University of Toronto**
 - i. If Yes, available relevant documents: **N/A**
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? **Yes**
 - (a) Reference: **See text of proposal**
4. The context of use for the proposed characters (type of use; common or rare): **Common**
 - (a) Reference: **The script was in common use among the Ismaili Muslim religious community of South and in other parts of the world. There is evidence that the script was used by communities related to the Ismaili community, including the Imamshahis of Gujarati.**
5. Are the proposed characters in current use by the user community?: **Yes. Members of the Ismaili community and specialists studying Ismaili literature require the script.**
 - (a) If Yes, where? Reference: **The script continues to be used wherever Ismaili Muslims of South Asian origin reside, with the largest user communities in Pakistan, India, Canada, the United States, the United Kingdom, Kenya, Tanzania, and Uganda.**
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?: **No**
 - (a) If Yes, is a rationale provided?: **N/A**
 - i. If Yes, reference: **N/A**
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**
 - (a) If Yes, is a rationale for its inclusion provided?: **N/A**
 - i. If Yes, reference: **N/A**
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? **No**
 - (a) If Yes, is a rationale provided?: **N/A**
 - i. If Yes, reference: **N/A**
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? **Yes**
 - (a) If Yes, is a rationale for its inclusion provided? **Yes**
 - i. If Yes, reference: **See text of proposal**
11. Does the proposal include use of combining characters and/or use of composite sequences? **Yes**
 - (a) If Yes, is a rationale for such use provided? **Yes**
 - i. If Yes, reference: **See text of proposal**
 - (b) Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? **N/A**
 - i. If Yes, reference: **N/A**
12. Does the proposal contain characters with any special properties such as control function or similar semantics? **Yes**
 - (a) If Yes, describe in detail (include attachment if necessary): **Virama**
13. Does the proposal contain any Ideographic compatibility character(s)? **No**
 - (a) If Yes, is the equivalent corresponding unified ideographic character(s) identified? **N/A**
 - i. If Yes, reference: **N/A**