Proposal to Encode the Khojki Script in ISO/IEC 10646

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

March 25, 2009

Contents

List of Tables

List of Figures

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]

Please fill all the sections A, B and C below. Please read Principles and Procedures Document (P & P) from
http://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: **Proposal to Encode the Khojki Script in ISO/IEC 10646**
2. Requester's name: **Anshuman Pandey (pandey@umich.edu)**
3. Requester type (Member Body/Liaison/Individual contribution): **Individual contribution**
4. Submission date: **March 25, 2009**
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: **As requested**

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: **67**
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 (provided by Shafique Virani), as well as several new
       characters not found in KhojkiJiwa and revisions to several glyphs. The font is a work-in-progress
       and the final font will be a new work that 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 pro-
   posed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the pro-
   posed 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 be-
   haviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equiv-
   alence 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 associ-
   ated 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.**

---

[1] 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, Mississauga**
            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 used primarily by the Ismaili religious community in South Asia.**
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: **In the United Kingdom, Canada, United States, Pakistan, and India.**
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**

1   Introduction

Purpose   This is a formal proposal to encode the Khojki script in the Universal Character Set (ISO/IEC 10646). The present proposal (N3596 L2/09-101) replaces the following document:

• L2/08-201: "Towards an Encoding for the Khojki Script in ISO/IEC 10646", May 2008

Acknowledgments   The proposal author is indebted to Professor Shafique Virani (University of Toronto, Mississauga) for providing Khojki specimens and for offering detailed comments on the forms of Khojki characters.

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.

2   Background

Name   The Khojki (𑈈𑈴𑈬𑈈𑈮 *khojakī*) script takes its name from the Sindhi form *khojā* of the Persian خواجه *khwājah* "master, lord". Literally, it means "of the master" (*khojā* + *kī*). The name appears in transliteration as Khojakī and Khojkī. The common form 'Khojki' is recommended as the formal name of the script in the UCS.

Overview   Khojki is the name of the writing system used specifically by the Nizari Ismaili community of South Asia for recording their religious literature. It is a Brahmi-based script of the Landa family, which is a class of mercantile scripts that is derived ultimately from Sharada. Khojki is a refined form of the Lohāṇakī variety of Landa used in Sind that was purposely developed as an ecclesiastical script. It is believed to have become distinct in the 16th century CE. Popular Ismaili tradition states that the script was invented by Pir Sadruddin, an Ismaili missionary actively working with the Hindu Lohana community of Sind. Khojki is one of two mercantile Landa scripts that were transformed into literary scripts in order to serve the purposes of religious communities; the other is Gurmukhi.

Developed in and used predominantly in Sind (in modern Pakistan), Khojki was used to represent Sindhi, Gujarati, Persian, and other regional languages. The majority of Khojki materials are manuscripts; however, hundreds of lithographs and books were published in the 20th century after Laljibhai Devraj produced metal types for Khojki in Germany for use at his Khoja Sindhi Printing Press in Bombay. Large collections of Khojki materials are housed at the Institute of Ismaili Studies in London and at Harvard University.

Although Khojki possesses a limited user base, manuscripts and books were used in Ismaili religious ceremonies not only in South Asia but in East and South Africa, where large diaspora communities were formed starting in the 19th century. While the position of Khojki as a living script is declining, it is still used by members of the Ismaili community and by specialists who study the history and literature of the the community. 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 religious and literary culture.

	1128	1129	112A	112B	112C
0	○̊ 11280	धा 11290	६ 112A0	꠰ 112B0	? 112C0
1	○̇ 11281	३: 11291	꠲ 112A1	꠱ 112B1	꠳ 112C1
2	ꠄ 11282	ꠅ 11292	꠆ 112A2	ꠇ 112B2	: 112C2
3	ꠈ 11283	ꠉ 11293	ꠊ 112A3	ꠋ 112B3	— 112C3
4	ꠌ 11284	ꠍ 11294	ꠎ 112A4	○) 112B4	═ 112C4
5	ꠏ 11285	ꠐ 11295	ꠑ 112A5	○ʃ 112B5	
6	ꠒ 11286	ꠓ 11296	ꠔ 112A6	○ɭ 112B6	
7	ꠕ 11287	ꠖ: 11297	ꠗ 112A7	○̆ 112B7	
8	ꠙ 11288	ꠚ 11298	ꠛ 112A8		
9	ꠜ 11289	ꠝ 11299	ꠞ 112A9	○̀ 112B9	
A	ꠟ 1128A	ꠠ 1129A	ꠡ 112AA	○̂ 112BA	
B	ꠢ 1128B	ꠣ 1129B	ꠤ 112AB	○̣ 112BB	
C	ꠥ 1128C	ꠦ 1129C	ꠧ 112AC	○̧ 112BC	
D	꠨ 1128D	꠩ 1129D	꠪ 112AD	○/ 112BD	
E	꠫ 1128E	꠬ 1129E	꠭ 112AE	○̈ 112BE	
F	꠮ 1128F	꠯ 1129F		○̈ 112BF	

Table 1: Glyph chart for Khojki

2

Various signs

11280 ̐ KHOJKI SIGN CANDRABINDU
11281 ̇ KHOJKI SIGN ANUSVARA

Independent vowels

11282 ꠂ KHOJKI LETTER A
11283 ꠃ KHOJKI LETTER AA
11284 ꠄ KHOJKI LETTER I
11285 ꠅ KHOJKI LETTER II
11286 ꠆ KHOJKI LETTER U
11287 ꠇ KHOJKI LETTER UU
11288 ꠈ KHOJKI LETTER E
11289 ꠉ KHOJKI LETTER AI
1128A ꠊ KHOJKI LETTER O
1128B ꠋ KHOJKI LETTER AU

Consonants

1128C ꠌ KHOJKI LETTER KA
1128D ꠍ KHOJKI LETTER KHA
1128E ꠎ KHOJKI LETTER GA
1128F ꠏ KHOJKI LETTER GGA
11290 ꠐ KHOJKI LETTER GHA
11291 ꠑ KHOJKI LETTER NGA
11292 ꠒ KHOJKI LETTER CA
11293 ꠓ KHOJKI LETTER CHA
11294 ꠔ KHOJKI LETTER JA
11295 ꠕ KHOJKI LETTER JJA
11296 ꠖ KHOJKI LETTER JHA
11297 ꠗ KHOJKI LETTER NYA
11298 ꠘ KHOJKI LETTER TTA
11299 ꠙ KHOJKI LETTER TTHA
1129A ꠚ KHOJKI LETTER DDA
1129B ꠛ KHOJKI LETTER DDHA
1129C ꠜ KHOJKI LETTER NNA
1129D ꠝ KHOJKI LETTER TA
1129E ꠞ KHOJKI LETTER THA
1129F ꠟ KHOJKI LETTER DA
112A0 ꠠ KHOJKI LETTER DDDA
112A1 ꠡ KHOJKI LETTER DHA
112A2 ꠢ KHOJKI LETTER NA
112A3 ꠣ KHOJKI LETTER PA
112A4 ꠤ KHOJKI LETTER PHA
112A5 ꠥ KHOJKI LETTER BA
112A6 ꠦ KHOJKI LETTER BBA
112A7 ꠧ KHOJKI LETTER BHA
112A8 ꠨ KHOJKI LETTER MA
112A9 ꠩ KHOJKI LETTER YA
112AA ꠪ KHOJKI LETTER RA
112AB ꠫ KHOJKI LETTER LA
112AC ꠬ KHOJKI LETTER LLA
112AD ꠭ KHOJKI LETTER VA
112AE ꠮ KHOJKI LETTER SHA
112AF ▨ <reserved>
112B0 ꠰ KHOJKI LETTER SA
112B1 ꠱ KHOJKI LETTER HA
112B2 ꠲ KHOJKI LETTER QA

Consonant Conjunct

112B3 ꠳ KHOJKI LETTER KSSA

Dependent vowel signs

112B4 ꠴ KHOJKI VOWEL SIGN AA

112B5 ꠵ KHOJKI VOWEL SIGN I
112B6 ꠶ KHOJKI VOWEL SIGN II
112B7 ꠷ KHOJKI VOWEL SIGN U
112B8 ▨ <reserved>
112B9 ꠹ KHOJKI VOWEL SIGN E
112BA ꠺ KHOJKI VOWEL SIGN AI
112BB ꠻ KHOJKI VOWEL SIGN O
112BC ꠼ KHOJKI VOWEL SIGN AU

Various signs

112BD ꠽ KHOJKI SIGN VIRAMA
112BE ꠾ KHOJKI SIGN NUKTA
112BF ꠿ KHOJKI SIGN SHADDA

Punctuation

112C0 ꠰ KHOJKI DANDA
112C1 ꠱ KHOJKI DOUBLE DANDA
112C2 ꠲ KHOJKI WORD SEPARATOR
112C3 ꠳ KHOJKI SECTION MARK
112C4 ꠴ KHOJKI DOUBLE SECTION MARK

Table 2: Names list for Khojki

3

ka	𑈈	KHOJKI LETTER KA	ک	U+06AA ARABIC LETTER SWASH KAF
kha	𑈉	KHOJKI LETTER KHA	ک	U+06A9 ARABIC LETTER KAF
ga	𑈊	KHOJKI LETTER GA	گ	U+06AF ARABIC LETTER GAF
ḡa	𑈋	KHOJKI LETTER GGA	ڳ	U+06B3 ARABIC LETTER GUEH
gha	𑈌	KHOJKI LETTER GHA	گھ	U+06AF ARABIC LETTER GAF + U+06BE HEH DO CHASHMEE
ṅa	𑈍	KHOJKI LETTER NGA	ڱ	U+06B1 ARABIC LETTER NGOEH
ca	𑈎	KHOJKI LETTER CA	چ	U+0686 ARABIC LETTER TCHEH
cha	𑈏	KHOJKI LETTER CHA	ڇ	U+0687 ARABIC LETTER TCHEHEH
ja	𑈐	KHOJKI LETTER JA	ج	U+062C ARABIC LETTER JEEM
j̄a	𑈑	KHOJKI LETTER JJA	ڄ	U+0684 ARABIC LETTER DYEH
jha	𑈒	KHOJKI LETTER JHA	جھ	U+062C ARABIC LETTER JEEM + U+06BE HEH DO CHASHMEE
ña	𑈓	KHOJKI LETTER NYA	ڃ	U+0683 ARABIC LETTER NYEH
ṭa	𑈔	KHOJKI LETTER TTA	ٹ	U+067D ARABIC LETTER TEH WITH THREE DOTS ABOVE DOWNWARDS
ṭha	𑈕	KHOJKI LETTER TTHA	ٺ	U+067A ARABIC LETTER TTEHEH
ḍa	𑈖	KHOJKI LETTER DDA	ڊ	U+068A ARABIC LETTER DAL WITH DOT BELOW
ḍha	𑈗	KHOJKI LETTER DDHA	ڍ	U+068D ARABIC LETTER DDAHAL
ṇa	𑈘	KHOJKI LETTER NNA	ڻ	U+06BB ARABIC LETTER RNOON
ta	𑈙	KHOJKI LETTER TA	ت	U+062A ARABIC LETTER TEH
tha	𑈚	KHOJKI LETTER THA	ٿ	U+067F ARABIC LETTER TEHEH
da	𑈛	KHOJKI LETTER DA	د	U+062F ARABIC LETTER DAL
ḏa	𑈜	KHOJKI LETTER DDA	ڏ	U+068F ARABIC LETTER DAL WITH THREE DOTS ABOVE DOWNWARDS
dha	𑈝	KHOJKI LETTER DHA	ذ	U+068C ARABIC LETTER DAHAL
na	𑈞	KHOJKI LETTER NA	ن	U+0646 ARABIC LETTER NOON
pa	𑈟	KHOJKI LETTER PA	پ	U+067E ARABIC LETTER PEH
pha	𑈠	KHOJKI LETTER PHA	ڦ	U+06A6 ARABIC LETTER PEHEH
ba	𑈡	KHOJKI LETTER BA	ب	U+0628 ARABIC LETTER BEH
ḇa	𑈢	KHOJKI LETTER BBA	ٻ	U+067B ARABIC LETTER BEEH
bha	𑈣	KHOJKI LETTER BHA	ڀ	U+0680 ARABIC LETTER BEHEH
ma	𑈤	KHOJKI LETTER MA	م	U+0645 ARABIC LETTER MEEM
ya	𑈥	KHOJKI LETTER YA	ي	U+064A ARABIC LETTER YEH
ra	𑈦	KHOJKI LETTER RA	ر	U+0631 ARABIC LETTER REH
la	𑈧	KHOJKI LETTER LA	ل	U+0644 ARABIC LETTER LAM
ṛa	𑈨	KHOJKI LETTER LLA	ڙ	U+0699 ARABIC LETTER REH WITH FOUR DOTS ABOVE
va	𑈩	KHOJKI LETTER VA	و	U+0648 ARABIC LETTER WAW
śa	𑈪	KHOJKI LETTER SHA	ش	U+0634 ARABIC LETTER SHEEN
sa	𑈫	KHOJKI LETTER SSA	س	U+0633 ARABIC LETTER SEEN
ha	𑈬	KHOJKI LETTER HA	ہ	U+06C1 ARABIC LETTER HEH GOAL
qa	𑈭	KHOJKI LETTER QA	ق	U+0642 ARABIC LETTER QAF

Table 3: Transliteration of Khojki characters and Sindhi-Arabic analogues

3　Characters Proposed

Number of Characters　　The 67 letters proposed here comprise the core Khojki character set.

Character Names　　The naming convention for Khojki characters follows that used for Devanagari. Characters for the Sindhi implosive consonants GGA, JJA, DDA, and BBA are based on the convention adopted for their Devanagari counterparts. There is no standard transliteration for these characters; eg. GGA is referred to as *gr* or *ṅg*, JJA is referred to as *dy*. Alignment with Devanagari names will faciliate recognition of analogous characters.

Allocation　　Khojki is currently allocated in the Supplementary Multilingual Plane (SMP) at the range U+11280..U+112CF.[1]

3.1　Character Set

Consonants　　There are 38 consonant letters:

𑊀	KHOJKI LETTER KA	𑊕	KHOJKI LETTER TTHA	𑋂	KHOJKI LETTER BBA
𑊁	KHOJKI LETTER KHA	𑊖	KHOJKI LETTER DDA	𑋃	KHOJKI LETTER BHA
𑊂	KHOJKI LETTER GA	𑊗	KHOJKI LETTER DDHA	𑋄	KHOJKI LETTER MA
𑊃	KHOJKI LETTER GGA	𑊘	KHOJKI LETTER NNA	𑋅	KHOJKI LETTER YA
𑊄	KHOJKI LETTER GHA	𑊙	KHOJKI LETTER TA	𑋆	KHOJKI LETTER RA
𑊅	KHOJKI LETTER NGA	𑊚	KHOJKI LETTER THA	𑋇	KHOJKI LETTER LA
𑊆	KHOJKI LETTER CA	𑊛	KHOJKI LETTER DA	𑋈	KHOJKI LETTER LLA
𑊇	KHOJKI LETTER CHA	𑊜	KHOJKI LETTER DDDA	𑋉	KHOJKI LETTER VA
𑊈	KHOJKI LETTER JA	𑊝	KHOJKI LETTER DHA	𑋊	KHOJKI LETTER SHA
𑊉	KHOJKI LETTER JJA	𑊞	KHOJKI LETTER NA	𑋋	KHOJKI LETTER SA
𑊊	KHOJKI LETTER JHA	𑊟	KHOJKI LETTER PA	𑋌	KHOJKI LETTER HA
𑊋	KHOJKI LETTER NYA	𑊠	KHOJKI LETTER PHA	𑋍	KHOJKI LETTER QA
𑊌	KHOJKI LETTER TTA	𑊡	KHOJKI LETTER BA		

Consonant Conjunct　　There is one consonant conjunct:

𑊋	KHOJKI LETTER KSSA

Vowels　　There are 10 independent vowels:

𑊰	KHOJKI LETTER A	𑊶	KHOJKI LETTER U	𑊹	KHOJKI LETTER O
𑊱	KHOJKI LETTER AA	𑊷	KHOJKI LETTER UU	𑊺	KHOJKI LETTER AU
𑊲	KHOJKI LETTER I	𑊸	KHOJKI LETTER E		
𑊳	KHOJKI LETTER II	𑊵	KHOJKI LETTER AI		

---

[1]　Unicode Roadmap Committee 2009.

Vowel Signs   There are 8 dependent vowel signs:

◌) KHOJKI VOWEL SIGN AA	◌ KHOJKI VOWEL SIGN U	◌) KHOJKI VOWEL SIGN O
◌ʃ KHOJKI VOWEL SIGN I	◌ KHOJKI VOWEL SIGN E	◌) KHOJKI VOWEL SIGN AU
◌ʆ KHOJKI VOWEL SIGN II	◌ KHOJKI VOWEL SIGN AI	

Various Signs   There are 5 various signs:

◌ KHOJKI SIGN CANDRABINDU	◌/ KHOJKI SIGN VIRAMA	◌ KHOJKI SIGN SHADDA
◌ KHOJKI SIGN ANUSVARA	◌ KHOJKI SIGN NUKTA	

Punctuation   There are 5 punctuation signs:

੧ KHOJKI DANDA	— KHOJKI SECTION MARK
੧੧ KHOJKI DOUBLE DANDA	= KHOJKI DOUBLE SECTION MARK
: KHOJKI WORD SEPARATOR	

3.2   Basis for Character Set and Glyph Shapes

3.2.1   Basis for Character Set

Given the absence of a standard for Khojki, this proposal intends to provide an idealized character set that may be used to encode any Khojki document completely. The character set proposed here is based upon the inventory of Khojki characters presented by George Stack's *A Grammar of the Sindhi Language* (as represented in George Grierson's *Linguistic Survey of India*: figures 1–3); Ali S. Asani in *The Harvard Collection of Ismaili Literature in Indic Languages* (figures 4–9); and Christopher S. Shackle and Zawahir Moir in *Ismaili Hymns from South Asia* (figures 10–11). Included in this character set are characters that are not enumerated in official Khojki primers, but are nonetheless attested in manuscripts and printed books, as well as in character charts. The presence of characters with which readers of Khojki are not necessarily familiar should not be a distraction, but rather should be understood as a letter of the script that at some historical period and within a particular region was identified as being part of the Khojki repertoire.

Notes on Characters

1. **LETTER II**   Generally, Khojki does not distinguish between independent forms of I and II. The character ੮ KHOJKI LETTER I is used to represent both. However, Figure 4 shows the glyph ੮ as for II. The character is included here in order to accommodate the attestation.

2. **LETTER UU**   Although the glyph ੭ (U) is generally used for also representing UU, Figure 4 shows the glyph ੭ used for UU. The character UU is, therefore, included in order to accommodate the attestation.

3. **LETTER JHA**   Traditional character inventories of Khojki, such as that shown in Figure 10, do not show an independent character for JHA. The character is, however, shown as part of the Khojki character set in other charts, such as those in Figure 2 and Figure 7, and is proposed for encoding in order to accommodate these attestations.

4. **LETTER LLA**   The glyph ੭ has been assigned to the character KHOJKI LETTER LLA. This glyph represents *ṛa* (/ɽ/), which is analogous to Devanagari ड़ U+095C DEVANAGARI LETTER DDDHA:

```
P Ism K 19                                    Bombay 1985 s./1929

Dash Avatār nāno pûro. 3rd edition.

Publisher: Rahemtulā Vīrjī, honorary secretary, The Recreation
Club Institute, Press Dept. at the Khoja Sindhi Printing Press;
500 copies; 20 pp.; 22 x 17; paper covers; in good condition;
same as P Ism K 31.

Does not use ":" to separate individual words; ထ = r̤;  ᷧ᷉ᷧ = ḍ
  ᷧ = ṇ.
```

The sound is an allophonic variant of /ḍ/ (DDA), which appears intervocalically. Some Khojki manuscripts use the glyph ᷧ (KHOJKI LETTER DDA) to represent the sound. The following specimen shows ထ as being analogous to ℓ U+0AB3 GUJARATI LETTER LLA (*la*), presumably to represent the sound /ɭ/:

```
vowel "r̤u" is indicated by subscript " ⌡ " ; (5) the following
conjuncts introduced: ᷧ᷉ = Gujarati ℓ᷉ ᷧ ;   ထ    = Gujarati ℓ  ;
   ᷧ᷉ = Gujarati ᷧ᷉ ;   ᷧ᷉ = Gujarati ᷧ᷉
```

Although ထ is assigned here to KHOJKI LETTER LLA, it may be justifiably named KHOJKI LETTER DDDHA based upon its phonological value. The sound /ɭ/ is not part of the Sindhi phonemic repertoire. It is likely that ထ was adapted to represent /ɽ/ because of its association with a retroflex character. But, as ထ is visually similar to the glyphs used to represent LLA in other Indic scripts, it is assigned to that character for Khojki as well.

5. **LETTER SHA**   Although the character ᷧ̈ SHA may be represented as ᷧ SA + ̈ NUKTA, it is proposed as an atomic character. Unlike other characters that are produced using NUKTA for the purpose of representing sounds not found in Sindhi, Khojki SHA represents /ʃ/, a sound native to Sindhi.

6. **LETTER QA**   In Khojki, the transcription of ق U+0642 ARABIC LETTER QAF (/q/) is generally represented as ᷧ̈ (KA + NUKTA), analogous to क़ U+0958 DEVANAGARI LETTER QA. However, some manuscript and printed sources show the character ᷧ used for representing /q/:

```
P Ism K 14                                    Bombay 1962 s./1905

Elam sār saṅgrah rāg mālā

Publisher: Lāljībhāî Devrāj, Khoja Siñdhī Čhāpakhānû; 147 pp. in
three paginations; 23 x 16; hardboard covers, binding in poor condi-
tion; 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 Khoja Siñdhī Printing Press pp. 3.1-6.

  ℯ᷉    = g̲ā̲f̲;  ᷧ᷉ = g, gr;  ᷧ᷉ = ᶜa̲i̲n̲;  ᷧ᷉ = 1?;   ∕  = vowel-less
consonant.
```

This glyph is also considered a variant of ᷧ KA:

```
P Ism K 11                                    Bombay 1959 s./1902

Ārtī:dhûā:vakhat:trejī:sāñjījā:ćogaḍīā:5

Publisher: Khojā Lāljībhāî Devrāj, Khojā Siñdhī Ćhāpākhānû for
the Imāmī Ismāīlī school (nisāl). 64 pp.; 16 x 12; covers miss-
ing; 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 pub-
lication of other works.

    ꢀ = z; ꢁ = j; ꢂ = k; ꢃ = ʿain; ꢄ = g; ꢅ = gain;
shaddah for doubled consonant or emphasis.
```

The following specimen from a printed book shows KA represented not only with ꢆ, but also with ꢇ. Both glyphs are used in the word *kalāma* (ꢇꢈꢉ / ꢆꢈꢉ) and are semantically identical:
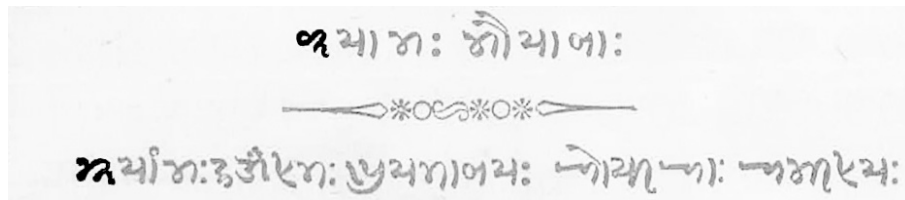
Thus, in Khojki, the letter QA may be represented as ꢊ and ꢇ. As ꢇ is used in some documents to represent the letter QA uniquely, it is included here in order to accommodate the attestation.

7. **LETTER KSSA**   Khojki does not have a character SSA to represent *ṣa* (/ṣ/), analogous to ष U+0937 DEVANAGARI CHARACTER SSA. It does, however, have the glyph ꢋ KSSA (Devanagari क्ष), which is a consonant conjunct produced with the sequence KA + VIRAMA + SSA in most Indic scripts. The absence of SSA in Khojki requires an alternate method of producing ꢋ. The easiest solution is to encode KSSA as an atomic character.

3.2.2 Basis for Glyph Shapes

The digitized Khojki characters proposed here are based on the forms of the metal fonts developed by Laljibhai Devraj in 1903. Devraj used the fonts to produce the first printed versions of Ismaili devotional literature (*ginān*) at his Khoja Sindhi Printing Press in Bombay. Although different publishers in Sindh, Gujarat, and Mumbai printed works in Khojki, the majority of publications were produced by Devraj. Devraj's font, therefore, is universally recognized by readers of Khojki. This digitized Khojki font contains characters not originally found in metal fonts, but additional characters have been designed with the spirit of Devraj's style, and it is hoped that this will appeal to the Ismaili community and will mirror the 'standard' appearance of Khojki to which they are accustomed.

3.3 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 have been identified in Khojki documents, 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.

1. **VOWEL SIGN UU**   Although ꢌ KHOJKI LETTER UU is proposed for encoding based upon attested use of the character, sources attesting a unique character for VOWEL SIGN UU have not been identified. The

general practice is to use ◌ KHOJKI VOWEL SIGN U for dependent forms of both U and UU. Space has been reserved at U+112B6 should a unique character for this sign be identified.

2. **LETTER SSA**   Although Khojki has a glyph for the consonant conjunct ☙ KSSA, it does not have the character SSA. For this reason, KSSA has been included as an atomic character. Space has been reserved at U+112AF should the character SSA be identified.

3. **Characters for *jña*, *tra*, and *dra***   Khojki charts traditionally show the characters ☙ *kṣa*, ◌ *jña*, ◌ *tra*, and ◌ *dra*. They are not atomic characters, but consonant conjuncts, and they should be handled as such. The exception is *kṣa* (see 3.2). 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 mask the consonant characters that constitute them.

4. **Pre-composed characters for consonant + NUKTA**   There are several Khojki letters that are composed of a base letter + NUKTA: ◌ A + NUKTA; ◌ GA + NUKTA (see section 4.8). It is not necessary to encode each of these uniquely. The only pre-composed *nukta* character proposed for encoding is ◌ KHOJKI LETTER SHA.

5. **Digits**   There are no unique Khojki digits. When digits appear in Khojki texts, they are generally Gujarati forms: ૦ ૧ ૨ ૩ ૪ ૫ ૬ ૭ ૮ ૯ (U+0AE6..U+0AEF).

6. **Number forms and unit marks**   Fraction signs and currency marks are found in Khojki documents. The forms and semantics of these characters are identical to the North Indic number forms proposed for inclusion in the "Common Indic Number Forms" block (U+A830..U+A83F). See section 4.11 for additional information.

7. **Verse marker**   The mark ◌ is written over verse numbers in long Khojki texts. As described in section 4.9, the mark is an extending character that spans all of the digits in a number. This specialized rendering behavior may pose difficulties for the implementation of this function. Therefore, the ◌ U+0305 COMBINING OVERLINE should be used for the Khojki verse marker. The drawback to this solution is that COMBINING OVERLINE would visually represent the verse marker, but it would not capture the semantic properties of the Khojki character.

8. **Various signs derived from Devanagari**   Authors of Khojki instructional materials at times attempted to fix the deficiencies of the script by borrowing orthographic practices from scripts such as Devanagari. The specimen below illustrates such uses. One such script primer, ◌◌◌ ◌◌◌◌ ◌◌◌◌ (*Sindhī pahelī copaḍī*), introduces VOWEL SIGN VOCALIC R and various diacritics to represent conjunct conjuncts with RA. These characters are not proposed at present until a determination can be made as to the necessity of encoding them.

9

```
P Ism K 17                              Bombay 1966 s./ 1910

Siṅdhī pahelī ćopaḍī. 1st edition.

Publisher: Mûkhī Lāljībhâî Devrāj, Manager of the Bombay Shia
Imami Ismaili Khoja Jamat Dept of Education, for the Siṅdhī
Textbook Committee at the Khoja Siṅdhī Printing Press; 2000
copies; 92 pp. in various paginations; 17 x 13; covers missing;
binding in fair condition; paper in good condition; saṅkaḍīûṅ
(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 famil-
iarize themselves with these forms before proceeding with
instruction. Special uses: (1) " ﹏ " 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 super-
script sign " ∧ " is used on the first member i.e. "ᵍᴿ" for "pr";
(3) Subscript " ◟ " with the consonant "r" indicates "r" as the
first member of the conjunct i.e. "ᵖᵘʳᵛᵃ" for "pûrva"; (4) the
vowel "ṛu" is indicated by subscript " ⌐ " ; (5) the following
conjuncts introduced: ꣲ = Gujarati શ્ય ;  ﻼ  = Gujarati ત્ય  ;
   શ્મ = Gujarati શ્મ ;  ꣲ  = Gujarati ત્ત
The script does not use ":" to separate individual words.
```

4 The Writing System

4.1 General Features

Classification The Khojki script is an abugida of the Brahmic type and is written from left to right.

Structure The structure of Sharada is identical to that of Devanagari. Consonant letters bear the inherent vowel *a* when unaccompanied by a vowel sign. The inherent vowel is suppressed by the *virāma* to produce the bare form of the consonant. A bare consonant followed by another consonant results in a consonant conjunct. The inherent vowel is changed by applying a dependent vowel sign to the consonant.

4.2 Inadequacies of the Script

The Khojki script has several incongruencies between glyphs and characters:

1. *Redundancy* One character represented by multiple glyphs

2. *Ambiguity* Multiple characters represented by a single or by multiple glyphs

3. *Deficiency* Ligature glyph forms composed of characters absent from the script

Redundancy Khojki was developed to represent the Sindhi language. It therefore has characters for sounds found in Sindhi, but not in neighboring languages such as Gujarati. These sounds are the four voiced implosive consonants: /ɠ/ (velar), /ʄ/ (palatal), /ɗ/ (dental), and /ɓ/ (labial). When the Khojki script was used outside of Sindh and by scribes who spoke other languages, the characters for the implosive consonants lost their phonetic value. With the deletion of phonetic properties, the glyphs for the implosive consonants were reassigned to consonants within the same articulation class, resulting in the use of multiple glyphs for a single character:

- ꣲ GGA /ɠ/ was used to represent /g/, already represented by ꣲ GA. ꣲ was also used for writing the consonant conjunct (GA + VIRAMA + RA) /gr/.

- **ᚹ** JJA /ʃ/ was used to represent /dʒ/, already represented as **ᚳ** KHOJKI LETTER JA. In some cases **ᚹ** was also used to write **ᚳ̤** (JA + VIRAMA + NUKTA) /z/.

- **ᚥ** DDDA /ɖ/ was used to represent /d/, already represented by **ᚱ** KHOJKI LETTER DA. The duplication of glyphs for DA led to the duplicate use of **ᚱ** for /dʰ/ (**ᚻ** KHOJKI LETTER DHA).

- **ᚻ** KHOJKI LETTER BBA /ɓ/ was used to represent /b/, already represented as (**ᚿ** BA). In turn, **ᚿ** was used synonymously with **ᚻ** to write BHA /bʰ/.

Other issues with redundancy are:

- **ᚻ** was used to represent not only /dʰ/ (DHA), but also /dʒʰ/ (**ᚼ** JHA).

- **ᚺ**, the glyph for the ligature *tra* (TA + VIRAMA + RA) was at times used to represent /t/ (**ᚽ** TA).

Ambiguity   In some cases, one glyph is used to represent multiple characters.

- The glyph **ᚂ** I is used for the independent form of both I and II.

- The glyph **ᚃ** II is also used for **ᚄ** YA.

- The glyph **ᚅ** is used for the independent form of both U and UU.

- The glyph **ᚆ** is used to represent both KA and QA (/q/, used to transcribe ق ARABIC LETTER QAF).

- In some manuscripts, the glyph **ᚇ** (NNA) represents /ɖ/ (DDA), /ɽ/ (LLA), and /ɳ/ (NNA). Furthermore, given allophonic variation between /ɖ/ an /ɽ/, it is difficult to determine precisely which of these sounds is being represented by **ᚇ** in such sources.

- The glyph **ᚈ** represents /ɖ/, but it was also used to represent /ɽ/.

- **ᚉ** is used for both O and U.

Deficiency   The script lacks unique glyphs to represent sounds not found in Sindhi:

- In order to represent Arabic sounds, new characters are created by adding the diacritic NUKTA to an existing character (see section 4.8).

- The consonant ligature **ᚊ**, which is composed of KA + VIRAMA + SSA; however, Khojki does not have the character SSA.

4.3   Supression of Inherent Vowel

As most Indic scripts, Khojki has the special sign *virāma*, which is written as a combining character with a consonant letter to supress the consonant's inherent vowel *a*. The **ᚋ** KHOJKI SIGN VIRAMA is written to the right of the consonant, not beneath as is typical in most Indic scripts, eg. Khojki **ᚌ** *k* and Devanagari क् *k*.

4.4   Consonant-Vowel Ligatures

The addition of dependent signs to consonant letters results in consonant-vowel ligatures for all consonants.

VOWEL SIGN I   Unlike the practice in other Indic scripts, KHOJKI VOWEL SIGN I is written to the right of the consonant, not to the left. The KHOJKI VOWEL SIGN I has two forms: **ᚍ** and **ᚎ**.

11

- ○ʃ is used with characters that have right descenders: ◌ → ◌ *ci*; ◌ → ◌ *vi*. Or that possess a final right stroke that terminates at the baseline: ◌ → ◌ *ki*; ◌ → ◌ *di*; ◌ → ◌ *hi*.

- ○⌐ is used with characters that have rounded right edges: ◌ → ◌ *chi*; ◌: → ◌ *ṅi*; ◌ → ◌ *ṭi*; ◌ → ◌ *ḍi*; ◌ → ◌ *ri*;

VOWEL SIGN II   The VOWEL SIGN II has two forms: ○◌ and ○◌:

- ○◌ is used with characters possessing right descenders: ◌ → ◌ *cī* and with characters with final strokes that terminate at the baseline: ◌ → ◌ *kī*.

- ○◌ is used with characters that have rounded right edges: ◌ → ◌ *chī*;

VOWEL SIGN U   With characters that possess a right descender, the ◌ VOWEL SIGN U is generally written as part of the descender: ◌ → ◌ *cu*. Otherwise, the vowel sign is written as an extension of the basic final stroke of a character: ◌ → ◌ *ku*, ◌ → ◌ *chu*, etc.

4.5   Consonant Conjuncts

Consonant clusters are typically limited to two consonants and are represented in four ways:

1. *Ligatures*   There are four consonant ligatures in Khojki:

   ◌ *kṣa*   =   ◌ KA   +   ○/ VIRAMA   +   *SSA

   ◌ *jña*   =   ◌ JA   +   ○/ VIRAMA   +   ◌: NYA

   ◌ *tra*   =   ◌ TA   +   ○/ VIRAMA   +   ◌ RA

   ◌ *dra*   =   ◌ DA   +   ○/ VIRAMA   +   ◌ RA

2. *Full-Form of Initial Consonant + Visible Virama*   The most common method of representing consonant conjuncts is by writing the full forms of the initial consonant and marking it with VIRAMA; eg. *pra* = ◌ PA + VIRAMA + ⃞ ZWNJ + RA

3. *Full-Form of Initial Consonant + Sub-Form of Final Consonant* This type of consonant conjunct is typically C + RA.

   ```
   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 form-
   ing thin paper covers; spine rounded and backed; black ink.

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

4. *Half-Form of Final Consonant*   In some cases, consonant conjuncts are written with the full-form of the initial consonant and a half-form of the second consonant.[2] For instance, the conjunct *kya* may be represented as ◌ (◌ YA → ◌ half-YA). See section 5.4 for discussion.

---

[2] Asani 1992: 72, 73.

```
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 form-
ing thin paper covers; spine rounded and backed; black ink.

ơ = ḍ, ṛ, ṇ; 𝌆 = dh; special form of conjuncts: ⲏ = br;
Ɫ = kr; ᵕ = sr; ⲏ = mr; ơⲟ] = ny; Ⲣⲟ) = ky; ⲏⲟ = ly; ⲟⲟ = ty;
ʒⲟ= jy; ᵕ = sv; ⲟⲟ] = gn.
```

4.6   Nasalization
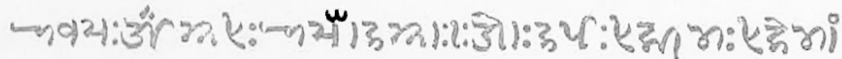
Nasalization is indicated by ̊ KHOJKI SIGN CANDRABINDU and ̇ KHOJKI SIGN ANUSVARA.

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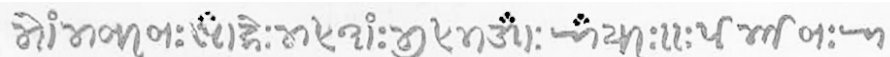
Geminated consonants are marked by ̆ KHOJKI SIGN SHADDA. The sign is written above the consonant that is doubled, eg. LA + LA is written 𝌆 (as below), not as a conjunct character. This practice is similar to that in Gurmukhi, where the ̆ U+0A71 GURMUKHI SIGN ADDAK is used to denote doubled consonants,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

4.8   Creation of New Characters

The Khojki script may be extended by producing characters to represent sounds not native to Sindhi. This is done by writing the sign *nukta* above a Khojki character whose phonetic value most closely matches that of the sound being accommodated. The *nukta* is written using the above-base combining character ̈ KHOJKI SIGN NUKTA:

The NUKTA may be written with vowel letters, vowel signs, and consonant letters. In Khojki, *nukta* is used primarily to represent Arabic letters and sounds, eg. 𝌆 A + NUKTA is used to represent ع U+0639 ARABIC LETTER AIN. The NUKTA has a variant, which is simply an inverted form: ̈.

The list of attested *nukta* letters in Khojki is:

KHOJKI			ARABIC		
	◌ّ	VOWEL SIGN A + NUKTA	ع	U+0639	ARABIC LETTER AIN
		LETTER A + NUKTA	ع	U+0639	ARABIC LETTER AIN
		VOWEL LETTER E + NUKTA	ع	U+0639	ARABIC LETTER AIN
qa		VOWEL LETTER KA + NUKTA	ق	U+0642	ARABIC LETTER QAF
k͟ha		VOWEL LETTER KHA + NUKTA	خ	U+062E	ARABIC LETTER KHAH
ġa		VOWEL LETTER GA + NUKTA	غ	U+063A	ARABIC LETTER GHAIN
za		VOWEL LETTER JA + NUKTA	ذ	U+0630	ARABIC LETTER THAL
			ز	U+0632	ARABIC LETTER ZAIN
			ض	U+0636	ARABIC LETTER DAD
			ظ	U+0638	ARABIC LETTER ZAH
ṭa		VOWEL LETTER TA + NUKTA	ط	U+0637	ARABIC LETTER TAH
fa		VOWEL LETTER PHA + NUKTA	ف	U+0641	ARABIC LETTER FEH
ḥa		VOWEL LETTER HA + NUKTA	ح	U+062D	ARABIC LETTER HAH

4.9   Punctuation

Word Boundaries   Khojki separates words using ꞉ KHOJKI WORD SEPARATOR. This character resembles characters such as ◌꞉ U+0903 DEVANGARI SIGN VISARGA. Khojki does not use VISARGA.

Sentence Boundaries   The ꞁ DANDA and ꞁꞁ DOUBLE DANDA are used to mark the end of sentences.
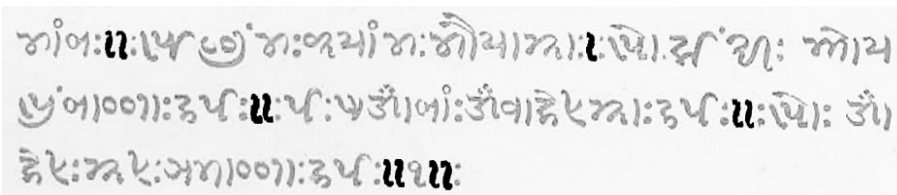
End of line is typically marked with ꞉ WORD SEPARATOR and ꞁꞁ DOUBLE DANDA:
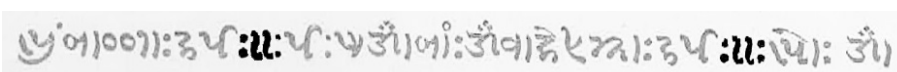
DOUBLE DANDA is also used to mark verse sections. Typically, DOUBLE DANDA is written with WORD SEPA-RATOR to the left and right of verse numbers:
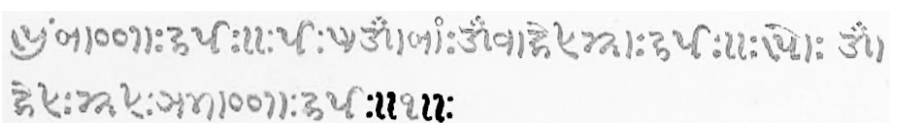
Section Marks   The section mark characters ︤ KHOJKI SECTION MARK and ︥ KHOJKI DOUBLE SECTION MARK appear frequently in Khojki manuscripts as punctuation that delimit the end of a section or another larger block of text. Both section marks are found in manuscripts. The KHOJKI SECTION MARK is generally

14

used to mark the end of a sentence, while KHOJKI DOUBLE SECTION MARK is used to delimit larger blocks of text, such as paragraphs. Both generally extend to the margin of the text-block.
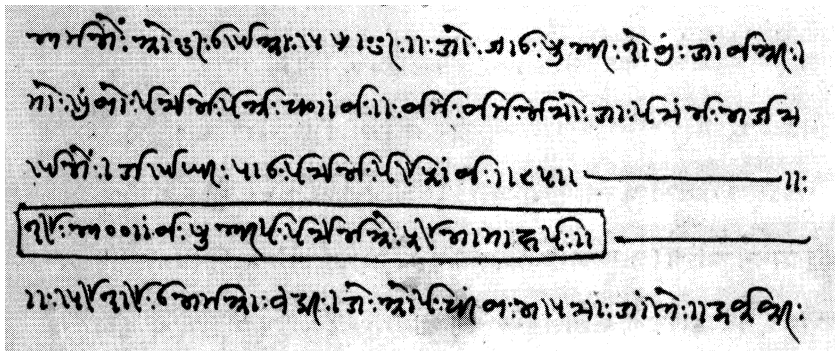
Verse Enumeration   The ⬚ KHOJKI VERSE MARKER is used to mark verse numbers in long Khojki texts. It is an above-base extending character that is written above numbers: ॥५२५॥ Its function may be considered similar to the Latin mark *vinculum*, which is written over numbers to indicate that they are to be considered a group.

Other Punctuation   Printed Khojki employs Latin punctuation marks, such as comma, which are already encoded in the UCS:
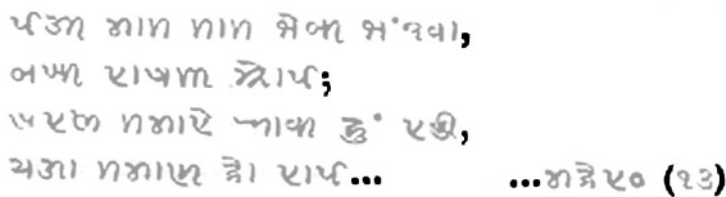
4.10   Abbreviations

Khojki has two abbreviations for Arabic benedictions. These are not proposed for encoding because they can be represented using other characters.

1. ﹾﻪﻧ = علیه السلام   *'alai-hi as-salām* 'peace be upon him'.

2. ﹾﻪﻧ = صلّی الله علیه و سلّم   *ṣallā 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; ૭ = û, o; ૩ૌ = z; ૨૭ = j; ﻭ = shaddah; ﹾﻪﻧ = abbreviation
for Arabic benediction ʿalai-hi as-salam "peace be upon him";
ﹾﻪﻧ = abbreviation for Arabic benediction ṣalla allahu ʿalai-hi
wa sallam "God bless him and grant him salvation."
```

4.11 Number Forms and Unit Marks

Number forms and unit marks are found in Khojki documents (see Figure 17). The most common characters are signs for writing fractions: eg. ‾ U+A833 NORTH INDIC FRACTION ONE SIXTEENTH, ‖ U+A831 NORTH INDIC FRACTION ONE HALF, etc.; and currency marks: eg.) U+A838 NORTH INDIC RUPEE MARK. These characters may be represented using those in the Common Indic Number Forms block (U+A830..U+A83F). See Pandey (2007) for more information on these characters.

In the specimen below, the parenthesis-like character represents) U+A838 NORTH INDIC RUPEE MARK. The number form represents the currency value of either '6 pai' or '8 ana'.

4.12 Variant Forms of Characters

1. ᴜ KHOJKI LETTER E is also represented as ᴐ.

2. ᴐ KHOJKI LETTER AI is also represented as ᴜᴜ.

3. 6 KHOJKI LETTER O is also represented as ᴐ.

4. ᴐ KHOJKI LETTER AU is also represented as ᴜᴐ and ᴜ6.

5. ᴎ KHOJKI LETTER KA is also written as ᴎ (KHOJKI LETTER QA).

6. ᴎ KHOJKI LETTER DHA is also written as 3.

7. ᴍ KHOJKI LETTER BA is also written as ᴎ.

5 Implementation

5.1 Encoding Model

The encoding model for Khojki may be based on the model implemented for Devanagari. There may be a requirement to accommodate an idiosyncratic method of rendering consonant conjuncts for the script (see section 5.4).

5.2 Collation

The collating order for Khojki is based on Sindhi. The script may be sorted according to both the Devanagari-based and Persian-based arrangement of letters.

Devanagari-Based Sort Order   Independent vowel letters are sorted before consonant letters. Vowels with special signs (CANDRABINDU and ANUSVARA) appear at the head of the vowel order.

The collating order for special signs and independent vowels in Khoji is:

aṁ	*aṃ*	*a*	*ā*	*i*	*ī*	*u*	*ū*	*e*	*ai*	*o*	*au*

Dependent vowel signs are sorted in the same position as their independent shape. Consonant-vowel combinations are sorted first by consonant letter and then by vowel sign. A consonant with *virāma* is sorted last.

ně	ṅ	ṅ	ṅ	ṅ	ṅ	ṅ	ṅ	ṅ	ṅ	ṅ	ṅ
kaṃ̌	*kaṃ*	*ka*	*kā*	*ki*	*kī*	*ku*	*ke*	*kai*	*ko*	*kau*	*k*

The pattern for consonants is as follows:

ka	kha	ga	ga	gha	ṅa	ca	cha	ja	ja	jha	ña	ṭa
ka	*kha*	*ga*	*g̱a*	*gha*	*ṅa*	*ca*	*cha*	*ja*	*j̱a*	*jha*	*ña*	*ṭa*

ṭha	*ḍa*	*ḍha*	*ṇa*	*ta*	*tha*	*da*	*ḏa*	*dha*	*na*	*pa*	*pha*	*ba*

ḇa	*bha*	*ma*	*ya*	*ra*	*la*	*ḷa*	*va*	*śa*	*sa*	*ha*

Persian-Based Sort Order   Since the Sindhi language is now commonly written in an extended form of the Arabic script, Khojki letters may be sorted according to the Arabic arrangement of characters as extended for Sindhi.

ba	*ḇa*	*bha*	*ta*	*tha*	*ṭa*	*ṭha*	*pa*	*pha*	*ja*	*j̱a*	*jha*	*ña*

ca	*cha*	*da*	*dha*	*ḏa*	*ḍa*	*ḍha*	*ra*	*sa*	*śa*	*ka*	*kha*	*ga*

g̱a	*gha*	*ṅa*	*la*	*ḷa*	*ma*	*na*	*ṇa*	*va*	*ha*	*ya*

5.3   Character Properties

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

```
11280;KHOJKI SIGN CANDRABINDU;Mn;0;NSM;;;;;N;;;;;
11281;KHOJKI SIGN ANUSVARA;Mn;0;NSM;;;;;N;;;;;
11282;KHOJKI LETTER A;Lo;0;L;;;;;N;;;;;
11283;KHOJKI LETTER AA;Lo;0;L;;;;;N;;;;;
11284;KHOJKI LETTER I;Lo;0;L;;;;;N;;;;;
11285;KHOJKI LETTER II;Lo;0;L;;;;;N;;;;;
11286;KHOJKI LETTER U;Lo;0;L;;;;;N;;;;;
11287;KHOJKI LETTER UU;Lo;0;L;;;;;N;;;;;
11288;KHOJKI LETTER E;Lo;0;L;;;;;N;;;;;
11289;KHOJKI LETTER AI;Lo;0;L;;;;;N;;;;;
1128A;KHOJKI LETTER O;Lo;0;L;;;;;N;;;;;
1128B;KHOJKI LETTER AU;Lo;0;L;;;;;N;;;;;
1128C;KHOJKI LETTER KA;Lo;0;L;;;;;N;;;;;
1128D;KHOJKI LETTER KHA;Lo;0;L;;;;;N;;;;;
1128E;KHOJKI LETTER GA;Lo;0;L;;;;;N;;;;;
1128F;KHOJKI LETTER GGA;Lo;0;L;;;;;N;;;;;
11290;KHOJKI LETTER GHA;Lo;0;L;;;;;N;;;;;
```

```
11291;KHOJKI LETTER NGA;Lo;0;L;;;;;N;;;;;
11292;KHOJKI LETTER CA;Lo;0;L;;;;;N;;;;;
11293;KHOJKI LETTER CHA;Lo;0;L;;;;;N;;;;;
11294;KHOJKI LETTER JA;Lo;0;L;;;;;N;;;;;
11295;KHOJKI LETTER JJA;Lo;0;L;;;;;N;;;;;
11296;KHOJKI LETTER JHA;Lo;0;L;;;;;N;;;;;
11297;KHOJKI LETTER NYA;Lo;0;L;;;;;N;;;;;
11298;KHOJKI LETTER TTA;Lo;0;L;;;;;N;;;;;
11299;KHOJKI LETTER TTHA;Lo;0;L;;;;;N;;;;;
1129A;KHOJKI LETTER DDA;Lo;0;L;;;;;N;;;;;
1129B;KHOJKI LETTER DDHA;Lo;0;L;;;;;N;;;;;
1129C;KHOJKI LETTER NNA;Lo;0;L;;;;;N;;;;;
1129D;KHOJKI LETTER TA;Lo;0;L;;;;;N;;;;;
1129E;KHOJKI LETTER THA;Lo;0;L;;;;;N;;;;;
1129F;KHOJKI LETTER DA;Lo;0;L;;;;;N;;;;;
112A0;KHOJKI LETTER DDDA;Lo;0;L;;;;;N;;;;;
112A1;KHOJKI LETTER DHA;Lo;0;L;;;;;N;;;;;
112A2;KHOJKI LETTER NA;Lo;0;L;;;;;N;;;;;
112A3;KHOJKI LETTER PA;Lo;0;L;;;;;N;;;;;
112A4;KHOJKI LETTER PHA;Lo;0;L;;;;;N;;;;;
112A5;KHOJKI LETTER BA;Lo;0;L;;;;;N;;;;;
112A6;KHOJKI LETTER BBA;Lo;0;L;;;;;N;;;;;
112A7;KHOJKI LETTER BHA;Lo;0;L;;;;;N;;;;;
112A8;KHOJKI LETTER MA;Lo;0;L;;;;;N;;;;;
112A9;KHOJKI LETTER YA;Lo;0;L;;;;;N;;;;;
112AA;KHOJKI LETTER RA;Lo;0;L;;;;;N;;;;;
112AB;KHOJKI LETTER LA;Lo;0;L;;;;;N;;;;;
112AC;KHOJKI LETTER LLA;Lo;0;L;;;;;N;;;;;
112AD;KHOJKI LETTER VA;Lo;0;L;;;;;N;;;;;
112AE;KHOJKI LETTER SHA;Lo;0;L;;;;;N;;;;;
112AF;<reserved>
112B0;KHOJKI LETTER SA;Lo;0;L;;;;;N;;;;;
112B1;KHOJKI LETTER HA;Lo;0;L;;;;;N;;;;;
112B2;KHOJKI LETTER QA;Lo;0;L;;;;;N;;;;;
112B3;KHOJKI LETTER KSSA;Lo;0;L;;;;;N;;;;;
112B4;KHOJKI VOWEL SIGN AA;Mn;0;NSM;;;;;N;;;;;
112B5;KHOJKI VOWEL SIGN I;Mc;0;L;;;;;N;;;;;
112B6;KHOJKI VOWEL SIGN II;Mc;0;L;;;;;N;;;;;
112B7;KHOJKI VOWEL SIGN U;Mn;0;NSM;;;;;N;;;;;
112B8;<reserved>
112B9;KHOJKI VOWEL SIGN E;Mn;0;NSM;;;;;N;;;;;
112BA;KHOJKI VOWEL SIGN AI;Mn;0;NSM;;;;;N;;;;;
112BB;KHOJKI VOWEL SIGN O;Mn;0;NSM;;;;;N;;;;;
112BC;KHOJKI VOWEL SIGN AU;Mn;0;NSM;;;;;N;;;;;
112BD;KHOJKI SIGN VIRAMA;Mc;9;NSM;;;;;N;;;;;
112BE;KHOJKI SIGN NUKTA;Mn;7;NSM;;;;;N;;;;;
112BF;KHOJKI SIGN SHADDA;Mn;0;NSM;;;;;N;;;;;
112C0;KHOJKI DANDA;Po;0;L;;;;;N;;;;;
112C1;KHOJKI DOUBLE DANDA;Po;0;L;;;;;N;;;;;
112C2;KHOJKI WORD SEPARATOR;Po;0;L;;;;;N;;;;;
112C3;KHOJKI SECTION MARK;Po;0;L;;;;;N;;;;;
112C4;KHOJKI DOUBLE SECTION MARK;Po;0;L;;;;;N;;;;;
```

5.4   Outstanding Issues

5.4.1   Encoding Model

1. **Half-form of final letter of a consonant conjunct**   Consonant conjuncts in Khojki are generally written with a visible VIRAMA. For instance, the conjunct *kya* is typically represented as 𑊬/𑈵:

𑈈 KA + ◌𑈵 VIRAMA + ⟦ZWNJ⟧ ZWNJ + 𑈯 YA

Alternately, this conjunct may also be written in a rather unconventional manner: the full-form of the initial consonant letter and the half-form of the second consonant letter:[3] 𑈈𑈯. Despite the visual disjunction, 𑈈𑈯 represents *kya*, not *kay*, and is semantically identical to 𑈈/𑈯. Given this, the expected character sequence would be

𑈈 KA + ◌𑈵 VIRAMA + ⟦ZWJ⟧ ZWJ + 𑈯 YA   (*kya*)

not

𑈈 KA + 𑈯 YA + ◌𑈵 VIRAMA + ⟦ZWJ⟧ ZWJ   (*kay*)

Here, the appearance of the conjunct is the reverse of the semantic value of the characters. The letter KA remains visually unchanged and unmarked by VIRAMA, but the elision of its inherent vowel is indicated by the half form of YA. Thus, the rendering behavior of VIRAMA in this case affects not the character it is written after (KA), but the one that follows it (YA): 𑈯 → 𑈯 (semantically YA, but visually half-YA).

This behavior cannot be produced by the typical method of producing conjuncts by applying VIRAMA to the consonant letter in question.

6   References

Asani, Ali S. 1992. *The Harvard Collection of Ismaili Literature in Indic Languages: A Descriptive Catalog and Finding Aid*. Boston: G. K. Hall & Co.

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.

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`

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.

Unicode Roadmap Committee, The Unicode Consortium. 2009. "Roadmap to the SMP." Authored by Michael Everson, Rick McGowan, and Ken Whistler. Revision 5.1.4 (February 4, 2009). `http://www.unicode.org/roadmaps/smp/smp-5-1-4.html`

---

[3] Asani 1992: 72, 73.

Roman characters.	Dēvanāgarī.	Khudāwāḍī.*	Shikārpurī.	Sakkar.	THAṬṬĀI.		Lārī.	Waṅgāī.	Rājāi.	Khwājās.	MAIMONS.		Sewhāṇī Bhāhiṇās.	Southern Lahnda.
					Luhānās.	Bhaṭiās.					Thaṭṭā.	Haidarābād.		
a	अ	m	m	m	″∧	m	m	⸮ʊ	ᴎ	⌐	⌐	⌐	)n	ʌ
ā	आ	m	m	m	″∿	m	m	⸮ʊ	ᴎ	⌐ⵏ	⌐	⌐	)n	ʌ
i	इ	†ᠸ	†ᠸ	†ᠸ	ᵒᵒᵢ	ʏ	⋺	ᶘ	⸮ɩ	⺭	⸳ɩ	⸳⸴	ᴄ	6
ī	ई	†ᠸ	†ᠸ	†ᠸ	ᵒᵒᵢ	ᴎ	ᴢ	ᶘ	⸳ɩ	⸮ⅈ	⸳ⅉ	⸳ⅉ	⸦	6
u	उ	m	m	m	ᠸ	ᠸ	ᠸ	ᠸ	᠆	ᔕ	᠆	᠆	ᴑ	ʌ
ū	ऊ	m	m	m	ᠸ	ᠸ	ᠸ	ᠸ	᠆	ᔕ	᠆	᠆	ᴑ	ʌ
ē	ए	m	m	m	ᴢ	ᴎ	⸉	ᴄ	⸮ᶘ	ᴃ	ᴇ	⸮	ɩn	6
ai	ऐ	m	m	m	″∿	m	ᵐᴢ	⸮ᶘᶘ	ᴎⵑ	⌐ᴃ	⌐⸮ⅉ	⌐⌐ᴎ	ɩn	ʌᶜ
ō	ओ	m	m	m	ᠸ	ᠸ	᠆	᠆	᠆	ᔕᠸ	᠆	᠆	ɩn	ʌ
au	औ	m	m	m	ᠸ	6	6	᠆	᠆	⌐ᔕ	ᴢᠸ	⌐ᠸ	ɩn	ᴍ

* The Khudāwāḍī is the character used at Haidarābād 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 **ⅿ** is used.

Figure 1: [Inventory of Khojki ('Khwājās') vowel letters from Grierson (1905: 15).

Roman characters.	Dēvanāgari.	Khudāwādi.	Shikārpuri.	Sakkar.	Thaṭṭāi. Lohānā.	Thaṭṭāi. Bhāṭiā.	Lāṛi.	Wāngaī.	Rājāi.	Khwājās.	Maimons. Thaṭṭa.	Maimons. Haidarābād.	Sewhāṇi Bhāibhirtā.	Southern Lahnda.
ka	क													
kha	ख													
ga	ग													
gga	ऽ													
gha	घ													
ṅa	ङ													
cha	च													
chha	छ													
ja	ज													
jja	ॼ													
jha	झ													
ña	ञ													
ṭa	ट													
ṭra	ड़													
ṭha	ठ													
ra	ड													
ḍḍa	꣪													
ḍra	꣪													
ḍha	ड													
ṇa	ण													

Figure 2: [Inventory of Khojki ('Khwājās') consonant letters from Grierson (1905: 16).

21

| Roman characters. | Dēvanāgarī. | Khudāwādī. | Shikārpurī. | Sakkar. | Thaṭṭāi. | | Lāṛāi. | Waṅgāi. | Rājjāi. | Khwājās. | Maimoṇs. | | Sewhāṇī Bhāabhirās. | Southern Lahndā. |
					Luhānās.	Bhāṭiās.					Thaṭṭa.	Haidarābād.		
ta														
tha														
da														
dha														
na														
pa														
pha														
ba														
bba														
bha														
ma														
ya														
ra														
la														
va														
śa														
sha														
sa														
ha														

Numerals	1	2	3		4	5	6	7	8	9	10

Figure 3: Inventory of Khojki ('Khwājās') consonant letters from Grierson (1905: 17).

22

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

Roman	Urdu	Sindhi	Gujarati	Khojki#
a	ا	ا	અ	—, —
ā	آ	آ	આ	—ا, —ا
i*	ِا	ِا	ઈ	[ୢ] (⤳ ?)
ī*	اِي	اِي	ઈ	[ୢ] (?)
u*	اُ	اُ	ઉ	[ ]
ū*	اُو	اُو	ઊ	[ ] (?)
o	او	او	ઓ	[,] (?)
e	اے	اے	એ	[ ] (?)
ai	اے	اے	ઐ	[,]
au	او	او	ઔ	[, ,]

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

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

Figure 4: Inventory of Khojki independent vowel letters (from Asani 1992: 678).

23

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

Roman	Urdu	Sindhi	Gujarati	Khojki#
bad	بَد	بَد	બદ	૪૬ [૪૬]
bād	باد	باد	બાદ	૪૧૬
din	دن	دن	દિન	દ્ન
dīn	دین	دین	દીન	ધ્ન
but*	بُت	بُت	બુત	શુન
būt*	بوت	بوت	બૂત	શુન
so	سو	سو	સો	ખો [૪૬]
be	بے	بی	બે	ખ [૪૬]
hai	بَے	ہَی	ૐ	[ૐ, ૨-૫-૪]
nau	نوُ	نوُ	નૌ	[નૌ, ન-૫-૬]

*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 5: Inventory of Khojki dependent vowel signs (from Asani 1992: 679).

24

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

Roman	Urdu	Sindhi	Gujarati	Khojki#
b	ﺏ	ﺏ	બ	ઝ [,]
ḅ		ﺏ		ઞ
bh	ﭖ	ﭖ	ભ	ઞ, [ ]
t	ﺕ	ﺕ	ત	ય
th	ﺕ	ﺕ	થ	[ ]
ṭ	ﻁ	ﺕ	ટ	ट, ट, (?)
ṭh	ﻁ	ﺕ	ઠ	ठ
ṣ	ﺙ	ﺙ	(સ)	( )
p	ﭖ	ﭖ	પ	(?)
ph	ﭖ	ﻑ	ફ	૪
j	ﺝ	ﺝ	જ	૩, , (?)
dy		ﺝ		(3?)

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

Figure 6: Inventory of Khojki consonants: *b* to *dy* (from Asani 1992: 680).

25

Roman	Urdu	Sindhi	Gujarati	Khojki‡
jh				
ny			म	२: (व्ह?)
ċ			य	ज, (ट?)
ċh			छ	२ए, (म?)
ḥ			(ड़)	(ह)
kh			(झ)	(च)
d			द	ट, २
dh			ध	२।, २, (ड?)
ḍ				ट
ḍ			ऽ	(२, ३, २?)
ḍh			ढ	२ (ड?)
ź			(ळ)	(३ं, २)
r			?	८
ṛ			ण	[ल]
ṛh				
z			?	३ं [य]
s			स	४

Figure 7: Inventory of Khojki consonants: *jh* to *s* (from Asani 1992: 681).

26

Roman	Urdu	Sindhi	Gujarati	Khojki#
sh	نٹ	نٹ	શ	...
ṣ	ۼ	ۼ	(સ)	(...)
ẓ	غٝ	غٝ	(ઝ)	(...)
ṭ	ط	ط	(ત,ટ)	(...)
ẓ	ظ	ظ	(ઝ)	(...)
ʿ	ع	ع	(અ)	... (...)
ġ	غ	غ	(ગ)	... (...)
f	ف	ف	ફ	...
q	ق	ق	(ક)	... (...)
k	ک	ک	(ક)	... [...]
kh	کﮬ	ک	ખ	...
g	گ	گ	ગ	... [...]
ṅ		گ		...
gh	گﮬ	گﮬ	ધ	... (...)
ng		گ	ડ	... (...)
l	ل	ل	લ	... [...]

Figure 8: Inventory of Khojki consonants: *sh* to *l* (from Asani 1992: 682).

27

Roman	Urdu	Sindhi	Gujarati	Khojki#
m	ρ	ρ	મ	ખ
n	ω	ω	ન	ન,બ
ṇ		♂	ણ	ખ (ખ:?)
w/v	و	و	વ	વ [વ]
h	১	১	ગ	ર [દ્]
y	ৎ	ৎ	ચ	વ,૦૦૧, (ﬤ?)
ksh			ક્ષ	ક્ષ , (શ્ર?)
gn			શ	ગ
tr			ત્ર	ત્ર (ત્ર,ટ?)
dhr			ધ્ર	૩

Nasalization is indicated by ñ or ṁ

Figure 9: Inventory of Khojki consonants: *m* to *dhr* (from Asani 1992: 683).

28

Table 1

Khojkī Script: Vowel System (early 20th century)

(a) Initial and Independent

a	ā	i	ī	u	ū	e	o	au	ay

(b) Medial and Final
(with the Consonant)

a	ā	i	ī	u	ū	e	o	au	ay

Table 2

Khojkī Script: Consonants (early 20th century)

k	kh	g	gh	ng
c	ch	j/jh	j/jh	dy
ṭ	ṭh	ḍ	ḍh	ny	ng
t	th	d	dh	n	ū
p	ph	b	bh	m	ḍ
y	r	l	v	s	sh
dr	tr	h	r/l		

	1		2		3		4		5
⟍	a	⟍⟍	lā	⟍⟍	to	⟍⟍	hā	⟍	r
⟍⟍	ā	⟍⟍	i	⟍⟍	u	⟍⟍	e	⟍⟍	g^2
⟍⟍	k	⟍⟍	kh	⟍⟍	g^1	⟍⟍	gh	⟍⟍	ṅ
⟍⟍	c	⟍⟍	ch	⟍⟍	j^1	⟍⟍	j^2	⟍⟍	ñ
⟍	ṭ	⟍	ṭh	⟍⟍	ḍ	⟍⟍	ḍh	⟍⟍	ṇ
⟍	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 10: 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
⟍⟍	ṅ	⟍	c	⟍⟍	ch	⟍⟍	j/jh	⟍⟍	j/jh
⟍⟍	ñ	⟍	ṭ	⟍	ṭh	⟍⟍	ḍ	⟍⟍	ḍh
⟍⟍	ṇ	⟍	t	⟍	tr	⟍⟍	th	⟍⟍	d
⟍	dr	⟍⟍	dh	⟍	n	⟍	p	⟍	ph
⟍⟍	b	⟍⟍	bh	⟍⟍	m	⟍⟍	y	⟍	r
⟍⟍	l	⟍	v	⟍⟍	s	⟍⟍	h		

Figure 11: Arrangement of Khojki characters based on the traditional Indic phonological pattern (from Shackle and Moir 1992: 36).

30

Roman	Devanagari	Gujarati	Sindhi	D MS 1815	Kx MS 1737	Khojaki Primer 1932
k						
kh						
g						
gh						
c						
ch						
j						
jh						
ṭ						
ṭh						
ḍ						

Figure 13: Comparison of Khojki with other scripts: *ka* to *ḍa* (from Khakee 1981).

Roman	Devanagari	Gujarati	Sindhi	D MS 1815	Kx MS 1737	Khojaki Primer 1932
a						
ā						
i						
ī						
u						
ū						
e						
o						
ai						
au ou						

Figure 12: Comparison of Khojki with other scripts: *a* to *au* (from Khakee Figure 13: Comparison of Khojki with other scripts: *a* to *au* (from Khakee 1981).

31

Roman	Devanagari	Gujarati	Sindhi	DMS 1815	Kx MS 1737	Khojaki Primer 1932
bh					[12]	
m						
y					[13]	
r						
ru						
l						
v						
v s					[14]	
s.					[14]	
s						
h						

Figure 15: Comparison of Khojki with other scripts: *bha* to *ha* (from Khakee 1981).

Roman	Devanagari	Gujarati	Sindhi	DMS 1815	Kx MS 1737	Khojaki Primer 1932
ḍh						
ṇ					[8]	
t						
th						
d						
dh					[9]	
n						
p					[10]	
f					[11]	
ph						
b						

Figure 14: Comparison of Khojki with other scripts: *ḍha* to *ba* (from Khakee 1981).

32

Roman	Devanagari	Gujarati	Sindhi	D MS 1815	Kx MS 1737	Khojaki Primer 1932
r	ॐ	५	‡	८	8 म	∅
ks	क्ष	क्ष	—	—	—	क्ष
gn	ग्न	ग्न	—	—	—	ग्न
tr	क	⅄	—	⅄	⅄	⅄
dv	—	ﻉ	—	ﻉ	—	—
dy	—	—	Ⅎ	—	15 ﻉ	—
ny	—	—	Ⅎ	—	16 ﻉ	—
z	—	लट	و و	—	17 —	—
dhr	—	—	—		ﻉ	ﻉ
sr	—	—	—	२	—	—

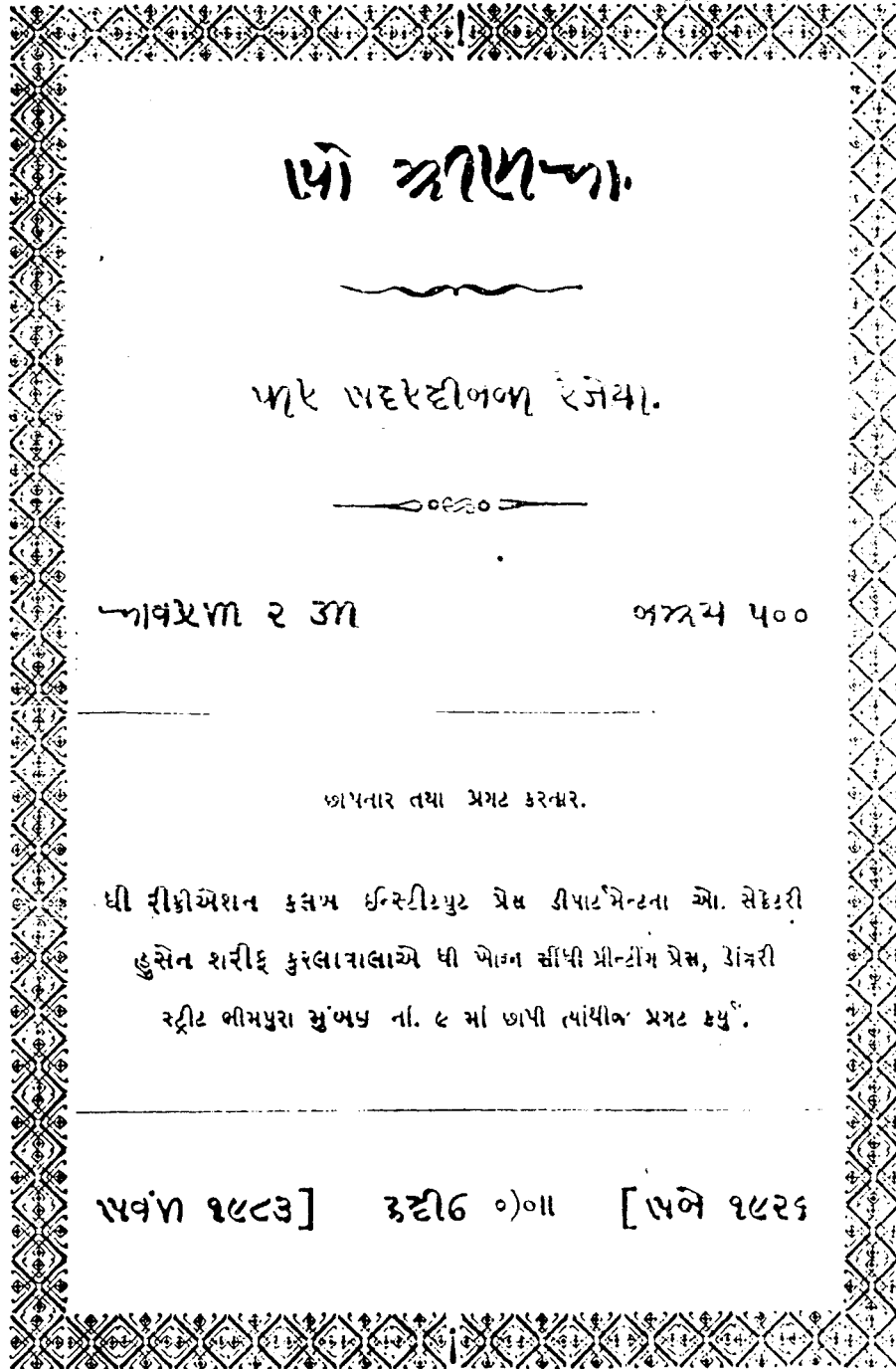Figure 16: Comparison of Khojki with other scripts: (*ra* to *sra* (from Khakee 1981).

33

Figure 17: Title page of the 𑈶𑈬 𑈈𑈮𑈦𑈬 (*So Kīrīā*) 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.

Figure 18: Pages from a Khojki book printed with the metal fonts cut by Laljibhai Devraj.

Ms Ism K 1   [Bhadneshwar] [192–?]

Foreign wove paper; watermark: E.R. & Co., Made in Austria,
chariot; 96 folios; 184; 20 x 16 (17 x 13); 12 – 16 lines;
stab/sewn binding laced on hard board with half-cloth covering
with decorative paper; black ink.

Scribe: Khojā bhāī Ladhû Devjī (65; 179–80); does not use short
"i" vowel; ... = ḍ ŗ ņ; ... = j; ... = h; ... = o; ... = pr.

Pp. 41–43 torn, no other significant damage; **tafsilo:** p. 182 [f.
91vl]; 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.

[**Garbis.**   _Ref:_   D 61-64 I 177 N 9,44,81,100
P 300]

1–2 (a.t.)	ejī:ād:gûr:samas:mûnivar:jā:n	
	10 verses, G, Pīr Shams	
2–4 (rāg ćāl)	ajī:ānañd:pāmeā:manke:sat:gûr	
	18 verses, G, Pīr Shams	
4–6 (same rāg as above)	gûr:āvea:bijī:rātke:avī	
	18 verses, G, Pīr Shams	
6–8 (same rāg as above)	gûr:āvatā:sarve:rātake:ganān	
	18 verses, G, Pīr Shams	
8–11 (same rāg as above)	ke:tame:amīras:pījo:dhanane:rātake	
	18 verses, G, Pīr Shams	

11–13 (same rāg)	mañhe:varaşete:abharat:nûrake
	18 verses, G, Pīr Shams
13–15 (same rāg)	tame:ramajo:dhanane:rātake
	19 verses, G, Pīr Shams
15–18 (same rāg)	tame:japajo:dhanane:rātake
	22 verses, G, Pīr Shams
18–20 (rāg ćaltī)	bhûlo:bhûlo:te:bhamaraņo
	22 verses, G, Pīr Shams
20–22 (rāg ṭhāl/ḍhāl?)	sat:mārg:samas:pīrake
	18 verses, G, Pīr Shams
22–25 (same rāg)	sate:pāvar:pīo:dhanane:rātke
	21 verses, G, Pīr Shams
25–26 (rāg ćaltī)	ćeto:ćeto:te:ćañćal:ćetīarelol
	18 verses, G, Pīr Shams
26–29 (rāg ṭhāl/ḍhāl?)	gûrû:āvea:sātmī:rātake
	18 verses, G, Pīr Shams
29–31 (rāg dohrā)	satgûr:samas:kaere:gafalo
	18 verses, G, Pīr Shams
31–32 (rāg ćaltī)	gûr:kahere:gafalo:sabhaŗo
	18 verses, G, Pīr Shams
33–35 (rāg ṭhāl/ḍhāl?)	tathī:āthami:avea:gamanā:lokake
	19 verses, G, Pīr Shams

56

57

Figure 19: Khojki text written by Ali S. Asani for his *The Harvard Collection of Ismaili Literature in Indic Languages* (1992: 56–57).