Proposal to encode the Khojki vowel sign VOCALIC R in Unicode

Anshuman Pandey

pandey@umich.edu pandey.github.io/unicode

May 25, 2021

1 Introduction

This is a proposal to encode a new character in the Khojki block of the Unicode standard:

GLYPH	CODE	CHARACTER NAME
ृ	11241	KHOJKI VOWEL SIGN VOCALIC R

The proposed character name aligns with that of analogous signs in other Indic scripts. The representative glyph is based on printed attestations. The c was described in "Final Proposal to Encode the Khojki Script in ISO/IEC 10646" (L2/11-021, p. 14), but it was not included in the repertoire at that time due to lack of access to materials containing usage of the character:

VOWEL SIGN VOCALIC R Authors of Khojki instructional books often attempted to extend the script by borrowing orthographic practices from Devanagari. As shown below, one such script primer, *Sindhī pahelī copadī*, introduces the character Q, the dependent form of $\frac{1}{24}$ U+090B DEVANAGARI LETTER VOCALIC R, for transcribing words of Sanskrit origin. The book also specifies diacritic marks for representing conjunct with RA.

In 2017, Srinidhi A and Sridatta A submitted "Proposal to encode two characters in Khojki" (L2/17-307), which presented attestations for the Khojki ε . The Script Ad Hoc reviewed L2/17-307 and raised the following comments (L2/18-039, p. 12):

- 1. Is the evidence based on just one source?
- 2. Provide a translation of the text in figure 2
- 3. How is the independent vowel represented?

The current proposal addresses these comments and provides additional information. Regarding #1, usage of the Khojki c_{i} is illustrated here using examples from three sources, along with transliterations and glosses. The figure mentioned in #2 and the requested translation is provided in the description of *Sindhī Pahelī Copadī* below. In response to #3, the three sources I've consulted that contain the dependent vowel sign c_{i} do not show a corresponding independent letter; it likely does not exist (see details below). Consequently, this document supersedes L2/17-307 and serves as sufficient justification for encoding this character in Unicode.

2 Description

The Sanskrit /r/ is known as the 'vocalic r' and is transliterated as r. In modern Indo-Aryan languages it is realized not as /r/, but as /ri/ (eg. Hindi) or /ru/ (eg. Sindhi, Gujarati). In northern Indic scripts with traditions for writing Sanskrit, the r is represented using an independent vowel letter and a corresponding dependent vowel sign, eg. \mathcal{R} U+090B DEVANAGARI LETTER VOCALIC R and \mathcal{Q} U+0943 DEVANAGARI VOWEL SIGN VOCALIC R. In scripts that lack these graphemes, the r is represented as the syllable ri or ru. In Khojki, r is represented as ru, eg. the independent γ would be $\mathcal{X} = \mathcal{K}$ U+11226 KHOJKI LETTER RA + \mathcal{Q} U+1122E KHOJKI VOWEL SIGN U.

In the early 20th century, the Khoja Sindhi Printing Press in Bombay began using \geq for transliterating the dependent *r* in Khojki. Its usage is described in the foreword of *Sindhī Pahelī Copadī* (1910: 9):

```
(૩) ડો કેસા છે "પ" બા બા ઝે ખાવા (.) બા જા બા તૈ ાગ
ગે ગેના શુ દા પા જા જાં ખાવા (ે) પૈક્ષના હ બ ગા પ દ્ર પ્
વા. ડો જ દે "પૂપ્વ " મેલુ શુ દા પા " પૂર્વ". દ્ર ખ
ને "પ્ર" ડો જાવાને જા દે જ્ર' બા બા ઝે ખાવા (ડ) બા જા બા
પા ખવા જાં ખાવા દે. ખાને પડા પા મે ડો ખક્ષ પ્લા
પા "પ્ર" ખક્ષ પ્લ 'ડો જા જ ખાળ દે મે દ પેદ્ર ખ્રબદ્ધ પ
ગ "પ્ર" ખક્ષ પ્લ 'ડો જા જ ખાળ દે મે દ પેદ્ર ખ્રબદ્ધ પ
ડા પા મે અલગા છું. ડો જ દ્રે દ્રુપા છું મેના છ ન ગા પ શ્રા
પા જ જાં કુ પા છું " ખાને " મુર્ધ છા " ના શ્રા પા જા
જાં છ ન ગા પ્ર " વૃષ્ણા " પ્રજા હવા.
```

Lines 4–9 state that c is realized as λru , as in λ

```
P Ism K 17
                                          Bombay 1966 s./ 1910
Sindhi paheli copadi. 1st edition.
Publisher: Mûkhî Laljîbhaî Devraj, Manager of the Bombay Shia
Imami Ismaili Khoja Jamat Dept of Education, for the Siñdhī
Textbook Committee at the Khoja Sifidhi Printing Press; 2000
copies; 92 pp. in various paginations; 17 x 13; covers missing;
binding in fair condition; paper in good condition; safkadiuf
(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)
```

The Khojki \geq occurs not only in *Sindhī Pahelī Copadī*, but also in other texts printed by the Khoja Sindhi Printing Press under the auspices of the Aga Khan (the spirital leader of the Khoja community), such as *Sīlsīle Īmāmat* (1910) and *Bahare Rahemat* (1911), which was composed by the Aga Khan III. Examples are given below with citations, the corresponding Sanskrit in Devanagari, and the English gloss:

ખુર્પાર્ટ્સાં ખરાજ છે.

Sindhī Pahelī Copaḍī : भूभेंट sṛṣṭi 'creation' = Devanagari सृष्टि sṛṣṭi

નજા રાજાવાઝી નાઉપજાં કા એપ જોહાવાર્ય, તે દુવ

Sindhī Pahelī Copadī : ܐܡ< 'immortal' = Devanagari अमृत amrta

બજાળા, પદાપસાર સુવરા, પરદેશગાષા, જાગ

Sindhī Pahelī Copaḍī : अञ्चल namrtā 'mortal' = Devanagari नमृता namrtā

જેરા લી પુરા ભા માગળતા જાબનાં જૂપ્તા બ ખાવ્ય

Sindhī Pahelī Copadī : भूभभ trptī 'satisfaction' = Devanagari तृप्ती trptī

વા લુદાર્પ જ્રૃપા જ્રષ્ટાને ન્ગાપ જાને ડો પૈર્ણાં ન્ગાપ્લ્લા

Sindhī Pahelī Copadī : $k_r p \bar{a}$ 'grace' = Devanagari कृपा $k_r p \bar{a}$

વાબા પાપે ખાજી મધ્યર દેગું, પંટને ગેજે પૈપ્ય સ્પ્ જેના પક્ષ ગહી દાપ્ટેજ છે વ્ર જેનાવ્છું, ગેજાં મેને પક્ષ

Sindhī Pahelī Copadī : अर्थे krṣṇa 'Krishna' = Devanagari कृष्ण krṣṇa

22, જેશરા હદા & વૃજ્યાના ભારા હસ્ટ જગા વદ્વાનાને હ

Sīlsīle Īmāmat : वृभ्भ vrtī 'conduct' = Devanagari वृत्ति vrtti

• ભવાગ – જાપગા જીગમાં વેગ ભૂગા સપા ક્રમા ક

Sīlsīle Īmāmat : व्रेण vrta 'action' = Devanagari वृत्त vrtta

Bahare Rahemat : ܐܘܪܟܪ amrta 'immortal' = Devanagari अमृत amrta

As Khojki is a liturgial script of the Ismaili communities of South Asia, it is important that the Unicode repertoire for Khojki provide all characters required for fully representation of records in digital plain text. The \geq KHOJKI VOWEL SIGN VOCALIC R should be encoded in the standard to further enable such support.

3 Collation

The $_{\bigcirc}$ KHOJKI VOWEL SIGN VOCALIC R should be sorted after $_{\bigcirc}$ U+1122F KHOJKI VOWEL SIGN U and before $\stackrel{\frown}{\circ}$ U+11230 KHOJKI VOWEL SIGN E. This pattern is based on the Brahmi order for vowels.

4 Input Methods

5 Character Data

Character Properties: UnicodeData.txt

11241;KHOJKI VOWEL SIGN VOCALIC R;Mn;0;NSM;;;;;N;;;;;

Linebreaking Properties: LineBreak.txt

11241;CM # Mn KHOJKI VOWEL SIGN VOCALIC R

Indic Syllabic Category: IndicSyllabicCategory.txt

11241 ; Vowel_Dependent # Mn KHOJKI VOWEL SIGN VOCALIC R

Indic Positional Category: IndicPositionalCategory.txt

11241 ; Bottom # Mn KHOJKI VOWEL SIGN VOCALIC R

6 References

- Aga Khan III. 1911. *Bahare Rahemat*. Bombay: Hāshām Boghā Māstar, Navā Nāgparā. MS Indic 2534. Houghton Library, Harvard University, Cambridge, Mass.
- Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, Liang Hai, and Richard Cook. 2018. "Recommendations to UTC #154 January 2018 on Script Proposals" (L2/18-039). https://www.unicode.org/L2/L2018/18039-script-adhoc-rec.pdf
- Asani, Ali S. 1992. The Harvard Collection of Ismaili Literature in Indic Languages: A Descriptive Catalog and Finding Aid. Boston: G. K. Hall & Co.

- Pandey, Anshuman. 2011. "Final Proposal to Encode the Khojki Script in ISO/IEC 10646" (L2/11-021) https://www.unicode.org/L2/L2011/11021-khojki.pdf
- Sīlsīle îmāmat. 1910. Bombay: Ismaili Literature Society. MS Indic 2534. Houghton Library, Harvard University, Cambridge, Mass.
- *Siñdhī pahelī ćopadī*. 1910. Bombay: Mûkhī Lāljībhāî Devrāj. MS Indic 2534. Houghton Library, Harvard University, Cambridge, Mass.
- Srinidhi A and Sridatta A. 2017. "Proposal to encode two characters in Khojki" (L2/17-307) https://www.unicode.org/L2/L2017/17307-two-khojki-letters.pdf

7 Acknowledgments

I'd like to thank Srinidhi A and Sridatta A for reminding me that c KHOJKI VOWEL SIGN VOCALIC R has not yet been encoded.

ISO/IEC JTC 1/SC 2/WG 2 PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS				
Please read Principles and Procedures Document (P & P) from <u>http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html</u> for				
guidelines and details before filling this form.				
Please ensure you are using the latest Form from .http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html				
See also <u>http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html</u> for latest <i>Roadmaps</i> .				
A. Administrative				
1. Title: Proposal to encode the Khojki vowel sign VOCALIC R in Unicode				
2. Requester's name: Anshuman Pandey <pandey@umich.edu></pandey@umich.edu>				
3. Requester type (Member body/Liaison/Individual contribution): Expert contribution 4. Submission date: 2021-05-25				
4. Submission date: 2021-05-25 5. Requester's reference (if applicable):				
6. Choose one of the following:				
This is a complete proposal: Yes				
(or) More information will be provided later:				
B. Technical – General				
1. Choose one of the following:				
a. This proposal is for a new script (set of characters):				
Proposed name of script:				
b. The proposal is for addition of character(s) to an existing block: Yes				
Name of the existing block: Khojki				
2. Number of characters in proposal:				
3. Proposed category (select one from below - see section 2.2 of P&P document):				
A-Contemporary X B.1-Specialized (small collection) B.2-Specialized (large collection)				
C-Major extinct D-Attested extinct E-Minor extinct				
F-Archaic Hieroglyphic or Ideographic G-Obscure or questionable usage symbols				
4. Is a repertoire including character names provided? Yes				
a. If YES, are the names in accordance with the "character naming guidelines"				
in Annex L of P&P document? Yes				
b. Are the character shapes attached in a legible form suitable for review? Yes				
5. Fonts related:				
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the				
standard?				
Anshuman Pandey				
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): Anshuman Pandey				
6. References: a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? <u>Yes</u>				
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?				
7. Special encoding issues:				
Does the proposal address other aspects of character data processing (if applicable) such as input,				
presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? Yes				
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 <u>http://www.unicode.org</u> for such information on other scripts. Also				
see Unicode Character Database (<u>http://www.unicode.org/reports/tr44/</u>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.				
in momation needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.				

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

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	No		
If YES explain			
2. Has contact been made to members of the user community (for example: National Body,			
user groups of the script or characters, other experts, etc.)?	No		
If YES, with whom?			
If YES, available relevant documents:			
3. Information on the user community for the proposed characters (for example:			
size, demographics, information technology use, or publishing use) is included?	Yes		
Reference: See text of proposal			
4. The context of use for the proposed characters (type of use; common or rare)	Common		
Reference: See text of proposal			
5. Are the proposed characters in current use by the user community?	Yes;		
If YES, where? Reference: The Khoja Ismaili religious community and schola	rs		
6. After giving due considerations to the principles in the P&P document must the proposed characters	be entirely		
in the BMP?	N/A		
If YES, is a rationale provided?			
If YES, reference:			
 Should the proposed characters be kept together in a contiguous range (rather than being scattered) Can any of the proposed characters be considered a presentation form of an existing 	? Yes		
character or character sequence?			
If YES, is a rationale for its inclusion provided?			
If YES, reference:			
9. Can any of the proposed characters be encoded using a composed character sequence of either			
existing characters or other proposed characters?	No		
If VES is a rationale for its inclusion provided?			
If VES_reference			
10. Can any of the proposed character(s) be considered to be similar (in appearance or function)			
to, or could be confused with, an existing character?			
If YES, is a rationale for its inclusion provided?			
If YES, reference:			
· · · · · · · · · · · · · · · · · · ·	Yes		
If YES, is a rationale for such use provided?	703		
If YES, reference: Combining characters; see text of proposal			
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provide	d2 N/A		
If YES, reference:	u		
12. Does the proposal contain characters with any special properties such as			
control function or similar semantics?	No		
If YES, describe in detail (include attachment if necessary)			
13. Does the proposal contain any Ideographic compatibility characters?	No		
If YES, are the equivalent corresponding unified ideographic characters identified?			
If YES, reference:			