

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  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 , the dependent form of  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 . 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  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  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.

As Khojki is a liturgical 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 𑀓 KHOJKI VOWEL SIGN VOCALIC R should be encoded in the standard to further enable such support.

3 Collation

The 𑀓 KHOJKI VOWEL SIGN VOCALIC R should be sorted after 𑀕 U+1122F KHOJKI VOWEL SIGN U and before 𑀔 U+11230 KHOJKI VOWEL SIGN E. This pattern is based on the Brahmi order for vowels.

4 Input Methods

On digital keyboards with touch-and-hold features, the 𑀓 KHOJKI VOWEL SIGN VOCALIC R should be added to the character palette for the 𑀒 U+11226 KHOJKI LETTER RA key. On other keyboards it may be placed on a SHIFT, CTRL, or ALT layer, mapped to the same key as U+11226 KHOJKI LETTER RA.

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 īmāmat. 1910. Bombay: Ismaili Literature Society. MS Indic 2534. Houghton Library, Harvard University, Cambridge, Mass.

Siñdhī pahelī ćopađī. 1910. Bombay: Mūkhī Lālĳībhāī Devrāĳ. 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 𑀓 KHOJKI VOWEL SIGN VOCALIC R has not yet been encoded.

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

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

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

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

A. Administrative

1. Title: **Proposal to encode the Khojki vowel sign VOCALIC R in Unicode**

2. Requester's name: *Anshuman Pandey <pandey@umich.edu>*

3. Requester type (Member body/Liaison/Individual contribution): *Expert contribution*

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: 1

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>	
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>	
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>		

4. Is a repertoire including character names provided? 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? 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:

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 <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

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