Proposal to encode two characters in Khojki

Srinidhi A and Sridatta A

Tumakuru, India

srinidhi.pinkpetals24@gmail.com, sridatta.jamadagni@gmail.com

August 30, 2017

1 Introduction

This is a proposal to encode two characters in the Khojki block.

Glyph	Code	Character name
◌ृ	U+1123F	KHOJKI VOWEL SIGN VOCALIC R
◌ः	U+11240	KHOJKI SIGN VISARGA

2 Description

These characters are used in printed texts to represent the words of Sanskrit origin.

VOWEL SIGN VOCALIC R: This character is used in the words that contain the vowel ṛ.

The proposal for Khojki (page 14 of L2/11-021) discusses the use of VOWEL SIGN VOCALIC R. But it was not proposed due to lack of primary sources. This document provides the attestations for this character.

7. **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 copaḍī*, introduces the character ृ , the dependent form of ऋ U+090B DEVANAGARI LETTER VOCALIC R, for transcribing words of Sanskrit origin. The book also specifies diacritic marks for representing conjunct with RA.

The character is derived from either U+0AC3 ◌ૃ GUJARATI VOWEL SIGN VOCALIC R or U+0943 ◌ृ DEVANAGARI VOWEL SIGN VOCALIC R.

VISARGA: It indicates the voiceless glottal fricative. Visarga ◌ः is confusable with U+1123A ꞉ KHOJKI WORD SEPARATOR. As Visarga is a combining mark; its character properties would differ from that of Separator. As in others Indic scripts independent encoding of Visarga in Khojki is recommended.

3 Character Data

3.1 Character Properties

In the format of `UnicodeData.txt`:

`1123F;KHOJKI VOWEL SIGN VOCALIC R;Mn;0;NSM;;;;;N;;;;;`

`11240;KHOJKI SIGN VISARGA;Mc;0;L;;;;;N;;;;;`

3.2 Syllabic Categories

In the format of `IndicSyllabicCategory.txt`:

`# Indic_Syllabic_Category=Vowel_Dependent`

`1123F; Vowel_Dependent # Mn KHOJKI VOWEL SIGN VOCALIC R`

`# Indic_Syllabic_Category=Visarga`

`11240; Visarga # Mc KHOJKI SIGN VISARGA`

3.3 Positional Categories

In the format of `IndicPositionalCategory.txt`:

`# Indic_Positional_Category=Bottom`

`1123F; Bottom # Mn KHOJKI VOWEL SIGN VOCALIC R`

`# Indic_Positional_Category=Right`

`11240; Right # Mc KHOJKI SIGN VISARGA`

Other properties are similar to the respective characters of Devanagari.

4 References

Mukhi Laljibhai Devraj. *Sindhī pahelī copaḍī.* Mumbai. Samvat 1966 [=1910 CE].

Pandey, Anshuman. *Final Proposal to Encode the Khojki Script in ISO/IEC 10646*. N3978 L2/11-021. January 28, 2011.

5 Attestations

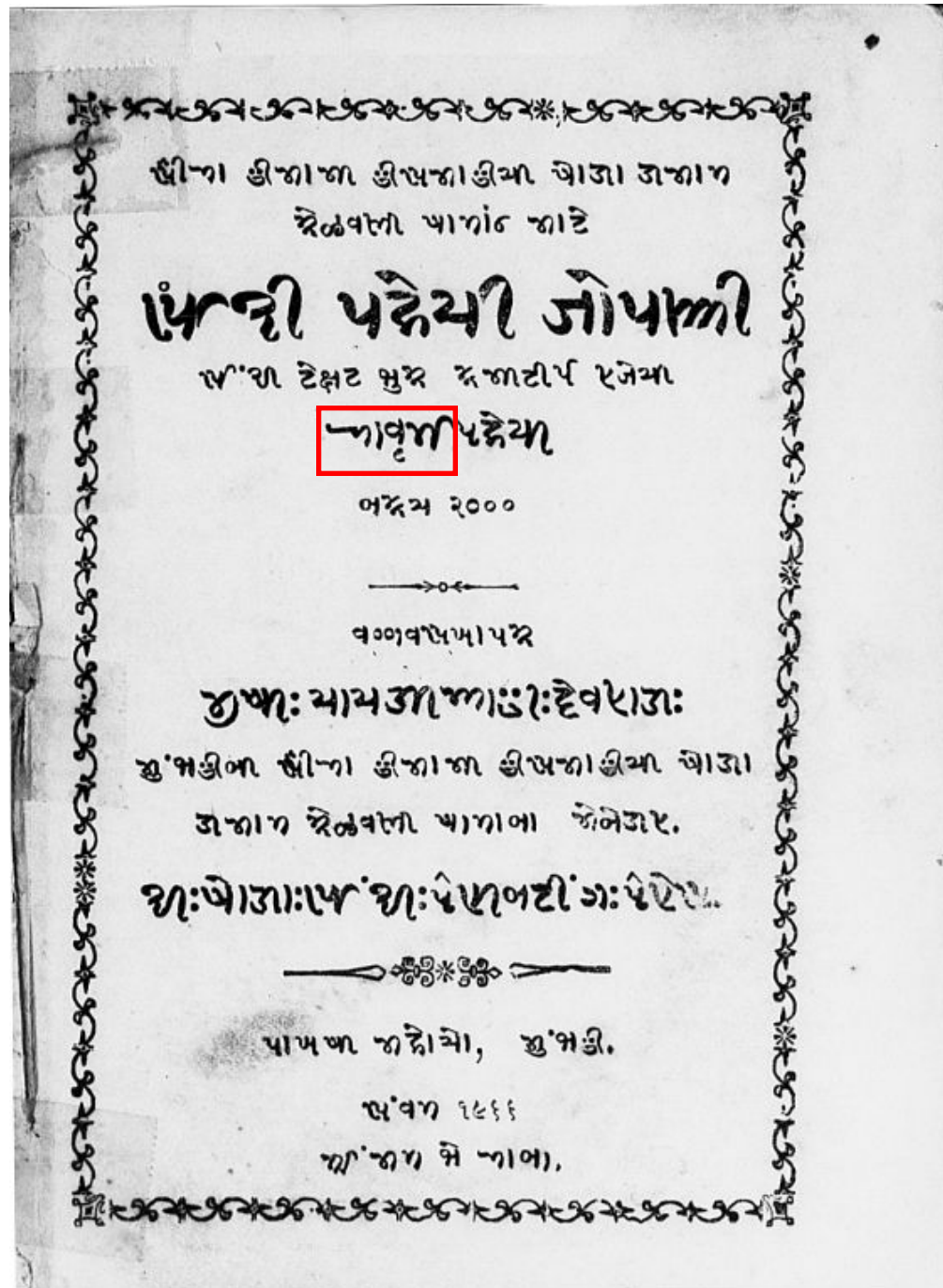All attestations provided here are from the book Sindhī pahelī copaḍī written by Mukhi Laljibhai Devraj.

Figure 1. Use of VOWEL SIGN VOCALIC R in the word *āvṛti.*

Figure 2. Use of VOWEL SIGN VOCALIC R in the words *kṛpāḷu* and *tṛṣṇā*

Figure 3. Use of VOWEL SIGN VOCALIC R in the word *sṛṣṭi*.

Figure 4. Use of VOWEL SIGN VOCALIC R in the word *amṛta*.

ખુદાવંદ શૌઆ નબાબે બાજ રાજ પચંદ નાવનાં બખા.
બાજ રાજ રૈબારૈબ ગો ગેર દાઝરનાં બાંધે છે.
દાઝરનાં ઘણુંઝ દુઃખ છે, બાજ ઓભાગભાઝ પાપનાં
રાજ ખાળબછે. ગૈ અમૈદ ચુગછ વેજાગું આવા પણ
કૈા ગે ચુગછ રાબ્બી રેંગબ ઓભાગનાં રાખધૈ, ગો
ગે ગબાર અમૈદ ચુગછ રાબું ખછી ઝધૈ. દૂર ગૈવા
નબગ ચુળવાછી બાઝનાં કૈા એૈર ગૌબવાર્પ, ગો દૂર
બાં પણ એૈરબ ગબર ખછી ઝાળબછે. એૈરા ઓભાગ
વાબું દૂર પાવાબા ઝાળય બબા ઝાળબછે. દૂબગબા
બાં બખા બાબરગબાં દુઃખબાંખ બજવાબ બબે ચુખ
ગૌબવવાબ બારાનાં બાબા બારબગ પ છે દે દેગૌધાં
ખરાબ ઓભાગબા દૂર રદેવું. બાપબા પાર ભૌખબા
ભૈબા દૌબબાં અગચુખબાં બાબબાં રરબાવેવું છે દેઃ—

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

A. Administrative

1.**Title:**	*Proposal to encode two characters in Khojki*
2. Requester's name:	*Srinidhi A and Sridatta A*
3. Requester type (Member body/Liaison/Individual contribution):	*Individual contribution*
4. Submission date:	*August 30, 2017*
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: *2*

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

A-Contemporary		B.1-Specialized (small collection)	**X**	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"   *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?

        Not necessary. Glyphs can be borrowed from Gujarati.

    b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):

        Not necessary

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 issue

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 the text of proposal

8. 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 UAX#44: 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.

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, 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? **No**

    Reference:

4. The context of use for the proposed characters type of use; common or rare) *common*

    Reference: *See the text of proposal*

5. Are the proposed characters in current use by the user community?

    If YES, where? Reference: *Details are not available*

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:

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**

    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 YES, is a rationale for its inclusion provided?

    If Yes, 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? *Yes*

    If YES, is a rationale for its inclusion provided? *Yes*

    If Yes, reference: *See the text of proposal.*

11. Does the proposal include use of combining characters and/or use of composite sequences? *Yes*

    If YES, is a rationale for such use provided? *Yes*

If Yes, reference:	*Visarga is a combining mark.*
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	*N/A*
If Yes, reference:	

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: