Proposal to encode archaic vowel signs O OO for Kannada

Shriramana Sharma, jamadagni-at-gmail-dot-com, India

2014-Feb-02

§1. Thanks

I thank Srinidhi of Tumkur, Karnataka, for alerting me to these old vowel signs used in Kannada and for providing the attestations, thus leading to this proposal.

§2. Introduction

The Kannada encoding currently has the following vowel signs for short ŏ and long ō:

0CCA	ಕೊ	KANNADA VOWEL SIGN O
0CCB	ಕೋ	KANNADA VOWEL SIGN OO

This is a proposal to encode two more vowel signs denoting the same short ŏ and long ō but consistently different in shape which are contrasted in old Kannada grammars:

0CBA	ಕಾ	KANNADA VOWEL SIGN ARCHAIC O
0CBB	ಕಾೕ	KANNADA VOWEL SIGN ARCHAIC OO

As shown below, these are found in archaic use and must be encoded for the textual representation of those earlier writings.

§3. Discussion

In L2/14-005* I have discussed the two systems followed in the Telugu and Kannada scripts as regards the vowel signs for O and OO. While the Kannada script currently employs a two-part vowel sign ◌ೆ + ◌ಾ = ◌ೊ for indicating O, earlier Kannada writings and printings have used a single-part vowel sign for O. In both cases, the length mark ◌ೕ is added for OO.

To elucidate this: in the old style, for KA ಕ, kŏ was written as ಕಾ, i.e. using a single-part vowel sign ◌ಾ. kō with the long vowel was written as ಕಾೕ.

In both Telugu and Kannada, the two systems are contrasted in the same text.

Attestations for the usage of the single-part vowel signs in Kannada follow.

---

* ... which I submit as a separate document since its scope also includes Telugu and since it is of the nature of a technical note discussing the two systems, various alternate notations etc in detail.

§4. Attestations

ಎ	ಏ	ಐ	ಒ	ಓ	ಔ	ಅಂ	ಅಃ	
ಕೆ	ಕೇ	ಕೈ	ಕಾ	ಕಾಃ	ಕೌ	ಕಂ	ಕಃ	ಕ
ಖೆ	ಖೇ	ಖೈ	ಖಾ	ಖಾಃ	ಖೌ	ಖಂ	ಖಃ	ಖ
ಗೆ	ಗೇ	ಗೈ	ಗಾ	ಗಾಃ	ಗೌ	ಗಂ	ಗಃ	ಗ
ಘೆ	ಘೇ	ಘೈ	ಘಾ	ಘಾಃ	ಘೌ	ಘಂ	ಘಃ	ಘ
ಜೆ	ಜೇ	ಜೈ	ಜಾ	ಜಾಃ	ಜೌ	ಜಂ	ಜಃ	ಙ
ಚೆ	ಚೇ	ಚೈ	ಚಾ	ಚಾಃ	ಚೌ	ಚಂ	ಚಃ	ಚ
ಛೆ	ಛೇ	ಛೈ	ಛಾ	ಛಾಃ	ಛೌ	ಛಂ	ಛಃ	ಛ
ಜೆ	ಜೇ	ಜೈ	ಜಾ	ಜಾಃ	ಜೌ	ಜಂ	ಜಃ	ಜ
ಝು	ಝೋ	ಝು	ಝಾ	ಝಾಃ	ಝೌ	ಝುಂ	ಝುಃ	ಝ
ಞೆ	ಞೇ	ಞೈ	ಞಾ	ಞಾಃ	ಞೌ	ಞಂ	ಞಃ	ಞ
ಟೆ	ಟೇ	ಟೈ	ಟಾ	ಟಾಃ	ಟೌ	ಟಂ	ಟಃ	ಟ
ಠೆ	ಠೇ	ಠೈ	ಠಾ	ಠಾಃ	ಠೌ	ಠಂ	ಠಃ	ಠ
ಡೆ	ಡೇ	ಡೈ	ಡಾ	ಡಾಃ	ಡೌ	ಡಂ	ಡಃ	ಡ
ಢೆ	ಢೇ	ಢೈ	ಢಾ	ಢಾಃ	ಢೌ	ಢಂ	ಢಃ	ಢ
ಣೆ	ಣೇ	ಣೈ	ಣಾ	ಣಾಃ	ಣೌ	ಣಂ	ಣಃ	ಣ

ಎ	ಏ	ಐ	ಒ	ಓ	ಔ	ಅಂ	ಅಃ	
ತೆ	ತೇ	ತೈ	ತಾ	ತಾೇ	ತೌ	ತಂ	ತಃ	
ಥೆ	ಥೇ	ಥೈ	ಥಾ	ಥಾೇ	ಥೌ	ಥಂ	ಥಃ	
ದೆ	ದೇ	ದೈ	ದಾ	ದಾೇ	ಧೌ	ದಂ	ದಃ	
ಧೆ	ಧೇ	ಧೈ	ಧಾ	ಧಾೇ	ಧೌ	ಧಂ	ಧಃ	
ನೆ	ನೇ	ನೈ	ನಾ	ನಾೇ	ನೌ	ನಂ	ನಃ	
ಪೆ	ಪೇ	ಪೈ	ಪಾ	ಪಾೇ	ಪೌ	ಪಂ	ಪಃ	
ಫೆ	ಫೇ	ಫೈ	ಫಾ	ಫಾೇ	ಫೌ	ಫಂ	ಫಃ	
ಬೆ	ಬೇ	ಬೈ	ಬಾ	ಬಾೇ	ಬೌ	ಬಂ	ಬಃ	
ಭೆ	ಭೇ	ಭೈ	ಭಾ	ಭಾೇ	ಭೌ	ಭಂ	ಭಃ	
ಮೆ	ಮೇ	ಮೈ	ಮೊ	ಮೋ	ಮೌ	ಮಂ	ಮಃ	
ಯೆ	ಯೇ	ಯೈ	ಯಾ	ಯಾೇ	ಯೌ	ಯಂ	ಯಃ	
ರೆ	ರೇ	ರೈ	ರಾ	ರಾೇ	ರೌ	ರಂ	ರಃ	
ಲೆ	ಲೇ	ಲೈ	ಲಾ	ಲಾೇ	ಲೌ	ಲಂ	ಲಃ	
ವೆ	ವೇ	ವೈ	ವಾ	ವಾೇ	ವೌ	ವಂ	ವಃ	

3

ఎ	ఏ	ఐ	ఒ	ఓ	ఔ	అం	అః
ಶೆ	ಶೇ	ಶೈ	ಶಾ	ಶಾೇ	ಶೌ	ಶಂ	ಶಃ
ಷೆ	ಷೇ	ಷೈ	ಷಾ	ಷಾೇ	ಷೌ	ಷಂ	ಷಃ
ಸೆ	ಸೇ	ಸೈ	ಸಾ	ಸಾೇ	ಸೌ	ಸಂ	ಸಃ
ಹೆ	ಹೇ	ಹೈ	ಹಾ	ಹಾೇ	ಹೌ	ಹಂ	ಹಃ
ಳೆ	ಳೇ	ಳೈ	ಳಾ	ಳಾೇ	ಳೌ	ಳಂ	ಳಃ
ಕ್ಷೆ	ಕ್ಷೇ	ಕ್ಷೈ	ಕ್ಷಾ	ಕ್ಷಾೇ	ಕ್ಷೌ	ಕ್ಷಂ	ಕ್ಷಃ

p 23 (೨೩):

ಅವನು ನನಗೆ ತಮ್ಮನು
ಕಾಣಿಯನ್ನು ನಾೇಡಲಕ್ಕಿ ಹಾೞಗಿದ್ದೆನು
ಈ ಕುದುರೆಯನ್ನು ನಾನು ನೂರು ರುಪಾಯಿಗೆ ಕಾಡುತ್ತೇನೆ
ಆ ಮಾತು ಆಗಿ ಇಂದಿಗೆ ನಾಲ್ಕು ದಿನ
ನಿನಗೆ ಹುಟ್ಟು ಹಿಡಿದದೆ

p 4:

೪. ಗೋವೃಕ್ಷ

 ವೃಕ್ಷವೆಂಬ ಹೆಸರಿನ ಗಿಡವು ಅಮೆರಿಕ ಖಂಡದ್ದಾಳ ಗೆ ದ್ದಾರಿಯುತ್ತದೆ. ಅಚರಾಳಗಿಂದ ಆಕಳ ಹಾಲಿನಂಥ

The samples on this page illustrate contrastive usage of the two forms of O/OO in the same text. The proposed forms are marked in blue and the existing forms are marked in red.

Canarese School Grammar

G Plebst, German Mission Press, Mangalore, 1859

p 6:

A Grammar of the Kannada language in English,

Ferdinand Kittel, Basel Mission Press, Mangalore, 1903

p 12 :

p 20:

At school the mark ꞌ used for the vowel ఒ and the mark ◌ used for the vowel ఎ are called ಗುಡಸು, ಗುಡಿಸು or ಗುಡುಸು, the mark ು used for ಉ is called ಕೊಮ್ಮು, the mark ಾ that descends from a consonant as the sign for the long vowels ಆ and ಈ (as in ಕಾ, etc., and in ವಾ and ಬಾ) is called ಇತ್ತ, and the name for the mark ೕ is ದೀರ್ಘ. ಕೊ ಕೋ ಗೊ ಗೋ etc. may be written also ಕೌ ಕೋ ಗೌ ಗೋ etc.

A grammar of the Carnataca language

John MacKerrel, College Press, Madras, 1820

http://books.google.co.in/books?id=Rr4JXT430CUC

pp 6-13:

etc ...

Ancient and Modern Alphabets of ... the Southern Peninsula of India

Henry Harkness, Royal Asiatic Society, 1837

http://www.archive.org/details/ancientmodernalp00harkrich

etc...

§5. Argument for disunification

In the script review doc L2/14-053 p 2 it has been suggested that encoding these characters might constitute a "duplicate encoding since they are semantically identical to the already encoded O's (no examples of contrast were provided)".

First it is not clear what "semantically identical" means. While there is linguistic identity in that both forms represent the same sounds, orthographically these are two different forms resulting from a divergent evolution between two different orthographic practices in the transitional period of Kannada and Telugu from the old Deccan (aka proto-Telugu-Kannada) script. Hence even in Telugu one sees this contrast as detailed in L2/14-005. As for Kannada the contrast is seen in p 5 and hence the comment about "no examples of contrast" may be since the earlier version of this doc had not highlighted the contrast.

Orthographic distinction was the basis for distinctive encoding in the case of 09B0 ব vs 09F0 ৰ denoting *ra* in different orthographies of Bengali and likewise for 0D08 ഈ vs 0D5F* ൟ (see L2/12-225 N4312) denoting *ī* in Malayalam. The same should be done here.

The review document also notes: "it is not clear that the distinction in the glyphs could not be done in an OpenType font". Smart font technologies like Graphite and OpenType can be used to accomplish very many modifications to the glyphs – even interchange As and Bs – but that does not mean it is advisable to do so. In this particular case, apart from the primary reason of orthographic distinctiveness discussed above, the canonical decomposition provided for the existing vowels applies only since they are two-part, while the single-part signs contrasted in the texts would not so decompose.

Hence distinctive encoding is warranted.

§6. Encoding model

While there is no problem in encoding just the single-part vowel sign for O ◌ా and adding 0CD5 ◌ೕ Kannada Length Mark for the long vowel, I propose to encode characters for both O and OO to retain symmetry with the existing Kannada vowel signs for O and OO. There is no constraint for space in the Kannada block.

Thus to achieve the old orthography, one only needs to add the appropriate vowel sign to the consonant or consonant cluster:

<p align="center">ಕ + ◌ా → ಕా   ಕ + ◌ాೕ → ಕాೕ</p>

Canonical Decomposition

Note that the proposed vowel sign for long OO would obviously require a canonical decomposition to the proposed one for short O followed by the length mark:

0CBB ಼ಾಀೕ KANNADA VOWEL SIGN ARCHAIC OO

≡ 0CBA ಼ಾಀ 0CD5 ಼ೕ

... which is parallel to the existing vowel sign for long OO:

0CCB ಼ೋೕ KANNADA VOWEL SIGN OO

≡ 0CCA ಼ೊ 0CD5 ಼ೕ

However, unlike the existing vowel sign for short O i.e. 0CCA ಼ೊ (which decomposes to 0CC6 ಼ೆ 0CC2 ಼ೂ), the newly proposed 0CBA ಼ಾಀ can not have any canonical decomposition as it is unitary in shape.

Choice of codepoints

The codepoints are chosen as 0CBA and 0CBB so as to fill up a "hole" in the Kannada block between 0CB9 KANNADA LETTER HA and 0CBC KANNADA SIGN NUKTA. This will also exactly position them in the charts beside the existing vowel signs for O and OO at 0CCA and 0CCB.

§7. Character Properties

The proposed glyphs are:

಼ಾಀ   ಼ಾಀೕ

0CBA   0CBB

The character properties are all identical to those of the existing vowel signs for O and OO except that the long vowel gets the correct canonical decomposition:

```
0CBA;KANNADA VOWEL SIGN ARCHAIC O;Mc;0;L;;;;;N;;;;;
0CBB;KANNADA VOWEL SIGN ARCHAIC OO;Mc;0;L;0CBA 0CD5;;;;N;;;;;
```

NamesList.txt should take a section header before these two characters as follows:

```
@    Archaic vowel signs
```

As for collation, normally a given text would use either the existing new-style or the newly proposed old-style characters. Even if both styles are encountered, they should be collated equivalent to each other.

§7. Official Proposal Summary Form

(Based on N3902-F)

A. Administrative

1. Title
Proposal to encode archaic vowel signs O OO for Kannada
2. Requester's name
Shriramana Sharma
3. Requester type (Member body/Liaison/Individual contribution)
Individual contribution
4. Submission date
2014-Feb-02
5. Requester's reference (if applicable)
6. Choose one of the following: This is a complete proposal (or) More information will be provided later
This is a complete proposal.

B. Technical – General

1. Choose one of the following:
1a. This proposal is for a new script (set of characters), Proposed name of script
No.
1b. The proposal is for addition of character(s) to an existing block, Name of the existing block
Yes. Kannada.
2. Number of characters in proposal
2 (two)
3. Proposed category
Category B1, specialized small (for these characters, though Kannada itself is "A, contemporary")
4. Is a repertoire including character names provided?
Yes.
4a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?
Yes.
4b. 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?
Shriramana Sharma.
b. Identify the party granting a license for use of the font by the editors (include address, e-mail etc.)
Shriramana Sharma and other contributors to the Lohit font project under the Open Font Licence.
6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes.
6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
Yes.
7. 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)?
Collation is discussed in the 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.
See detailed proposal.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.
This is a revised version of a document submitted earlier this month revised in response to feedback.
2a. 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.
2b. If YES, with whom?
Narasimha Murthy, Kannada Ganaka Parishad, Bangalore. Srinidhi, Tumkur. Narasimha Bhat, Shirsi.

2c. If YES, available relevant documents
None specifically. The matter was discussed in person or via email.
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?
Those who would like to use the old-style orthography for the Kannada vowel signs O/OO.
4a. The context of use for the proposed characters (type of use; common or rare)
Rare. For recording heritage texts.
4b. Reference
See detailed proposal.
5a. Are the proposed characters in current use by the user community?
No.
5b. If YES, where?
6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?
Yes.
6b. If YES, is a rationale provided?
Yes.
6c. If YES, reference
They belong in the Kannada block which is in the BMP.
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
Yes.
8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
No.
8b. If YES, is a rationale for its inclusion provided?
8c. If YES, reference
9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?
Yes. The proposed long vowel sign OO is equivalent to the proposed short vowel sign O followed by length mark. No existing characters can be used to compose these however.
9b. If YES, is a rationale for its inclusion provided?
Yes. To retain symmetry with the existing encoded characters.
9c. If YES, reference
See detailed proposal.
10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?
No.
10b. If YES, is a rationale for its inclusion provided?
10c. If YES, reference
11a. Does the proposal include use of combining characters and/or use of composite sequences?
The characters by nature are Indic combining marks.
11b. If YES, is a rationale for such use provided?
11c. If YES, reference
11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
12a. Does the proposal contain characters with any special properties such as control function or similar semantics?
No.
12b. If YES, describe in detail (include attachment if necessary)
13a. Does the proposal contain any Ideographic compatibility character(s)?
No.
13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?
13c. If YES, reference:

–o–o–o–