L2/04-364 2004-10-22

Universal Multiple-Octet Coded Character Set International Organization for Standardization Organisation Internationale de Normalisation Международная организация по стандартизации

**Doc Type: Working Group Document** 

Title: Proposal to add six characters for Kannada to the BMP of the UCS

**Source:** Michael Everson

**Status: Individual Contribution** 

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2004-10-22

This document requests six additional characters to be added to the UCS and contains the proposal summary form. Four of these signs (the first and last pair) are rarely used in Kannada and provide support for Sanskrit. The JIHVAMULIYA and UPADHMANIYA are script-specific characters in Kannada in the same way that ANUSVARA and VISARGA are; the four characters belong to a group called *ayogavāha*. The punctuation marks are also chiefly used in the context of Sanskrit texts and Kannada poetry. These have a thin, slender appearance quite unlike the thick-stemmed Devanagari; opinion is strong on the INDIC@UNICODE.ORG list that a disunification of these from Devanagari is required. (It is likely that this is true for all of the Indic scripts, but it makes sense to meet Kannada requrements in this single proposal here.)

0CE2	.∵. .∵.	KANNADA VOWEL SIGN VOCALIC L
0CE3	.∵ .∵	KANNADA VOWEL SIGN VOCALIC LL
0CE4	Ţ	KANNADA DANDA
0CE5	11	KANNADA DOUBLE DANDA
0CF1	X	KANNADA SIGN JIHVAMULIYA
0CF2	00	KANNADA SIGN UPADHMANIYA

# **Unicode Character Properties**

```
0CE2;KANNADA VOWEL SIGN VOCALIC L;Mn;0;NSM;;;;N;;;;
0CE3;KANNADA VOWEL SIGN VOCALIC LL;Mn;0;NSM;;;;N;;;;
0CE4;KANNADA DANDA;Po;0;L;;;;N;;;;
0CE5;KANNADA DOUBLE DANDA;Po;0;L;;;;N;;;;
0CF1;KANNADA SIGN JIHVAMULIYA;Lo;0;L;;;;N;;;;
0CF2;KANNADA SIGN UPADHMANIYA;Lo;0;L;;;;N;;;;
```

# Bibliography

Imprimerie Nationale. 1990. Les caractères de l'Imprimerie Nationale. Paris: Imprimerie Nationale Éditions. ISBN 2-11-081085-8

Kittel, Ferdinand. 1993 (1894). *A Kannada-English dictionary = Kannaḍa-imgliṣ ḍikṣnari*. New Delhi & Madras: Asian Educational Services. ISBN 1-07019915-1

Kittel, Ferdinand. 1993 (1903). A grammar of the Kannada language: comprising the three dialects of the language (ancient, medieval and modern). New Delhi & Madras: Asian Educational Services. ISBN 81-206-0056-8

Kôno Rokurô, Chino Eiichi, & Nishida Tatsuo. 2001. The Sanseido Encyclopaedia of Linguistics. Volume 7: Scripts and Writing Systems of the World [= Gengogaku dai ziten (bekkan) sekai mozi ziten]. Tokyo: Sanseido Press. ISBN 4-385-15177-6

Reichsdruckerei. 1951. Alphabete und Schriftzeichen des Morgen- und Abendlandes, zum allgemeinen Gebrauch mit besonderer Berücksichtigung des Buchgewerbes. Unter Mitwerkung von Fachgelehrten zusammengestellt in der Reichsdruckerei. Berlin: Reichsdruckerei.

# **Figures**

#### KANARESISCH ದ daZ Allgemeines. Die kanaresische Sprache ဗ a<u>h</u> gehört zu den Drawidasprachen; sie wird rhಧ dhaಆ ā 00 in der vorderindischen Küstenlandschaft Kanāra gesprochen. Die Schrift läuft von ₹ Кa ನ nalinks nach rechts und hat viel Ligaturen. 8 i ಖ khaಸ paVokale. Außer den im Alphabet aufge-4 ī führten Vokalen gibt es noch Vokalzeichen, 7 phaga -ಭ die über, unter oder neben die Konsonanten സ geschrieben werden, z. B. beim Konsonant $\cap$ (g) in folgender Weise: $\boldsymbol{u}$ ghaಘ ಬ baಊ $\bar{u}$ ಜ $\dot{n}a$ ಭ bha ヹ qa ಯ ? ध caಮ ma ₹₹® gā gai ಯೂ $\bar{r}$ ಚ cha ክ gi ಯ ya l ഷ ಜ ja he gi ರ ra $g ar{l}$ **が** gu Į ч လူ jhaභ ra ™ gū e X ña ಲ la Konsonanten. Konsonantenverdopplungen ē ಟ బ taವ vaund -verbindungen werden gebildet, indem unter die Konsonanten verkürzte oder um-ಠ tha හ ai ಕ śα gebildete Konsonanten gesetzt werden, z. B.: ಡ daಒ $\vec{Q}_{l} = rka$ usw. 0 ಷ şа દ્વ ō ಢ dha ズ sa 0 ಔ au ಚಿ naಹ ha 7 8 9 0 2 3 4 5 1 ತ ಳ ṁ taLiteratur. la Q Grammar of the Kannada Language in English by Dr. Kittel. Mangalore 1903. laķ ह्य tha ಅ

**Figure 1.** From the Reichsdruckerei 1951. Illustration of both LETTERS and VOWEL SIGNS VOCALIC L and VOCALIC LL as well as JIHVAMULIYA and UPADHMANIYA.

g klri	हूं klrî	<b>⊀</b> kĕ	ಕೇ kê	ಕ್ಕೆ kai	ಕೊ kŏ	ಕೋ kô	₹® kau
aj khlri	ມູ້ khlṛî	3 khě			ൂറെ khŏ	ಖೋ khô	
ಗ್ಗ glṛi	ಗ್ಲ glṛì	ಗೆ gĕ	ಗೇ gê			ಗೋ gô	ಗೌ gau
ಘ್ನ ghlri	ಘ್ಲ ghlrî	ಘೆ ghě	ಘೇ ghê		ಘೂ ghŏ	ಘೋ ghô	ಘೌ ghau
🙇 ilri	æ nlrî	කුපී ně	නා <sup>3</sup> ල niê		ක්ග nŏ	ಜ್ೋ nô	කුණ nau
👼 člṛi	æ člŗî	ಚೆ Čě	ಚೇ čê	ಚೈ čai	ಚೊ čŏ	ಚೋ čô	ಚೌ čau
g chlri	ಛ್ಞ čhlṛî	ಛೆ čhě	ಛೇ čhê		မို <b>ာ</b> čhŏ	ಛೋ čhô	ಛೌ čhau
क्षु jiri	ಜ್ಞ jlṛî	ಜೆ jě	ಜೇ jê		ಜೊ jŏ	ಜೋ jô	ಜೌ jau
the jblri	ಝ್ಸ್ jhlrî	ఝి jhĕ			ಝೊ jhŏ	ထုံး၈e jhô	jhau
a niri		කුපී ńĕ			ක්තී úč	are no	

Figure 2. From Kittell 1993 (1903). Illustration of VOWEL SIGNS VOCALIC L and VOCALIC LL.

సత్యద ధర్ములో నిత్యవు బోగాధినే తోగాత్తిన మనసింగే నేగాగస్టుతోం తత్వద ఆథులో విజిత్రది నేంటలు రత్తియ మనసింగే తిళిదిందిం! వృత్త్రిబోగామ్టేయ జిత్రది బరిదిరే ముత్తు రిగాటరి మాతాడిందిం రిత్తురితిలరవన్ ఒత్తి ఫట్రియింగ్ ఇడే ఆత్రిలాయ తోగాంటి ఇదిందిం!! corps 17

Kittel, A grammar of the Kannada language

**Figure 3.** From the Imprimerie Nationale 1990. Illustration of DANDA and DOUBLE DANDA in a poetic context. The stroke is thin and narrow (and slightly tapering in the original document) and quite different from the Devanagari dandas, which typically are based on the thick vertical common to many Devanagari letters.

ಕೊಳ್ಳುಳ್ಳೆನ್ಡೊಡನೊಡನೆಯೆ ತುಳ್ಳುವ ಯುವತಿಯರ ಕರಪರಿಚ್ಯುತಸಲಿಂ | ಗಳ್ಳೊಳೆ ಬಿಡಿಕೆದಿಯವ ಕಿಂ ಜಳ್ಳದಿನ್ ಆದುದು ಸರೋವರಂ ಕಪಿಶಜಳಂ || When the pond took the waters that fell from the hands of the young women who said "take! take!" and all at once sprinkled, it became reddish-brown water on account of the filaments (of lotus flowers) that became loose and were scattered about. ತಮ್ಮಾ, ಇಲ್ಲಿಗೆ ಬಾ, brother, come here! ತಂಗೀ. ಊಟಕ್ಕೆ ನಡೆ, sister, go to dine! ಮಗಳೇ, ಸುಖದಿನ್ದ ಇರು, daughter, be joyful! ದೇವಾ, ನನ್ನನ್ನು ಉದ್ದರಿಸು, God, save me!

**Figure 4.** From Kittell 1993 (1903). Another illustration of DANDA and DOUBLE DANDA in a poetic context.

③ (hと転写する) は気音を表わし、ヴィサルガ (visarga) と呼ばれる。この記号は、サンスクリット語からの借用語の転写のみに用いられる (乙)(2) duḥka [苦」).
○は ka、kha の前の気音、%は pa、pha の前の気音を示すが、今日では一般には用いない。この2つの記号は、クシャーナ朝以降のグプタ (Gupta) 文字のそれ ((∑ hka、oo hpaa) に近似した形を保っている。カンナダ数字は、インド数字の系統をひいている (表4)

Figure 5. From Kôno 2001. Illustration of JIHVAMULIYA and UPADHMANIYA along with ANUSVARA and VISARGA

The other nine letters are: —  $\omega$  & & & & o. Of these  $\omega$  & & & o. The first of the Kannada country (desiya), and this letter  $\varphi$  is the kula.

The four letters o & X oo, together with the above-mentioned forty-eight, form the Samskrita alphabet of 52 letters.

Figure 6. From Kittell 1993 (1903). JIHVAMULIYA and UPADHMANIYA along with ANUSVARA and VISARGA.

**32**8

 $\bigcirc$ 

0

 M

00

and labial pronunciation (Smd. 24), are often, for the sake of more expeditious writing, replaced by the sign of the unmodified sonne. With regard to me and contains mode of writing has been adopted also for this Dictionary; but the alphabetical places of the two letters have been retained, so that, e.g., wow follows with and would follow with the containing means of the containing of th

o m. 2. = 90 (am) 8 q. v., of which it is the grammatical form (smd. 114. 115).

0

 The eighteenth letter of the Alphabet, (called visarga, Smd. 19. 20. 44. 45). It is an unmodified H

(avargiya) sibilant, and one of the four yôgavâhas

X

X h. The nineteenth letter of the Alphabet. It is a guttural sibilant, used before v<sup>4</sup> (Smd. 25. 26), and one of the four yogavahas (26. 44. 45). It is called

X

H

jihvāmūliya (26), the tongue-root sibilant, and according to its sign which has the shape of the thunderbolt, vairākriti. See 영주다리자다.

00

OO Th. or Co (i. e. the K. consonant Co, in Hià. and Abh. P. M88.). or Co (in Vôpadêva's grammar). The twentieth letter of the Alphabet. It is a kind of sibilant, used before w (Smd. 25. 26), and one of the four yôgaváhas (26. 44. 45). It is called upadhmántys (26), the to-be-breathed one, and according to its sign which has the shape (of a female's breast 26, or that)

RH

of an elephant's two frontal bones, gajakumbhåkriti. See ತೇಜದ್ಪಣ್ಣ, ನಿದ್ದಾವಕ, ಭದ್ಪ, ಭದ್ಪಕ, ಭದ್ಪಕಾಪ, ಭದ್ಪವಕ, ಭದ್ಯವಕ್ಷಕ್ಕೆ ಕದ್ದದೇ ಪಾದ್ಯೋ ಮೂಲಕ, etc. (instead of ಕನ್ನಡ ಪಾದೋ). See ಅರ್ಥವಿಸರ್ಗ.

**Figure 7.** From Kittel 1993 (1894). Illustration of JIHVAMULIYA and UPADHMANIYA as the 19th and 20th Kannada letters in their alphabetical order following ANUSVARA and VISARGA.

ಱ್

(Smd. 26-28). It is the final of many words (58. 59. 78. 79. 275. 284. 286), and the initial of some; it changes into 3' (215. 282. 285), into 3' (206), into 15' (218), and into coto' (228); replaces the Samskrita of e. g. in cap (in Padmarája's Sánandačaritra there is also τος στοκοφού, στος, πτος; and since about one or two hundred years is represented by of in the common Kannada language. (See also ατο 2, ατό). It is occasionally elided (Smd. 222. 246. 247). About its representation of the upadhmāniya see OO. 2, it forms an auxiliary in α2.

⇔ ra. 1. The letter ⇔ pronounced with a short a.

R

ದಪ್ಪು rappe. = ಜೆಪ್ಪ. (My.).

ಜಾಮ್ನ rambhě. = ರಮ್ಮ. (Šmd. 160 Mdb. Cm.).

Cರಿವರ್ ravake. = ಜಎರ್. A bodice (worn by females, V. 5, after 14; My.; Tu.; Tš. ಜಯಕ, ಜಎಕ); a quilted jacket, armour, mail (ವಾರವಾಣ, ಕಂಡುಕಿ Nn. 97). See ಪೂ., ಎಸುನಿ.

ದಿವಿಕೆ ravike. = ಅವರೆ. (Bp. 14, 15; My.).

CCC ra-la. A compound letter used to represent the letter α or a term with the letter α (8md. 26. 27. 28. 29. 31. 32. 51. 78).

ಆರಂತನ್ನ raia-anta. A term with final ್. (śmd. 52.54).

**Figure 8.** From Kittel 1993 (1894). Showing LETTER RRA, the 40th of the Kannada letters, related but distinct from the UPADHMANIYA.

## A. Administrative

#### 1. Title

Proposal to add six characters for Kannada to the BMP of the UCS.

#### 2. Requester's name

Michael Everson

#### 3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

#### 4. Submission date

2004-10-22

5. Requester's reference (if applicable)

### 6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

### **B.** Technical – General

### 1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

No.

Proposed name of script

1b. The proposal is for addition of character(s) to an existing block

Yes.

1b. Name of the existing block

Kannada

2. Number of characters in proposal

6

3. Proposed category (see section II, Character Categories)

Category A

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 2.

4b. Is a rationale provided for the choice?

Yes.

#### 4c. If YES, reference

Brahmic combining characters.

5a. Is a repertoire including character names provided?

Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson. TrueType.

**6b.** If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: Michael Everson. Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes, see bibliography below.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

8. 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 below.

9. 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/Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

Yes, see Unicode properties below.

# C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

Nο

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.)?

Vec

2b. If YES, with whom?

Discussion on the indic@unicode.org list has shown support for the encoding of the characters.

2c. 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. Schlars and poets writing Sanskrit in Kannada script.

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

Rarely used Kannada characters.

4b. Reference

See examples above.

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

Yes.

5b. If YES, where?

In Sanskrit and older Kannada texts.

6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?

Yes

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

All Kannada points are in the BMP.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

No.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

Yes. All Indic scripts share a similar structure, but they are disunified on grounds of glyph identity. Devanagari has a LETTER KA, VOWEL SIGN VOCALIC L, AVAGRAHA, DANDA, and in Vedic texts uses a JIHVAMULIYA. Likewise, Kannada has all of these. They should not be unified with their Devanagari counterparts.

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?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

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 (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

Yes

11b. If YES, is a rationale for such use provided?

Nο

11c. If YES, reference

 $12a.\ Is\ a\ list\ of\ composite\ sequences\ and\ their\ corresponding\ glyph\ images\ (graphic\ symbols)\ provided?$ 

No

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

13b. If YES, describe in detail (include attachment if necessary)

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

14c. If YES, reference