

Title:	Encoded Representation of Devanagari Candra A
Doc. Type:	Expert contribution
Source:	Peter Constable, Microsoft
Date:	2007-01-24
Action:	For consideration by UTC
References:	email from SS Thattey (Gov't of India) to Unicode Consortium, 2007-1-23
Distribution:	UTC Members

Synopsis

There is a need for an encoded representation of a Devanagari text element, “candra a”. Precedent suggests that this be encoded as a new, atomic character. However, existing implementations are able to display the requisite graphic form using a sequence of existing characters. UTC needs to determine what the encoded representation of candra a shall be.

Details

SS Thattey (sthattey@barc.gov.in) has brought to our attention the need to represent a text element used in Marathi documents. This text element is used in writing loan words from languages such as English to represent the vowel sound /æ/ (e.g. the “a” in “apple”).

Others have mentioned a letter they deemed to be missing from Unicode. For instance, Gautam Sengupta [1] mentions that Marathi uses “a CANDRA mounted on A”, and that Unicode does not represent this missing “DEVANAGARI LETTER CANDRA A”.

The following images illustrate the text element in question:

FIRST PART

MARATHI ALPHABET

मराठी अक्षर माला

Vowels

(स्वर)

अ	A
आ	Ā
इ	I
ई	Ī
उ	U
ऊ	Ū
ऋ	Ṛ
ए	E
ऐ	AI
ओ	O
औ	AU
अं	AM
अः	AH
अँ	Ä
ऑ	Ö

used for
writing
foreign
sounds

Figure 1. CANDRA A in a list of vowel letters of the Marathi alphabet ([2], p. 13)

Signs of Vowels (स्वरचिह्ने)

Vowel	Sign	Usage	Vowel	Sign	Usage
अ A	ए E	˘	केस KES
आ A	।	राम RAM	ऐ AI	ˆ	कैद KAID
इ I	ि	शिव SIV	ओ O	े	लोक LOK
ई I	ी	गीत GIT	औ AU	ी	कौल KAUL
उ U	ु	चुप CHUP	अं AM̐	·	कंस KAMS
ऊ U	ू	दूध DŪDH	अः Aḥ	:	पुनः PUNAH
ऋ R̄	ॠ	नृप NR̄P	अँ A̎	ँ	फँन
			ऑ O̎	ँ	कॉफी

Figure 2. CANDRA A in a table of vowel representations ([2], p. 19)

There is little question that an encoded representation for CANDRA A is needed. The question is what that encoded representation should be: a new atomic character, or a sequence of existing characters.

The practice for encoding of independent vowels that are formed from a basic vowel letter (such as the LETTER A) plus some diacritic sign has been to encode these as separate, atomic characters. This includes at least the following cases:

- U+0904 DEVANAGARI LETTER SHORT A
- U+0907 DEVANAGARI LETTER I
- U+090D DEVANAGARI LETTER CANDRA E
- U+090E DEVANAGARI LETTER SHORT E
- U+0910 DEVANAGARI LETTER AI
- U+0911 DEVANAGARI LETTER CANDRA O
- U+0912 DEVANAGARI LETTER SHORT O
- U+0913 DEVANAGARI LETTER O

- U+0914 DEVANAGARI LETTER AU

All of these are potentially decomposable to sequences of other characters (e.g. <0905, 0946> could be seen as a decomposition of 0904). However, none of these characters has a decomposition mapping. To deal with confusability, TUS 5.0 added text stating that such sequences should not be used (see Table 9-1 and surrounding text on p. 299).

Precedent suggests that CANDRA A can be treated the same way and encoded as a new atomic character. The counterargument is that there are existing implementations that follow the guidance in TUS 5.0, not supporting sequences that would be confusable with existing characters in TUS 5.0, but that *do* support other sequences that are not in the scope of that guidance since the composite text element is not encoded as an atomic character in TUS 5.0. Such implementations could support the sequence <0905, 0945>, which would have the visual appearance of CANDRA A. This is true, for instance, in the case of Devanagari support in Windows Vista:

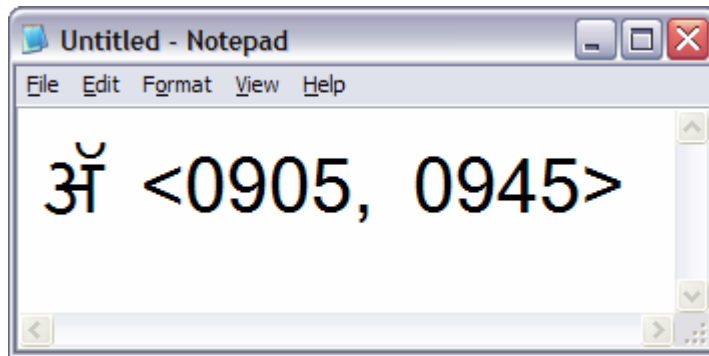


Figure 3. CANDRA A displayed using a character sequence in Windows Vista

The question before UTC, then, is whether to encode a new atomic character DEVANAGARI LETTER CANDRA A and specify that the sequence <0905, 0945> should not be used, or to specify that this sequence is considered to be the encoded representation for Devanagari candra a.

References

- [1] Sengupta, Gautam. 2006. "Multilingualism on the Internet: an Indian Perspective." Presentation at the Joint UNESCO and ITU Global Symposium on Promoting the Multilingual Internet, Geneva, 9–11 May, 2006.
- [2] Sanjay. 2002. *Learn Marathi in 30 days*. (National integration language series.) Chennai: Balaji Publications.