Comments on Proposal to Encode the Saurashtra Script in the UCS
Vide: ISO/IEC JTC1/SC2/EG2 N3607   LN/03-231

by
Peri Bhaskararao
bhaskar@aa.tufs.ac.jp

[Section numbers below refer to the section numbers in the original proposal]
Section: B6b.
There are two fonts ('Kuber' and 'SURESH') available at www.palkar.org As of 2003.08.23, a couple of important glyphs are missing in these fonts. For instance, the glyphs for AB63 (as in the modified chart given at the end of this write-up) should have a subscript flourish. It is not there in either of the fonts. AB62 needs be a full-bodied letter. However, it is available as a subscript in both the fonts. The Xenotype Saurashtra font list (as seen in a PDF file) has both the glyphs. AB0B and AB37 are missing in Kuber font. They are available in Suresh font.

C2c.
This subsection asks 'whether any contact has been made to members of the user community, and whether any relevant documents available'. The answer given was: N/A.
The proposal mentions that a contact was made with Mr. Kubendiran. However, no document is enclosed. The Kuber font which is named after Mr. Kubendiran (and is used in a magazine edited by him) seems to be defecient as mentioned under B6b above.

There is another Saurashtra person who has been recently appointed as the Honorary Director of the Saurashtra Sabha Library and Museum in. He is Mr. O.S.Subramanian. He has been trained in linguistics and contributed an appendix to the Saurashtra-English dictionary of Prof. N.Ucida. I was told that he has his opinions about the so-called 'upaakSara' as well as the need for a kind of 'ja-phala' for Saurashtra script. Mr. Subramanian does not seem to be aware of the current proposal.

When I started asking about the originally proposed 'Aayatam' which was later changed to 'UpaakSara', I received the following replies from one of the proposers:

-------------*Some of the answers from Mr.Jeyakumar, the first Proposer-------------*
As I already enquired some of the sources in Madurai, Madras (Tamil Nadu)
and other local communitiy heads, I was not able to get any significant information about this AAYTAM
character. Please help me to sort out this issue.

The Name of this Sign is "UPAKSHARA"

I got this information from Mr. K.G.Sethuraman. He is a retired professor and has authored several
books related to Saurashtra. He showed this information in his Book "Tamilnattil Saurashtirar
Muzhuvaralaaru". A Tamil book which explains about the history of Saurashtra in TamilNadu published
in 1977. This year he has published a new Trilingual Dictionary (Saurashtra-Tamil-English). In this he
has given several examples using this sound.

It will be straightforward and easier, if 'upakshara' sign is encoded in Saurashtra.

From the above scenario, it seems that a thorough discussion about various issues of the script with some informed members of the Saurashtra community is still very much needed.

C4a. The context of use for the proposed characters : Rare

C5a. Are the proposed characters in current use by the user community + C5b. If YES, where:

Yes + In Tamil Nadu, Andhra Pradesh, Gujarat states of India. A monthly magazine called BhaaSaabhimaani is published in Madurai, India.

----------Comments about the above items----------

Saurashtra script is not currently used in Gujarat state. I was told that the number of persons who can claim proficiency in the script (reading and writing) is just about 3 or 4. The rest of the 'users' use it for religious or cultural purposes. It is used on Wedding Invitation cards, occasionally in a magazine, and on sign-boards. The magazine BhaaSaabhimaani is not published in Saurashtra language. It is printed in Tamil script with occasional use of Saurashtra letters. Figure 5 given in the original proposal itself will vouch for this. In this figure, the whole explanation is in Tamil language and Tamil script. The folio heading of the magazine is mostly in Tamil script and Tamil language. Just the name 'BhaaSaabhimaani' is transliterated into Saurashtra script (that too after giving it in Tamil script first).

D. Proposal: *The following comments on this section are in continuation with the above comments:*

However, the Saurashtra script has two important functions. One is ethnic identity. The script could identify the group as a distinct group. All the members of this community are bilinguals in Saurashtra language and the regional dominant language such as Tamil and Telugu. Usage of the script gives them a distinct identity. Usage of the script on sign-boards also contributes to this feature of identity. The other function is cultural.

The script is used for writing sacred and ritually important materials such wedding invitations. Perhaps it may be used for writing horoscopes etc. also. Another important cultural aspect of the script is the availability of several manuscripts written in this script. Knowledge of the script gives the members of the community access to these materials. However, the glyph shapes of various characters differ considerably in these manuscripts. I just received a copy of a 'Saurashtra Primer' published by Litho-press process in 1876 where the number of characters and glyph shapes are drastically different from what we see in the proposal. The proposal mentions 'there are a number [of] TrueType implementations of the modern Saurashtra script in circulation on the internet'. However, I could see just the two fonts (mentioned earlier) on the internet. Even then, both of them are very similar in their general shapes. One provides more glyphs of C+matra type. An interesting point is: "how is this 'modern' Saurashtra script defined?" Is it defined by the very few fonts available on the internet? Who has defined the usage of AB4D (halanta) sign to be compulsorily used after the first C in a consonant cluster which is entirely different from all the other Indian Indic scripts except Tamil? How much printed or written body of literature is available in the 'modern' Saurashtra script that vouches for these glyph shapes?

The proposal further says "To date, no computer implementation of the older version of the script has been created". As explained elsewhere in this comment, there seems to be not one but several older stages of the script which need to be taken into consideration in future. Precisely for this purpose, we need to have a wider character encoding schema for this script exactly parallel to the Devanagari code chart.

The glyph shapes given in the proposal (as taken from the available fonts and the printed materials) probably reflect their current usage. But if one goes about 50 years into the past, the glyph shapes and the number of characters start differring. This was clear from my perusal of the 1876 publication mentioned above as well as a couple of other books that Prof. Ucida showed me recently.

Thus it is important to keep in mind the history of the script while encoding the current stage. Looking at the rich variation in the script over its history, proposals for adding some older characters will certainly arise. An example from Telugu and Kannada scripts would be apt here. Corresponding to Tamil U+0BB4, Telugu U+0C34 and U+0CB4 is kept blank. However, Devanagari has the character U+0934. It should be recalled that ISCII Devanagari (on which Uncode coding was primarily based), does not just encode Devanagari as used for Hindi, Marathi, Sanskrit, Nepali or Newari. It encodes what is called 'pari-vardhita-devanaagarii' (=extended Devanagari). Extended Devanagari was created to cater to most of the characters that are found in Dravidian languages as well as Bengali, in addition to the regular Devanagari characters. Now, after going through various linguistic materials on Telugu and Kannada I came to the conclusion that we need to provide characters for Telugu U+0C34 and U+0CB4. These two characters are used in writing Inscriptional Telugu and Kannada. I will be making a proposal in this regard soon. Similarly, we need to look into historical aspects of Saurashtra script and provide space for characters that are used in their historical documents. Saurashtra script is an Indic script. Just like Tulu and Grantha scripts it was primarily used for writing down Saurashtra and Sanskrit materials (as well as

Telugu materials). Hence, I propose the positioning of characters parallel to Devanagari character set. This is in line with many other Indian scripts. I was told that one of the very few users of the Saurashtra script would like to propose for inclusion of an Anusvaara in the chart. Having a chart symmetrical with Devanagari chart would facilitate any future additions.

UpaakSara: (spelt as 'upakshara' in the Proposal) [I will refer to it as H hereafter]: There has been some good amount of discussion on this topic. First, the proposers wanted to call it 'Aayatam' perhaps in analogy with the Tamil 'aayatam'. However, Tamil Aayatam has its own specific characteristics. Its phonetic shape has not been finally resolved. Its occurrence is very limited. The proposers later dropped this name and took up the term 'Upakshara'. As the discussion in *C2c* above indicates, it is not clear who coined this word. As far as its meaning in Sanskrit goes, it means 'sub (upa) letter (akSara)'. This is obviously someone's recent coinage.

H need not and should not be given the status of a character at this stage. The reasons are the following:
1. The character-set of an Indic script has two groups of characters. The major group contains characters which are direct representations of Graphemes. The other group contains characters representing punctuation etc. A grapheme is an abstract element that is represented by one or more 'allographs'. A glyph is a physical token of an allograph. (Please see my document in Current Documents L2/02-402 at http://www.unicode.org/L2/L2002/02402-indic.pdf for a longer discussion on this.)
2. There are three basic premises in Graphemics (and in phonemics and morphology). They are: Contrast, Complementation, and Free Variation. If two elements occur in the same environment and bring about a difference in meaning, then those two elements are in 'Contrastive distribution'. They cannot be interchanged without changing the meaning. For instance, /p/ and /b/ contrast in English (but not in Native Tamil vocabulary). If two elements occur in mutually exclusive environments, then they are said to be in 'complementary distribution' or 'predictable environments'. For instance the aspirated [ph] and the unaspirated [p] in English (but in Hindi, they contrast). Their occurrence is derived by a rule. They form a set of allophones of the same phoneme (say English /p/). If two elements occur in contexts where one can replace the other freely, then they are in 'free variation'.
3. I went through the whole of the 'Saurashtra-English Dictionary' by Prof. N.Ucida. I met Prof. Ucida and discussed with him the issue. Here I discuss the graphemic issue of H (not its phonology). All the occurrences of the four sequences of <nH>, <mH>, <rH>, and <lH> are just surface phenomena of underlying sequences of {nh}, {mh}, {rh} and {lh}. {h} is represented at AB39 in the character chart given at the end of this comment. Thus, <h> and <H> are allographs of the underlying grapheme {h}. There is no single instance of <nh>, <mh>, <rh>, or <lh>. There is also not a single case of <XH> where <X> is a letter other than <n>, <m>, <r>, or <l>. This is classic case of allographic distribution. There are several more parallel allographic cases in Devanagari and other Indic scripts.

Formulaically:
Let N = \<n\>,\<m\>, \<r\>, or \<l\>
Let X ≠ \<n\>,\<m\>, \<r\>, or \<l\>
Rewrite {Nh} → \<NH\>
    and {Xh} → \<Xh\>

For this reason, I do not recommend setting up of 'H' as a character. However, a glyph of 'H' will be necessary for rendering the typography which can be included in the glyph set of the font.

Character Chart: I explained above the need for having the Saurashtra Code Chart which is parallel to the Devanagari chart. This will be highly necessary for any future expansion of the code points as there are several other intricate letters in Saurashtra manuscripts. By adapting the general Devanagari pattern, all the necessities of the Saurasthra script can be satisfactorily addressed. The proposed character chart follows. The letter names of the above letters are the same as that of Devanagari letter names given in Unicode 4.0, except the word 'Devanagari' needs to be replaced by the word 'Saurashtra'.

Row AB : SAURASHTRA

	AB0	AB1	AB2	AB3	AB4	AB5	AB6	AB7
0		ꢀ	ꢁ	ꢂ	◌ꢃ		ꢄ	
1			ꢅ		◌ꢆ		ꢇ	
2	ꢈ	ꢉ	ꢊ	ꢋ	◌ꢌ		ꢍ	
3	ꢎ	ꢏ	ꢐ	ꢑ	◌ꢒ		ꢓ	
4		ꢔ	ꢕ		◌ꢖ			
5	ꢗ	ꢘ	ꢙ	ꢚ				
6	ꢛ	ꢜ	ꢝ	ꢞ	◌ꢟ		ꢠ	
7	ꢡ	ꢢ	ꢣ	ꢤ	◌ꢥ		ꢦ	
8	ꢧ	ꢨ	ꢩ	ꢪ	◌ꢫ		ꢬ	
9	ꢭ	ꢮ		ꢯ			ꢳ	
A	ꢰ	ꢱ	ꢲ		◌ꢴ		ꢵ	
B	ꢶ	ꢷ	ꢸ		◌ꢹ		ꢺ	
C	ꢻ	ꢼ	ꢽ		◌ꢾ		ꢿ	
D		ꣀ	ꣁ		◌ ꣠		ꣃ	
E	꣄	ꣅ	ꢦ	◌ꢧ			꣍	
F	꣎	꣏	ꢨ	◌ꢩ			꣙	