

Feedback on L2/17-209 on Takri SSA

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

2017-Jul-18

The document has provided compelling evidence of the use of the shape 𑂔 currently encoded as U+1168B TAKRI LETTER KHA to be in fact – i.e. as per script evolution – the Takri letter SSA i.e. the one nominally denoting the phoneme *ṣa*. Though 𑂔 has also been used to denote *kha*, the document has shown that this practice of using the nominal written form of *ṣa* for *kha* is also seen in Devanagari. It suggests that therefore the shape 𑂔 should be re-encoded as U+116B8 TAKRI LETTER SSA and the shape of U+1168B TAKRI LETTER KHA should be changed to 𑂕.

In fact the original proposal of Takri L2/09-111 has also provided attestations for 𑂕 denoting *kha* on pp* 21, 36 and 40 but seems to have been led by a larger number of letter charts showing 𑂔 for *kha* to instead encode 𑂔 as LETTER KHA and not encode a separate LETTER SSA.

Of course, in retrospect one sees that the original proposal should have considered the following:

Looking at the epigraphical charts on pp 21 and 23, given that the columns from left to right are a temporal progression, 𑂔 has been used for *ṣa* all along (barring the most recent column which doesn't at all show very representative glyphs IMO) and has been used for *kha* only in the most recent periods, whereas 𑂕 was used for *kha* in all earlier periods. How can 𑂕 naturally evolve into 𑂔 which has all along been an entirely separate letter? Obviously this is a case of conflation.

Such reassignment of the nominal written form of one phoneme to another over the source of time is seen in other scripts. For instance in Bengali, based on the shape one can say that “BENGALI LETTER BA” 𑂗 “originally” would have denoted *va*. This is based on the observation that generally in the Brahmi-derived Indic scripts, there is an additional inner stroke or cusp compared to VA in BA as in Devanagari व व, Gujarati વ વ, Telugu వ వ, Malayalam വ വ etc. However the phonemes *va* and *ba* themselves seem to have merged in the Bengali language due to which the simpler 𑂗 was pronounced as [ba] and became identified and encoded as the LETTER BA, even though it still is pronounced as [va] in cluster-final position. Likewise Malayalam vowelless 𑂗 derived from TA 𑂗 came to be pronounced as [l] due to the users' linguistic habits and is thus encoded as CHILLU L.

Now in the case of *ṣa* and *kha*, it is well known that the Mādhyandina school of the Śukla Yajur Veda which is common in central and western North India renders the phoneme *ṣa* when not followed by a vowel as [kha], as in *puruṣa* (person) being rendered as [purukha], *pūṣā* (sun god) as [pūkhā] etc, but *kṛṣṇa* and *ayakṣmā* etc are not modified since here *ṣa* is followed by a consonant.

It is very well known to scholars in other cases also that there is an interplay between the Vedic traditions and the local languages by which the pronunciation patterns seen in the one are

* Page numbers referring to the original proposal are those of the PDF and not of the internal page numbering.

reflected in the other. Thus the later habit seen in Takri of using the letter originally intended for *ṣa* for the sound of *kha* (and possibly merging the phonemes) is quite understandable.

It is also certain that **𑂔** and **𑂕** are two different characters, of which the former is the sign originally used for *kha* but later replaced in writing by the latter which has however been consistently used for *ṣa*.

One also notes that in the related script Gurmukhi: 1) there is no character encoded for SSA; 2) the shape of U+0A16 KHA **ਖ** is quite evidently the same as what was consistently used for SSA and in later periods for KHA in Takri; 3) there is no phoneme *ṣa* in the Panjabi language. No doubt, similar script-language interactions were at work in Gurmukhi-Panjabi creating this situation.

However I am not sure whether the solution which L2/17-209 proposes for Takri — that of changing the glyph of KHA to **𑂔** and encoding a new character SSA with the shape **𑂕** — is in line with Unicode practice. As I understand it, it is written forms and not phonemes that are encoded with a particular codepoint and a particular name. As such, having allotted the codepoint U+1168B to the written form **𑂕**, how appropriate is it to encode a new character with the same shape at a different codepoint just because it may denote a different phoneme? Is it not tantamount to re-encoding or changing the codepoint of an already encoded character?

It is true that **𑂕** should ideally have been encoded with the name SSA, but so should **𑂆** have been encoded as Kannada LLLA and not FA. While FA is an entirely mistaken name for **𑂆**, KHA is not an entirely mistaken name for **𑂕** since many letter charts in references do identify it as *kha*. If KHA is an appropriate name for Gurmukhi **ਖ** though at some point in history even that character probably denoted *ṣa*, so is it for Takri **𑂕** which is in effect the very same written form.

As such, I feel that the proposed solution is not in line with Unicode principles. My suggestion would be to instead go the way of my L2/12-225 which proposed an alternate historic representation of II in Malayalam as a separate character.

In effect, **𑂔** should be encoded as U+116B8 TAKRI LETTER TRADITIONAL KHA, and appropriate annotations added to the new character to indicate that it denoted *kha* in some earlier writings and to the existing character U+1168B to indicate that it is also used to denote *ṣa*.