Proposed Changes to Gurmukhi 4

Document Number: L2/06-030
Submitter’s Name: Sukhjinder Sidhu (Punjabi Computing Resource Centre)
Submission Date: 2006-01-27

This document aims to clarify the use of C1 conjuncts in Gurmukhi and also aims to highlight problems in rendering engines which may have been caused by ambiguity in the standard.
A. Conjuncts

Although Devanagari and Gurmukhi are closely related, they differ in the extent to which they use conjuncts. Gurmukhi as used for the Punjabi language tends to use conjoined forms of Ra, Ha, Va and Ya only (the later two being rare if not obsolete in modern usage). Devanagari tends to represent all conjuncts faithfully and as such has many half forms and conjunct ligatures.

Before Gurmukhi was ‘standardised’ in the form used today, it also employed extensive use of conjuncts. Most of these were listed in document L2/05-344 that was submitted to the UTC at the last meeting. However all the conjuncts were subjoined or post-base and it included no details of consonants that change form in the first half of a conjunct (C1 conjuncts).

There are several manuscripts in the British Library that employ the C1 half-form of Sa (ਸ):

Foll. 65b – 68a, सत-शिक्षा, Santa-šikshā
“ਸਤਿਸ਼ਕਾ ਉਪਰੂਪਾ॥ ॥

Foll. 106a – 172a, अष्टावक्र, Aṣṭāvakra
“अष्टावक्र रिहायल आतेमी अष्टावक्र स्वाम्य घरनਹੈ॥

In most Gurmukhi text, the first example would be rendered ‘ਸਤਸਾ’ with a subjoined Va. There are reports of C1 half-forms of other consonants, but no direct evidence for these has been found yet.

There is also evidence for some use of the reph form of Ra (for example, in Mahan Kosh) although this is not widespread:

तर्तेलेक → उतर्तेल (tartarīka) – typically rendered उतर्तेल
धुर्यासन → पुर्यासन (dhuryāsana) – typically rendered पुर्यासन

The use of a reph is non-standard and appears to be used only when transliterating from Devanagari. The Gurmukhi reph form of Ra is visually distinct and is not the same as the Devanagari reph:

In Gurmukhi a reph would not be predominant. A subjoined or post-base consonant (C2 conjunct) is always to be preferred:

 ora + ’ ਚ + ਜ  → ਚਜ [Post-base Ya]
 or + ’ ਚ + ਵ  → ਚਵ [Subjoined Va]

For evidence, see section G.
B. Half-form and Reph Recommendations

Neither C1 half-forms nor the reph form of Ra are standard behaviour in Gurmukhi – in fact, the use of these forms would be confusing to most modern Gurmukhi readers. As such, the following recommendations are made to the UTC:

List in Chapter 9 that archaic Gurmukhi may use C1 half-forms and indicating Sa (ਸ) as an example.

Clarify that consonants that change form in the second half of a conjunct (C2 subjoined or post base consonants) always take priority over C1 half-forms. C1 half-forms can be specifically shown by using ZWJ.

\[\begin{array}{c}
\text{ਐ + \u0a3c + ਰ} \rightarrow \text{ਐ} \\
\text{ਐ + \u0a3c + ZWJ + ਰ} \rightarrow \text{ਐ} \\
\text{ਐ + \u0a3c + ZWNJ + ਰ} \rightarrow \text{ਐ} \\
\end{array}\]

Indicate that the reph form of Ra may also be used in Gurmukhi. However, clarify that unlike Devanagari, it is not the preferred form and should only be shown when specifically requested. The rules for differing forms of Ra in Devanagari are:

\[\begin{array}{c}
\text{ਰ + '◌੍' + क} \rightarrow \text{क} \quad \text{[Reph]} \\
\text{ਰ + '◌੍' + ZWJ + क} \rightarrow \text{क} \quad \text{[Half-form]} \\
\text{ਰ + '◌੍' + ZWNJ + κ} \rightarrow \text{र} \quad \text{[Visible Halant]} \\
\end{array}\]

For Gurmukhi, it is requested that the following is the preferred form:

\[\begin{array}{c}
\text{ਰ + '◌੍' + Cons} \rightarrow \text{ਰ} \quad \text{C2 conjunct} \\
\text{ਰ + '◌੍' + ZWJ + Cons} \rightarrow \text{ਰ} \quad \text{C1 half-form (as with Devanagari) or reph form [font dependent]} \\
\text{ਰ + '◌੍' + ZWNJ + Cons} \rightarrow \text{ਰ} \quad \text{Visible Halant (as with Devanagari)} \\
\end{array}\]

Such reph behaviour may be useful for specialist applications but is not suitable for mainstream Gurmukhi. Rendering engines and font technology should merely provide the ability to implement such features.

C. Independent Vowel Behaviour

We recommend that the UTC clarify that although Ura (U+0A73) and Iri (U+0A72) are designed for singular representation, they should be able to combine with Nukta (U+0A3C). The following sequences should combine as shown:

\[\begin{array}{c}
\text{ਾ + '◌੍' } \rightarrow \text{ਾ} \\
\text{ਾ + '◌੍' } \rightarrow \text{ਾ} \\
\text{ਇ + '◌੍' } \rightarrow \text{ਇ} \\
\end{array}\]

In addition, some dialects of Punjabi are peculiar in that when they are written, Pairin Haha (the subjoined form of ‘Ha’) can attach to independent vowels. This behaviour initially appears strange when compared to other Indic scripts, but the purpose of the subjoined Ha (in this instance) is to indicate a high tone in Punjabi.

\[\begin{array}{c}
\text{਼ਰਿਤ} \rightarrow \text{਼ਰਿਤ} \\
\text{ਜ਼ਾ} \rightarrow \text{ਜ਼ਾ} \\
\end{array}\]

This peculiar behaviour should be made clear in chapter 9 of the Unicode Standard. This applies to all independent vowels and may also apply to Ura (U+0A73) and Iri (U+0A72) for representation purposes. The behaviour also applies to the recently accepted GURMUKHI SIGN UDAAAT (U+0A51) which should also have the ability to join with independent vowels.

For evidence, see section G.
D. Nasal Sign Placement

In some cases, the nasal signs (Bindi, Tippi and possibly Adhak Bindi) are represented to the left of GURMUKHI SIGN I:

\[ \ddot{Y} \ddot{I} \ddot{I} \]  
\[ \ddot{Y} \ddot{I} \ddot{I} \]  
\[ \ddot{Y} \ddot{I} \ddot{I} \] (This form has not been found, although it is not inconceivable that it was also used).

However, this does pose a problem because of the way independent vowels are handled. Because the independent sign ‘\( \ddot{Y} \ddot{I} \)’ is pre-composed, there is no way to place a nasal sign before the ‘\( \ddot{Y} \ddot{I} \)’ vowel as shown:

\[ \ddot{Y} \ddot{I} \ddot{I} \]

The only way this form can be represented in Unicode is:

\[ \ddot{Y} \ddot{I} \ddot{I} \]

I recommend that the UTC enables the use of VARIANT SELECTOR 1 (U+FE00) to enable this function. This can be applied universally to Gurmukhi nasal signs:

\[ \ddot{Y} + \ddot{I} + \text{VS1} + \ddot{I} \]  
\[ \ddot{Y} + \ddot{I} + \text{VS1} + \ddot{I} \]  
\[ \ddot{Y} + \ddot{I} + \text{VS1} + \ddot{I} \]

The representative images are not positioned correctly due to deficiencies in current rendering techniques. The nasal sign should be placed fully on the base character, and the vowel sign should connect to the top line.

E. Collation Issues

The character GURMUKHI SIGN UDAAT was recently accepted for encoding by the UTC. This character will need an entry into the collation tables for Gurmukhi. It should sort as if it is the subjoined form of /ha/. The existing Unicode Collation Charts should be altered to:

<table>
<thead>
<tr>
<th>( \ddot{Y} \ddot{I} \ddot{I} )</th>
<th>0A38</th>
<th>0A39</th>
<th>0A51</th>
<th>0A15</th>
<th>0A16</th>
<th>...</th>
</tr>
</thead>
</table>

It should sort as the equivalent of ‘Virama + Ha’.

F. Feedback on Public Review Issues

The feedback here has already been submitted using the e-mail form on the Unicode web site.

PRI 81

The Gurmukhi entries in the Provisional Named Sequences are seemingly of no particular relevance or use.

GURMUKHI PARI CA;0A4D 0A1A  
GURMUKHI PARI TA;0A4D 0A24  
GURMUKHI PARI NA;0A4D 0A28  
GURMUKHI PARI VA;0A4D 0A35  
# GURMUKHI HALF YA;0A2F 0A4D  
# GURMUKHI PARI YA;0A4D 0A2F
They should be removed because they are incorrectly transliterated (Pari instead of Pairin), inconsistent (why is Pari used and then Half instead of the Punjabi word ‘Adha’?) and incomplete (some conjuncts, such as Pairin Haha, Rara have been recognised in the standard for a while now and are not listed).

In addition, conjuncts can take both subjoined and post-base forms, so Half Ya could also be Pairin Ya. It would be inappropriate to use or prefer one form over the other. This was pointed out by me in document L2/05-167 section B1 and since then the two entries have been commented out. This case is not restricted to just Ya.

PRI 82

I recommend the ordering LEFT TOP BOTTOM RIGHT. This seems to be the most logical order for scripts written from left to right, top to bottom. I see no reason why this same principle could not be applied to other Indic scripts.

A top-to-bottom and left-to-right approach is already used on Microsoft Windows.

Telugu: చువనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనवనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనवనवనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనवనవనవనవనవనవనवనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవనవన

Malayalam: നി നി നി നി

However, no cases in other Indic scripts seem to have yet used other combinatinos (e.g. left + bottom).
G. Evidence for Usage

Mahan Kosh - Encyclopedia of Sikh Literature, page 2041. Shows the Reph form of Ra used when transliterating from Devanagari. Also shows the C1 form of Sa.

Use of the conjunct “svā”. This is typically represented as ‘ਸਹ’ using a subjoined Va.

Use of the conjunct ‘ṣṭa’. This is typically represented as ‘ਸ੍ਟ’ with a subjoined Ṭa.

The use of a subjoined Ha with independent vowels.

Panjābī bhāṣā sarōt atē sarūp, Butta Singh Brar, p. 107

ਪੰਜਾਬੀ ਭਾਸ਼ਾ ਸਰੋਤ ਅਤੇ ਸਰੂਪ (ਲੇਖਕ: ਬੂਟਾ ਸਿੰਘ ਬਰਾੜ)