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-śikshā

"ਗੁਰਤੇ ਬਿ**ਸ੍ਵਾ**ਸ ਰਖਲਾ ॥ ੩ ॥"

Foll. 106a – 172a, अश्टावक्र, Aśṭāvakra

"ਅਥ ਕਵਿਤ ਦਿਆਲ ਅਨੇਮੀ ਅ**ਸ੍ਟਾ**ਵਦਕੂ ਭਾਖਾ ਬਰਨਤੇ ॥"

In most Gurmukhi text, the first example would be rendered 'ਬਿਸ੍ਵਾਸ' 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:

तर्तरीक   →   ਤ**ਰ੍ਤ**ਰੀਕ (tartarīka) – typically rendered ਤਰੁਰੀਕ

धुर्यासन   →   ਧੁ**ਰ੍ਯਾ**ਸਨ (dhuryāsana) – typically rendered ਧੁਰਜਾਸਨ

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:

ਰ + '੍' + ਯ   → ਰਜ   [Post-base Ya]
ਰ + '੍' + ਵ   → ਰੂ   [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.

ਸ + ੍ + ਵ	→ ਸ੍ਵ
ਸ + ੍ + ZWJ + ਵ	→ ਸ੍ਵ
ਸ + ੍ + ZWNJ + ਵ	→ ਸ੍ਵ

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:

र + 'ि' + क	→ र्क	[Reph]
र + 'ि' + ZWJ + क	→ र्क	[Half-form]
र + 'ि' + ZWNJ + क	→ र्क	[Visible Halant]

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

ਰ + 'ਿ' + Cons	→ C2 conjunct
ਰ + 'ਿ' + ZWJ + Cons	→ C1 half-form (as with Devanagari) or reph form [font dependent]
ਰ + 'ਿ' + ZWNJ + Cons	→ Visible Halant (as with Devanagari)

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:

ੳ + 'ਿ'	→	ੳ਼
ਅ + 'ਿ'	→	ਅ਼
ੲ + 'ਿ'	→	ੲ਼

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.

ਤਿਆੜਾ	→	ti'áɖā
ਜਾਅ	→	jā'á

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 UDAAT (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 Gᴜʀᴍᴜᴋʜɪ Sɪɢɴ II:

ਂੀ

ਁੀ

ਁੀ (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 'ਈ' is pre-composed, there is no way to place a nasal sign before the 'ੀ' vowel as shown:

ਗੁਸਾਂਈ

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

ਗੁਸਾਈਂ

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

ਸ + ੀ + VS1 + ਂ → ਸਂੀ

ਕ + ੀ + VS1 + ਁ → ਕਁੀ

ਦ + ੀ + VS1 + ਁ → ਦਁੀ

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 Gᴜʀᴍᴜᴋʜɪ Sɪɢɴ Uᴅᴀᴀᴛ 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:

	ਸ	ਹ	ੑ	ਕ	ਖ	
...	0A38	0A39	0A51	0A15	0A16	...

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: ల్లై -> ల ఄ ఄ

Malayalam: കൊ -> ന കെ ാ

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.

IV. Foll. 65b—68a. संतशिक्षा Santa-śikshā. Rules of conduct for the guidance of Vaishṇavas.

Heading: ਅਥ ਸੰਤਸਿਖਜਾ ਗ੍ਰੰਥ ਲਿਖਤੇ ॥

Beg. ਪਹਲੇ ਗੁਰਾਨ ਨਿਰੰਕਾਰ ਜਾਨਲਾ ॥੧॥ ਸੰਤੋਖ ਰਖਲਾ ॥੨॥ ਗੁਰਾਤੇ ਬਿਸ੍ਵਾਸ ਰਖਲਾ ॥੩॥ ਖਿਮਾ ਰਖਲੀ ॥੪॥ ਗਰੀਬੀ ਰਖਲੀ ॥੫॥ ਬੈਠਲਾ ਸ਼ੁਭ ਅੰਕ ਸੰਕੋਚਕੇ ॥੬॥ ਭੋਜਨ ਥੋੜ੍ਹਾ ਖਾਲਾ ॥੭॥

Use of the conjunct "svā". This is typically represented as 'ਸ੍ਵਾ' using a subjoined Va.

I. Foll. 1—36. अमृतसरमाहात्म्य Amritsar-māhātmya. A poem in 8 chapters (adhyāya) in praise of the sacred city of Amritsar, said to be taken from the Bhavishyapurāṇa.

Heading: ਅਥ ਅੰਮ੍ਰਤਸ੍ਰ ਮਹਾਤਮ ਲਿਖਤੇ ॥

Beg. ਦੋ : ॥

ਗੁਰ ਗਨੈਸ ਅਰ ਸਾਰਦਾ ਬੰਦੇ ਹਿਤ ਚਿਤ ਲਾਇ ॥
ਜਾਕੀ ਕ੍ਰਿਪਾ ਕਟਖਤੇ ਉਤਮ ਮਤਿ ਕੋ ਪਾਇ ॥੧॥

Colophon: ਇਤਿ ਸ੍ਰੀ ਭਵਿਖਤਪੁਰਾਨੇ ਉਮਾਮਹੇਸ ਸੰਬਾਦੇ ਅੰਮ੍ਰਤਸ੍ਰਮਹਾਤਮੇ ਅਸਟਮੋਧਿਆਇ ॥

Use of the conjunct 'sṭa'. This is typically represented as 'ਸ੍ਟ' with a subjoined Ṭa.

ਧੁਨੀ ਵਿਉਂਤ

ਧੁਨੀ ਵਿਉਂਤ ਦੇ ਪੱਖ ਤੋਂ ਸੁਰ (tone) ਮਾਝੀ ਦੀ ਉੱਘੀ ਵਿਸ਼ੇਸ਼ਤਾ ਹੈ। ਮਾਝੀ ਵਿਚ ਤਿੰਨ ਸੁਰਾਂ— ਉੱਚੀ, ਨੀਵੀਂ ਅਤੇ ਪੱਧਰੀ ਉਚਾਰੀਆਂ ਜਾਂਦੀਆਂ ਹਨ। ਸੁਰਾਂ ਦੀ ਹੋਂਦ ਨਾਦੀ-ਮਹਾਂਪ੍ਰਾਣ ਧੁਨੀਆਂ /ਘ,ਝ,ਢ,ਧ,ਭ/ ਅਤੇ ਸੁਰਯੰਤਰੀ ਧੁਨੀ /ਹ/ ਆਦਿ ਦੀਆਂ ਵੱਖ-ਵੱਖ ਸਥਿਤੀਆਂ ਉੱਤੇ ਨਿਰਭਰ ਕਰਦੀ ਹੈ। ਮਾਝੀ ਵਿਚ ਸ਼ਬਦ ਦੀਆਂ ਤਿੰਨੇ ਸਥਿਤੀਆਂ ਵਿਚ /ਹ/ ਧੁਨੀ ਸੁਰ ਵਿਚ ਬਦਲ ਜਾਂਦੀ ਹੈ। ਮਿਸਾਲ ਵਜੋਂ 'ਹੋਰ' ਦਾ ਮਾਝੀ ਉਚਾਰਨ 'ਓਰੁ', 'ਤੁਹਾਡਾ' ਦਾ 'ਤਿਆੜਾ' ਅਤੇ 'ਜਾਹ' ਦਾ 'ਜਾਆੂ' ਹੈ। ਇਸ ਦੇ ਮੁਕਾਬਲੇ ਮਲਵਈ ਵਿਚ /ਹ/ ਦਾ ਵਿਅੰਜਨੀ ਸਰੂਪ ਕਈ ਥਾਵਾਂ ਉੱਤੇ ਕਾਇਮ ਰਹਿੰਦਾ ਹੈ। ਖ਼ਾਸ ਕਰਕੇ ਸ਼ਬਦਾਂ ਦੇ ਮੱਧ ਵਿਚ ਤਾਂ ਮਲਵਈ /ਹ/ ਧੁਨੀ ਨੂੰ ਵਿਅੰਜਨੀ ਸਰੂਪ ਵਿਚ ਹੀ ਉਚਾਰਦੇ ਹਨ।

The use of a subjoined Ha with independent vowels.
Panjābī bhāśā sarōt atē sarūp, Butta Singh Brar, p. 107
ਪੰਜਾਬੀ ਭਾਸ਼ਾ ਸਰੋਤ ਅਤੇ ਸਰੂਪ (ਲੇਖਕ: ਬੂਟਾ ਸਿੰਘ ਬਰਾੜ)