

Proposed Changes to Gurmukhi

Document Number: L2/05-088

Submitter's Name: Sukhjinder Sidhu (Punjabi Computing Resource Centre)

Submission Date: 21 April 2005

Abstract

This document proposes changes in the rendering behaviour of Gurmukhi and the possible addition of new code points. These recommendations should be implemented in chapter 9 of the Unicode standard.

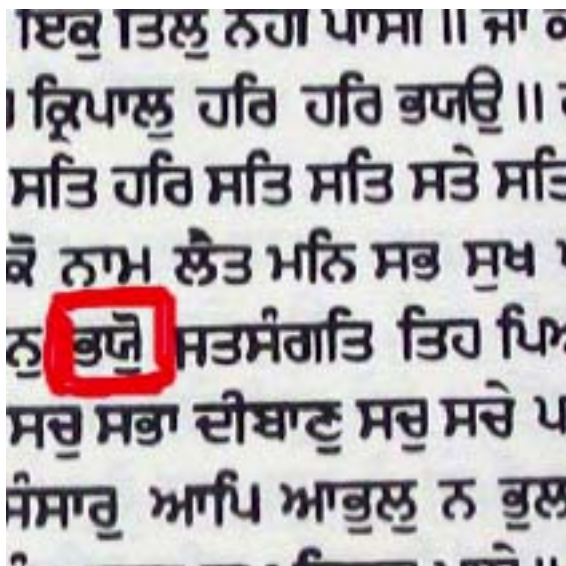
Thanks to everyone who has contributed both time and effort to the research involved in making this document.

A. General Rendering Changes

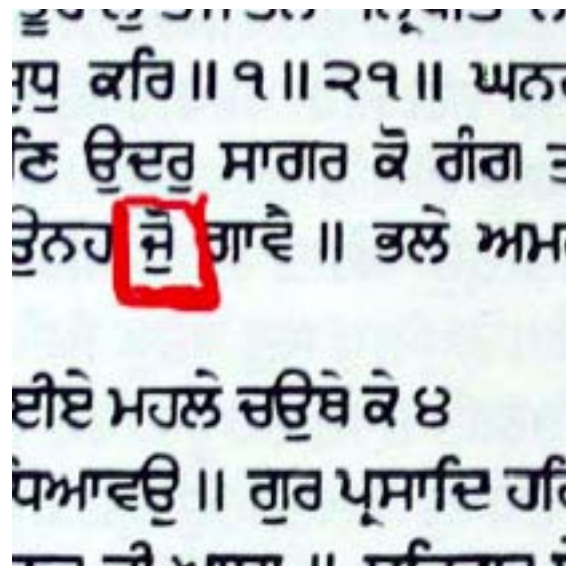
A1. Double Vowel Signs

Older Gurmukhi (for example, in the Sikh holy book – the Sri Guru Granth Sahib) is known to use two vowel signs on one consonant. This behaviour is restricted to Hora (Vowel Sign OO (ੌ) – U+0A4B) and Aunkar (Vowel Sign U (ॄ) – U+0A41). This particular combination represents “the metrical shortening of ‘ō’ or lengthening of ‘u’ depending on context.”¹

The additional vowel sign is added to a syllable and lengthens or shortens the vowel based on the original vowel sign. It is designed to keep the meaning of the original word intact, while indicating how the vowel should be pronounced in poetry.²



A1.1 SGGS page 1386 (੧੩੮੬)



A1.2 SGGS page 1396 (੧੩੯੬)

Example

Umāhā (ਉਮਾਹਾ) becomes Ūmāhā (ॄਮਾਹਾ)

Gōbind (ਗੋਬਿੰਦ) becomes Gobind (ਗੁੋਬਿੰਦ)

Both examples maintain the original meaning of the word while altering the pronunciation.

Proposed Changes

Unicode rendering engines should render the sequence of U+0A4B and U+0A41 (in that order) on one consonant. They should not force U+0A41 to stand alone. If the order is reversed, U+0A4B should stand alone.

The independent form (ੳ) must also be rendered when entering Letter OO (U+0A13) and Vowel Sign U (U+0A41).

If this is an unacceptable solution, I recommend the addition of two new vowel signs to accommodate this phenomenon:

ੳ – U+0A11 – GURMUKHI LETTER SHORT OO OR LONG U

ੌ – U+0A49 – GURMUKHI VOWEL SIGN SHORT OO OR LONG U

¹ Jeevan Deol, Research Fellow in Indian History, St. John's College, University of Cambridge

² Page 405, Gurbani Vyakaran (Gurbani Grammar), Prof. Sahib Singh – See Appendix F1

0A11;GURMUKHI LETTER SHORT OO OR LONG U;Lo;0;L;;;;;N;;;;;
0A49;GURMUKHI VOWEL SIGN SHORT OO OR LONG U;Mn;0;NSM;0A4B 0A41;;;;;N;;;;;

These code points correspond to the independent and dependent forms of Devanagari Candra O although they have no relation to this character.

The reasoning behind using the name 'Short OO or Long U' as opposed to 'O or UU' is that it more accurately conveys the use of the actual character.

GURMUKHI VOWEL SIGN SHORT OO OR LONG U may also be constructed as shown:

U+0A49 (ॐ) = U+0A4B (ॐ) + U+0A41 (ॐ)

See F2 for the proposal summary.

A2. Positioning of Tippi and Bindi

In some instances, the positioning of Tippi (U+0A70 – ੰ) and Bindi (U+0A02 – ਂ) occurs before a selected vowel sign. For example, a Bindi and Tippi can occur before Bihari (U+0A40 - ਾੀ). They can also occur in the middle of an independent vowel sign. This does not occur in modern Gurmukhi. The phenomenon of having Tippi and Bindi in front of a vowel sign is synonymous with it being on the outside of a vowel sign.

Example

ਥਿਤੀ ਕੀ ਦਾਸੀ ਈਧਨ ਹਸਤੀ

ਗਣੀ ਗਨੀ ਗਲੀ ਗੁੰਦਾਈ ਗੁਸਾਈ ਗੁਣੀ ਗੁਰਸਿਖੀ ਗੋਸਾਈ ਚੰਗਿਆਈ ਚੰਦੀ ਜਾਈ ਜਾਂਹੀ ਠਾਈ ਤਾਈ

Proposed Changes

I currently do not recommend any changes for encoding this stylistic difference. This point is merely highlighted as an area that needs further investigation.

B. Rendering Clarification

B1. Ura, Aira and Iri

At present, there is a disparity as to how rendering engines handle the use of Ura (U+0A73), Aira (U+0A05) and Iri (U+0A72). Some rendering engines allow the addition of vowel signs to these 'vowel sign bearers' while others do not. There appears to be no coherent approach as to what forms a valid "renderable" sequence. For example, should ੲ (U+0A72) + ਿ (U+0A3F) render as 'ਇ' or 'ਇਿ'?

There were two approaches discussed and these are listed below. In a purely technical aspect, B2 is easier and less disruptive method to implement, however, B3 better represents how the users of Gurmukhi wish their own script – a way they find easier and natural. It may be unfeasible to implement the recommendations in B3 and that is the only reason why B2 is considered.

Only **one** of these methods needs to be implemented.

B2. Clarification of Ura, Aira and Iri usage

Some rendering engines allow the addition of vowel signs to Ura (U+0A73), Aira (U+0A05) and Iri (U+0A72) as if they were standard consonants. This is not the case and they should not be treated as standard consonants.

Allowing this behaviour causes normalisation issues (there is no recognised decomposition between the pre-composed code points and the combinations with Ura, Aira and Iri as vowel sign bearers) and major rendering issues when transporting data from one system to another.

Ura, Aira, and Iri must not accept any vowel signs and should only be used when they need to be represented on their own. The pre-composed independent vowels should be used when the user wishes to add a vowel sign to Ura, Aira or Iri.

Although this same problem does not generally affect Aira (Gurmukhi Letter A – U+0A05), it should also be clarified as applying to Aira.

Proposed Changes

A clarification is required in the Unicode Standard. In Unicode 4.0, chapter 9.3, Ura, Aira and Iri are mentioned under the heading 'ordering':

"Ordering. U+0A72 GURMUKHI IRI and U+0A73 GURMUKHI URA are the first and third "letters" of the Gurmukhi syllabary, respectively. They are used as bases or bearers for some of the independent vowels, while U+0A05 GURMUKHI LETTER A is both the second "letter" and the base for the remaining independent vowels."

This section is misleading because it does not clearly specify that Ura and Iri as encoded in Unicode are designed merely to represent their standalone forms and not to form bases for independent vowels. This section needs to be modified to clarify the aforementioned issues.

B3. Recognise Decompositions

The Unicode Standard does not specifically prevent the addition of vowel signs onto independent vowels. This is generally not a problem with other Indic scripts because most rendering engines prevent the addition of dependent vowel signs on independent vowel code points. For example, with Devanagari:

अ + ा = आ

अ + ा ≠ अा

When using Gurmukhi there are difficulties because of the way independent vowels are constructed in Unicode. Gurmukhi uses three vowel bearer characters (Ura, Aira and Iri) to form independent vowels. On their own, neither of these characters represent any particular meaning. They must combine with an appropriate vowel sign before they inherit any meaning:

ਉ (Ura) + ੁ (Onkar – U) = ਉ
 ਅ (Aira) + ੋ (Mukta – A) = ਅ
 ਏ (Iri) + ੇ (Lavan – EE) = ਏ

The problem with the current Unicode Standard is it treats Ura (U+0A73) and Iri (U+0A72) as standard consonants and Aira (U+0A05) as an independent vowel. Some rendering engines (for example, Uniscribe) do not allow the addition of any vowel signs to these characters. Other rendering engines (for example, ICU and Pango) allow the addition of vowel signs to Ura and Iri but not to Aira. Some rendering engines may even allow vowel signs to be added to any of the three characters.

Allowing this current rendering mix to continue causes massive compatibility problems. On one system a particular combination might be valid (ਏ + ਿ = ਇ) while on others it may be invalid (ਏ + ਿ = ਇ). Preventing the combinations of vowel signs with the bearer characters is also completely unnatural to a Gurmukhi user – it does not allow them to use Gurmukhi as they were taught and they are dumbfounded as to why they have to enter a combined character. Using independent vowels as currently encoded in Unicode is a completely alien concept for Gurmukhi users.

Proposed Changes

I propose a solution which allows any vowel sign to be attached to Ura, Aira and Iri. These ‘constructed’ independent vowels can then be converted to the pre-composed code points when normalising text. This is the only method that allows *any* text already created using the current rules of the Unicode Standard to be valid on all systems.

ਉ + ੁ = ਉ
 ਉ + ੊ = ਉ
 ਉ + ੋ = ਉ

 ਏ + ਿ = ਇ
 ਏ + ੀ = ਏ
 ਏ + ੇ = ਏ

 ਅ = ਅ
 ਅ + ੋ = ਅ
 ਅ + ੀ = ਅ
 ਅ + ੇ = ਅ

ਅ is equivalent to Aira + Mukta (inherent ‘a’)

If the characters proposed in A1 are encoded, the following forms must also be recognised:

ਉ + ੋ + ੁ = ਉ
 ਉ + ੊ = ਉ
 ਉ + ੁ = ਉ

B3. Chapter 9.3 Ura and Iri Correction

In chapter 9.3, there is a mistake under the heading ‘ordering’. It indicates that Iri is the first letter of the syllabary and Ura is the third. This is incorrect. The Gurmukhi ‘alphabet’ begins Ura, Aira and Iri and as such, Ura is the first letter and Iri the third. The text should read:

“**Ordering.** U+0A73 GURMUKHI URA and U+0A72 GURMUKHI IRI are the first and third “letters” of the Gurmukhi syllabary, respectively.”

C. Conjunct Additions

These subjoined consonants should always be shown. Their use can be inhibited by adding a ZWNJ after Virama.

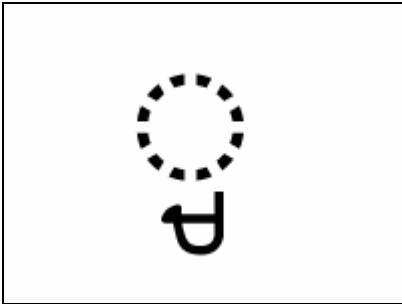
C1. Ga (U+0A17)



C1.1 Subjoined form of Ga

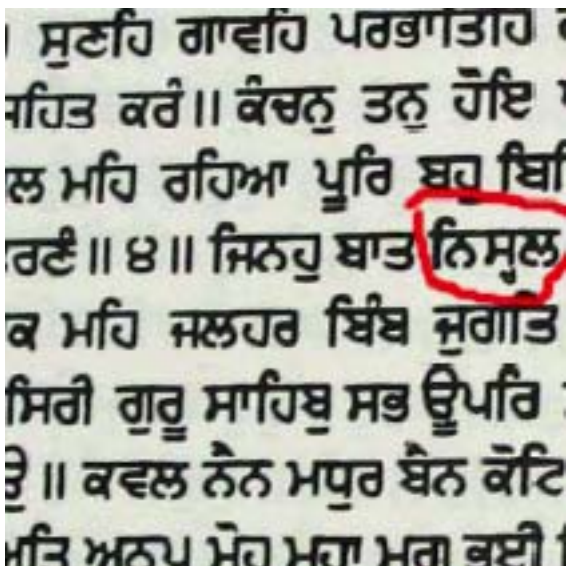
Virama (U+0A4D) + **Ga** (U+0A17)

C2. Ca (U+0A1A)



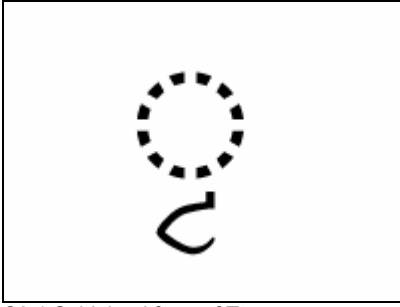
C2.1 Subjoined form of Ca

Virama (U+0A4D) + **Ca** (U+0A1A)



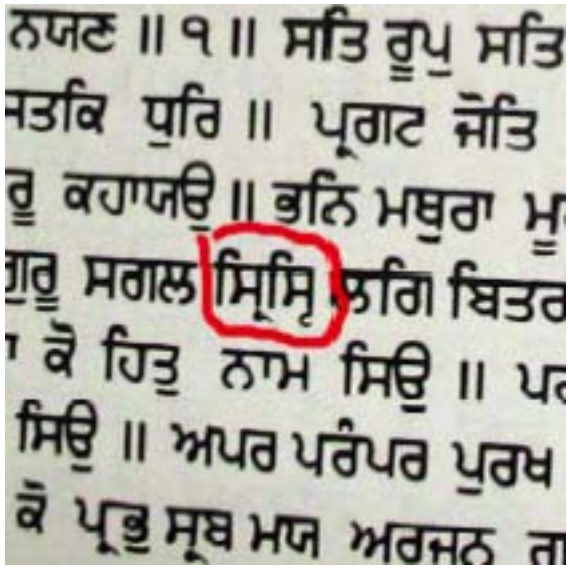
C2.2 SGGS page 1402 (੧੪੦੨)

C3. Tta (U+0A1F)



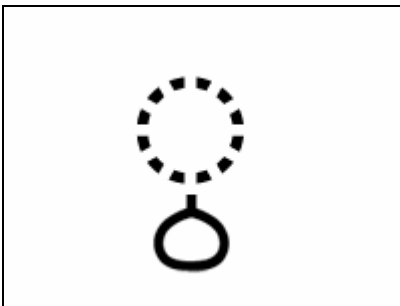
C3.1 Subjoined form of Tta

Virama (U+0A4D) + Tta (U+0A1F)



C3.2 SGGS page 1408 (ੴ੦੮)

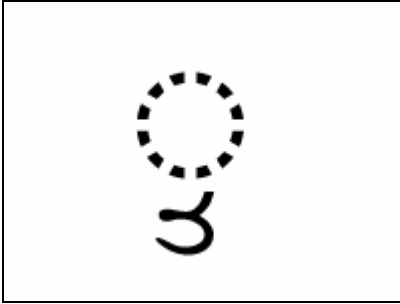
C4. Ttha (U+0A20)



C4.1 Subjoined form of Ttha

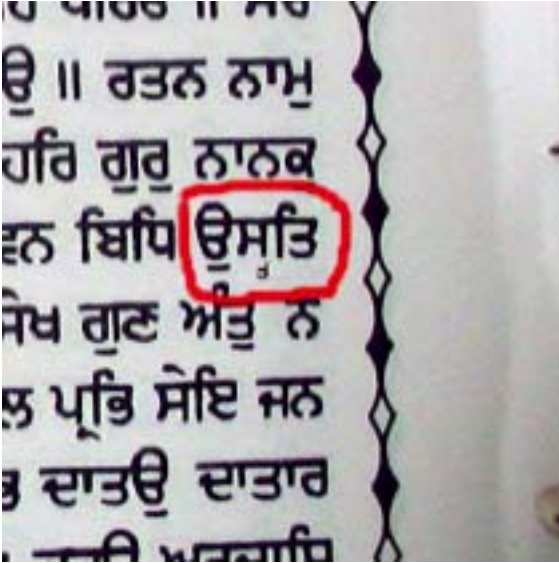
Virama (U+0A4D) + Ttha (U+0A20)

C5. Ta (U+0A24)



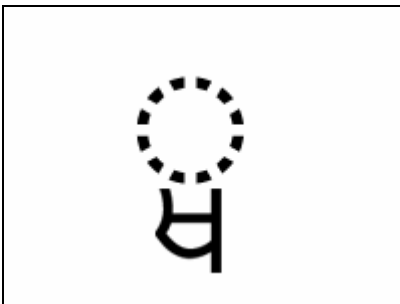
C5.1 Subjoined form of Ta

Virama (U+0A4D) + Ta (U+0A24)



C5.2 SGGS page 1386 (ੴ੮੬)

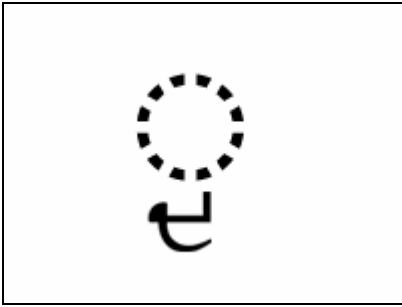
C6. Tha (U+0A25)



C6.1 Subjoined form of Tha

Virama (U+0A4D) + Tha (U+0A25)

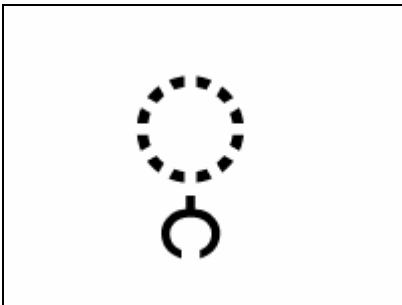
C7. Da (U+0A26)



C7.1 Subjoined form of Da

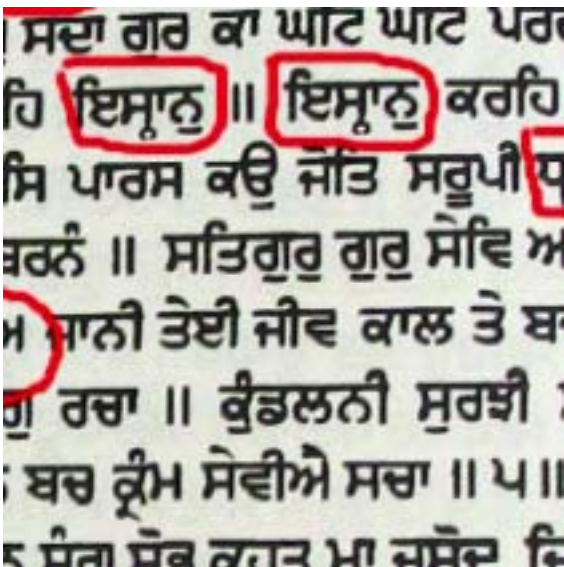
Virama (U+0A4D) + Da (U+0A26)

C8. Na (U+0A28)



C8.1 Subjoined form of Na

Virama (U+0A4D) + Na (U+0A28)



C8.2 SGGS page 1402 (ੴ੦੨)

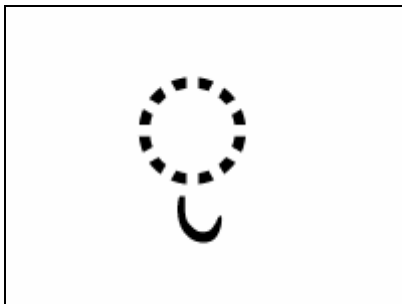
C9. Ya (U+0A2F)

There are three distinct conjoined forms of Ya. Two of these are subjoined consonants and one is a post-base consonant. They are identical in meaning but their use differs based on time period and writing style.

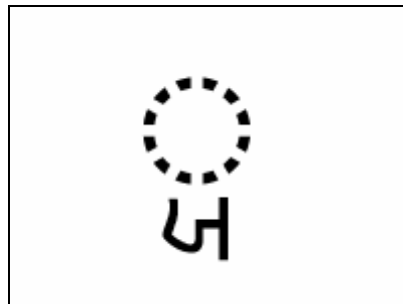
The post-base form (9.1) is recognised in chapter 9 of the Unicode 4 standard. This is the form used in modern Gurmukhi. The subjoined forms are used in older texts.



C9.1 Post-base form of Ya (Modern)

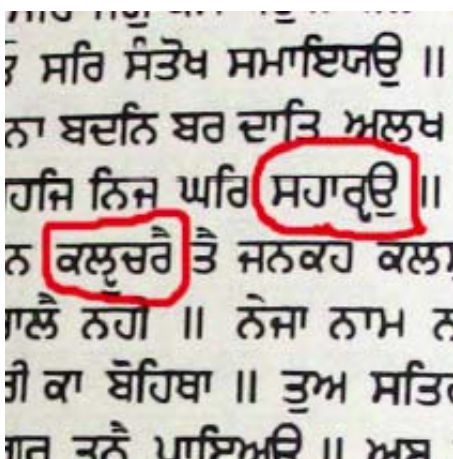


C9.2 Subjoined form of Ya (Old, SGGS)



C9.3 Subjoined form of Ya (Old, Other)

Virama (U+0A4D) + Ya (U+0A2F)



C9.4 SGGS page 1408 (੧੪੦੮)

C9.1 is used in modern contexts and C9.2 is the older form used in the Sikh holy book. Use of C9.3 appears to occur more often in older texts and manuscripts than C9.2, but because of the prevalence and duplication of the Sikh holy book, it is not as well known. It is not inconceivable that more than one form is used at the same time – especially in texts which wish to use the older scribal forms for quotations, but maintain the modern post-base Ya for the main body of text.

Proposed Changes

I recommend that the post-base form of Ya remains encoded as at present (Virama + Ya). I recommend that the subjoined forms are encoded using a “double conjunct”. That is Cons + Virama + Ya + Virama + Ya. This should have no compatibility problems with Gurmukhi because reinforced (geminate) consonants are shown using Addak (e.g. ੱਯ = ਯਯ).

The font or rendering engine can then select which subjoined form of Ya to use (the post-base form always being encoded as Virama + Ya).

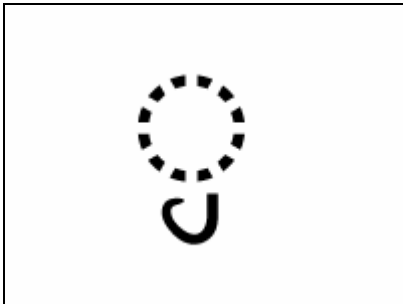
Example

Cons (Ka - U+0A15) + Virama (U+0A4D) + Ya (U+0A2F) = ਕ + ੍ + ਯ = ਕਯ

Cons (Ka - U+0A15) + Virama (U+0A4D) + Ya (U+0A2F) + Virama (U+0A4D) + Ya (U+0A2F) = ਕ + ੍ + ਯ + ੍ + ਯ = ਕ੍ਯ (or alternative subjoined form)

C10. Ha (U+0A39)

There are two subjoined forms of Ha. One form (10.1) is already recognised in the Unicode specification. The second form – referred to by some as ‘Udaat’ – is visually similar to Virama (Halant) but should not be considered the same.

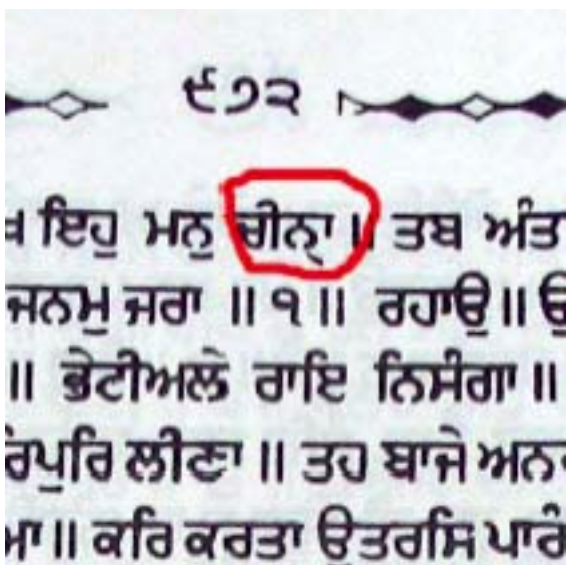


C10.1 Subjoined form of Ha (Modern)

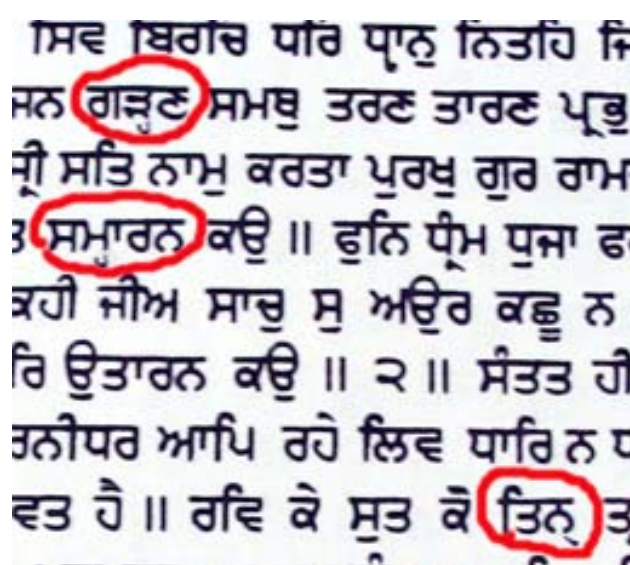


C10.2 Subjoined form of Ha (Old)

Virama (U+0A4D) + Ha (U+0A39)



C10.3 SGGS page 972 (੯੭੨)



C10.4 SGGS page 1404 (੧੪੦੪)

C10.3 shows that although this character is commonly referred to as a Virama (Halant), it is not so because it can carry vowel signs and does not suppress the inherent vowel. C10.4 shows the requirement to show **both** forms of subjoined Ha in one document. I have been unable to determine how to show either variant based on context.

Proposed Changes

Because of the requirement to show both forms in one document I recommend that the C10.2 form of Ha (Udaat) should be encoded by adding a “double conjunct”. That is Cons + Virama + Ha + Virama + Ha. This should have no compatibility problems with Gurmukhi because reinforced (geminate) consonants are shown using Addak (e.g. ੱਹ = ਵਹ).

Example

Cons (Ka - U+0A15) + Virama (U+0A4D) + Ha (U+0A39) + Virama (U+0A4D) + Ha (U+0A39) =
 ਕ + ੱ + ਹ + ੱ + ਹ = ਕ੍ਹ

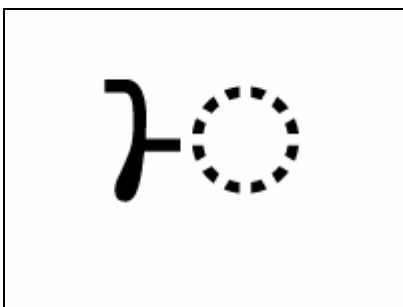
D. C1 Conjunct Additions

This section deals with 'C1' conjoined characters – i.e. characters that take on a half-form in the initial section of a conjunct. These are used in several nineteenth century manuscripts. They are to be formed by entering Cons + Virama + ZWJ.

They must only take on their half-form when ZWJ is entered after Virama. This prevents the introduction of extra complexity for most users of Gurmukhi who will have never seen or used these half-forms before.

Note: I do NOT recommend that the proposals in this section are encoded. See below.

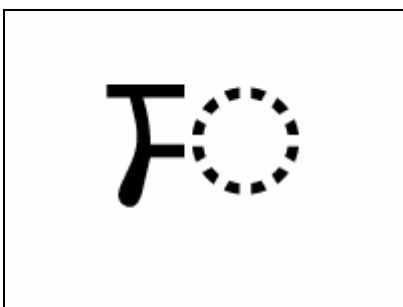
D1. Ma (U+0A2E)



D1.1 Half-form of Ma

Ma (U+0A2E) + Virama (U+0A4D) + ZWJ (U+200D)

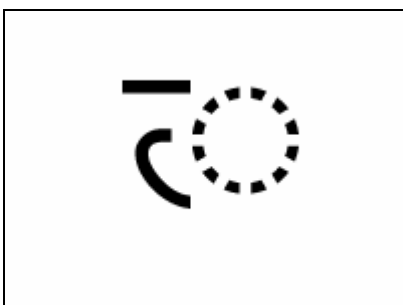
D2. Sa (U+0A38)



D2.1 Half-form of Sa

Sa (U+0A38) + Virama (U+0A4D) + ZWJ (U+200D)

D3. Ha (U+0A39)



D3.1 Half-form of Ha

Ha (U+0A39) + Virama (U+0A4D) + ZWJ (U+200D)

Proposed Changes

I do **NOT** recommend that these changes are implemented because the relevant evidence to back up these half-forms is not presently available. They are only included for completeness and will require further research before a final decision is made.

E. Summary

This is a summary of behaviour that has been suggested for implementation.

A1. Double Vowel Signs

U+0A4B + U+0A41 = ੱ

U+0A13 + U+0A41 = ੳ

ੳ - U+0A11 - GURMUKHI LETTER SHORT OO/LONG U

ੳ̣ - U+0A49 - GURMUKHI VOWEL SIGN SHORT OO/LONG U

B3. Constructing Independent Vowels

ੳ + ੱ = ੳ

ੳ + ੳ = ੳ

ੳ + ੱ̣ = ੳ̣

ੲ + ਿ = ੲਿ

ੲ + ੀ = ੲੀ

ੲ + ੇ = ੲੇ

ਅ = ਅ

ਅ + ਾ̣ = ਅਾ

ਅ + ੈ̣ = ਅੈ

ਅ + ੌ̣ = ਅੌ

ੳ + ੱ̣ + ੱ = ੳ̣

ੳ + ੱ̣ = ੳ̣

ੳ + ੱ = ੳ

C1. Subjoined Ga

C2. Subjoined Ca

C3. Subjoined Tta

C4. Subjoined Ttha

C5. Subjoined Ta

C6. Subjoined Tha

C7. Subjoined Da

C8. Subjoined Na

C9. Subjoined Ya

Cons (Ka - U+0A15) + Virama (U+0A4D) + Ya (U+0A2F) = ਕ + ੍ + ਯ = ਕਯ

Cons (Ka - U+0A15) + Virama (U+0A4D) + Ya (U+0A2F) + Virama (U+0A4D) + Ya (U+0A2F) =
ਕ + ੍ + ਯ + ੍ + ਯ = ਕ੍ਯ (or alternative subjoined form)

C10. Subjoined Ha

Cons (Ka - U+0A15) + Virama (U+0A4D) + Ha (U+0A39) + Virama (U+0A4D) + Ha (U+0A39) =
ਕ + ੍ + ਹ + ੍ + ਹ = ਕ੍ਹ

F. Appendix

F1. Gurbani Vyakaran – Page 405

This is an extract from Gurbani Vyakaran (Gurbani Grammar) by Professor Sahib Singh. Published by Singh Brothers, Amritsar – 10th Edition 1994. This is an English version of the original Punjabi text kindly translated by Serjinder Singh.

Hora (ੌ) or Aunkar (ੁ)?

We find in the Sri Guru Granth Sahib several such words that have both 'ੌ' and 'ੁ' attached to a single character. The character with 'ੁ' is counted as one Matra (short vowel, also called Laghoo Matra in Sanskrit) and the one with 'ੌ' is counted as two Matras (long vowel, also called Deeragh Matra in Sanskrit). In poetry sometimes a character with a single Matra has to be lengthened to be a one with two Matras, and on other occasions a bi-Matric letter has to be shortened to be one with as single Matra.

In such a situation, in order to keep the real form of the word in view for the readers, one finds both the Matras having been used.

1. The words where in actual practice (ੁ) is used with some character of a word, but to maintain the metric rhythm, it is lengthened to two Matras (ੌ). While reading, one should read (ੌ) instead of (ੁ). See for instance,

ਮੋ ਕਉ ਧਾਰਿ ਕ੍ਰਿਪਾ ਮਿਲੀਐ ਗੁਰ ਦਾਤੇ ਹਰਿ ਨਾਨਕ ਭਗਤਿ ਉਮਾਹਾ ਰਾਮ ॥੪॥੨॥੮॥ ਜੈਤਸਰੀ ਮਃ ੪, ਪੰ: ੬੯੯

Original word 'ਉਮਾਹਾ' changed to 'ਉਮਾਹਾ'. Here with the letter 'ਓ' we have both 'ੁ' and 'ੌ'.

Again:

ਸੰਤਾ ਕਉ ਮਤਿ ਕੋਈ ਨਿੰਦਹੁ ਸੰਤ ਰਾਮੁ ਹੈ ਏਕੋ ॥
ਕਹੁ ਕਬੀਰ ਮੈ ਸੋ ਗੁਰ ਪਾਇਆ ਜਾ ਕਾ ਨਾਉ ਬਿਬੇਕੋ ॥੪॥੫॥ ਸੂਹੀ ਕਬੀਰ ਜੀ, ਪੰ: ੭੯੩

From 'ਏਕੁ' we have 'ਏਕੋ' and from 'ਬਿਬੇਕੁ' we have 'ਬਿਬੇਕੋ'

Other examples are:

ਮੁੱਲਿ ਅਮੁੱਲੁ ਨ ਪਾਈਐ ਵਣਜਿ ਨ ਲੀਜੈ ਹਾਟਿ ॥੧॥ ਮਾਰੂ ਕੀ ਵਾਰ ਮਃ ੩ ਪੰ: ੧੦੮੭
ਸੁਖ ਸਾਗਰੋ ਪਾਇਆ ਸਹਜ ਸੁਭਾਇਆ ਜਨਮ ਮਰਨ ਦੁਖ ਹਾਰੇ ॥੪॥੧॥ ਕੇਦਾਰਾ ਮਃ ੫ ਛੰਤ, ਪੰ: ੧੧੨੩
ਮੈ ਅਨਦਿਨੋ ਸਦ ਸਦਾ ਹਰਿ ਜਪਿਆ ਹਰਿ ਨਾਉ ॥੨॥੩॥੧੦॥ ਸਾਰੰਗ ਮਃ ੪, ਪੰ: ੧੨੦੧

2. The words that originally are of the form with 'ੌ' but to maintain the metric rhythm, the bi-Matric nature of 'ੌ' is changed by attaching 'ੁ' to the same character. While reading these, they should be read as if these words are ending in the vowel sign 'ੁ'. For example:

ਪ੍ਰਭ ਕਿਰਪਾਲ ਦਇਆਲ ਗੁੰਬਿੰਦ ॥੧॥੧੫॥ ਗੌਡ ਮਃ ੫, ਪੰ: ੮੬੬
ਦਇਆ ਧਾਰੀ ਗੋਵਿੰਦ ਗੁੱਸਾਈ ॥੪॥੨੦॥ ਰਾਮਕਲੀ ਮਃ ੫ ਪੰ: ੮੬੧
ਉਨਿਮੋ ਭਗਵੰਤ ਗੁੱਸਾਈ ॥੪੫॥ ਰਾਮਕਲੀ ਮਃ ੪ ਪੰ: ੮੬੭

ਗੁਰਮੁਖਿ ਤੋਲਿ ਤੁੱਲਾਵੈ ਤੋਲੈ ॥੨੨॥ ਓਅੰਕਾਰੁ ਪੰ: ੯੩੨
ਜਿਹ ਠਾਕੁਰੁ ਸੁਪ੍ਰਸੰਨੁ ਭਯੋ ਸਤ ਸੰਗਤਿ ਤਿਹ ਪਿਆਰੁ ॥੬॥ ਸਵਯੇ ਸ੍ਰੀ ਮੁਖਵਾਕ ਮਃ ੫ ਪੰ: ੧੩੮੬
ਸਾਈ ਸੁੱਗਾਗਣਿ ਸਾਈ ਭਾਗਣਿ ਜੈ ਪਿਰਿ ਕਿਰਪਾ ਧਾਰੀ ॥੧॥੬॥ ਰਾਮਕਲੀ ਵਾਰ ਮਃ ੫ ਪੰ: ੯੫੯
ਗੋਪੀ ਕਾਨੁ ਨ ਗਊ ਗੁੱਆਲਾ ॥੭॥੧੫॥ ਮਾਰੂ ਮਃ ੧ ਸੋਹਲੇ ਪੰ: ੧੦੩੫

F2. Proposal Summary

ISO/IEC JTC 1/SC 2/WG 2

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646³

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title: Proposal to add two Gurmukhi characters to the BMP of the UCS

2. Requester's name: Sukhjinder Sidhu

3. Requester type (Member body/Liaison/Individual contribution): Individual contribution

4. Submission date: 2005-04-21

5. Requester's reference (if applicable): _____

6. Choose one of the following:

This is a complete proposal: Yes

or, More information will be provided later: No

B. Technical - General

1. Choose one of the following:

a. This proposal is for a new script (set of characters): No

Proposed name of script: _____

b. The proposal is for addition of character(s) to an existing block: Yes

Name of the existing block: Gurmukhi

2. Number of characters in proposal: _____ 2 _____

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary _____ B.1-Specialized (small collection) _____ B.2-Specialized (large collection) _____

C-Major extinct D-Attested extinct _____ E-Minor extinct _____

F-Archaic Hieroglyphic or Ideographic _____ G-Obscure or questionable usage symbols _____

4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document): _____ 3 _____

Is a rationale provided for the choice? _____

If Yes, reference: _____ Combining Characters _____

5. Is a repertoire including character names provided? _____

a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? _____ Yes _____

b. Are the character shapes attached in a legible form suitable for review? _____ Yes _____

6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? _____ Bhupinder Singh, Kulbir Thind. TrueType _____

If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: _____

7. References:

a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? _____ Yes _____

b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? _____ Yes _____

8. Special encoding issues:

Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

_____ Yes, see above _____

9. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

³ Form number: N2652-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain _____	_____No_____
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? Jeevan Deol, Serjinder Singh, Kulbir Thind and others. Unicode Indic mailing list. If YES, available relevant documents: _____	_____Yes_____
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference: ___Restricted to ancient poetic texts, including religious texts. See above._____	_____Yes_____
4. The context of use for the proposed characters (type of use; common or rare) Reference: ___See above_____	_____Rare_____
5. Are the proposed characters in current use by the user community? If YES, where? Reference: _____	_____No_____
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? _____ If YES, reference: ___Gurmukhi block is entirely in BMP_____	_____Yes_____ _____Yes_____
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	_____
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? _____ If YES, reference: _____	_____No_____
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? _____ If YES, reference: _____	_____Yes_____ _____Yes_____ _____See Above_____
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? _____ If YES, reference: See above. Visually similar to existing characters but has different properties.	_____Yes_____ _____Yes_____ _____Yes_____
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? _____ If YES, reference: ___See Above_____	_____Yes_____ _____Yes_____
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? _____ If YES, reference: ___See Above_____	_____Yes_____
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary) _____	_____No_____
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? _____ If YES, reference: _____	_____No_____