

Proposed Changes to Gurmukhi

Manvir Singh, Irvanjit Singh

July 1, 2020

Introduction

This document suggests the follow two changes to the Unicode Gurmukhi script:

1. Support the rendering of subscript Gurmukhi numerals ੦ through ੯
2. Rendering change for 0965 (DEVNAGRI DOUBLE DANDA)

Furthermore, the issue of alternative forms YAYYA (ਯਯ) and the existing HALF YAYYA is discussed.

Abbreviations

SGGS - Siri Guru Granth Sahib (the primary Sikh holy scripture), contains writings (Gurbani) of several Sikh Gurus and various other Saints

SDGS - Sri Dasam Granth Sahib, contains the writings (Gurbani) of the tenth Sikh Guru, Guru Gobind Singh

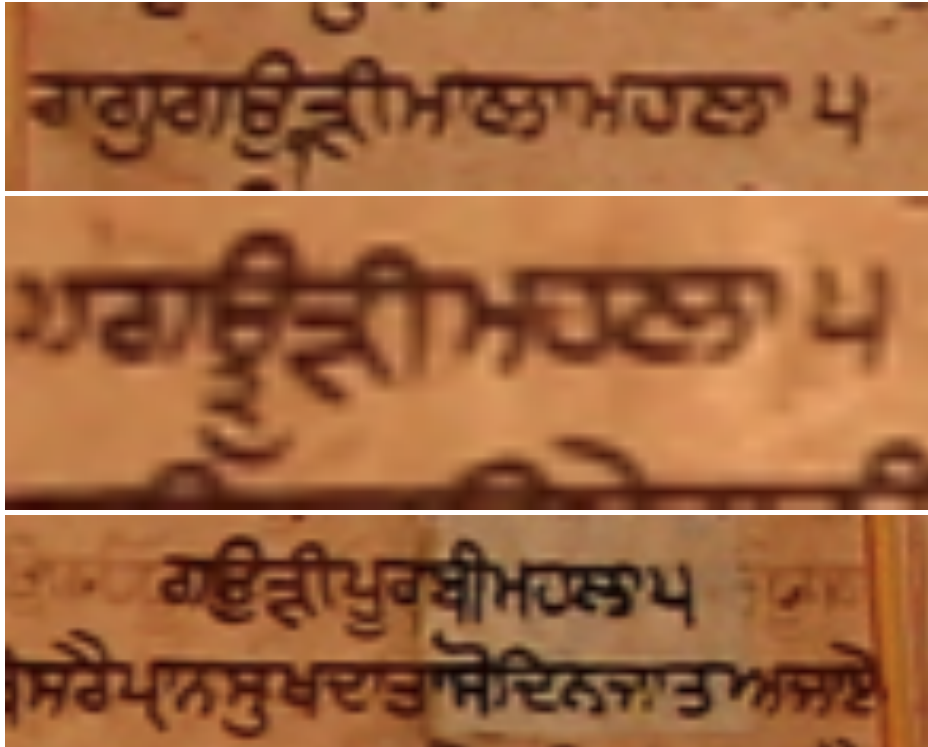
SSGS - Sri Sarbloh Granth Sahib, also contains the writings (Gurbani) of the tenth Sikh Guru, Guru Gobind Singh

1 - Support the rendering of subscripted Gurmukhi numerals ੦ through ੯

Historical handwritten manuscripts and modern printed versions of SGGS both include the use of subscript numerals. These are used under various headings for musical measure or author and indicate meta information pertaining to the heading. The meaning of the number is contextual to the specific part of the text that it is under. For example, if the subscript is

found underneath the name of a musical measure, it may indicate a specific subclassification of that measure. It may also indicate the Ghar (ਘਰ).

Attestations



(the number last picture here is slightly cut off because of the tape/restoration done on the right side)

ਗਗੁ ਗਉੜੀ ਚੇੜੀ ਮਹਲਾ ੫

ਗਗੁ ਗਉੜੀ ਪੂਰਬੀ ਮਹਲਾ ੫

Implementation

A good solution for subscripts would allow for multiple digits to be represented, for the character to be renderable as a zero width character (so that it can be placed underneath other character), and to be renderable independently of any base character (similar to how superscripts are used in citations), and to also support superscripts (used in various Punjabi texts)

One option would be to encode all of the numbers as separate code points.

Another would be to use some combination of variation selector(s) to render subscript or superscript versions of each number, and also allow for multiple digits to be rendered. Furthermore, this could be extended to provide a general-case solution for some of the issues raised in [L2/13-227](#), in terms of being able to make the letters used for musical notations into superscripts.

2 - Rendering change for 0965 (DEVNAGRI DOUBLE DANDA)

There is currently a significant inaccuracy in the way that the DOUBLE DANDA character renders in Gurmukhi text where multiple words are separated without whitespace (or by ZWSP).

Historical handwritten manuscripts of all Sikh scriptures (SGGS, SDGS, SSGS) as well as many other Sikh manuscripts all feature a style of writing without spaces, similar to the Latin scriptio continua. This style of writing in the context of Gurmukhi is called “larivaar”. Whitespace is only used to separate a text heading from its body. Sentences (or verses) are separated by the DOUBLE DANDA, and the onus of parsing the verse into separate words is left to the reader. In all historical larivaar manuscripts, the first vertical line of the DOUBLE DANDA is attached to the last letter of the current verse. If the verse is followed by another verse within the same text body, the second vertical line is attached to the first letter of the proceeding verse. If there is no proceeding verse, then a standalone vertical line is drawn. Thus, the DOUBLE DANDA would effectively be rendered as a single DANDA attached to the beginning and end of adjacent verses. This method of rendering the DOUBLE DANDA disappeared during the advent of lithography and the printing presses in India and is one of the many examples of western/British influence on the Gurmukhi script.

Attestations

Examples are provided of both handwritten and lithographed manuscripts (in which the DOUBLE DANDA was changed to how it is currently rendered in modern fonts)

SGGS:

ੴ ਸਤਿਨਾਮੁ ਕਰਤਾ ਪੁਰਖੁ ਨਿਰਭਉ ਨਿਰਵੈਰੁ
 ਅਕਾਲ ਮੂਰਤਿ ਅਜੂਨੀ ਸੈਭੰ ਗੁਰ ਪ੍ਰਸਾਦਿ ॥ ਜਪੁ ॥
 ਆਦਿ ਸਚੁ ਜੁਗਾਦਿ ਸਚੁ ॥ ਹੋਭੀ ਸਚੁ ਨਾਨਕ ਹੋਸੀ ਭੀ ਸ
 ਚੁ ॥ ੧ ॥ ਸੋਚੈ ਸੋਚਿ ਨ ਹੋਵਈ ਜੇ ਸੋਚੀ ਲਖ ਵਾਰ ॥ ਚੁਪੈ ਚੁਪਿ ਨ ਹੋਵਈ ਜੇ ਲਾਇ ਰਹਾ ਲਿਵਤਾਰ

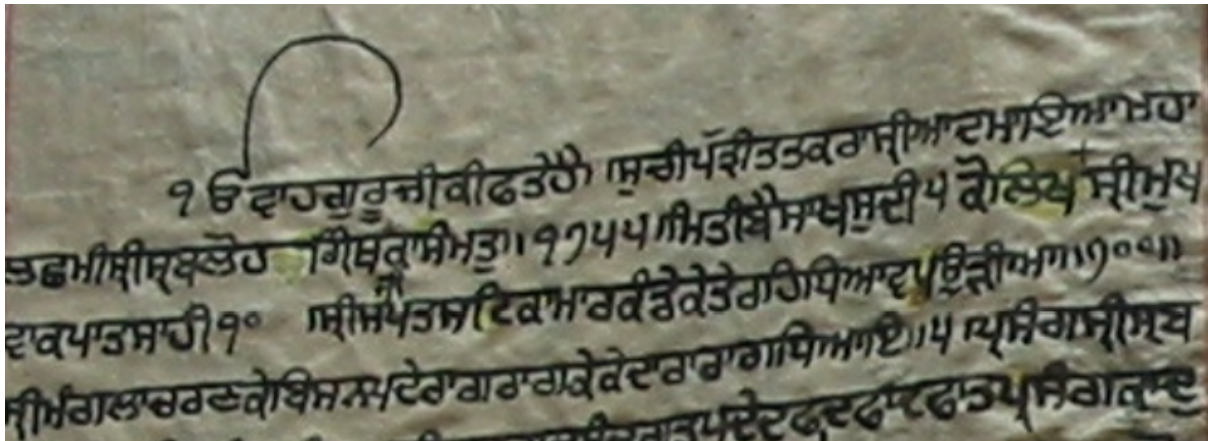
ੴ ਸਤਿਨਾਮੁ ਕਰਤਾ ਪੁਰਖੁ ਨਿਰਭਉ ਨਿਰਵੈਰੁ
 ਅਕਾਲ ਮੂਰਤਿ ਅਜੂਨੀ ਸੈਭੰ ਗੁਰ ਪ੍ਰਸਾਦਿ ॥
 ਜਪੁ ॥

ਆਦਿ ਸਚੁ ਜੁਗਾਦਿ ਸਚੁ ॥ ਹੋਭੀ ਸਚੁ ਨਾਨਕ ਹੋਸੀ ਭੀ ਸਚੁ ॥ ੧ ॥ ਸੋਚੈ ਸੋਚਿ
 ਨ ਹੋਵਈ ਜੇ ਸੋਚੀ ਲਖ ਵਾਰ ॥ ਚੁਪੈ ਚੁਪਿ ਨ ਹੋਵਈ ਜੇ ਲਾਇ ਰਹਾ ਲਿਵਤਾਰ
 ॥ ਭਖਿਆ ਭਖਨ ਓਤਰੀ ਜੇ ਬੰਨਾ ਪਰੀਆ ਭਾਰ ॥ ਸਹਸ ਸਿਆਣਪਾਲਖ

SDGS:

ੴ ਸਤਿਗੁਰ ਪ੍ਰਸਾਦਿ ॥ ਜਪੁ ॥ ॥ ਸ੍ਰੀ ਮੁਖ ਵਾਕ ਪਾਤਿਸਾਹੀ ॥ ੧੦ ॥
 ਛਪੇ ਛੰਦਾ ॥ ॥ ਗੁਰ ਪ੍ਰਸਾਦਿ ॥ ॥ ਚਕ੍ਰ ਚਿਹਨ ਅਰੁ ਬਰਨ ਜਾਤ ਅਰੁ ਪਾਤਿਸਾਹੀ ਨਿਯਮਿ ॥
 ਗੁਪਤਿ ਗੁਪਤਿ ਅਰੁ ਰੇਖ ਭੇਖ ਕੋਠਿ ਕਹਿਨ ਸਕਤ ਕਿਹਾ ॥ ਅਚਲ ਮੂਰਤਿ ਅਨਭਉ ਪ੍ਰਕਾਸ ਅਮਿ
 ਤੇ ਜਿ ਕਹੈ ॥ ਕੋਇ ਇੰਦ੍ਰ ਇੰਦ੍ਰਾਣਿ ਸਾਹਸਾ ਚਾਇ ਗਇਐ ॥ ਤ੍ਰਿਭੁਵਣ ਮਹੀ ਪਸੁਰਨਰ
 ਅਸੁਰ ਨੇਤਨੇ ਤਬ ਨ ਰਿਣ ਕਹਤਾ ॥ ਸੂਰਬਨਾਮ ਕਥੈ ਕਵਨ ਕਰਮ ਨਾਮ ਬਰਨਤ ਸੁਮਤਾ ॥
 ਭੁਜੰਗ ਪ੍ਰਯਾਤ ਛੰਦਾ ॥ ਨਾਮ ਸਤ੍ਰ ਅਕਾਲੇ ॥ ਨਾਮ ਸਤ੍ਰ ਕ੍ਰਿਪਾਲੇ ॥ ਨਾਮ ਸਤ੍ਰ ਅਰੂਪੇ ॥ ਨ
 ਆਦਿ ਅਨਪੇ ॥ ੨ ॥ ਨਾਮ ਸਤ੍ਰ ਅਭੇਖੇ ॥ ਨਾਮ ਸਤ੍ਰ ਅਲੇਖੇ ॥ ਨਾਮ ਸਤ੍ਰ ਅਕਾਏ ॥ ਨਾਮ ਸਤ੍ਰ ਅ

SSGS:



Implementation

Although this is a font issue, it needs to be addressed in the Unicode standard as well as the default unicode fonts since it is a fundamental change. Every other vowel and consonant letter in the Gurmukhi block behaves as expected in terms of rendering with all letters adjoined with a line on top. If two words are separated by whitespace, they render as such (ex. ਮੇਰੇ ਮਨ) and without whitespace (ਮੇਰੇਮਨ). The ਚ and ਮ are neatly joined together. The DOUBLE DANDA should behave in the same way when rendered with Gurmukhi characters. This is an important change to make since many Sikhs who are studying the Sikh scriptures are learning to read it in the larivaar form, and future generations will become more and more accustomed to seeing the DOUBLE DANDA separated, which is a departure from the written tradition. Furthermore, unlike other stylistic departures of traditional Gurmukhi, the solution for this issue might require more nuance since the rendering behaviour of the DOUBLE DANDA would change depending on whether it is followed by whitespace on both sides, characters on both sides, or characters on one side and whitespace on the other.

One solution would be to use the reserved code points 0A64 and 0A65 to encode new characters for DOUBLE DANDA that are independent of Devnagri.

Alternatively, a note could be added in the standard so as to support the various rendering edge-cases.

The edge cases are as follows:

1. If the double danda is placed immediately before any of the base characters of the Gurmukhi alphabet, the right-side vertical line should attach to the base character
2. If the double danda is placed immediately after any of the base characters of the Gurmukhi alphabet, the left-side vertical line should attach to the base character
3. If the double danda is placed immediately before whitespace or any character apart from the base characters of the Gurmukhi alphabet, the right-side line should be a straight vertical line.
4. If the double danda is placed immediately after whitespace or any character apart from the base characters of the Gurmukhi alphabet, the left-side line should be a straight vertical line.

Issues regarding YAYYA (ਯ) and HALF YAYYA (ਯ਼)

The Gurmukhi script contains a total of five variations of the character OA2F ਯ GURMUKHI LETTER YA. They are as follows:

- 1) ਯ (encoded as OA2F ਯ GURMUKHI LETTER YA)
- 2) ਯ਼ (encoded as OA75 ਯ਼ GURMUKHI SIGN YAKASH)
- 3) ਯ (rendered using ਯ + ਯ਼ + ਯ = ਯ਼ HALF YAYYA)
- 4) ਯ (not supported in unicode, YAYYA OPEN TOP)
- 5) ਯ (not supported in unicode, HALF YAYYA OPEN TOP)

Of these variations, 1) and 2) are found in SGGS, while all five appear in SDGS and SSGS. There is also debate as to whether 3) 4) and 5) are different stylistic variations of the same character, or unique characters. Reviews of handwritten manuscripts have been done in the past and are still being done to conclusively determine whether these characters are unique. Some prominent Sikh scholars have published compilations of Gurbani, including SDGS, with all five of these alternatives have been used. Therefore, many Sikhs hold the view that in order to faithfully reproduce this published material digitally in unicode, there needs to be support for

these alternative HALF YAYYAs, whether that is through fonts or an encoded solution.

Unlike other half-letters in Gurmukhi, the HALF-YAYYA can be written with multiple vowels. This can be problematic, and the character sequence is not intuitive.

Ex: ਕਿਯੋ is rendered as (ਕ + ੍ + ਯ + ਿ + ੋ) even though the ੋ is applied to the ੍ਯ and the ਿ is applied to the ਕ.

Furthermore, some combinations cannot be rendered.

Ex. ਕੀਯੋ

Since the HALF YAYYA does not behave like other half characters in Gurmukhi, it would make sense to encode it separately. This would resolve the issues mentioned above. Furthermore, it provides application developers a mechanism for rendering the alternative versions using fonts. If a developer wants to render all five versions, then they can create custom implementations in their fonts (i.e. using ZWJs other other characters) to support this, while developers who would like to only use HALF YAYYA in one form can continue to do so.

Attestations

Some modern publications of SDGS includes all of the YAYYA characters, as shown below:

ਨ ਕੋਊ ॥ ਯੋਂ ਉਪਮਾ ਉਪਜੀ ਜੀਯ ਮੈ ਜਿਵ ਸੇਵਤ ਸਿੰਘ ਜਗਾਵਤ ਕੋਊ ॥੧੦੯੧॥ ਸ੍ਰੈਯਾ ॥ ਸਯਾਮ ਕੇ ਬਾਨ
ਲਗਯੋ ਉਰ ਮੈ ਗਡ ਕੈ ਸੋਊ ਪੰਖਨ ਲਉ ਸੁ ਗਯੋ ਹੈ ॥ ਸ੍ਰਉਨ ਕੇ ਸੰਗ ਭਰਯੋ ਸਰ ਅੰਗ ਬਿਲੋਕ ਤਬੈ ਹਰਿ ਕੋਪ
ਭਯੋ ਹੈ ॥ ਤਾ ਛਬਿ ਕੋ ਜਸੁ ਉੱਚ ਮਹਾਂ ਕਬਿ ਨੇ ਕਹਿ ਕੈ ਇਹ ਭਾਂਤਿ ਦਯੋ ਹੈ ॥ ਮਾਨਹੁ ਤੱਛਕ ਕੋ ਲਰਿਕਾ ਖਗਰਾਜ
ਲਖਯੋ ਗਹਿ ਲੀਲ ਲਯੋ ਹੈ ॥੧੦੯੨॥ ਸ੍ਰੈਯਾ ॥ ਸ੍ਰੀ ਬ੍ਰਿਜਨਾਥ ਸਰਾਸਨ ਲੈ ਰਿਸ ਕੈ ਸਰੁ ਰਾਜਨ ਬੀਚ ਕੱਸਾ ॥

They are also being used in the Nihung Santhia app published by the Nihung Santhia organization. The contents of the app was sourced from a compilations of prayers that was a result of a meticulous review process during which several authenticated, early, handwritten manuscripts of SDGS were reviewed.