

# Proposal for a Gurmukhi Script Root Zone Label Generation Ruleset (LGR)

---

*LGR Version:* 3.0

*Date:* 2018-07-27

*Document version:* 2.4

*Authors:* Neo-Brahmi Generation Panel [NBGP]

## 1. General Information/ Overview/ Abstract

This document lays down the Label Generation Ruleset for Gurmukhi script. Three main components of the Gurmukhi Script LGR i.e. Code point repertoire, Variants and Whole Label Evaluation Rules have been described in detail here. All these components have been incorporated in a machine-readable format in the accompanying XML file named "Proposal-LGR-Guru-20180727.xml".

In addition, a document named "Gurmukhi-Test-Labels-20180727.txt" has been provided. It provides a list of labels which can produce variants as laid down in Section 6 of this document and it also provides valid and invalid labels as per the Whole Label Evaluation laid down in Section 7.

## 2. Script for which the LGR is proposed

ISO 15924 Code: Guru

ISO 15924 Key N°: 310

ISO 15924 English Name: Gurmukhi

Latin transliteration of native script name: gurmukhī

Native name of the script: ਗੁਰਮੁਖੀ

Maximal Starting Repertoire [MSR] version: 3

### 3. Background on Script and Principal Languages Using It

It is commonly accepted that Gurmukhi script is a member of the BRAHMI family. Brahmi is an Indic script that was developed in the Indian subcontinent and adapted to the local needs. According to an opinion, the Brahmi script was introduced between the 8th and the 6th century BC. It does not concern us here whether the script was foreign or local, but it has now been established, on the basis of its name, that the Indians did have a system of writing which must have been borrowed freely from local script.

Alternatively, when the Indo-Bactrians established themselves in the Gandhara region of Western Indian subcontinent, the Kharoshthi script contemporary with Brahmi was developed and used in the Punjab, Gandhara and Sindh between 300 BC and 3rd century AD. But even then, Brahmi which in its development in the Punjab had undergone several changes, was commonly used along with Kharoshthi. There are coins of the Bactrian kings and inscriptions of the Kushan rulers having both scripts on them. Brahmi was, of course, more popular on account of its simple curves alternating with straight strokes. Hence, in due course, it replaced Kharoshthi and became the single script with composite features brought about by various local and neighborly influences. With the growth of literary and cultural activity during the Gupta period (4th-5th century AD) the Brahmi script improved further and became more expansive and common.

Immediately later, it developed, especially in Northern India, fine curves and embellished flourishes with a small headline over each letter and became rather ornamental. This stage of Indian script was called Kutil (meaning curved) which had the widest use in Northern India. With the rise of regional languages taking the place of Sanskrit and Prakrit, regional scripts grew in number. Ardhanagari (west), Sharda (Kashmir) and Nagari (beyond Delhi) came into use and later both Sharda and Devanagari started their inroads into the land of the five rivers. This is evident from the coins of the Ghaznavids and Ghorids minted at Lahore and Delhi. It is also known that the common (non-Brahman and non-official) people used a number of scripts for their temporal and commercial requirements. Of these scripts, Landa and Takri characters were most prevalent.

### 3.1. The Evolution of the Script

Like most of the North Indian writing systems, the Gurmukhi script is a descendant of the Brahmi script. The Proto-Gurmukhi letters evolved through the Gupta script from 4th to 8th century, followed by the Sharda script from 8th century onwards and finally adapted their archaic form in the Devasesha stage of the later Sharda script, dated between the 10th and 14th centuries.

The development of Gurmukhi script is indicated by the following diagram (where “-Kalin” means “in the time of”).

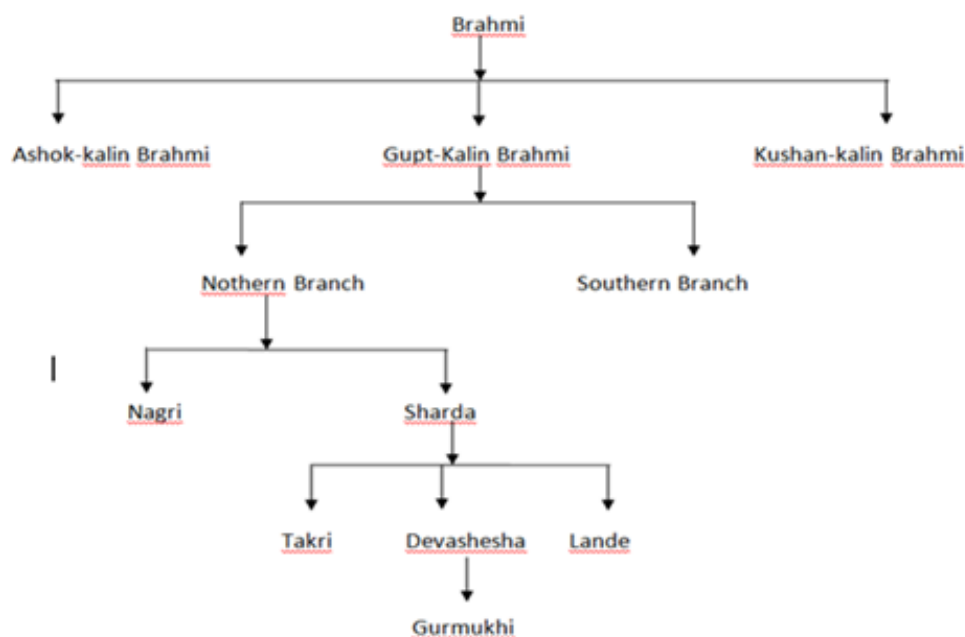


Figure 1: Evolution of Gurmukhi

The traditional accounts such as the references found in Janamsakhi literature say that the Gurmukhi script was invented by the second Sikh Guru, Guru Angad Dev. However, it would be correct to say that the script was standardized rather than invented. Newton [100] writes that at least 21 Gurmukhi characters are found in ancient manuscripts: 6 from 10th century, 12 from 3rd century B.C. Apparently, the first Sikh Guru, Guru Nanak Dev also used the Gurmukhi script for his writings. Scholars have tried to establish the relationship of Gurmukhi script with different scripts, such as Devanagari [101], Ardhanagari [102], Siddham/Siddhamatrika [103], Sharda [104] and Brahmi (generally).

Some ascribe it to Landa [105] [106] and others to Takri, a branch of Sharda used in the Chamba area. The fact is that it is derived from, or at least allied to, all these and others mentioned above in their historical perspective. Regionally and contemporarily compared, Gurmukhi characters have direct similarities with Gujarati, Landa, Nagari, Sharda, and Takri: they are either exactly the same or essentially alike. Internally, A (ਅ), HA (ਹ), CA (ਚ), DA (ਦ), NNA (ਣ), LA (ਲ) letters of Gurmukhi had undergone some minor orthographical changes before 1610 A.D. A major change occurred in NGA (ਙ) and NYA (ਞ) letters. BA (ਬ) letter was invented later. Further changes came in the forms of A (ਅ), HA (ਹ) and LA (ਲ) letters in the first half of the nineteenth century.

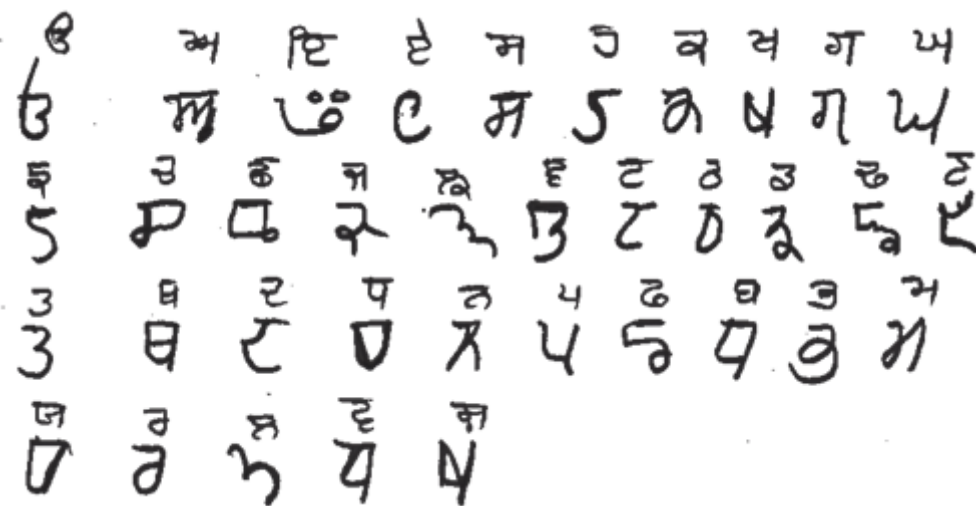


Figure 2: Pictorial depiction of Proto-Gurmukhi (13th century) with current glyphs displayed above each character

Another reform carried out is the separation of lexical units of the sentence which previously formed one jumbled unit; lately punctuation marks borrowed from English have been incorporated besides the full stop. In place of the full stop, dandi has been used which existed traditionally.

There are two major theories on how the Proto-Gurmukhi script emerged in the fifteenth century. Singh [102] while quoting Al-Biruni's Al-Hind [107], says that the script evolved from Ardhanagari. Al-Biruni writes that the Ardhanagari script was used in Bathinda (more exactly, Bhātiya), including Sindh and western parts of the Punjab in the tenth century. For some time Bathinda remained the capital of the kingdom of Bhatti Rajputs,

who ruled North India before the Muslims occupied the country. Resulting from its connection with the Bhattis, the Ardhanagari script was also called Bhatachhari. According to Al-Biruni [107] Ardhanagari was a mixture of Nagari, used in Ujjain and Malwa, and Siddhamatrika is a variant of the Sharda script used in Kashmir. Singh [103] also traced the origins of Gurmukhi to the Siddhamatrika.

Bedi [108] has suggested that the Gurmukhi script was developed during the tenth-fourteenth centuries from the Devasesha stage of the Sharda script. His argument suggests that regional differences started to appear in the Sharda script used in Punjab, partly Himachal Pradesh and Kashmir from the tenth century. The regional form of Sharda, used in Punjab, started to appear in the form of Gurmukhi in the 14th century. Bedi [108] called this stage Pritham Gurmukhi or Proto-Gurmukhi.

The Sikh Gurus adopted the proto-Gurmukhi script to write the Guru Granth Sahib, the primary religious scripture of the Sikhs. Other contemporary scripts used in Punjab were Takri and the Landa alphabets. Takri was a script that developed through the Devasesha stage of the Sharda script and is found mainly in the Hill States such as Chamba, where it is called Chambyali, and in Jammu, where it is called Dogri. The local Takri variants attained the status of official scripts in some of the Punjab, Hills States and were used for both administrative and literary purposes. When Himachal Pradesh was established, the local Takri was replaced by Devanagari.

Meanwhile, the mercantile scripts of Punjab known as Landa were normally not used for literary purposes. In Punjab, there were at least ten different variants of Landa. Landa alphabets were used for household and trade purposes.

The letters no doubt existed before the period of the Guru. But Sikh Gurus not only modified and re-arranged certain letters but also shaped them into a script. They gave new shape and new order to the alphabet and made it precise and accurate. They fixed one letter for each of Punjabi phonemes; use of vowel-symbols was made obligatory; the letters used to construct conjuncts were not adopted; and only those letters were retained which depicted sounds of the then spoken language. There was some re-

arrangement of the letters also in alphabetical order: e.g., SA ( ਸ ) and HA ( ਹ ) were shifted to the first line and URA ( ਊ ) was given the first place in the new alphabet.

### **“Gurmukhi” Etymology**

The word Gurmukhi is the compound form of Guru and Mukh. It is commonly translated as from the mouth of Guru. However, the term is used for the Punjabi script and has somewhat different connotations. The opinion given by traditional scholars is that as the Sikh holy writings, before they were scribed, were uttered by the Gurus, they came to be known as Gurmukhi (the Utterance of the Guru) and consequently, the script that was used for the scribing the utterance was also given the same name. However the prevalent view among Punjabi linguists is that as in the early stages, the Gurmukhi letters were primarily used by Gurmukhs (those living a life as instructed by the Guru) or the Sikhs devoted to the Guru, the script came to be associated with them. The script associated with Gurmukhs came to be known as Gurmukhi.

Now Gurmukhi is the name of the script used in writing primarily for the Punjabi language. It was once used secondarily for the Sindhi language, but is no longer.

## **3.2. Languages considered**

Punjabi (EGIDS 2) is the only language currently using the Gurmukhi script.

## **3.3. The structure of written Gurmukhi**

Punjabi is written using the Gurmukhi script. It is an alphasyllabary with the akshar as its core. All scripts derived from Brahmi are Abugidas, i.e. syllabary driven systems. The main features of Abugidas are:

- The consonant has an implicit /ə/ vowel which is also known as the schwa.
- The inherent vowel can be modified by the addition of other vowels or muted by a diacritic termed as a Virama.
- Vowels can be handled as full vowels with a vocalic value.
- When two or more consonants join together they form ligatures. In Gurmukhi script, ligatures are formed only with following /h, r and v/ consonants.

The writing system of Gurmukhi could be summed up as composed of the following:

### 3.3.1. The Consonants

In Gurmukhi, all consonants contain an implicit vowel schwa /ə/ [109]. In Punjabi, the /ə/ vowel is called mukta. The word mukta is derived from the word *mukt* that means free. So mukta means free from any vowel sign [110]. As an example, the word ਕਰ is made up of three phonemes /k/, /ə/ and /r/, but /ə/ does not appear in the word ਕਰ as it is inherited in the letter ਕ. Hence mukta is, in a sense, “free” from any vowel sign. But Gurumukhi consonants are also used without any modification to represent consonant sounds without following /ə/ vowel. As a result, Gurmukhi script is of semi-syllabic nature, in that a Punjabi consonant letter by itself may stand for a consonant sound as well as for the consonant plus following /ə/ vowel.

Punjabi is a tone language; but each tone is not represented by its own distinct letter or symbols in the Gurmukhi script. Nevertheless, in Punjabi the same sequence of vowel and consonant segments can represent different words depending on the pitch of voice or tone used in pronouncing it.

In the traditional classification, consonants are categorized according to their phonetic properties; there are 7 groups (vargas) representing points of articulation, and one non-varga group, which comes last in display. *Varga* in general means a category of consonants that are all pronounced at the same point of articulation. However, the first so-called varga group in the Gurmukhi alphabet actually consists of three vowel carriers, as well as two consonants. In this first group, both the consonants represent fricatives, one dental and another glottal. The next five groups each lay out the stops and nasal of the varga systematically, each displaying five consonants classified as per their manner of articulation. In each varga, the first four consonants are classified on the basis of Voicing and Aspiration, and the last consonant is the corresponding Nasal.

As a final complication, the fourth consonant in each of these five vargas is traditionally classified (following its historic use) as a voiced aspirated consonant; but it is in fact used to denote tone.

Punjabi does not now contain voiced aspirated consonants [111]. Instead, the pronunciation of these five, once voiced aspirated, consonants corresponds to tonally marked syllables. When any of these letters comes in initial position it is to be pronounced as a unvoiced unaspirated consonant of that varga with a low tone [112]; in middle position it is to be pronounced as a voiced unaspirated consonant of that varga with a high or low tone, depending on the length of the preceding or the following vowel; at the end of a word, it is to be pronounced as a voiced unaspirated consonant of that varga with high tone. So these letters can be pronounced only in two tones, either a high tone or a low tone.

After the varga groups, the next five consonants do not have a single point and manner of articulation. So they do not correspond to a single varga. They are categorized as a non-varga group. The last group has six letters. All the letters in this group have a bindi (dot) placed in their foot. So they are categorized as *pairin bindi* letters, meaning “having dot in the foot”.

Varga	Vowel carriers			Fricatives	
	For back vowels: u, ū, o	For low vowels: a, ā	For front vowels: i, ī, e	Dental: [s]	Glottal: [h]
<b>Mul Varga</b>	ੳ U+0A73	ਅ U+0A05	ੲ U+0A72	ਸ U+0A38	ਹ U+0A39

Table 1: Mul varga

Varga	Unvoiced		Voiced		Nasal
	-Asp	+Asp	-Asp	+Asp*	
<b>Velar</b>	ਕ U+0A15 k	ਖ U+0A16 kh	ਗ U+0A17 g	ਘ U+0A18 (gh)	ਙ U+0A19 ṅ



<b>Palatal</b>	ਚ U+0A1A c	ਛ U+0A1B ch	ਜ U+0A1C j	ਝ U+0A1D (jh)	ਞ U+0A1E ñ
<b>Retroflex</b>	ਟ U+0A1F ṭ	ਠ U+0A20 ṭh	ਡ U+0A21 ḍ	ਢ U+0A22 (dh)	ਣ U+0A23 ṇ
<b>Dental</b>	ਤ U+0A24 t	ਥ U+0A25 d	ਦ U+0A26 th	ਧ U+0A27 (dh)	ਨ U+0A28 n
<b>Bi-labial</b>	ਪ U+0A2A p	ਫ U+0A2B ph	ਬ U+0A2C b	ਭ U+0A2D (bh)	ਮ U+0A2E m

Table 2: Varga classification of consonants

\*Traditionally these letters indicate voiced aspirates but in Punjabi they are used to represent low + high tones on adjacent syllables.

<b>Non Varga</b>	ਯ U+0A2F y	ਰ U+0A30 r	ਲ U+0A32 l	ਵ U+0A35 v	ੜ U+0A5C rr
------------------	------------------	------------------	------------------	------------------	-------------------

Table 3: Non-Varga consonants

<b>Pairin Bindi Varga</b>	ਸ਼ U+0A36 ṣ	ਖ਼ U+0A59 x	ਗ਼ U+0A5A γ	ਜ਼ U+0A5B z	ਫ਼ U+0A5E f	ਲ਼ U+0A33 ḷ
---------------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

Table 4: Pairin bindi consonants

### 3.3.2. The Implicit Vowel Killer: Virama

In Gurmukhi and Devanagari, consonants have an implicit schwa /ə/ included in them. In Hindi, a special sign called halant "ँ" (U+094D) is needed to indicate that this implicit vowel is stripped off, so to create conjuncts, halant is used with the consonants in Devanagari. Unlike Devanagari, Gurmukhi consonants are also used to represent consonant sounds where /ə/ is not included in them.

In Gurmukhi, virama "ँ" (U+ 0A4D) is used in place of halant "ँ" (U+094D). In Gurmukhi, virama is not used with any consonant that represents only the consonant sound instead

of consonant plus vowel sound. Therefore, in Punjabi virama is only used to create a conjunct where the letter HA ਹ (U+0A39), RA ਰ (U+0A30) or VA ਵ (U+0A35) is the second element in a conjunct. When /h, r and v/ phonemes occur as the second member of a consonant cluster, the virama joins these consonants in the foot of their preceding consonants and creates a conjunct. In these consonant clusters, HA (ਹ), RA (ਰ) and VA (ਵ) letters change their shape to pairin haha (ੜ), pairin rara (ੜ) and pairin vava (ੜ). In practice, the three letters assume a smaller shape which is subjoined to the preceding consonant. For example, in the word ਸ੍ਰੀ, ਸ and ਰ occurs as consonant conjuncts, wherein ਸ is followed by ੜ, ਰ and ੀ i.e. ਸ + ੜ + ਰ + ੀ => ਸ੍ਰੀ (srī). Similar pattern is followed when RA (ਰ) and VA (ਵ) occur as consonant clusters. By contrast, in the word ਸਰੀ (sarī), ਸ and ਰ do not occur as consonant conjuncts as ਸ is followed by ਐ; they prohibit the formation of consonant conjunct, hence ਰ does not here appear in the foot of ਸ. Therefore, the word ਸਰੀ consists phonetically of ਸ + ਐ + ਰ + ੀ.

The words that show examples of pairin haha (ੜ) and pairin vava (ੜ) are as follows:

In the word ਮਨ੍ਹਾ (manhā), ਮ is followed by ਨ, ੜ, ਹ and ਾ i.e. ਮ + ਨ + ੜ + ਹ + ਾ. Here ਨ and ਹ occur as consonant conjunct. And in the word ਸ੍ਵਰ (svar), ਸ is followed by ੜ, ਵ and ਰ i.e. ਸ + ੜ + ਵ + ਰ. So in this word ਸ and ਵ occur as consonant conjunct.

### 3.3.3. Vowels

Punjabi has ten vowels /ਅ(ə), ਆ(a), ਇ(I), ਈ(i), ਉ(U), ਊ(u), ਏ(e), ਐ(ɛ), ਓ(o) and ਔ(ɔ)/. The vowels are represented by nine matras (vowel signs) + three matra vahaks (vowel carriers). Of these vowels, three /ਅ(ə), ਇ(I), ਉ(U)/ are short vowels and seven (ਆ(a), ਈ(i), ਊ(u), ਏ(e), ਐ(ɛ), ਓ(o) and ਔ(ɔ)/ are long vowels. Separate symbols exist for all vowels, when they occur at the initial position of syllables. To indicate a vowel sound after a consonant other than the implicit /ə/, a vowel sign (Matra) is attached to the consonant. Since the consonant has a built-in schwa, there are equivalent matras for all vowels except the ਅ [113]. Punjabi has ten vowels but it has signs for only nine of them.

The correlation is shown as below:

ਅ	ਆ	ਇ	ਈ	ਉ	ਊ	ਏ	ਐ	ਓ	ਔ
Mukta [i.e. zero] (without any vowel sign) ਐ	ਾ a	ਿ I	ੀ i	ੁ U	ੂ u	ੇ e	ੈ ε	ੋ o	ੌ ੌ

Table 5: Vowels with corresponding Matras

### 3.3.4. Suprasegmental signs; bindi, tippi and addak

Gurmukhi script has three suprasegmental signs; bindi, tippi and addak, which are used with Gurmukhi vowel signs to represent the suprasegmental phonemes' nasality, gemination and stress. These suprasegmental signs are called lagakhars in Punjabi [114]. Every vowel in Punjabi has a nasalized version. Bindi and Tippi are allographic variants of the nasal meaning that in Gurmukhi, both bindi and tippi signs are used to nasalize vowels. Addak is used to represent gemination and stress. The following subsections describe the usage of these signs.

#### 3.3.4.1. The Bindi (◌̣-U+0A02)

The bindi (◌̣) represents a homorganic nasal. Bindi is used with all long vowels/ਆ, ਈ, ਉ, ਏ, ਐ, ਓ, ਔ/ and the short vowel ਉ as in words - ਆਂਚਲ (āñchal), ਜਾਈ (jāīṃ), ਏਂਜਲ (ēñjal), ਐਂਗਲ (aiṅgal), ਓਂਕਾਰ (ōṅkār), ਐਂਕੜ (auṅkar), ਉਂਗਲ(uṅgal), ਊਂਘ (ūṅgh) and with the matras of long vowels/ ਾ, ੀ, ੇ, ੈ, ੋ, ੌ / except the matra ( ੁ ) as in the words - ਹਾਂ(hāṃ), ਟੀਂ (tīṃ), ਪੇਂਟ (paint), ਦੈਂਤ (daint), ਤੋਂ (tōṃ), ਜੋਂ (jaṃ).

#### 3.3.4.2. The Tippi (◌̣-U+0A70)

Tippi (◌̣) is used to nasalize short vowels /ə/ and /I/ at all places and /U and u/ after a consonant. So tippi comes with the matras of /ə/ and /I/ i.e. mukta (without any vowel sign) and ਿ with vowel carriers as ਅੰ and ਇੰ as in words ਅੰਗ (aṅg) and ਇੰਡੀਆ (india) and with consonants as ਸੰ and ਸਿੰ as in words ਸੰਦ (sand) and ਸਿੰਘ (siṅgh). Matras of /U and u/ i.e. ( ੁ, ੂ ) after a consonant can be followed by tippi as in words- ਖੁੰਬ (khumb), ਗੁੰਦ (gūnd).

In addition to this, tippi is also used in gemination for nasal consonants ਙ, ਵ, ਨ and ਮ. The rules for placement of bindi and tippi are:

1. The initial forms of *u, uu* vowels i.e. ਉ and ਊ can be followed by bindi as ਉਂ, ਊਂ. In addition to this the forms of *u, uu* vowels after any other vowel i.e. ਉ and ਊ can be followed by bindi as in words ਆਉਂਦਾ (āundā) and ਜਾਊਂ (jāūm).
2. Matras of *U, u* (◌ੁ, ◌ੁ) after a consonant can be followed by tippi – ਖੁੰਬ (khumb), ਗੁੰਦ (gūnd).
3. All other short vowels / matras (mukta, ਿੰ) can be followed by tippi as in words – ਅੰਗ (aṅg), ਇੰਡੀਆ (india), ਸੰਦ (sand), ਚਿੰਤਾ (chintā).
4. All other long vowels/mātrās (ਆ, ਈ, ਏ, ਐ, ਓ, ਔ/ ਾ, ੀ, ੇ, ੈ, ੋ, ੌ) can be followed by bindi as in words – ਆਂਦਰ (āndar), ਸਾਈਂ (sāiṁ), ਜਾਏਂ (jāēṁ), ਐਂਠ (aiṅṭh), ਸਿਓਂਕ (siōṅk), ਐਂਤਰਾ (auntrā)/ਰਾਂ (hāṁ), ਟੀਂ (ṭiṁ), ਪੇਂਟ (paint), ਦੈਂਤ (daint), ਤੋਂ (tōṁ), ਜੋਂ (jaūṁ).

#### 3.3.4.3. The Addak (◌ੳ -U+0A71)

Addak is used to mark the gemination of the following consonant. In Punjabi, addak usually comes with mukta, aunkar (◌ੁ) and sihari (ਿੰ), the vowel signs of /ə, u and i/ short vowels and geminates the consonant which follows it. Actually gemination of consonants occurs only when their preceding vowels are short vowels. For example in ਟੱਪਾ (ṭappā), ਗਿੱਲਾ (gillā) and ਮੁੱਕਾ (mukkā), the geminated /ਪ/, /ਲ/ and /ਕ/ consonants have /ə, I and U/ short vowels as their preceding vowels which are represented by mukta(zero vowel sign), sihari (ਿੰ) and aunkar (◌ੁ)vowel signs. In addition to this, addak is also used to write English source words having English vowel /ɛ/. For example, set, jet and web are written in Gurmukhi as ਸੈੱਟ (set), ਜੈੱਟ (jet) and ਵੈੱਬ (web).

#### **We now look at some of the exceptions.**

Addak does not precede HA (ਹ), NGA (ਙ), NYA (ਵ), NNA (ਣ), RRA (ੜ), KHHA (ਖ), GHHA (ਗ) and LLA (ਲ) letters. In these letters, NGA (ਙ) and NYA (ਵ) are nasal consonants so these are stressed or doubled by the nasal sign tippi. The rest of these letters cannot be pronounced with stress or elongation. So, addak is not used before any of the above mentioned letters. Addak is also not used with the last letter of the word, as there is no

letter after it which has to be geminated. Addak is used with geminated consonants and the sign is placed on the preceding syllable. Addak cannot be used at the beginning of a word.

#### 3.3.4.4. Nukta (◌̣ - U+0A3C)

Termed as *pairin bindi* in Punjabi, nukta is used with the following consonants: ਸ /s/, ਖ /kh/, ਗ /g/, ਜ /j/, ਫ /ph/ and ਲ /l/ to represent the phonemes of words of Sanskrit and Perso-Arabic sources. ਸ਼ /ś/ is used to represent the phoneme of Sanskrit source words. ਲ਼ /ḷ/ is used to represent Punjabi's retroflex /ḷ/ phoneme and ਖ਼ /x/, ਗ਼ /ɣ/, ਜ਼ /z/, ਫ਼ /f/ are used to represent Perso-Arabic sources words.

When pairin bindi is adjoined to SA (ਸ), KHA (ਖ), GA (ਗ), JA (ਜ), PHA (ਫ) and LA (ਲ) letters, these are written as:

ਸ਼(U+0A38+U+0A3C), ਖ਼(U+0A16+U+0A3C), ਗ਼(U+0A17+U+0A3C), ਜ਼(U+0A1C+U+0A3C), ਫ਼(U+0A2B+ U+0A3C), ਲ਼(U+0A32+ U+0A3C)

These letters are called pairin bindi letters. All the letters are combinations of Consonant+Nukta. But in Gurmukhi, these letters can also be written as a single unit as ਸ਼ (U+0A36), ਖ਼ (U+0A59), ਗ਼ (U+0A5A), ਜ਼ (U+0A5B), ਫ਼ (U+0A5E) and ਲ਼ (U+0A33). Thus

ਸ਼ (U+0A36)= ਸ਼(U+0A38+U+0A3C)

ਖ਼ (U+0A59)= ਖ਼(U+0A16+U+0A3C)

ਗ਼ (U+0A5A)= ਗ਼(U+0A17+U+0A3C)

ਜ਼ (U+0A5B)= ਜ਼(U+0A1C+U+0A3C)

ਫ਼ (U+0A5E)= ਫ਼(U+0A2B+ U+0A3C)

ਲ਼ (U+0A33)= ਲ਼(U+0A32+ U+0A3C)

Unlike the combinations, the single-unit cannot be part of an IDN. See Section 4.1.1. (Item ii).

#### 3.3.4.5. Visarga (◌̣: U+0A03)

The visarga is used in Sanskrit. It is rarely found in old Punjabi writings as “Sri Guru Granth Sahib” or “Mahan Kosh” where it acts like a Sanskrit visarga where a voiceless 'h' sound is pronounced after the vowel. But its use is not common now, and seems to be used in Punjabi only to mark abbreviations.

### 3.3.5. Zero Width Non-joiner (U+200C) and Zero Width Joiner (U+200D)

The Zero Width Non-joiner (ZWNJ) is an invisible character used in certain cases (after virama) where default conjunct formation is to be explicitly restricted and the virama joining the two consonants participating in the conjunct formation needs to be explicitly shown. However, ZWJ and ZWNJ are not used in modern Gurmukhi as virama is only used to create a conjunct with the letters HA ਯ (U+0A39), RA ਰ (U+0A30) or VA ਵ (U+0A35). So there are not many conjunct combinations in Gurmukhi and also Virama is not explicitly shown in modern Gurmukhi.

One of the usage of the ZWNJ and ZWJ has been for encoding in Unicode the Gurmukhi text from holy scriptures. Some of the character combinations, such as using two vowel signs with a single consonant or some vowel and vowel sign combinations which are not used in modern Gurmukhi but present in older text are encoded using ZWJ and ZWNJ. But they not used in modern Gurmukhi.

Excluding ZWJ and ZWNJ does not affect the usage of Gurmukhi Script in modern Gurmukhi, therefore it has no affect the usage of Gurmukhi Script in the domain name system.

## 4. Overall Development Process and Methodology

Under the Neo-Brahmi Generation Panel, there are many different scripts belonging to separate Unicode blocks. Each of these scripts will be assigned a separate LGR; however Neo-Brahmi GP will ensure that the fundamental philosophy behind building those LGRs are all in sync with all other Brahmi-derived scripts. This is the Gurmukhi LGR, which caters to the Punjabi language written using the Gurmukhi script.

### 4.1. Guiding Principles

#### 4.1.1. External Limits on Scope:

The code point repertoire for the root zone being a very special case, at the top of protocol hierarchies, the set of characters available for selection as a part of the Root Zone code

point repertoire is already constrained by various protocol layers beneath it. The following three main protocols/standards act as successive filters:

*i. The Unicode Chart:*

Out of all the characters that are needed by the given script, if the character in question is not encoded in Unicode, it cannot be incorporated in the code point repertoire. Such cases are quite rare, given the elaborate and exhaustive efforts at character inclusion made by Unicode consortium.

*ii. IDNA Protocol:*

Unicode being the character encoding standard for providing the maximum possible representation of a given script/language, it has encoded as far as possible all the possible characters needed by the script. However the domain name being a specialized case, it is governed by an additional protocol known as IDNA (Internationalized Domain Names in Applications). The IDNA protocol excludes some characters in the Unicode repertoire from being part of domain names.

For example: the Gurmukhi letters ਞ (U+0A36), ਝ (U+0A59), ਞ (U+0A5A), ਜ਼ (U+0A5B), ਢ (U+0A5E), ਝ (U+0A33) are not allowed to be a part of domain name. But their decomposed forms, i.e. Gurmukhi letters ਞ (U+0A38), ਝ (U+0A16), ਞ (U+0A17), ਜ਼ (U+0A1C), ਢ (U+0A2B), ਝ (U+0A32) followed by Gurmukhi Sign Nukta (pairin bindi) “◌” (U+0A3C) can be used instead.

IDNA Protocol also excludes invisible characters Zero Width Non-Joiner (U+200C) and Zero Width Joiner (U+200D), as they require a CONTEXTJ rule. These are required in certain cases where a typical visual shape of an akshar is desired, such as two vowel signs attached with a consonant. But such cases do not occur in modern Gurmukhi text.

Also as Virama is not displayed in Gurmukhi, unlike in Devanagari, we do not have issues like we face in Devanagari, where use inability to use ZWNJ can pose issues where two words need to be joined together where previous word ends with an explicit Halant.

*iii. Maximal Starting Repertoire:*

Since the Root-zone LGR is a repertoire of the characters to be used for creation of root-zone TLDs, which in turn are an even more specialized case of domain names, the ROOT LGR procedure introduces additional exclusions on IDNA allowed set of characters.

To sum up, the restrictions start off with admitting only such characters as are part of the code-block of the given script/language. This is further narrowed down by the IDNA Protocol and finally an additional filter in the form of Maximal Starting Repertoire restricts the character set associated with the given language even more.

#### 4.1.2. No Punctuation Marks:

The TLDs being identifiers, punctuation marks present in Brahmi-based languages such as Dandi “|” and double Dandi “||” will not be included.

#### 4.1.3. No Symbols and Abbreviations:

Gurmukhi sign Addak Bindi<sup>ˆ</sup> (U+ 0A01) will not be included as it is not used in modern Punjabi.

#### 4.1.4. No Rare and Obsolete Characters:

There are characters which have been added to Unicode to accommodate the forms used in Medieval writings such as those of Sri Guru Granth Sahib, e.g. Gurmukhi signs Yakash “<sup>ˆ</sup>” (U+ 0A75), and Visarga : (U+ 0A03). Such characters will not be included. This is in compliance with the letter principle as laid down in the Root Zone LGR procedure.

#### 4.1.5. No Stress Markers of Medieval Punjabi:

Medieval Punjabi stress markers, and the tone marker sign Uddat “<sup>ˆ</sup>” (U+ 0A51), will not be included. This is also in compliance with the Letter principle as laid down in the Root Zone LGR procedure.

#### 4.1.6. No Vowel Carriers

The vowel carriers URA, <sup>ˆ</sup> (U+0A73) and IRI, <sup>ˆ</sup> (U+0A72), cannot occur without a matra in a word. They are used as vowel carriers and thus always need to be followed by some



matra when used in text. However, where they occur with a matra they will be identical with one of the independent vowels (ੳ (U+ 0A09), ਊ (U+ 0A0A), ਇ (U+ 0A07), ਈ (U+ 0A08), ਏ (U+ 0A0F), ਓ (U+ 0A13); this is also not allowed in Unicode. Thus ਓ (U+0A73) + ੱ (U+0A41), which looks the same as ਊ (U+ 0A09), will create confusion and hence will not be allowed in the LGR. These characters can occur as single character words, but in TLD, single character labels are not allowed, so these letters will not be added.

## 5. Repertoire

### 5.1. Code Points

Sr. No.	Unicode Code Point	Glyph	Character Name	Indic Syllabic Category	Reference
1.	0A02	ੰ	GURMUKHI SIGN BINDI	Bindi	[0], [105], [112]
2.	0A05	ਅ	GURMUKHI LETTER A = aīra	Vowel/Vowel Carrier	[0], [105], [112]
3.	0A06	ਆ	GURMUKHI LETTER AA	Vowel	[0], [105], [112]
4.	0A07	ਇ	GURMUKHI LETTER I	Vowel	[0], [105], [112]
5.	0A08	ਈ	GURMUKHI LETTER II	Vowel	[0], [105], [112]
6.	0A09	ਉ	GURMUKHI LETTER U	Vowel	[0], [105], [112]
7.	0A0A	ਊ	GURMUKHI LETTER UU	Vowel	[0], [105], [112]
8.	0A0F	ਏ	GURMUKHI LETTER EE	Vowel	[0], [105], [112]
9.	0A10	ਐ	GURMUKHI LETTER AI	Vowel	[0], [105], [112]

Sr. No.	Unicode Code Point	Glyph	Character Name	Indic Syllabic Category	Reference
10.	0A13	ਊ	GURMUKHI LETTER OO	Vowel	[0], [105], [112]
11.	0A14	ਐ	GURMUKHI LETTER AU	Vowel	[0], [105], [112]
12.	0A15	ਕ	GURMUKHI LETTER KA	Consonant	[0], [105], [112]
13.	0A16	ਖ	GURMUKHI LETTER KHA	Consonant	[0], [105], [112]
14.	0A17	ਗ	GURMUKHI LETTER GA	Consonant	[0], [105], [112]
15.	0A18	ਘ	GURMUKHI LETTER GHA	Consonant	[0], [105], [112]
16.	0A19	ਙ	GURMUKHI LETTER NGA	Consonant	[0], [105], [112]
17.	0A1A	ਚ	GURMUKHI LETTER CA	Consonant	[0], [105], [112]
18.	0A1B	ਛ	GURMUKHI LETTER CHA	Consonant	[0], [105], [112]
19.	0A1C	ਜ	GURMUKHI LETTER JA	Consonant	[0], [105], [112]
20.	0A1D	ਝ	GURMUKHI LETTER JHA	Consonant	[0], [105], [112]
21.	0A1E	ਞ	GURMUKHI LETTER NYA	Consonant	[0], [105], [112]
22.	0A1F	ਟ	GURMUKHI LETTER TTA	Consonant	[0], [105], [112]
23.	0A20	ਠ	GURMUKHI LETTER TTHA	Consonant	[0], [105], [112]
24.	0A21	ਡ	GURMUKHI LETTER DDA	Consonant	[0], [105], [112]
25.	0A22	ਢ	GURMUKHI LETTER DDHA	Consonant	[0], [105], [112]

Sr. No.	Unicode Code Point	Glyph	Character Name	Indic Syllabic Category	Reference
26.	0A23	ਠ	GURMUKHI LETTER NNA	Consonant	[0], [105], [112]
27.	0A24	ਤ	GURMUKHI LETTER TA	Consonant	[0], [105], [112]
28.	0A25	ਥ	GURMUKHI LETTER THA	Consonant	[0], [105], [112]
29.	0A26	ਦ	GURMUKHI LETTER DA	Consonant	[0], [105], [112]
30.	0A27	ਧ	GURMUKHI LETTER DHA	Consonant	[0], [112], [105]
31.	0A28	ਨ	GURMUKHI LETTER NA	Consonant	[0], [105], [112]
32.	0A2A	ਪ	GURMUKHI LETTER PA	Consonant	[0], [105], [112]
33.	0A2B	ਫ	GURMUKHI LETTER PHA	Consonant	[0], [105], [112]
34.	0A2C	ਬ	GURMUKHI LETTER BA	Consonant	[0], [105], [112]
35.	0A2D	ਭ	GURMUKHI LETTER BHA	Consonant	[0], [105], [112]
36.	0A2E	ਮ	GURMUKHI LETTER MA	Consonant	[0], [105], [112]
37.	0A2F	ਯ	GURMUKHI LETTER YA	Consonant	[0], [105], [112]
38.	0A30	ਰ	GURMUKHI LETTER RA	Consonant	[0], [105], [112]
39.	0A32	ਲ	GURMUKHI LETTER LA	Consonant	[0], [105], [112]
40.	0A35	ਵ	GURMUKHI LETTER VA	Consonant	[0], [105], [112]
41.	0A38	ਸ	GURMUKHI LETTER SA	Consonant	[0], [105], [112]

Sr. No.	Unicode Code Point	Glyph	Character Name	Indic Syllabic Category	Reference
42.	0A39	ੜ	GURMUKHI LETTER HA	Consonant	[0], [105], [112]
43.	0A3C	ੜ	GURMUKHI SIGN NUKTA = pairin bindi	Nukta	[0], [105], [112]
44.	0A3E	ੜ	GURMUKHI VOWEL SIGN AA = kanna	Matra	[0], [105], [110], [112]
45.	0A3F	ੜ	GURMUKHI VOWEL SIGN I = sihari	Matra	[0], [105], [112]
46.	0A40	ੜ	GURMUKHI VOWEL SIGN II = bihari	Matra	[0], [105], [112]
47.	0A41	ੜ	GURMUKHI VOWEL SIGN U = aunkar	Matra	[0], [105], [112]
48.	0A42	ੜ	GURMUKHI VOWEL SIGN UU = dulainkar	Matra	[0], [105], [112]
49.	0A47	ੜ	GURMUKHI VOWEL SIGN EE = lavan	Matra	[0], [105], [112]
50.	0A48	ੜ	GURMUKHI VOWEL SIGN AI = dulanvan	Matra	[0], [105], [112]
51.	0A4B	ੜ	GURMUKHI VOWEL SIGN OO = hora	Matra	[0], [105], [112]
52.	0A4C	ੜ	GURMUKHI VOWEL SIGN AU = kanaura	Matra	[0], [105], [112]
53.	0A4D	ੜ	GURMUKHI SIGN VIRAMA	Virama	[0], [105], [112]
54.	0A5C	ੜ	GURMUKHI LETTER RRA	Consonant	[0], [105], [112]

Sr. No.	Unicode Code Point	Glyph	Character Name	Indic Syllabic Category	Reference
55.	0A70	ੳ	GURMUKHI TIPPI	Tippi	[0], [105], [112]
56.	0A71	ੲ	GURMUKHI ADDAK	Addak	[0], [105], [112]

Table 6: Code point repertoire

## 5.2. Code points excluded from repertoire

Code points that occur in MSR-3 but are excluded because they are either not in common use or used for some special purpose only (e.g. as vowel carrier).

Sr. No.	Unicode Code Point	Glyph	Character Name	Note
1.	0A03	:	GURMUKHI SIGN VISARGA	Limited or declining use
2.	0A51	ੲ	GURMUKHI SIGN UDAAT	Limited or declining use
3.	0A72	ੲ	GURMUKHI IRI	Does not occur alone
4.	0A73	ੲ	GURMUKHI URA	Does not occur alone
5.	0A75	ੲ	GURMUKHI SIGN YAKASH	Limited or declining use

Table 7: List of excluded characters

## 5.3. Syllable formation rules for Gurmukhi:

The syllable is a basic unit of speech studied on both the phonetic and phonological levels of analysis. It is very easy for a native language speaker to count the number of syllables in a sequence; however, the orthographic syllable recognized for text processing need not correspond exactly with a phonological syllable. This section details the syllable-formation rules as applicable to Gurmukhi. The definition represents a vowel, consonant, or a conjunct.

### Variables involved:

C → Consonant, which may or may not include a single nukta

M → Matra

- V → Vowel  
 B → Bindi  
 D → Tippi  
 H → Halant / Virama  
 A → Addak

**Operators used:**

Symbol	Function
	Alternative
[]	Optional
{}	Zero or One occurrence
()	Sequence Group

The definition is a combination of 2 rules:

Rule 1: V[A|B|D]

Rule 2: {CH}C[M][A|B|D]

**Rule 1: V[A|B|D]**

Sl. No.	Examples	Definition
1	ਅ, ਆ, ਇ	V (Vowel) is a syllable
2	ਇੰ, ਉਂ, ਏਂ	V+ (A/B/D) is a syllable

**Rule 2: {C(H|A)}C[M][B|D]**

Sl. No.	Examples	Definition
1	ਕ, ਛ, ਧ	Consonant is a syllable where it has inherent 'ਓ' vowel
2	ਸ਼, ਲੱਗ	Zero or one Consonant + Virama/Addak sequence followed by consonant is a syllable
3	ਸ਼ੈ	Zero or one Consonant+ Virama sequence followed by a consonant followed by a matra or vowel sign is a syllable

4	ਸ੍ਰੰ	Zero or one Consonant+ Virama sequence followed by a consonant followed by bindi/tippi is a syllable
5	ਗ੍ਰਾਂ	Zero or one Consonant+ Virama sequence followed by a consonant followed by a matra and bindi/tippi is a syllable

Examples of combination of the rules:

1. ਕਰਸੀ (*karasī*) - C + CD + CM has following syllables:

ਕ	C
ਰੰ	CD
ਸੀ	CM

2. ਪਰਿੰਦਾ (*parindā*) - CV + CMD + CM has following syllables:

ਪ	CV
ਰਿੰ	CMD
ਦਾ	CV

3. ਅੰਦਰ (*andar*) - VD + CV+ C

ਅੰ	Vm
ਦਰ	CvC

## 6. Candidate Variants

There are no characters/character sequences in Gurmukhi that can be created by using the characters permitted in the [MSR] and that look exactly alike. However, Gurmukhi has ample cases of confusable characters in both Gurmukhi and Devanagari scripts. We have categorized these confusable character pairs in three groups.

**Group 1:** Visually similar Gurmukhi characters (Table 8)

**Group 2:** Visually similar Gurmukhi character combinations, due to the presence of dots and other characters (Table 9)

**Group 3:** Cross-script variants

No cases belonging to Group 1 and Group 2 are proposed as variants, as there is another panel (String similarity assessment panel) entrusted to deal with such cases.

ਚ (0A1A)	ੜ (0A30)
ਟ (0A1F)	ਦ (0A26)
ੜ (0A22)	ਦ (0A26)
ੜ (0A22)	ਫ (0A2B)
ਤ (0A24)	ਭ (0A2D)
ਬ (0A2C)	ਥ (0A25)
ੋ (0A47)	ੌ (0A4B)

Table 8: List of Group1 characters

Code Point Sequence	Code Point
ਖ (0A16 + 0A3C)	ਖ (0A16)
ਗ (0A17 + 0A3C)	ਗ (0A17)
ਫ (0A2B + 0A3C)	ਫ (0A2B)
ੳ (0A13 + 0A02)	ੳ (0A13)
ਈ (0A08 + 0A02)	ਈ (0A08)



ਐ (0A10 + 0A02)	ਐ (0A10)
ਐ (0A14 + 0A02)	ਐ (0A14)
ਐ (0A17 + 0A70)	ਐ (0A30 + 0A40)
ਐ (0A28 + 0A41)	ਐ (0A20)

Table 9: List of Group2 characters

## 6.1 Cross-script Variants

A "Whole Label confusable" is the case where one label in one script can be composed in such a way that it can resemble another entire label in a different script. Where the similarity is so close as to reach identical appearance, cross-script variants can be defined. Every individual LGR under NBGP is supposed to provide a set of cross script variants it identifies with all other scripts under NBGP.

The Gurmukhi script has a major set of possible cross-script variants only with the Devanagari script. Cases listed in Table 10 are of the variants that are proposed to be cross-script variants between Devanagari and Gurmukhi. Similarly, Table 11 has the cases proposed to be cross-script variants between Gurmukhi and Bengali.

It is to be noted that none of the combinations listed in Table 10 and Table 11 are termed to be equivalents of each other semantically or otherwise. They are only grouped based on possible visual confusability.

NBGP has ensured that Devanagari, Bengali and Gurmukhi LGR teams propose a same set of cross-script variants by meeting face-to-face on many occasions as well as through mail communications. The same set of cross-script variants (with Gurmukhi) is supposed to be found in the Bengali and Devanagari LGR documents.

<b>Devanagari</b>	<b>Gurmukhi</b>
ॐ U+0902	ੴ U+0A02
इ U+0907	ਙ U+0A19
उ U+0909	ਤ U+0A24
ग U+0917	ਗ U+0A17
घ U+0918	ਬ U+0A2C
ट U+091F	ਟ U+0A1F
ठ U+0920	ਠ U+0A20
ढ U+0922	ਢ U+0A2B
प U+092A	ਧ U+0A27
भ U+092D	ਮ U+0A2E
म U+092E	ਸ U+0A38
व U+0935	ਕ U+0A15
ह U+0939	ਵ U+0A35
ॐ U+093A	ੴ U+0A02
ॠ U+093C	ੴ U+0A3C
ि U+093F	ਿ U+0A3F
ी U+0940	ੀ U+0A40
ँ U+0945	ੴ U+0A71

ॐ U+0946	ॐ U+0A47
ॐ U+0946	ॐ U+0A4B
ॐ U+0947	ॐ U+0A47
ॐ U+0947	ॐ U+0A4B
ॐ U+0948	ॐ U+0A48
ॐ U+0956	ॐ 0A41
ॐ U+0957	ॐ 0A42
ष्टि U+092A U+094D U+091F U+093F	ष्टि U+0A07
ष्टी U+092A U+094D U+091F U+0940	ष्टी U+0A08
ष्टे U+092A U+094D U+091F U+0947	ष्टे U+0A0F
त्त U+0924 U+094D U+0924	त्त U+0A1C

Table 10: Proposed Cross-script Devanagari-Gurmukhi Variants

Gurmukhi	Bangla
म U+0A38	म U+09AE
फि U+0A3F	फि U+09BF

Table 11: Proposed Cross-script Gurmukhi-Bangla Variants

## 7. Whole Label Evaluation Rules (WLE)

This section provides the Whole Label Evaluation rules for text written in the Gurmukhi script. The rules have been drafted in such a way that they can be easily translated into the LGR specification.

Below are the symbols used in the WLE rules, for each of the "Indic Syllabic Category" as mentioned in the Table 6: Code point repertoire. In addition, we have created a few more symbols related to matras and vowels for the explanation of the rules.

A	→	Addak
B	→	Bindi
C	→	Consonant
C1	→	{ख (U+0A16), ऋ (U+0A17), ञ (U+0A1C), ढ (U+0A2B), ळ (U+0A32), ण (U+0A38)}
C2	→	{र (U+0A30), ञ (U+0A35), ञ (U+0A39)}
C3	→	C - {ऌ(U+0A19), ऍ(U+0A1E), ऎ(U+0A23), ञ(U+0A39), ञ(U+0A5C)}
D	→	Tippi
H	→	Virama
M	→	Matra
M1	→	{ँ(U+0A3F), ँ(U+0A41)} (Short matras)
M2	→	M - M1 (Long matras)
N	→	Nukta
V	→	Vowel
V1	→	{अ (U+0A05), इ (U+0A07), उ (U+0A09)} (Short Vowels)
V2	→	V - V1 (Long Vowel)

7.1. N: must be preceded only by C1

7.2. H: must be preceded by C or N and followed by C2 only

7.3. M: must be preceded by C or N

7.4. B: must be preceded by specific V or M

The specific Vs are:

- V2
- ँ (U+0A09)

The specific Ms are:

- M2 - {ँ (U+0A42)}

## 7.5. D: must be preceded by, C, N or a specified set of V or M

The specific Vs are:

- a. V1- { ॢ (U+0A09)}

The specific Ms are:

- a. M1
- b. { ॣ (U+0A42)}

## 7.6. A: must be preceded by C, N or specific V or M and followed by C3

The specific Vs are:

- a. V1
- b. ॠ (U+0A10)

The specific Ms are:

- a. M1
- b. ॡ (U+0A48)

## 8. Contributors

Name	Address
Dr. Gurpreet Singh Lehal	Professor, Department of Computer Science, Punjabi University, Patiala
Dr. Harvinder Pal Kaur	Assistant Professor, Research Centre for Punjabi Language Technology, Punjabi University, Patiala
Dr. Boota Singh Brar	Professor, Punjabi University Regional Centre, Bathinda
Dr. Paramjit Singh Sidhu	Retd. Professor, School of Punjabi Studies, Guru Nanak Dev University, Amritsar

## 9. References

[NBGP] Neo-Brahmi Generation Panel

- [MSR] Integration Panel, "Maximal Starting Repertoire — MSR-3 Overview and Rationale", 28 March 2018  
<https://www.icann.org/en/system/files/files/msr-3-overview-28mar18-en.pdf>
- [0] Gurmukhi Unicode chart (Accessed on 21 May 2018)  
<https://unicode.org/charts/PDF/U0A00.pdf>
- [100] Newton, E. P., 1961, Panjabi Grammar, Patiala: Language Department Punjab.
- [101] Ojha, Gauri Shankar Hira Chand, 1962, Bharti Prachin Lipi Mala (Ed. Jagdish Chand & Others), Patiala: Bhasha Vibhag Punjab.
- [102] Singh, G.B., 1950, Gurmukhi Lipi Da Janam te Vikas, Chandigarh: Punjab University.
- [103] Singh, Pritam, 1958, Gurmukhi Lipi Di Utpati te Vikas, Ludhiana: Lahore Book Shop.
- [104] Diringer, David, 1948, The Alphabet: A Key to the History of Mankind, London: Hutchinson Scientific and Technical Publication.
- [105] Omniglot, <https://www.omniglot.com/writing/punjabi.htm> (Accessed on 10th Nov. 2017)
- [106] Unicode 10.0.0, "South and Central Asia-I - Official Scripts of India", Page 475-479, <http://www.unicode.org/versions/Unicode10.0.0/ch12.pdf> (Accessed on 13th Nov. 2017)
- [107] Al-Biruni, 2000, Al-Hind, (Ed. Kyamu Din Ahmad and Trn. Gurcharn Singh Arshi), New Delhi: National Book Trust, India.
- [108] Bedi, Tarlochan Singh, 1999, Gurmukhi Lipi da Janam te Vikas, Patiala: Punjabi University.
- [109] A start in Punjabi, Lesson-09, "Gurmukhi Orthography-I"  
<http://pt.learnpunjabi.org/av.aspx?l=9> (Accessed on 10th Nov. 2017)
- [110] Gurmukhi Alphabet :: Lesson 11, "Gurmukhi Vowel Signs Group-1 Mukta and Kanna", <http://elearnpunjabi.com> (Accessed on 10th Nov. 2017)
- [111] A start in Punjabi, Lesson-10, "Gurmukhi Orthography-II"  
<http://pt.learnpunjabi.org/av.aspx?l=10> (Accessed on 10th Nov. 2017)
- [112] A reference Grammar of Punjabi,  
[http://pt.learnpunjabi.org/assets/A%20Reference%20Grammar\\_Final.pdf](http://pt.learnpunjabi.org/assets/A%20Reference%20Grammar_Final.pdf)  
 (Accessed on 10th Nov. 2017)
- [113] Bhardwaj, Mangat Rai, 1995, Colloquial Panjabi: A Complete Language Course, Routledge, London.
- [114] Brar, Boota Singh, 2016, Panjabi Viakarn: Sidhant ate Vihar, Ludhiana: Chetna Parkashan.