

# Proposal for a Devanagari Script Root Zone Label Generation Ruleset [LGR]

---

*LGR Version: 2.0*

*Date: 2013-23-03*

*Document version: 1.0*

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

## 1 General Information/ Overview/ Abstract

This document lays down the Label Generation Ruleset for Devanagari script. Three main components of the Devanagari 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 "Proposed-LGR-Deva-20170323.xml".

## 2 Script for which the LGR is proposed

ISO 15924 Code: Deva

ISO 15924 Key N°: 315

ISO 15924 English Name: Devanagari (Nagari)

Latin transliteration of native script name: dévanâgarī

Native name of the script: देवनागरी

Maximal Starting Repertoire (MSR) version: MSR-2

## 3 Background on Script and Principal Languages Using It

The script called Nāgarī or Devanāgarī is written from left to right. Historically it derives from the Brāhmī alphabet of the Ashokan inscriptions. Devanāgarī is currently used for 11 out of 22 official languages of India (Boro/Bodo, Dogri, Hindi, Kashmiri, Konkani, Maithili, Marathi, Nepali, Sanskrit, Santhali and Sindhi) and around 45 other languages especially the related Indo-Aryan languages: Bagheli, Bhili, Bhojpuri, Himachali dialects, Magahi, Newari and Rajasthani and its dialects: Marwari, Mewati, Shekhawati, Bagri, Dhundhari,

Harauti and Wagri. Closely associated with Sanskrit and Prakrit, it is an alternative script for Kashmiri (by Hindu speakers), Sindhi and Santhali. It is growing popular in use by speakers of tribal languages of Arunachal Pradesh, Bihar and Andaman & Nicobar Islands. The script is also used in Fiji to represent Fiji Hindi. Hindi is also used in Mauritius, Malaysia, England, Canada, South Africa, Indonesia as well as emigrant communities around the world.

Devanāgarī is used by over 120 languages both in India and in South-east-Asia.

A detailed tabular list of the main languages using Devanagari is provided in what follows

### 3.1 The Evolution of the Script

It is well-known that Devanāgarī has evolved from the parent script Brāhmī, with its earliest historical form known as Aśokan Brāhmī, traced to the 4th century B.C. Brāhmī was deciphered by Sir James Prinsep in 1837. The study of Brāhmī and its development has shown that it has given rise to most of the scripts in India as well as other countries viz. Sri Lanka, Myanmar, Kampuchea, Thailand, Laos, and Tibet to name a few.

The evolution of Brāhmī into present-day Devanāgarī involved intermediate forms, common to other scripts such as Gupta and Śāradā in the north and Grantha and Kadamba in the South. Devanāgarī can be said to have developed from the Kutila script, a descendant of the Gupta script, in turn a descendent of Brāhmī. The word kutila, meaning 'crooked', was used as a descriptive term to characterize the curving shapes of the script, compared to the straight lines of Brāhmī. This inheritance is the reason for some of the characters across the scripts that will be considered under the Neo-brahmi GP to look similar to each other despite belonging to totally different code blocks.

A look at the development of Devanāgarī from Brāhmī gives an insight into how the Indic scripts have come to be diversified: the handiwork of engravers and writers who used different types of strokes leading to different regional styles. The development of the script is outlined below. Figure 1 illustrates the stages in the evolution of the script<sup>1</sup>.

Period	Description
300 BCE	Mauryan : Early Brahmi form the Asokan edicts. Some scholars believe that Brahmi itself evolved from "karoshti" a script written right to left.
200 CE	Kushan/Satavahana Dynasties.
400 CE	Gupta Dynasty
600 CE	Yasodharman

<sup>1</sup> [http://www.acharya.gen.in:8080/sanskrit/script\\_dev.php](http://www.acharya.gen.in:8080/sanskrit/script_dev.php)

800 CE	Origins of the present day Nagari Script. Vardhana dynasty in the North and Pallava period in the South.
900 CE	The period of the Chalukyas and Rashtrakutas
1100 CE	Continuation of the Chalukya Rule
1300 CE	Yadavas in the north and Kakatiyas in the south.
1500 CE	The Vijayanagar empire.

Table 1: Evolution of Devanagari

300 BCE	†	ε	୪		୮	୩
200 CE	‡	E	⊗	J	୨	୩
400 CE	†	E	୨	I	୩	୩
600 CE	‡	E	୩	I	୨	୩
800 CE	‡	ε	୩	I	୨	अ
900 CE	‡	୩	म	I	र	अ
1100 CE	‡	୩	म	र	स	अ
1300 CE	क	ज	म	र	स	अ
Modern	क	ज	म	र	स	अ

Figure 1: Pictorial depiction of Evolution of Devanagari

### 3.2 Tabular representation of Languages using Devanāgarī

The languages chosen are represented majorly in the scripts which are part of MSR 2 and belong to the EGIDS scale of 1 to 4.

Language	Native name
Bhojpuri	भोजपुरी

Bodo	बड़ो
Chhattisgarhi	छत्तीसगढ़ी
Fiji Hindi	फिजीबात, Fiji Baat
Hindi	हिन्दी
Kashmiri	कॉशुर, Kāśur, Koshur
Konkani	कोंकणी
Magahi	मगही
Maithili	मैथिली, मगधेली
Marathi	मराठी
Nepal Bhasa	नेपालभाषा
Nepali	नेपाली
Rajasthani	राजस्थानी
Sanskrit	संस्कृतम्, संस्कृतावाक्
Santali/Santhali	संथाली
Sindhi	सिंधी

Table 2: Languages considered under Devanagari LGR

### 3.3 The structure of written Devanāgarī

Devanāgarī is an alphasyllabary and the heart of the writing system is the syllable or akshar. It is this unit which is instinctively recognized by users of the script. To understand the notion of akshar, a brief overview of the writing system is provided in this Section and the akshar itself will be treated in depth in Section 3.4.

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

### 3.3.1 The Consonants

Devanāgarī consonants have an implicit schwa /ə/ included in them. As per traditional classification they are categorized according to their phonetic properties. There are 5 (Varg) groups and one non-Varg group. Each Varg contains five consonants classified as per their properties. The first four consonants are classified on the basis of Voicing and Aspiration and the last is the corresponding nasal.

Varg	Unvoiced		Voiced		Nasal
	-Asp	+Asp	-Asp	+Asp	
<b>Velar</b>	क	ख	ग	घ	ङ
<b>Palatal</b>	च	छ	ज	झ	ञ
<b>Retroflex</b>	ट	ठ	ड	ढ	ण
<b>Dental</b>	त	थ	द	ध	न
<b>Bi-labial</b>	प	फ	ब	भ	म

Table 3: Varga classification of consonants

<b>Non-Varg</b>	य	र	ल	ळ	व	श	ष	स	ह
-----------------	---	---	---	---	---	---	---	---	---

Table 4: Non-varga consonants

### 3.3.2 The Implicit Vowel Killer: Halanta<sup>2</sup>

All consonants have an implicit vowel sign (schwa) within them. A special sign is needed to denote that this implicit vowel is stripped off. This is known as the Halanta (◌्). The Halanta thus joins two consonants and creates conjuncts which can be generally from 2 to 4 consonant combinations. In rare cases it can join upto 5 consonants. However the notion of maximum number of consonants joining to form one akshar is not empirical. It is just an observation drawn from the words that have been observed till date. Given the confluence of languages happening in the Internet age, the possibility that one may want a domain

<sup>2</sup> Unicode (cf. Unicode 3.0 and above) prefers the term Virama. In this report both the terms have been used to denote the character that suppresses the inherent vowel.

(TLD or generic) which may have more than the observed maximum cannot be ruled out. Hence, in the LGR work, this limit will not be enforced.

### 3.3.3 Vowels

Separate symbols exist for all Vowels which are pronounced independently either at the beginning or after a vowel sound. To indicate a Vowel sound other than the implicit one, a Vowel modifier (Mātrā) is attached to the consonant. Since the consonant has a built in schwa, there are equivalent Mātrās for all vowels excepting the अ.

The correlation is shown as under:

अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ	ओ	औ
	ा	ि	ी	ु	ू	ृ	े	ै	ो	ौ

Table 5: Vowels with corresponding Matras

In addition to show sounds borrowed from English, some languages using Devanāgarī such as Hindi, Marathi, and Konkani also admit 2 vowels and their corresponding Mātrās as in

ँ            ॉ

ँण्ड /and/    ॉर /or/

Marathi replaces the ँ by अँ

### 3.3.4 The Anuswāra (ं)

The Anuswāra represents a homo-organic nasal. It replaces a conjunct group of a Nasal Consonant+Halanta+Consonant belonging to that particular varg. Before a Non-varg consonant the anuswāra represents a nasal sound. Modern Hindi, Marathi and Konkani prefer the anuswāra to the corresponding Half-nasal:

सन्त vs. संत /sənt/    saint    चम्पा vs. चंपा /tʃəmpa/

### 3.3.5 Nasalization: Chandrabindu (ँ)

Chandrabindu/Anunasika denotes nasalization of the preceding vowel as in आँख (eye) /ākh/ eye. Present-day Hindi users tend to replace the chandrabindu by the anuswāra

### 3.3.6 Nukta (◌̣)

Mainly used in Hindi, the nukta sign is placed below a certain number of consonants to represent words borrowed from Perso-Arabic. It can be adjoined to क ख ग ज फ to show that words having these consonants with a nukta are to be pronounced in the Perso-Arabic style.

e.g. फ़िरोज़ /firoz/

It is also placed under ङ and ढ in Hindi to indicate flapped sounds

With the exception of flaps, users of modern-day Hindi hardly use the nukta characters today.

### 3.3.7 Visarg (◌:) and Avagrah (◌ː)

The Visarg is frequently used in Sanskrit and represents a sound very close to /h/. दुःख /du:kh/ sorrow, unhappiness

The Avagrah ː creates an extra stress on the preceding vowel and is used in Sanskrit texts.

It is rarely used in other languages using Devanāgarī. In case of LGR, the Avagraha is not part of the repertoire as it is barred in the Maximal Starting Repertoire.

## 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 Devanagari LGR which caters to multiple languages written using Devanagari belonging to EGIDS scale 1 to 4.

## 5 Repertoire

This section details the code-point repertoire that the NGBP proposes to be included in the Devanagari LGR.

One of the major sources of reference to the justification for inclusion of the code-point is the Indian National Standard 'Enhanced Inscript Keyboard layouts' [INSCRIPT]" laying down the language specific keyboard layouts for all the scheduled languages of India. It is officially published and notified in the Gazette of India. The standard specifies key-layouts for each of the scheduled languages of India. The standard among other things provides a comprehensive language-wise list of various characters as used by the scheduled (of which the set of languages under the NGBP ambit is a sub-set) languages of the India. The [INSCRIPT] standard carves out a sub-set of characters applicable to each of the languages out of the respective code-page of the script used by that language.

### 5.1 Code point Repertoire:

Sr. No.	Unicode Code Point	Character	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference
1.	0901	ँ	DEVANAGARI SIGN CANDRABINDU	Mn	Chandrabindu / Anunasika	[INSCRIPT]
2.	0902	ं	DEVANAGARI SIGN ANUSVARA	Mn	Anusvara / Bindu	[INSCRIPT]
3.	0903	ः	DEVANAGARI SIGN VISARGA	Mc	Visarga	[INSCRIPT]
4.	0905	अ	DEVANAGARI LETTER A	Lo	Vowel	[INSCRIPT]
5.	0906	आ	DEVANAGARI LETTER AA	Lo	Vowel	[INSCRIPT]
6.	0907	इ	DEVANAGARI LETTER I	Lo	Vowel	[INSCRIPT]
7.	0908	ई	DEVANAGARI LETTER II	Lo	Vowel	[INSCRIPT]
8.	0909	उ	DEVANAGARI LETTER U	Lo	Vowel	[INSCRIPT]
9.	090A	ऊ	DEVANAGARI LETTER UU	Lo	Vowel	[INSCRIPT]



10.	090B	ऋ	DEVANAGARI LETTER VOCALIC R	Lo	Vowel	[INSCRIPT]
11.	090D	ॠ	DEVANAGARI LETTER CANDRA E	Lo	Vowel	[INSCRIPT]
12.	090F	ए	DEVANAGARI LETTER E	Lo	Vowel	[INSCRIPT]
13.	0910	ऐ	DEVANAGARI LETTER AI	Lo	Vowel	[INSCRIPT]
14.	0911	ऑ	DEVANAGARI LETTER CANDRA O	Lo	Vowel	[INSCRIPT]
15.	0913	ओ	DEVANAGARI LETTER O	Lo	Vowel	[INSCRIPT]
16.	0914	औ	DEVANAGARI LETTER AU	Lo	Vowel	[INSCRIPT]
17.	0915	क	DEVANAGARI LETTER KA	Lo	Consonant	[INSCRIPT]
18.	0916	ख	DEVANAGARI LETTER KHA	Lo	Consonant	[INSCRIPT]
19.	0917	ग	DEVANAGARI LETTER GA	Lo	Consonant	[INSCRIPT]
20.	0918	घ	DEVANAGARI LETTER GHA	Lo	Consonant	[INSCRIPT]
21.	0919	ङ	DEVANAGARI LETTER NGA	Lo	Consonant	[INSCRIPT]
22.	091A	च	DEVANAGARI LETTER CA	Lo	Consonant	[INSCRIPT]
23.	091B	छ	DEVANAGARI LETTER CHA	Lo	Consonant	[INSCRIPT]
24.	091C	ज	DEVANAGARI LETTER JA	Lo	Consonant	[INSCRIPT]
25.	091D	झ	DEVANAGARI LETTER JHA	Lo	Consonant	[INSCRIPT]
26.	091E	ञ	DEVANAGARI LETTER NYA	Lo	Consonant	[INSCRIPT]
27.	091F	ट	DEVANAGARI LETTER TTA	Lo	Consonant	[INSCRIPT]
28.	0920	ठ	DEVANAGARI LETTER TTHA	Lo	Consonant	[INSCRIPT]
29.	0921	ड	DEVANAGARI LETTER DDA	Lo	Consonant	[INSCRIPT]
30.	0922	ढ	DEVANAGARI LETTER DDHA	Lo	Consonant	[INSCRIPT]

31.	0923	ण	DEVANAGARI LETTER NNA	Lo	Consonant	[INSCRIPT]
32.	0924	त	DEVANAGARI LETTER TA	Lo	Consonant	[INSCRIPT]
33.	0925	थ	DEVANAGARI LETTER THA	Lo	Consonant	[INSCRIPT]
34.	0926	द	DEVANAGARI LETTER DA	Lo	Consonant	[INSCRIPT]
35.	0927	ध	DEVANAGARI LETTER DHA	Lo	Consonant	[INSCRIPT]
36.	0928	न	DEVANAGARI LETTER NA	Lo	Consonant	[INSCRIPT]
37.	092A	प	DEVANAGARI LETTER PA	Lo	Consonant	[INSCRIPT]
38.	092B	फ	DEVANAGARI LETTER PHA	Lo	Consonant	[INSCRIPT]
39.	092C	ब	DEVANAGARI LETTER BA	Lo	Consonant	[INSCRIPT]
40.	092D	भ	DEVANAGARI LETTER BHA	Lo	Consonant	[INSCRIPT]
41.	092E	म	DEVANAGARI LETTER MA	Lo	Consonant	[INSCRIPT]
42.	092F	य	DEVANAGARI LETTER YA	Lo	Consonant	[INSCRIPT]
43.	0930	र	DEVANAGARI LETTER RA	Lo	Consonant	[INSCRIPT]
44.	0932	ल	DEVANAGARI LETTER LA	Lo	Consonant	[INSCRIPT]
45.	0933	ळ	DEVANAGARI LETTER LLA	Lo	Consonant	[INSCRIPT]
46.	0935	व	DEVANAGARI LETTER VA	Lo	Consonant	[INSCRIPT]
47.	0936	श	DEVANAGARI LETTER SHA	Lo	Consonant	[INSCRIPT]
48.	0937	ष	DEVANAGARI LETTER SSA	Lo	Consonant	[INSCRIPT]
49.	0938	स	DEVANAGARI LETTER SA	Lo	Consonant	[INSCRIPT]
50.	0939	ह	DEVANAGARI LETTER HA	Lo	Consonant	[INSCRIPT]
51.	093A	'	DEVANAGARI VOWEL SIGN OE	Mn	Matra	[INSCRIPT]
52.	093B	†	DEVANAGARI VOWEL SIGN OOE	Mc	Matra	[INSCRIPT]

53.	093C	◌̣	DEVANAGARI SIGN NUKTA	Mn	Nukta	[INSCRIPT]
54.	093E	◌ा	DEVANAGARI VOWEL SIGN AA	Mc	Matra	[INSCRIPT]
55.	093F	◌ि	DEVANAGARI VOWEL SIGN I	Mc	Matra	[INSCRIPT]
56.	0940	◌ी	DEVANAGARI VOWEL SIGN II	Mc	Matra	[INSCRIPT]
57.	0941	◌ु	DEVANAGARI VOWEL SIGN U	Mn	Matra	[INSCRIPT]
58.	0942	◌ू	DEVANAGARI VOWEL SIGN UU	Mn	Matra	[INSCRIPT]
59.	0943	◌ृ	DEVANAGARI VOWEL SIGN VOCALIC R	Mn	Matra	[INSCRIPT]
60.	0944	◌ॄ	DEVANAGARI VOWEL SIGN VOCALIC RR	Mn	Matra	[INSCRIPT]
61.	0945	◌ं	DEVANAGARI VOWEL SIGN CANDRA E = candra	Mn	Matra	[INSCRIPT]
62.	0947	◌े	DEVANAGARI VOWEL SIGN E	Mn	Matra	[INSCRIPT]
63.	0948	◌ै	DEVANAGARI VOWEL SIGN AI	Mn	Matra	[INSCRIPT]
64.	0949	◌ॉ	DEVANAGARI VOWEL SIGN CANDRA O	Mc	Matra	[INSCRIPT]
65.	094B	◌ो	DEVANAGARI VOWEL SIGN O	Mc	Matra	[INSCRIPT]
66.	094C	◌ौ	DEVANAGARI VOWEL SIGN AU	Mc	Matra	[INSCRIPT]
67.	094D	◌्	DEVANAGARI SIGN VIRAMA	Mn	Halant / Virama	[INSCRIPT]
68.	094F	◌ाँ	DEVANAGARI VOWEL SIGN AW	Mc	Matra	[INSCRIPT]
69.	0956	◌ँ	DEVANAGARI VOWEL SIGN UE	Mn	Matra	[INSCRIPT]
70.	0957	◌ं	DEVANAGARI VOWEL SIGN UUE	Mn	Matra	[INSCRIPT]

71.	0972	अँ	DEVANAGARI LETTER CANDRA A	Lo	Consonant	[INSCRIPT]
72.	0973	अं	DEVANAGARI LETTER OE	Lo	Consonant	[INSCRIPT]
73.	0974	आँ	DEVANAGARI LETTER OOE	Lo	Consonant	[INSCRIPT]
74.	0975	औ	DEVANAGARI LETTER AW	Lo	Consonant	[INSCRIPT]
75.	0976	अु	DEVANAGARI LETTER UE	Lo	Consonant	[INSCRIPT]
76.	0977	अू	DEVANAGARI LETTER UUE	Lo	Consonant	[INSCRIPT]
77.	0979	ज़	DEVANAGARI LETTER ZHA	Lo	Consonant	[INSCRIPT]
78.	097A	ष	DEVANAGARI LETTER HEAVY YA	Lo	Consonant	[INSCRIPT]
79.	097B	ग़	DEVANAGARI LETTER GGA	Lo	Consonant	[INSCRIPT]
80.	097C	ज़	DEVANAGARI LETTER JJA	Lo	Consonant	[INSCRIPT]
81.	097E	ड़	DEVANAGARI LETTER DDDA	Lo	Consonant	[INSCRIPT]
82.	097F	ब़	DEVANAGARI LETTER BBA	Lo	Consonant	[INSCRIPT]

Table 6: Code point repertoire

Apart from the above individual code-points, the Neo-brahmi Generation Panel also proposes below sequences which enable conditional inclusion of the "DEVANAGARI LETTER RRA" in the repertoire.

Sr. No.	Unicode Code Points	Sequence	Character Names	Unicode General Category (gc)	Reference
1.	0931	य्य	DEVANAGARI LETTER RRA	Lo	[INSCRIPT]
	094D		DEVANAGARI SIGN VIRAMA	Mn	
	092F		DEVANAGARI LETTER YA	Lo	
2.	0931	र्र	DEVANAGARI LETTER RRA	Lo	[INSCRIPT]
	094D		DEVANAGARI SIGN VIRAMA	Mn	
	0939		DEVANAGARI LETTER HA	Lo	

Table 7: Sequences

## 5.2 Structural Formation of Devanagari:

All the languages written in Brahmi derived scripts follow a particular way of formation of its words, known as "akshar". This akshar construct can be modelled in a mathematical representation in the form of an Augmented Backus Naur Formalism. In the next section there is detailed analysis of how this formalism is applicable to representation of "Hindi" language when written in Devanagari Script. This needs slight changes for different languages written in Devanagari in terms of

- Character addition/deletion (e.g. Nukta [U+093C] character is applicable for Hindi but not Marathi)

- Presence or absence of a particular rule (e.g. Eyelash Ra construct is required in Marathi but not in Hindi).

## 5.3 AUGMENTED BACKUS-NAUR FORMALISM (ABNF) for Hindi:

The formalism described in this section is based on the work done for securing the IDN registrations under various IDN ccTLD zones operated by India, typically referred to as ".bharat framework". The ".bharat framework" policies are in place for all the scheduled languages of India.

Here the part of the work as applicable to Hindi is discussed in details only for informational purpose. Similar work is already carried out and put in implementation for all other scheduled languages of India. The Whole Label Evaluation Rules (WLEs) as applicable to Devanagari at large are discussed in Point 7.

### 5.3.1 Declaration of variables

Dash	→ Hyphen -
Digit	→ Indo-Arabid digits [0-9]
C	→ Consonant
M	→ Matra
V	→ Vowel
D	→ Anusvara / Bindu
B	→ Chandrabindu / Anunasika
X	→ Visarga
Y	→ Avagraha
H	→ Halant / Virama
N	→ Nukta

### 5.3.2 ABNF Operators:

Symbol	Function
	Alternative
[ ]	Optional
*	Variable Repetition
( )	Sequence Group

Table 8: ABNF Symbol functions

In what follows the Vowel Sequence and the Consonant Sequence pertinent to Devanagari, when used to write Hindi, are given.

### 5.3.3 The Vowel Sequence

A vowel sequence is made up of a single vowel. It may be followed but not necessarily (optionally) by an Anusvara (D), Chandrabindu (B) or a Visarga (X). The number of D, B or X which can follow a V in Devanagari may be restricted to one.

The possibility of a Visarga following a Chandrabindu or Anusvara is ruled out, since it is used only in Vedic and in Bangla script.

The vowel sequence in Hindi is therefore  $V [ D | B | X ]$

Examples :

Sequence Description	Sequence	Example	Example Decomposition
Vowel	V	अ /a/	
Vowel + Anusvara	V[D]	अं /am̐/	अ ँ
Vowel + Chandrabindu	V[B]	अँ /am̐̃/	अ ँँ
Vowel + Visarga	V[X]	अः /aḥ/	अ ः

Table 9

### 5.3.4 Consonant Sequence

A consonant sequence admits the following shapes:

1. A single consonant (C)

(The consonant shall be treated as coterminus with the Consonant along with the Nukta sign wherever such a case is pertinent.)

Examples:

Sequence Description	Sequence	Example	Example Decomposition
Consonant	C	क /ka/	
Consonant + Nukta	C[N]	क /ḱa/	क ँ

Table 10

2. A consonant optionally followed by dependent vowel sign/Matra[M] or Anusvara[D] Chandrabindu[B] or visarga[X] or Halant [H]

C[M|D|X|H]

Examples:

Sequence Description	Sequence	Example	Example Decomposition
Consonant + Matra	C[M]	कि /ki/	क ि
Consonant + Anusvara	C[D]	कं /kaṁ/	क ँ
Consonant + Chandrabindu	C[B]	कँ /kaṁ̄/	क ँ̄
Consonant + Visarga	C[X]	कः /kaḥ/	क ः
Consonant + Halanta	C[H]	क् /k/ (Pure Consonant)	क ्

Table 11

2.a. A CM sequence can be optionally followed by D, B or X

(CM)[D|B|X]

Example:

Sequence Description	Sequence	Example	Example Decomposition
Consonant + Matra + Anusvara	CM[D]	की /kīm/	क ि ँ
Consonant + Matra + Chandrabindu	CM[B]	कौ /kām̄/	क ि ँ̄
Consonant + Matra + Visarga	CM[X]	कीः	क ि ः

		/kīḥ/	
--	--	-------	--

Table 12

3. A sequence of consonants (up to 4) joined by Halant \*3(CH)C

Example:

Sequence Description	Sequence	Example	Example Decomposition
Consonant + Halanta + Consonant + Halanta + Consonant + Halanta + Consonant	CHCHCHC	न्क्रय /nkrya/	न क र य

Table 13

**Subsets:**

3.a. The combination may be followed by M, D, B or X

Example:

Sequence Description	Sequence	Example	Example Decomposition
Consonant + Halanta + Consonant + Matra	CHC[M]	क्की /kkī/	क् क ी
Consonant + Halanta + Consonant + Anusvara	CHC[D]	क्कं /kkaṁ/	क् क ं
Consonant + Halanta + Consonant + Chandrabindu	CHC[B]	क्कँ /kkaṃ̣/	क् क ँ
Consonant + Halanta + Consonant + Visarga	CHC[X]	क्कः /kkaḥ/	क् क ः

Table 14

3.b. \*3(CH)CM may be followed by a D, B or X

Example:

Sequence Description	Sequence	Example	Example Decomposition
Consonant + Halanta + Consonant + Matra + Anusvara	CHCM[D]	क्कीं /kkīm/	क् क ी ं



Consonant + Halanta + Consonant + Matra + Chandrabindu	CHCM[B]	क्कीं /kkīm/	क् क् की ँ
Consonant + Halanta + Consonant + Matra + Visarga	CHCM[X]	क्कीः /kkīḥ/	क् क् की ः

Table 15

The final canonical structure of the consonant sequence in Hindi IDN can be defined in ABNF as: \*3(C[N]H)C[N][H|D|B|X|M[D|B|X]]

It should be noted that in the case of Sanskrit the possible sequence could be:

\*4(CH)C [H|D|B|X|M[D|B|X]]

as exemplified in the word: कार्त्स्न्य /kārtsnya/

Although this is possible, it is not used in neo-Prakrits unless a potential client chooses it as a possible IDN.

### 5.3.5 Sequence

A sequence can be made up by Consonant-sequence or Vowel-sequence.

- A Consonant-sequence can optionally be followed by Avagraha[Y].
- A Vowel-sequence can optionally be followed by Avagraha[Y].

### 5.3.6 ABNF Applied to the Hindi IDN

The formalism can be applied to create/validate IDN labels. So a valid IDN label can be defined as follows.

Vowel-sequence → V [D | B | X]

Consonant-sequence → \*3(C[N]H)C[N] [H|D|B|X|M[D|B|X]]

Sequence → consonant-sequence[Y] | vowel-sequence[Y]

IDN-label → (sequence | digit) \* ([dash] (sequence | digit))

This is the background work based on which the NBGP LGR is based on. However, in case of LGR, there are different constraints in terms of

- base code point repertoire
- freedom to be language specific as opposed to being script focused

In LGR work, the base code point repertoire is defined by the Maximal Starting Repertoire Version 2, which bars some of the characters like Avagraha and digits which are part of the above base work.

In the above work, there was a freedom to be language specific as the zones of individual languages were different. In case of LGR, root zone being the single zone, a different approach of clubbing the needs of all the languages using Devanagari into a single ruleset is taken.

## 6 Variants

<TBD>

## 7 Whole Label Evaluation Rules (WLE)

In Devanagari Whole Label evaluation rules, there are two major building blocks. First is the Vowel sequence and the second is Consonant Syllable. Here while defining the WLEs for Devanagari LGR, we are changing the way of representing the rule for clear reading as well as facilitating translation in the LGR specification.

Also, as these WLEs are supposed to all accommodative for the languages using Devanagari, the section in the ABNF construct that restricts the conjunct depth

C	→	Consonant
M	→	Matra
V	→	Vowel
D	→	Anusvara / Chandrabindu
X	→	Visarga
H	→	Halant / Virama
N	→	Nukta
S	→	Eyesh Reph (C1HC2) where C1 is 0931 (ऋ - DEVANAGARI LETTER RRA) H is 094D (् - DEVANAGARI SIGN VIRAMA) C2 is either - 092F (य - DEVANAGARI LETTER YA) or 0939 (ह - DEVANAGARI LETTER HA)

Below are the specific context rules:

1. N: must be preceded only by either of specific set of Cs viz.
  - a. क (U+0915)
  - b. ख (U+0916)

- c. ञ (U+0917)
  - d. ज्ञ (U+091C)
  - e. ङ (U+0921)
  - f. ञ (U+0922)
  - g. ञ (U+092B)
2. H: must be preceded by C or N
  3. X: must be preceded by either of V, C, N or M
  4. D: must be preceded by either of V, C, N or M (Can be combined with rule for X)
  5. M: must be preceded either by C or N
  6. V: Can **NOT** be preceded by H

## 8 Contributors

Neo-brahmi Generation Panel members.

## 9 References

[INSCRIPT] Bureau of Indian Standards (BIS), "Enhanced Inscript Keyboard layouts" (IS 16350: 2016)

[NBGP] Neo-Brahmi Generation Panel

*<This is a paid resource managed by Bureau of Indian Standards. NBGP will try to get a copy of the same and then share the same with IP>*

## Appendix