

Proposal for a Devanagari Script Root Zone Label Generation Rule-Set (LGR)

LGR Version: 3.0

Date: 15th January 2018

Document version: 1.9

Authors: Neo-Brahmi Generation Panel [NBGP]

1 General Information/ Overview/ Abstract

This document lays down the Label Generation Rule Set 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-20180115.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: dévanâgarî

Native name of the script: देवनागरी

Maximal Starting Repertoire [MSR] version: 2

3 Background on Script and Principal Languages Using It

The script called Nagari or Devanagari is written from left to right. Historically it derives from the Brahmi alphabet of the Ashokan inscriptions. Devanagari is currently used for 11

out of 22 scheduled 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, Newar and Rajasthani and its dialects: Marwari, Mewati, Shekhawati, Bagri, Dhundhari, Harauti and Wagdi. 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, Chattisgarh, Jharkhand, Madhya Pradesh and Andaman & Nicobar Islands. The script is also used in Fiji to represent Fiji Hindi. Hindi is also a language of communication in Mauritius, Malaysia, England, Canada, South Africa, Indonesia as well as emigrant communities around the world. Nepali is the official language of Nepal as well as one of the official languages of the state of Sikkim in India. It is spoken by over 30 million people.

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

3.1 The Evolution of the Script

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

The evolution of Brahmi into present-day Devanagari involved intermediate forms, common to other scripts such as Gupta, and its two generates – Siddaṃ and Śāradā in the north and Grantha and Kadamba in the South. Devanagari can be said to have developed from the Kutila script, a descendant of the Gupta script, in turn a descendent of Brahmi. 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 Brahmi. 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 Devanagari from Brahmi 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 led to different regional styles. The development of the script is outlined below. Figure 1: Pictorial depiction of Evolution of Devanagari illustrates the stages in the evolution of the script¹.

| Period | Description |
|---------|---|
| 300 BCE | Mauryan: Early Brahmi form in 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 |
| 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 | ∪ | | ∩ | 𑀓 |
| 200 CE | ‡ | E | ∠ | J | ∪ | 𑀓 |
| 400 CE | † | E | ∪ | I | 𑀓 | 𑀓 |
| 600 CE | ‡ | E | ∪ | I | ∪ | 𑀓 |
| 800 CE | ‡ | ε | ∪ | I | ∪ | 𑀓 |
| 900 CE | ‡ | 𑀓 | 𑀓 | I | 𑀓 | 𑀓 |
| 1100 CE | ‡ | 𑀓 | 𑀓 | 𑀓 | 𑀓 | 𑀓 |
| 1300 CE | ‡ | 𑀓 | 𑀓 | 𑀓 | 𑀓 | 𑀓 |
| Modern | क | ज | म | र | स | अ |

Figure 1: Pictorial depiction of Evolution of Devanagari

¹ http://www.acharya.gen.in:8080/sanskrit/script_dev.php

3.2 Languages considered

Devanagari script is used by over 120 languages which makes it one of the most used script in the world. The languages using Devanagari as their primary script belong to varying geo-political scenarios as given below.

- designated as official (scheduled) languages of some countries
- used by communities living in urban areas
- used by communities living in rural yet accessible areas
- used by communities living in far-flung areas which are not easily connected either by roads or by communication mechanisms.

The information about the languages which are part of official (scheduled) languages of some countries was easily available. The information of languages which are used by communities living in urban areas was also easily obtainable. There was some effort needed to cover the languages which are spoken by communities living in rural yet accessible areas. However it was quite difficult to cover rest of the languages being spoken by the communities living in remote tribal areas which are generally not connected by road or by communication means. Defining the scope of language coverage was hence essential to limit the scope of the work to be undertaken for analysis of Devanagari LGR.

NBGP decided to employ “Expanded Graded Intergenerational Disruption Scale” [EGIDS] which is designed to measure status of the languages of the world in terms of endangerment or development. The EGIDS consists of 13 levels with each higher number on the scale representing a greater level of disruption to the intergenerational transmission of the language. NBGP decided to accommodate all the languages belonging to EGIDS Scale 1 to 4 for its analysis which represents languages in one form or the other are still in usage. Following are the descriptions² of those scales.

| Scale | Label | Description |
|-------|------------|---|
| 1 | National | The language is widely used between nations in trade, knowledge exchange, and international policy. |
| 2 | Provincial | The language is used in education, work, mass media, and government at the national level. |

² <https://www.ethnologue.com/about/language-status>

| | | |
|---|---------------------|--|
| 3 | Wider Communication | The language is used in education, work, mass media, and government within major administrative subdivisions of a nation. |
| 4 | Educational | The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education. |

Languages belonging to Level 5 and onwards are not in modern usage.

Below is the tabular representation of the languages that have been considered for the Devanagari LGR.

| EGIDS Scale 1 | EGIDS Scale 2 | EGIDS Scale 3 | EGIDS Scale 4 |
|---------------|---------------|---------------|----------------------|
| Hindi | Konkani | Bhatri | Bhojpuri |
| Nepali | Maithili | Halbi | Chhattisgarhi |
| | Marathi | Kinnauri | Dogri |
| | Sindhi | Kukna | Kashmiri |
| | | Panchpargania | Limbu |
| | | Sadri | Magahi |
| | | Wagdi | Sanskrit |
| | | | Santhali |
| | | | Tamang, Eastern |
| | | | Avadhi |
| | | | Newar |
| | | | Saraiki ³ |

Table 2: Languages considered under Devanagari LGR

Despite of being classified under EGIDS Scale 5, Boro language is also considered under the Devanagari LGR as it is one of the scheduled languages of India and is widely spoken.

³ Though listed in EGIDS scale 4, Saraiki is not covered by the NBGP. As per ethnologue, Devanagari script is "no longer in use" by the Saraiki community.

Ref: <https://www.ethnologue.com/language/skr>

Apart from the above-mentioned languages, Braj, Dhundari, Mundari, Kharia have also been considered for the analysis.

3.2.1 Case of Sanskrit:

Sanskrit is generally perceived as an archaic language used only in ancient religious texts. However, it is worth noting that there is a quite vibrant and active user community of Sanskrit in India which practices Sanskrit on day to day basis. Sanskrit is still taught in schools under various State and Central educational boards. There is increasing use of Sanskrit on social media as well. The same is reflected in EGIDS scale where Sanskrit is categorized in Scale 4 indicating status of the language as “Educational”.

3.3 The structure of written Devanagari

Devanagari is an alphasyllabary and the heart of the writing system is the 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 5.4.

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

3.3.1 The Consonants

Devanagari consonants have an implicit schwa⁴ /ə/ included in them. As per traditional classification they are categorized according to their phonetic properties (especially in terms of place plus manner of articulation). There are 5 Varga groups (classes) and one non-Varga group. Each Varga, which corresponds to Stops, 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.

| Varga | Unvoiced | | Voiced | | Nasal |
|--------------|-------------|-------------|-------------|-------------|-------------|
| | -Asp | +Asp | -Asp | +Asp | |
| Velar | क U+0915 | ख U+0916 | ग U+0917 | घ U+0918 | ङ U+0919 |

⁴ Although representing the implicit vowel as /a/ is more correct orthographically, the schwa /ə/, although not part of the orthographic system has been used since the /a/ would be misunderstood and read as अ/आ/।.

| | | | | | |
|------------------|-------------|-------------|-------------|-------------|-------------|
| Palatal | च U+091A | छ U+091B | ज U+091C | झ U+091D | ञ U+091E |
| Retroflex | ट U+091F | ठ U+0920 | ड U+0921 | ढ U+0922 | ण U+0923 |
| Dental | त U+0924 | थ U+0925 | द U+0926 | ध U+0927 | न U+0928 |
| Bi-labial | प U+092A | फ U+092B | ब U+092C | भ U+092D | म U+092E |

Table 3: Varga classification of consonants

| | | | | | | | | | |
|------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Non-Varga | य U+092F | र U+0930 | ल U+0932 | ळ U+0933 | व U+0935 | श U+0936 | ष U+0937 | स U+0938 | ह U+0939 |
|------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|

Table 4: Non-Varga consonants

3.3.2 The Implicit Vowel Killer: Halant⁵

All consonants have an implicit vowel (schwa) within them. A special sign is needed to denote that this implicit vowel is stripped off. This is known as the Halant "्" (U+094D).

The Halant thus joins two consonants and creates conjuncts, which can be generally from 2 to 4 consonant combinations. In rare cases it can join up to 5 consonants. However the notion of maximum number of consonants joining to form one akshar is 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 generic Top Level Domain [gTLD] which may have more than the observed maximum cannot be ruled out. Hence, in the LGR work, this limit will not be enforced⁶.

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 sign (Matra) is attached to the consonant. Since the consonant has a built in schwa, there are equivalent Matras for all vowels excepting the अ.

The correlation is shown as under:

⁵ 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.

⁶ This can be the case when a foreign language word, which admits a large number of consonants, is transliterated into Devanāgarī

| Vowel | Corresponding vowel sign (Matra) |
|-----------------------|--|
| अ U+0905 | |
| आ U+0906 | ा U+093E |
| इ U+0907 | ि U+093F |
| ई U+0908 | ी U+0940 |
| उ U+0909 | ु U+0941 |
| ऊ U+090A | ू U+0942 |
| ऋ U+090B | ृ U+0943 |
| ए U+090F | े U+0947 |
| ऐ U+0910 | ै U+0948 |
| ओ U+0913 | ो U+094B |
| औ U+0914 | ौ U+094C |
| अ U+0973 | ँ U+093A |
| आ U+0974 | ं U+093B |
| ए/अ U+090D/ U+0972 | ँ U+0945 |

| | |
|-------------|-------------|
| ॠ U+0960 | ॡ U+0944 |
| ॢ U+0911 | ॣ U+0949 |
| । U+0975 | ॥ U+094F |
| ० U+0976 | ॠ U+0956 |
| ॠ U+0977 | ॡ U+0957 |

Table 5: Vowels with corresponding Matras

Marathi uses ॢ (U+0972) instead of ॣ (U+090D).

3.3.4 The Anusvara (ँ - U+0902)

The Anusvara represents a homorganic nasal. It replaces a conjunct group of a Nasal Consonant+Halant+Consonant belonging to that particular varga. Before a non-varga consonant the Anusvara represents a nasal sound. Modern Hindi, Marathi and Konkani prefer the Anusvara to the corresponding Half-nasal:

सन्त vs. संत /sənt/ saint

चम्पा vs. चंपा /tʃəmpa/

U+0938 U+0928 U+094D U+0924 vs. U+0938 U+0902 U+0924

U+091A U+092E U+094D U+092A U+093E vs. U+091A U+0902 U+092A U+093E

3.3.5 Nasalization: Candrabindu (ँ - U+0901)

Candrabindu denotes nasalization of the preceding vowel as in आँख /ãkh/ eye (U+0906 U+0901 U+0916). Present-day Hindi users tend to replace the candrabindu by the Anusvara.

3.3.6 Nukta (ँ - U+093C)

The nukta sign is placed below a certain number of consonants to represent sounds found only in words borrowed from Perso-Arabic. It is pre-dominantly used in this manner in Bodo, Hindi, Kashmiri, Maithili, Santhali and Sindhi. It can be adjoined to "क" (U+0915), "ख"

(U+0916), "ग" (U+0917), "ज" (U+091C) and "फ" (U+092B) to show that words having these consonants with a nukta are to be pronounced in the Perso-Arabic style.

e.g. फ़िरोज़ /firoz/ (U+092B U+093C U+093F U+0930 U+094B U+091C U+093C)

It is also placed under "ड" (U+0921) and "ढ" (U+0922) to indicate flapped sounds

बढ़ /bədh/ (U+092C U+0922 U+093C)

Central Hindi Directorate, Ministry of HRD, Government of India Web Publication [109] "DEVANĀGARĪ ALPHABET AND ITS ROMANIZATION" clearly states such a use of Nukta in Hindi.

In Bodo it is adjoined to "ड" (U+0921) [110]. In Maithili it is adjoined to क (U+0915), ज (U+091C), "ड" (U+0921) and "ढ" (U+0922) [111]. In Sindhi, it is adjoined to "ख" (U+0916), "ग" (U+0917), "ज" (U+091C), "फ" (U+092B) "ड" (U+0921) and "ढ" (U+0922) [104].

In Kashmiri, it can also be adjoined to "च" (U+091A), "छ" (U+091B) and "ज" (U+091C) [108] to indicate the laterally released affricates.

चाय / čāy/ 'tea' (U+091A U+093C U+093E U+092F)

छल / čhal/ 'wash; Imperative ' (U+091B U+093C U+0932)

पोज़ / póz/ 'fact' (U+092A U+094A U+091C U+093C)

Normally a Nukta is appended to Consonants. However, Santali language uses Nukta in a unique way. The nukta is adjoined to following vowels and vowel signs

- a. आ (U+0906)
- b. ओ (U+0913)
- c. ा (U+093E)
- d. ो (U+094B)

3.3.7 Visarga (◌ः - U+0903) and Avagraha (s - U+093D)

The Visarga is frequently used in Sanskrit and represents a sound very close to /h/. दुःख /du:kh/ sorrow, unhappiness (U+0926 U+0941 U+0903 U+0916).

The Avagraha "s" (U+093D) creates an extra stress on the preceding vowel and is used in Sanskrit texts. It is rarely used in other languages using Devanagari. 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.

4.1 Guiding Principles

The NBGP adopts following broad principles for selection of code-points in the code-point repertoire across the board for all the scripts within its ambit.

4.1.1 Inclusion principles:

4.1.1.1 *Modern usage:*

Every character proposed should be in the everyday usage of a particular linguistic community. The characters which have been encoded in the Unicode for transcription purposes only or for archival purposes will not be considered for inclusion in the code-point repertoire.

4.1.1.2 *Unambiguous use:*

Every character proposed should have unambiguous understanding among the linguistic about its usage in the language.

4.1.2 Exclusion principles:

The main exclusion principle is that of Acknowledgement to Environmental Limitations. These comprise of protocols or standards which are pre-requisites to the Label Generation Rulesets. All further principles are in fact subsumed under these limitations but have been spelt out separately for the sake of clarity.

4.1.2.1 Acknowledgement to Environment Limitations:

The code point repertoire for root zone being a very special case, up the ladder in the protocol hierarchies, the canvas of available characters for selection as a part of the Root Zone code point repertoire is already constrained by various protocol layers beneath it. 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 character inclusion efforts 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 introduces exclusion of some characters out of Unicode repertoire from being part of the domain names.

Example: Devanagari Letter Qa "क़" (U+0958) is not allowed to be a part of domain name. Its decomposed form, i.e. Devanagari Letter Ka followed by Devanagari Sign Nukta "क़" (U+0915) + "ँ" (U+093C) can be used instead.

iii. Maximal Starting Repertoire:

The Root-zone LGR being a repertoire of the characters which are going to be used for creation of the 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.

Example: Devanagari Sign Avagraha "s" (U+093D) even if allowed by IDNA protocol, is not permitted in the Root Zone Repertoire as per the [MSR].

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.

4.1.2.2 No Punctuation Marks:

The TLDs being identifiers, punctuation markers present in Brahmi based languages such as Danda "।" (U+0964) and double Danda "॥" (U+0965) will not be included.

4.1.2.3 No Symbols and Abbreviations:

Abbreviations, weights and measures and other such iconic characters like Isshar "८" (U+09FA), Abbreviation sign "॰" (U+0970) etc. will not be included.

4.1.2.4 No Rare and Obsolete Characters:

There are characters which have been added to Unicode to accommodate rare forms especially like DEVANAGARI LETTER VOCALIC RR "ठ" (U+0960) and DEVANAGARI LETTER VOCALIC LL "ड" (U+0961) as well as their matra forms "ँ" (U+0944) and "ं" (U+0963). All such characters will not be included. This is in consonance with the Conservatism principle as laid down in the Root Zone LGR procedure.

4.1.2.5 No Stress Markers of Classical Sanskrit and Vedic:

Stress markers for classical Sanskrit e.g. DEVANAGARI STRESS SIGN UDATTA "◌̄" (U+0951) and DEVANAGARI STRESS SIGN ANUDATTA "◌̎" (U+0952) will not be included. This is also in consonance with the Letter principle as laid down in the Root Zone LGR procedure.

5 Repertoire

Section 5.1 provides the section of the [MSR] applicable to the Devanagari script on which the Devanagari code-point repertoire is based.

Section 5.2 details the code-point repertoire that the Neo-Brahmi Generation Panel [NBGP] proposes to be included in the Devanagari LGR.

5.1 Devanagari section of Maximal Starting Repertoire [MSR] Version 2

| | 090 | 091 | 092 | 093 | 094 | 095 | 096 | 097 |
|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | ॐ | ऐ | ठ | र | ी | ॐ | ऋ | ० |
| 1 | ँ | ऑ | ड | र | ु | ँ | ऌ | ॰ |
| 2 | ं | ओ | ढ | ल | ॡ | ॢ | ॣ | अँ |
| 3 | ः | ओ | ण | ळ | ॣ | ॥ | ॥ | अँ |
| 4 | ऐ | औ | त | ळ | ॥ | ॥ | ॥ | आँ |
| 5 | अ | क | थ | व | ॥ | ॥ | ॥ | औँ |
| 6 | आ | ख | द | श | ॥ | ॥ | ॥ | अु |
| 7 | इ | ग | ध | ष | ॥ | ॥ | ॥ | अु |
| 8 | ई | घ | न | स | ॥ | ॥ | ॥ | रु |
| 9 | उ | ड | न | ह | ँ | ख | ३ | ज़ |
| A | ऊ | च | प | ं | ो | ग | ४ | य |
| B | ऋ | छ | फ | ा | ो | ज़ | ५ | ग |
| C | ऌ | ज | ब | ॥ | ौ | ड़ | ६ | ज़ |
| D | ँ | झ | भ | ॥ | ॥ | ॥ | ॥ | ॥ |
| E | ऐ | ज | म | ा | ि | फ़ | ८ | ड |
| F | ए | ट | य | ि | ौ | य़ | ९ | ब |

Color convention⁷:

All characters that are included in the [MSR]
- Yellow background

PVALID in IDNA2008 but excluded from the
[MSR] - Pinkish background

Not PVALID in IDNA2008, or are ineligible
for the root zone (digits, hyphen) - White
background

Figure 2: Devanagari Code Page from [MSR]

⁷ This document needs to be printed in color for this to be read correctly.

5.2 Code Point Repertoire:

For each of the code points, language references have been given in the last column titled "Reference". For the entire coverage of Devanagari code points, references of Hindi, Marathi, Sanskrit, Sindhi and Kashmiri have been given. Though only five representative languages have been chosen for referencing, they together cover all the code-points required for all the languages that NBGP has considered as given in 3.2.

| Sr. No. | Unicode Code Point | Glyph | Character Name | Unicode General Category (gc) | Indic Syllabic Category | Example languages using the code-point (Not exhaustive list) | Language with lowest EGIDS scale using the code point | Reference |
|---------|--------------------|-------|-----------------------------|-------------------------------|-------------------------|---|---|---------------------------------|
| 1. | 0901 | ◌ु | DEVANAGARI SIGN CANDRABINDU | Mn | Candrabin du | Bodo, Hindi, Kashmiri, Konkani, Maithili, Marathi, Nepali, Santali and Sanskrit | 1 Hindi, Nepali | [0], [101], [102], [103] |
| 2. | 0902 | ◌ं | DEVANAGARI SIGN ANUSVARA | Mn | Anusvara (Bindu) | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103] |
| 3. | 0903 | ◌ः | DEVANAGARI SIGN VISARGA | Mc | Visarga | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103] |
| 4. | 0905 | अ | DEVANAGARI LETTER A | Lo | Vowel | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104] |

| | | | | | | | | |
|-----|------|---|-----------------------------------|----|-------|---|--------------------|--|
| 5. | 0906 | आ | DEVANAGARI LETTER AA | Lo | Vowel | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104] |
| 6. | 0907 | इ | DEVANAGARI LETTER I | Lo | Vowel | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104] |
| 7. | 0908 | ई | DEVANAGARI LETTER II | Lo | Vowel | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104] |
| 8. | 0909 | उ | DEVANAGARI LETTER U | Lo | Vowel | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104] |
| 9. | 090A | ऊ | DEVANAGARI LETTER UU | Lo | Vowel | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104] |
| 10. | 090B | ऋ | DEVANAGARI LETTER VOCALIC R | Lo | Vowel | Hindi, Marathi, Sanskrit | 1 Hindi | [0], [101], [102], [103] |
| 11. | 090D | ॠ | DEVANAGARI LETTER CANDRA E | Lo | Vowel | Hindi | 1 Hindi | [0], [101] |
| 12. | 090E | ऐ | DEVANAGARI LETTER SHORT E | Lo | Vowel | Kashmiri | 4 Kashmiri | [0], [105], [108] |
| 13. | 090F | ए | DEVANAGARI LETTER E | Lo | Vowel | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 14. | 0910 | ऐ | DEVANAGARI LETTER AI | Lo | Vowel | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |

| | | | | | | | | |
|-----|------|---|----------------------------------|----|-----------|---|--------------------|--|
| 15. | 0911 | ॐ | DEVANAGARI LETTER CANDRA O | Lo | Vowel | Hindi, Konkani, Marathi | 1 Hindi | [0], [100], [108] |
| 16. | 0912 | औ | DEVANAGARI LETTER SHORT O | Lo | Vowel | Kashmiri | 4 Kashmiri | [0], [105], [108] |
| 17. | 0913 | ओ | DEVANAGARI LETTER O | Lo | Vowel | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 18. | 0914 | औ | DEVANAGARI LETTER AU | Lo | Vowel | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 19. | 0915 | क | DEVANAGARI LETTER KA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 20. | 0916 | ख | DEVANAGARI LETTER KHA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 21. | 0917 | ग | DEVANAGARI LETTER GA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 22. | 0918 | घ | DEVANAGARI LETTER GHA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104] |
| 23. | 0919 | ङ | DEVANAGARI LETTER NGA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103] |
| 24. | 091A | च | DEVANAGARI LETTER CA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |

| | | | | | | | | |
|-----|------|---|------------------------------|----|-----------|---|--------------------|--|
| 25. | 091B | छ | DEVANAGARI LETTER CHA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 26. | 091C | ज | DEVANAGARI LETTER JA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 27. | 091D | झ | DEVANAGARI LETTER JHA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104] |
| 28. | 091E | ञ | DEVANAGARI LETTER NYA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103] |
| 29. | 091F | ट | DEVANAGARI LETTER TTA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 30. | 0920 | ठ | DEVANAGARI LETTER TTHA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 31. | 0921 | ड | DEVANAGARI LETTER DDA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 32. | 0922 | ढ | DEVANAGARI LETTER DDHA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104] |
| 33. | 0923 | ण | DEVANAGARI LETTER NNA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104] |
| 34. | 0924 | त | DEVANAGARI LETTER TA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |

| | | | | | | | | |
|-----|------|---|--------------------------|----|-----------|---|--------------------|--|
| 35. | 0925 | थ | DEVANAGARI LETTER THA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 36. | 0926 | द | DEVANAGARI LETTER DA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 37. | 0927 | ध | DEVANAGARI LETTER DHA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 38. | 0928 | न | DEVANAGARI LETTER NA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 39. | 092A | प | DEVANAGARI LETTER PA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 40. | 092B | फ | DEVANAGARI LETTER PHA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 41. | 092C | ब | DEVANAGARI LETTER BA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 42. | 092D | भ | DEVANAGARI LETTER BHA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 43. | 092E | म | DEVANAGARI LETTER MA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 44. | 092F | य | DEVANAGARI LETTER YA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |

| | | | | | | | | |
|-----|------|---|---------------------------------|----|-----------|--|--------------------|--|
| 45. | 0930 | र | DEVANAGARI LETTER RA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 46. | 0932 | ल | DEVANAGARI LETTER LA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 47. | 0933 | ळ | DEVANAGARI LETTER LLA | Lo | Consonant | Bodo, Konkani, Marathi, Nepali, Sanskrit | 1 Nepali | [0], [102], [103] |
| 48. | 0935 | व | DEVANAGARI LETTER VA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 49. | 0936 | श | DEVANAGARI LETTER SHA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 50. | 0937 | ष | DEVANAGARI LETTER SSA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104] |
| 51. | 0938 | स | DEVANAGARI LETTER SA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 52. | 0939 | ह | DEVANAGARI LETTER HA | Lo | Consonant | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [104], [105], [108] |
| 53. | 093A | . | DEVANAGARI VOWEL SIGN OE | Mn | Matra | Kashmiri | 4 Kashmiri | [11], [105], [108] |
| 54. | 093B | † | DEVANAGARI VOWEL SIGN OOE | Mc | Matra | Kashmiri | 4 Kashmiri | [11], [105], [108] |

| | | | | | | | | |
|-----|------|----|--|----|-------|---|--------------------|--|
| 55. | 093C | ◌̣ | DEVANAGARI SIGN NUKTA | Mn | Nukta | Bodo, Hindi, Kashmiri, Maithili, Santhali, Sindhi | 1 Hindi | [0], [101], [105], [108] |
| 56. | 093E | ◌ा | DEVANAGARI VOWEL SIGN AA | Mc | Matra | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103] |
| 57. | 093F | ◌ि | DEVANAGARI VOWEL SIGN I | Mc | Matra | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103] |
| 58. | 0940 | ◌ी | DEVANAGARI VOWEL SIGN II | Mc | Matra | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103] |
| 59. | 0941 | ◌ु | DEVANAGARI VOWEL SIGN U | Mn | Matra | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103] |
| 60. | 0942 | ◌ू | DEVANAGARI VOWEL SIGN UU | Mn | Matra | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103] |
| 61. | 0943 | ◌ृ | DEVANAGARI VOWEL SIGN VOCALIC R | Mn | Matra | Hindi, Marathi, Sanskrit | 1 Hindi | [0], [101], [102], [103] |
| 62. | 0945 | ◌ँ | DEVANAGARI VOWEL SIGN CANDRA E = candra | Mn | Matra | Hindi, Konkani, Marathi, Sanskrit | 1 Hindi | [0], [101], [100], [108] |
| 63. | 0946 | ◌ॅ | DEVANAGARI VOWEL SIGN SHORT E | Mn | Matra | Kashmiri | 4 Kashmiri | [0], [105], [108] |
| 64. | 0947 | ◌े | DEVANAGARI VOWEL SIGN E | Mn | Matra | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [105], [108] |

| | | | | | | | | |
|-----|------|----|--------------------------------------|----|--------------------|---|-----------------------|--|
| 65. | 0948 | ै | DEVANAGARI VOWEL SIGN AI | Mn | Matra | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103] |
| 66. | 0949 | ँ | DEVANAGARI VOWEL SIGN CANDRA O | Mc | Matra | Hindi, Konkani, Marathi | 1 Hindi | [0], [100], [108] |
| 67. | 094A | औ | DEVANAGARI LETTER SHORT O | Mc | Matra | Kashmiri | 4 Kashmiri | [0], [105], [108] |
| 68. | 094B | ो | DEVANAGARI VOWEL SIGN O | Mc | Matra | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [105], [108] |
| 69. | 094C | ौ | DEVANAGARI VOWEL SIGN AU | Mc | Matra | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [105], [108] |
| 70. | 094D | ् | DEVANAGARI SIGN VIRAMA | Mn | Halant / Virama | Most of the languages given in section 3.2 | 1 Hindi, Nepali | [0], [101], [102], [103], [105], [108] |
| 71. | 094F | ँ | DEVANAGARI VOWEL SIGN AW | Mc | Matra | Kashmiri | 4 Kashmiri | [0], [105], [108] |
| 72. | 0956 | ँ | DEVANAGARI VOWEL SIGN UE | Mn | Matra | Kashmiri | 4 Kashmiri | [11], [105], [108] |
| 73. | 0957 | ँ | DEVANAGARI VOWEL SIGN UUE | Mn | Matra | Kashmiri | 4 Kashmiri | [11], [105], [108] |
| 74. | 0972 | अँ | DEVANAGARI LETTER CANDRA A | Lo | Vowel | Konkani, Marathi, | 2 Konkani, Marathi | [9], [100], [108] |
| 75. | 0973 | अ | DEVANAGARI LETTER OE | Lo | Vowel | Kashmiri | 4 Kashmiri | [11], [105], [108] |
| 76. | 0974 | अँ | DEVANAGARI LETTER OOE | Lo | Vowel | Kashmiri | 4 Kashmiri | [11], [105], [108] |

| | | | | | | | | |
|-----|------|----|------------------------------|----|-----------|----------|------------|-----------------------|
| 77. | 0975 | औ | DEVANAGARI LETTER AW | Lo | Vowel | Kashmiri | 4 Kashmiri | [11], [105], [108] |
| 78. | 0976 | अु | DEVANAGARI LETTER UE | Lo | Vowel | Kashmiri | 4 Kashmiri | [11], [105], [108] |
| 79. | 0977 | अू | DEVANAGARI LETTER UUE | Lo | Vowel | Kashmiri | 4 Kashmiri | [11], [105], [108] |
| 80. | 097B | ग | DEVANAGARI LETTER GGA | Lo | Consonant | Sindhi | 2 Sindhi | [8], [104] |
| 81. | 097C | ज | DEVANAGARI LETTER JJA | Lo | Consonant | Sindhi | 2 Sindhi | [8], [104] |
| 82. | 097E | ड | DEVANAGARI LETTER DDDA | Lo | Consonant | Sindhi | 2 Sindhi | [8], [104] |
| 83. | 097F | ब | DEVANAGARI LETTER BBA | Lo | Consonant | Sindhi | 2 Sindhi | [8], [104] |

Table 6: Code point repertoire

Apart from the above individual code-points, the Neo-Brahmi Generation Panel also proposes some specific sequences which enable conditional inclusion of the "DEVANAGARI LETTER RRA" in the repertoire for enabling inclusion of "Eyelash Reph"⁸ construct.

| Sr. No. | Unicode Code Points | Sequence | Character Names | Unicode General Category (gc) | Example languages using the code-point (Not exhaustive list) | Reference |
|---------|----------------------|----------|---|-------------------------------|--|-----------------|
| 1. | 0931 094D 092F | य | DEVANAGARI LETTER RRA DEVANAGARI SIGN VIRAMA DEVANAGARI LETTER YA | Lo Mn Lo | Konkani, Marathi, Nepali | [106], [107] |

⁸ Unicode uses term "Eyelash Ra" instead. Since the construct that is formed by this sequence is a special form of Reph (which is otherwise formed by Normal Ra U+0930), the term "Reph" is used.

| | | | | | | |
|----|------|---|------------------------|----|--------------------------------|-----------------|
| 2. | 0931 | ह | DEVANAGARI LETTER RRA | Lo | Konkani, Marathi, Nepali | [106], [107] |
| | 094D | | DEVANAGARI SIGN VIRAMA | Mn | | |
| | 0939 | | DEVANAGARI LETTER HA | Lo | | |

Table 7: Sequences

5.3 Code points not included:

Following code points have not been included in the repertoire.

| Sr. No. | Unicode Code Point | Glyph | Character Name | Reason for exclusion |
|---------|--------------------|-------|----------------------------------|---|
| 1. | U+0904 | अे | DEVANAGARI LETTER SHORT A | Usage unknown. Not required explicitly by any language. |
| 2. | U+090C | ल्ल | DEVANAGARI LETTER VOCALIC L | Not in modern usage. Excluded as per conservatism principle. |
| 3. | U+0929 | न्न | DEVANAGARI LETTER NNNA | Not required in any spoken language. Required only for transcribing Dravidian alveolar n. |
| 4. | U+0934 | ळ | DEVANAGARI LETTER LLLA | Not required in any spoken language. Required only for transcribing Dravidian l. |
| 5. | U+0944 | ृ | DEVANAGARI VOWEL SIGN VOCALIC RR | Not in modern usage. Excluded as per conservatism principle. |
| 6. | U+0979 | ज़ | DEVANAGARI LETTER ZHA | Not required in any spoken language. Required only in transliteration of Avestan. |
| 7. | U+097A | ष | DEVANAGARI LETTER HEAVY YA | Usage unknown. Not required explicitly by any language. |

5.4 Structural Formation of Devanagari:

All the languages written in Brahmi derived scripts follow a particular way of formation of its words, known as "akshar". In the next section there are detailed akshar formation rules as applicable to representation of "Hindi" language when written in Devanagari Script. These rules need slight additions 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 Reph construct is required in Marathi, Konkani and Nepali but not in Hindi).

It is worth noting that the rules required for accommodation of additional languages in Devanagari ruleset apart from those required for Hindi are never in conflict with one another.

In section 0, the Whole Label Evaluation (WLE) rules are given which cover all the languages under the purview of the NBGP for Devanagari script.

5.5 Akshar formation rules for Hindi:

This section details the Akshar formation rules as applicable to Hindi. The first section lists the categories of the characters in the form of variables. In the rules, instead of their descriptive names, the variable names are used. The second section lists four operators along with their functions which are assumed while specifying the rules. The following two sections describe the two major categories of the Akshar formations first of which begins with the vowels and the second one with the consonants.

5.5.1 Variables involved

| | |
|-------|----------------------------|
| Dash | → Hyphen - |
| Digit | → Indo-Arabic digits [0-9] |
| C | → Consonant |
| M | → Matra |
| V | → Vowel |
| B | → Anusvara (Bindu) |
| D | → Candrabindu |
| X | → Visarga |
| H | → Halant / Virama |
| N | → Nukta |

5.5.2 Operators used:

| Symbol | Function |
|--------|----------|
|--------|----------|

| | |
|----|---------------------|
| | Alternative |
| [] | Optional |
| * | Variable Repetition |
| () | Sequence Group |

Table 8: Symbol functions

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

5.5.3 The Vowel Sequence

A vowel sequence begins with a vowel. It may be optionally followed by an Anusvara (B), Candrabindu (D) or a Visarga (X). The number of B, D or X which can follow a V in Devanagari are restricted to one.

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

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

Examples:

| Sequence Description | Sequence | Example | Constituting characters |
|----------------------|----------|---------------------------|-------------------------|
| Vowel | V | अ /a/ U+0905 | |
| Vowel + Anusvara | V[B] | अं /am/ U+0905 U+0902 | अ ँ U+0905 U+0902 |
| Vowel + Candrabindu | V[D] | अँ /am̐/ U+0905 U+0901 | अ ँँ U+0905 U+0901 |
| Vowel + Visarga | V[X] | अः /ah̐/ U+0905 U+0903 | अ ः U+0905 U+0903 |

Table 9

5.5.4 Consonant Sequence

A consonant sequence begins with a consonant. It may be optionally followed by a Nukta (N), Matra (M), Anusvara (B), Candrabindu (D), Visarga (X) or a Halant (H). The number of

instances of these characters occurring after a consonant is restricted to one. There is a possibility of further extension of the Consonant sequence after the N, M and H. Each of these has been discussed in the following sections:

1. A single consonant (C)

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

Examples:

| Sequence Description | Sequence | Example | Constituting characters |
|----------------------|----------|----------------------------|-------------------------|
| Consonant | C | क /ka/ U+0915 | <single character> |
| Consonant + Nukta | C[N] | क ः /kaḥ/ U+0915 U+093C | |

Table 10

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

C [M|B|D|X|H]

Examples:

| Sequence Description | Sequence | Example | Constituting characters |
|-------------------------|----------|----------|-------------------------|
| Consonant + Matra | C[M] | कि /ki/ | क ि U+0915 U+093F |
| Consonant + Anusvara | C[B] | कं /kaṁ/ | क ँ U+0915 U+0902 |
| Consonant + Candrabindu | C[D] | कँ /kaṃ/ | क ँ U+0915 U+0901 |
| Consonant + Visarga | C[X] | कः /kaḥ/ | क ः U+0915 U+0903 |

| | | | |
|--------------------|------|----------------------------|---------------------|
| Consonant + Halant | C[H] | क् /k/ (Pure Consonant) | क् U+0915 U+094D |
|--------------------|------|----------------------------|---------------------|

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 | Constituting characters |
|---------------------------------|----------|------------|-------------------------------|
| Consonant + Matra + Anusvara | CM[B] | कीं /kīm̐/ | क ि ं U+0915 U+0940 U+0902 |
| Consonant + Matra + Candrabindu | CM[D] | काँ /kāṃ/ | क ा ँ U+0915 U+093E U+0901 |
| Consonant + Matra + Visarga | CM[X] | कीः /kīḥ/ | क ि ः U+0915 U+0940 U+0903 |

Table 12

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

Example:

| Sequence Description | Sequence | Example | Constituting characters |
|---|----------|----------------|---|
| Consonant + Halant + Consonant + Halant + Consonant + Halant + Consonant | CHCHCHC | न्क्रय /nkrya/ | न् क् र् य U+0928 U+094D U+0915 U+094D U+0930 U+094D U+092F |

Table 13

Subsets:

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

Example:

| Sequence Description | Sequence | Example | Constituting characters |
|--|----------|-----------------|--|
| Consonant + Halant + Consonant + Matra | CHC[M] | क्की /kkī/ | क् क् की U+0915 U+094D U+0915 U+0940 |
| Consonant + Halant + Consonant + Anusvara | CHC[B] | क्कं /kkam/ | क् क् कं U+0915 U+094D U+0915 U+0902 |
| Consonant + Halant + Consonant + Candrabindu | CHC[D] | क्कँ /kkam̐/ | क् क् कँ U+0915 U+094D U+0915 U+0901 |
| Consonant + Halant + Consonant + Visarga | CHC[X] | क्कः /kkaḥ/ | क् क् कः U+0915 U+094D U+0915 U+0903 |

Table 14

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

Example:

| Sequence Description | Sequence | Example | Constituting characters |
|--|----------|---------------|--|
| Consonant + Halant + Consonant + Matra + Anusvara | CHCM[B] | क्कीं /kkīm/ | क् क् कीं U+0915 U+094D U+0915 U+0940 U+0902 |
| Consonant + Halant + Consonant + Matra + Candrabindu | CHCM[D] | क्कीँ /kkīm̐/ | क् क् कीँ U+0915 U+094D U+0915 U+0940 U+0901 |
| Consonant + Halant + Consonant + Matra + Visarga | CHCM[X] | क्कीः /kkīḥ/ | क् क् कीः U+0915 U+094D U+0915 U+0940 U+0903 |

Table 15

These are the basic akshar rules on which the overall Devanagari LGR is based. As languages other than Hindi are considered, some additional language specific characters

and rules are introduced. There are some additional finer aspects to these rules as one takes into account the digits, punctuations and special standalone characters like Avagraha. Those aspects are not discussed here as the [MSR] on which the LGRs are supposed to be based, excludes those characters.

6 Variants

There are no characters/character sequences in Devanagari which can be created by using the characters permitted as per the [MSR] and look exactly alike. However, Devanagari has ample cases of confusingly similar variants. The NBGP categorizes these confusingly similar variants in two groups.

Group 1: Confusing due to pure visual similarity

Group 2: Confusing due to deviation from normally perceived character formations by larger linguistic community

As advised by ICANN, no cases belonging to Group 1 are proposed, as there is another panel (String similarity assessment panel) entrusted to deal with such cases. The "Table 19: Visually confusable variants" in "Appendix A: Variants based on pure visual similarity" lists them.

Cases which belong to Group 2, however, are proposed to be considered as variants. These cases are not of mere visual similarity as they involve some deviations from the widely accepted norms of Devanagari Akshar formations. These can cause confusion even to a careful observer and hence being proposed as variants. Following is the brief description of these variants followed by variants in Table 16 and Table 17.

6.1 Vowel/Vowel sign followed by Nukta:

Santhali language has a unique requirement for Nukta character "◌̣" (U+093C) positioning which is not common in other Devanagari based languages. Santhali requires the Nukta character to follow certain Vowels and Matras. Complete representation of these Santhali combinations necessitated the Whole Label Evaluation rules (given in the Section 6.20) to be opened up for these specific cases. A regular non-Santhali user mostly cannot even anticipate possibility of such a combination and can confuse it for something else.

This gives rise to a possibility of creation of certain labels which can be deceptively similar to a majority of the Devanagari user-base. Being a unique case of homographic similarity, following variants are being proposed.

| Variant 1 | Variant 2 |
|-------------|--------------------|
| आ U+0906 | आ U+0906 U+093C |
| ओ U+0913 | ओ U+0913 U+093C |
| ा U+093E | ा U+093E U+093C |
| ो U+094B | ो U+094B U+093C |

Table 16: Proposed Variants - Set 1

6.2 Unique Vowels and Vowel Signs required for Kashmiri

Kashmiri when written in Devanagari script requires a unique set of Vowels and Vowel signs which only a Kashmiri speaker can understand. Majority of Devanagari users who are not conversant with Kashmiri can easily confuse them with some of the Vowels / Vowel signs which look similar to the Kashmiri ones. There are also cases where a Kashmiri Vowel / Vowel signs can be confused with certain Akshar formations. Hence they are being proposed as variants.

| Variant 1 | Variant 2 |
|-------------|--------------------|
| अ U+0973 | अ U+0905 U+0902 |
| आ U+093A | आ U+0902 |
| ई U+0974 | ई U+0906 U+0902 |
| ऀ U+093B | ऀ U+093E U+0902 |
| ऐ U+090E | ऐ U+0910 |
| ँ U+0946 | ँ U+0947 |
| औ U+0975 | औ U+0914 |
| ँ U+094F | ँ U+094C |

Table 17: Proposed Variants - Set 2

6.3 Halant ending (Only a discussion, not proposed as variants):

Another case of deceptive similarity to a majority of the Devanagari user-base is of a word ending in Halant "ँ" (U+094D) vis-à-vis the same word without the final Halant. As the function of Halant is of a vowel killer, coming at the end, many users tend to ignore the phonetic effect of its presence/absence. Majority of the users would pronounce both the words in the same way, thereby creating a perception of (false) equivalence. However, there also exists a section of community which clearly requires the final Halanta to achieve the peculiar phonetic effect of a truncated implicit vowel sound in the end. This section of community makes a clear distinction between two words (with and without the final Halant). It is for this reason; the final Halant is being accommodated in the Whole Label Evaluation rules for Devanagari.

As the final Halant is clearly visible, this is being considered as a variant belonging to Group 1 (i.e. of pure visual confusability) and hence not being proposed as variant case. Eventually, if needed, future NBGP revision may assess if this case needs to be considered as a variant depending on the prevalent experience.

6.4 Variant Disposition:

As variants mentioned in both (Table 16 and Table 17) categories are of confusingly similar, albeit of a peculiar nature, it is proposed that they be considered of "blocking" nature.

There is no preference among these variants. Whichever label containing either of these variants is chosen earlier, the other one equivalent variant label should be blocked.

6.5 Cross-script Variants:

A cross-script variant, also sometimes referred to as "Whole Label confusable", is the variant 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.

Every individual LGR under NBGP is supposed to provide a set of cross script variants it identifies with all other scripts under NBGP.

Devanagari script has a major set of possible cross-script variants only with the Gurmukhi script. Cases listed in Table 18 are of the variants that are proposed to be cross-script variants between Devanagari and Gurmukhi.

It is to be noted that none of the combinations listed in Table 18 are termed to be equivalents of each other semantically or otherwise. They are only grouped based on possible visual confusability.

| Devanagari | Gurmukhi |
|-------------------|-----------------|
| ं U+0902 | ਂ 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 |
| त U+0924 | ਜ U+0A1C |
| प U+092A | ਧ U+0A27 |

| | |
|--------------------|--------------------|
| भ U+092D | 𑂢 U+0A2E |
| म U+092E | 𑂣 U+0A38 |
| य U+092F | 𑂤 U+0A27 |
| र U+0930 | 𑂥 U+0A15 |
| व U+0935 | 𑂦 U+0A15 |
| ह U+0939 | 𑂧 U+0A35 |
| ि U+093F | 𑂨 U+0A3F |
| ी U+0940 | 𑂩 U+0A40 |
| ँ U+0945 | 𑂪 U+0A71 |
| े U+0947 | 𑂫 U+0A47 |
| ै U+0948 | 𑂬 U+0A48 |

| | |
|--|---|
| ॠ U+0956 |  OA41 |
| ॡ U+0957 |  OA42 |
| ष्टि U+092A U+094D U+091F U+093F | ਠਿ U+0A07 |
| ष्ठी U+092A U+094D U+091F U+0940 | ਠੀ U+0A08 |
| ष्टे U+092A U+094D U+091F U+0947 | ਠੇ U+0A0F |
| ष्ट U+092A U+094D U+091F | ਠ U+0A72 |
| त् U+0924 U+094D U+0924 | ਜ U+0A1C |

Table 18: Proposed Cross-script Devanagari-Gurmukhi Variants

In addition to above cases, Devanagari and Gurmukhi scripts have a possible set of cross-script variants which look similar but not similar enough to be recommended as cross-script variants. The "Table 20: Devanagari Cross-script Variants" in "Appendix B: Cross-script Variants" lists them.

7 Whole Label Evaluation Rules (WLE)

This section provides the WLEs that are required by all the languages mentioned in section 3.2 when written in Devanagari 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.

| | | |
|---|---|--|
| C | → | Consonant |
| M | → | Matra |
| V | → | Vowel |
| B | → | Anusvara (Bindu) |
| D | → | Candrabindu |
| X | → | Visarga |
| H | → | Halant / Virama |
| N | → | Nukta |
| S | → | Eyelash Reph (C1 H C2) |
| | | where |
| | | C1 is 0931 (ॠ - 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 WLE rules:

1. N: must be preceded only by either of specific set of Cs, Vs and Ms

The specific Cs are:

- a. क (U+0915)
- b. ख (U+0916)
- c. ग (U+0917)

- d. च (U+091A)
- e. छ (U+091B)
- f. ज (U+091C)
- g. ङ (U+0921)
- h. ढ (U+0922)
- i. फ (U+092B)

The specific Vs are:

- a. ञ (U+0906) (Required in Santhali language)
- b. ण (U+0913) (Required in Santhali language)

The specific Ms are:

- a. ण (U+093E) (Required in Santhali language)
- b. ो (U+094B) (Required in Santhali language)

2. H: must be preceded by C or CN
3. M: must be preceded by C or CN
4. X: must be preceded by either of V, C, N or M
5. B: must be preceded by either of V, C, N or M
6. D: must be preceded by either of V, C, N or M
7. V: Can **NOT** be preceded by H (details in "Case of V preceded by H")

Case of Eyelash Reph:

In the WLE rules, there is no specific mention of the Eyelash Reph for two reasons:

1. As the U+0931 is added as a part of permissible sequences in Table 7: Sequences, it gets permitted only with the specific sequences.

2. The last characters of both the sequences of which the U+0931 is part, are consonants. As the Eyelash-Reph can take all the combinations as that of a consonant, no specific handling in terms of context rule is required.

Case of V preceded by H:

There could be cases involving multi-word domains where V may need to be allowed to follow an H

e.g. आम्रअचार /a:m əcha:r/ (U+0906 U+092E U+094D U+0905 U+091A U+093E U+0930)
(meaning: *Mango pickle*)

This is the case where two different words are joined together first of which ends in an H and the second word begins with a V. Some sections of the linguistic community require the explicit presence of H for full representation of the sound intended. However, by and large, the form of the first word without an H is considered enough for full representation of the sound intended for the first word.

This is a unique situation necessitated by the lack of hyphen, space or the Zero Width Non-joiner character in the permissible set of characters in the Root zone repertoire. Otherwise, V is never required to be allowed to follow an H. Permitting this may create a perceptive similarity among two labels (with and without H) for majority of the linguistic community, hence this is explicitly prohibited by the NBGP.

In future if required, depending on the prevailing requirements by the community, the future NBGP may consider revisiting this rule.

8 Contributors

NBGP Co-chairs: Dr. Udaya Narayan Singh, Mr. Mahesh D Kulkarni and Dr. Ajay Data

Following is the full list of NBGP members with their Language expertise.

| Name | Language Expertise |
|--------------------------------|-----------------------------------|
| Udaya Narayana Singh | Bengali, Maithili, Hindi, English |
| Ajay Data | Hindi |
| Mahesh D. Kulkarni | Marathi, Hindi |
| Anupam Agrawal | Hindi, Bengali |
| Akshat S. Joshi | Hindi, Marathi |
| Abhijit Dutta | Bengali, Hindi |
| Neha Gupta | Hindi |
| Nishit Jain | Hindi |
| Prabhakar Pandey | Hindi |
| Raiomond Doctor | English, Hindi, Marathi, Gujarati |
| N. DeivaSundaram | Tamil |
| Shantaram S. Warde Walawalikar | Konkani |
| Bal Krishna Bal | Nepali |
| Ganesh Murmu | Santali |
| Balaram Prasain | Nepali |
| Rajib Chakraborty | Bangla (Bengali) |
| Gurpreet Singh Lehal | Panjabi |
| Saroja Bhate | Sanskrit |
| Shambhu Kumar Singh | Maithili |
| SwarnaPrabha Chainary | Bodo |
| Ghanashyam Nepal | Nepali |

| | |
|-----------------------|--|
| Kalyan Vasudeo Kale | Marathi |
| Shashi Pathania | Dogri |
| Santhosh Thottingal | Malayalam, Sourashtra, Tamil |
| Uma Maheshwar G | Telugu |
| Girish Chandra Mishra | Odia |
| K. C. Tikayat ray | Odia |
| Debajit Sharma | Assamese |
| Basanta Kumar Panda | Odia |
| Arvind Bhandari | Gujarati |
| Harish Chowdhary | Hindi |
| Chitrita Chatterjee | Multiple languages represented by members of IAMAI |
| U.B. Pavanaja | Kannada |
| Hempal Shrestha | Nepali, Newari |
| Suraj Adhikari | Nepali |
| Gangadhar Panday | Telugu |
| Vinay Murarka | Hindi |
| Mukesh Saini | Hindi |
| Jay Paudyal | Hindi |
| Pawan Chitrakar | Nepali |
| Nirajan Parajuli | Nepali |
| Uttam Shrestha Rana | Nepali |
| Dev Dass Manandhar | Nepali, Newari |
| Bhim Dhoj Shrestha | Nepali, Newari |
| Rajiv Kumar | Hindi |
| Shubham Saran | Hindi |

| | |
|---------------------------|-----------|
| Anivar A. Aravind | Malayalam |
| Shanmugam R | Tamil |
| Prasad PK | Malayalam |
| Sinnathambi Shanmugarajah | Tamil |

In addition, following members externally gave inputs to NBBP for the respective languages/scripts.

| Name | Language/Script Expertise |
|-----------------------------|----------------------------------|
| Ajit Kumar | Awadhi, Braj Language |
| Basil Baa | Sadri Language |
| Basil Kiro | Kharia Language |
| Biswa Limbu | Limbu Language |
| Devendra Kumar Devesh | Bhojpuri Language |
| Dinbandhu Mahto | Panchpargania Language |
| Dr. Birendra Kumar Soy | Mundari Language |
| Dr. Dinesh Kumar Shrivastav | Magahi Language |
| Dr. Harvinder Kaur | Gurmukhi Script |
| Dr. Laxmi Prasad Khatiwada | Nepali Language |
| Jagannath Singh | Panchpargania Language |
| Narendra Kumar Negi | Kinnauri Language |
| Prateek Harshwal | Wagdi and Dhundhari Language |
| Urmila Harshwal | Wagdi Language |
| Rayem Olem Dungdung | Sadri Language |
| Tej Man Angdembe | Limbu Language |
| Amar Tumyahang | Limbu Language |

| | |
|-----------------------|-----------------|
| Amrit Yonjan | Tamang Language |
| Indra Kumar Tamang | Tamang Language |
| Dipika Sangma Narzary | Bodo Language |
| Devdass Manandhar | Newar |
| Dr K.P. Lekhwani | Sindhi |
| Harihar Vaishnav | Halbi |

Full Updated list of NBGP members is available at :

<https://community.icann.org/display/croscomlgrprocedure/Neo-Brahmi+GP>

9 References

[MSR] Integration Panel, "Maximal Starting Repertoire — MSR-2 Overview and Rationale", 14 April 2015 <https://www.icann.org/en/system/files/files/msr-2-overview-14apr15-en.pdf>

[EGIDS] Expanded Graded Intergenerational Disruption Scale, <https://www.ethnologue.com/about/language-status> (Accessed on 13th Nov. 2017)

[NBGP] Neo-Brahmi Generation Panel

[gTLD] generic Top Level Domain

[0] The Unicode Standard 1.1, <http://www.unicode.org/versions/Unicode1.1.0/> (Accessed on 12th Dec. 2017)

[8] The Unicode Standard 5.0, <http://www.unicode.org/versions/Unicode5.0.0/> (Accessed on 12th Dec. 2017)

[9] The Unicode Standard 5.1, <http://www.unicode.org/versions/Unicode5.1.0/> (Accessed on 12th Dec. 2017)

[11] The Unicode Standard 6.0, <http://www.unicode.org/versions/Unicode6.0.0/> (Accessed on 12th Dec. 2017)

- [100] DEVANĀGARĪ VIP TEAM: VARIANT ISSUES REPORT, 3rd Oct. 2011,
<https://archive.icann.org/en/topics/new-gtlds/devanagari-vip-issues-report-03oct11-en.pdf> (Accessed on 10th Oct. 2017)
- [101] Omniglot, "Hindi", <https://www.omniglot.com/writing/hindi.htm> (Accessed on 10th Oct. 2017)
- [102] Omniglot, "Marathi", <https://www.omniglot.com/writing/marathi.htm> (Accessed on 10th Oct. 2017)
- [103] Omniglot, "Sanskrit", <https://www.omniglot.com/writing/sanskrit.htm> (Accessed on 10th Oct. 2017)
- [104] Omniglot, "Sindhi", <https://www.omniglot.com/writing/sindhi.htm> (Accessed on 10th Oct. 2017)
- [105] Omniglot, "Kashmiri", <https://www.omniglot.com/writing/kashmiri.htm> (Accessed on 10th Oct. 2017)
- [106] Unicode 10.0.0, "South and Central Asia-I, Page 456 (R5 and R5a) ",
<http://www.unicode.org/versions/Unicode10.0.0/ch12.pdf> (Accessed on 13th Nov. 2017)
- [107] Unicode Indic Group, "Devanagari Eyelash Ra",
<http://unicode.org/~emuller/iwg/p8/utcdoc.html> (Accessed on 13th Nov. 2017)
- [108] M.K. Raina, "How to read and write Kashmiri in Devanagari?",
<http://www.koshur.org/pdf/Let%20Us%20Learn%20Kashmiri.pdf> (Accessed on 12th Dec. 2017)
- [109] Central Hindi Directorate-Ministry of HRD-Govt. of India, "DEVANĀGARĪ ALPHABET AND ITS ROMANIZATION",
http://hindinideshalaya.nic.in/english/hindi_organ/devnagarithesymbols.html (Accessed on 12th Dec. 2017)
- [110] Omniglot, "Bodo", <https://www.omniglot.com/writing/bodo.htm> (Accessed on 12th Dec. 2017)
- [111] Omniglot, "Maithili", <https://www.omniglot.com/writing/maithili.htm> (Accessed on 12th Dec. 2017)

10 Books, articles and webographies consulted

Following is a thematically sorted set of documents, books, articles and webographies consulted in the drafting of this report

10.1 WRITING SYSTEMS

1. Dillinger. D., *The Alphabet. A Key to the History of Mankind*. 3rd Edition in 2 Volumes. Hutchison. London. 1968.

10.2 DEVANĀGARĪ

1. Agrawala, V. S. (1966). *The Devanāgarī script*. In: *Indian Systems of Writing*. (Pp. 12-16) Delhi: Publications Division.
2. Agyeya, Sacchidanand Hiranand Vatsyayan. 1972. *Bhavanti*. Delhi: Rajpal and Sons.
3. Beames, John. 1872-79. *A Comparative Grammar of the Modern Aryan Languages of India*. 3 vols. London, Trubner and Co. [Reprinted by Munshiram Manoharlal, New Delhi, 1966.]
4. Bhatia, Tej K. 1987. *A History of the Hindi Grammatical Tradition: Hindi-Hindustani Grammar, Grammarians, History and Problems*. Leiden/New York: E. J. Brill.
5. Bright, W. (1996). *The Devanāgarī script*. In P. Daniels and W. Bright (eds), *The World's Writing Systems*. (Pp. 384-390). New York: Oxford University Press.
6. Cardona, George. 1987. *Sanskrit*. In *The World's Major Languages*. Bernard Comrie (ed.). London: Croom Helm. 448-469.
7. Dwivedi, Ram Awadh. 1966. *A Critical Survey of Hindi Literature*. Delhi: Motilal Banarsidass.
8. Faruqi, Shamsur Rahman. 2001. *Early Urdu Literary Culture and History*. Delhi: Oxford University Press.
9. Guru, Kamta Prasad. 1919. *Hindi Vyakaran*. Varanasi: Nagari Pracharini Sabha. (1962 edition).
10. Kachru, Yamuna. 1965. *A Transformational Treatment of Hindi Verbal Syntax*. London: University of London Ph.D. dissertation (Mimeographed).
11. Kachru, Yamuna. 1966. *An Introduction to Hindi Syntax*. Urbana: University of Illinois, Department of Linguistics.

12. Kalyan Kale and Anjali Soman, 1986. Learning Marathi. Shri Vishakha Prakashan, Pune :
13. McGregor, R. S. (1977). Outline of Hindi Grammar. 2nd ed. Delhi: Oxford University Press.
14. McGregor, R. S. 1972. Outline of Hindi Grammar with Exercises. Delhi: Oxford University Press.
15. McGregor, R. S. 1974. Hindi Literature of the Nineteenth and Early Twentieth Centuries. Wiesbaden: Harrassowitz.
16. McGregor, R. S. 1984. Hindi Literature from Its Beginnings to the Nineteenth Century. Wiesbaden: Harrassowitz.
17. Pandey, P. K. (2007). Phonology-orthography interface in Devanāgarī for Hindi. *Written Language and Literacy*, 10 (2): 139-156. 2007.
18. Rai, Amrit. 1984. A House Divided. The Origin and Development of Hindi/Hindavi. Delhi: Oxford University Press.
19. Sharad, Onkar. 1969. Lohiya ke Vicar. Allahabad: Lokbharati Prakashan.
20. Singh, A. K. (2007). Progress of modification of Brāhmī alphabet as revealed by the inscriptions of sixth-eighth centuries. In P.G. Patel, P. Pandey and D. Rajgor (eds), *The Indic Scripts: Paleographic and Linguistic Perspectives*. (Pp. 85-107). New Delhi: DK Printworld.
21. Sproat, R. (2000). *A Computational Theory of Writing Systems*. Cambridge University Press.
22. Tiwari, Pandit Udaynarayan. 1961. Hindi Bhasha ka Udgam aur Vikas [The Origin and Development of the Hindi Language]. Prayag: Leader Press.
23. Verma, M. K. 1971. *The Structure of the Noun Phrase in English and Hindi*. Delhi: Motilal Banarsidass.

10.3 INDIC COMPUTING SPECIFIC

1. IS 10401: 8-bit code for information interchange. 1982
2. IS 10315: 7-bit coded character set for information interchange. 1985
3. IS 12326: 7-bit and 8-bit coded character sets-Code extension techniques. 1987

4. ISO 15919, Information and documentation - Transliteration of Devanāgarī and related Indic scripts into Latin characters. 2001
5. ISO 2375: Procedure for registration of escape sequences. 2003
6. ISO 8859: 8-bit single-byte coded graphic character sets - Parts 1-13. 1998-2001
7. IDN POLICY http://mit.gov.in/sites/upload_files/dit/files/India-IDN-Policy.pdf

11 Appendix A: Variants based on pure visual similarity

| Variant 1 | Variant 2 |
|-------------|---------------------|
| क U+0915 | क़ U+0915 U+093C |
| ख U+0916 | ख़ U+0916 U+093C |
| ग U+0917 | ग़ U+0917 U+093C |
| च U+091A | च़ U+091A U+093C |
| छ U+091B | छ़ U+091B U+093C |
| ज U+091C | ज़ U+091C U+093C |
| ड U+0921 | ड़ U+0921 U+093C |
| ढ U+0922 | ढ़ U+0922 U+093C |
| फ U+092B | फ़ U+092B U+093C |

Table 19: Visually confusable variants

12 Appendix B: Cross-script Variants

Devanagari script has a major set of possible cross-script variants with the Gurmukhi script. The Table 20 lists them.

In addition to Gurmukhi,

It is to be noted that none of the combinations listed in Table 20 are termed to be equivalents of each other semantically or otherwise. They are only grouped based on possible visual confusability.

At first they may not look exactly the same, however, in the given context e.g. in browser bar as a part of a domain name, or as a single word where there is no surrounding text from the same script for distinguishing, they can create visual confusion.

A label can be considered to have a cross-script variant label only if "all" the constituent characters/aksharas have an equivalent confusable in the other script. If there is even one single character/akshara which does not have an equivalent visual confusable in other script, it essentially provides a visually distinguishability and hence a non-confusable string.

| Devanagari | Variant | From script |
|-------------------|----------------|--------------------|
| ॐ U+0903 | ॐ U+0A83 | Gujarati |
| ॐ U+0903 | ॐ U+0C03 | Telugu |
| ॐ U+0903 | ॐ U+0C83 | Kannada |
| ॐ U+0903 | ॐ U+0D03 | Malayalam |
| ॐ U+0903 | ॐ U+0A28 | Sinhala |

| | | |
|-------------|-------------|-----------------|
| ठ U+0920 | ਠ U+0A28 | Gurmukhi |
| ਠ U+0920 | ਰ U+0A30 | Gurmukhi |
| ड U+0921 | ਡ U+0A21 | Gurmukhi |
| ड U+0921 | ਤ U+0A24 | Gurmukhi |
| ढ U+0922 | ਢ U+0A22 | Gurmukhi |

Table 20: Devanagari Cross-script Variants