# Introduction
# to
# Label Generation Ruleset
# (LGR)

Neo-Brahmi scripts perspective

By,
Akshat Joshi
C-DAC GIST
24th May 2017
Kathmandu F2F Meeting

# Why the name Neo-Brāhmī ?

## Background:

- ICANN began this “IDN Variant TLDs” initiative with six generation panels, based on specific scripts from major writing systems.
  - Latin, Greek, Cyrillic, CJK, Arabic and **Devanagari**
- Devanagari was representative of the **Brāhmī** family scripts
- The Brāhmī script
  - progenitor of all scripts used to write Modern Indo-Aryan languages
  - Dravidian
  - to a lesser extent scripts of the Tibeto-Burman and Munda families.
  - Also adopted by a large number of cultures in Southeast Asia to transcribe their languages: Burmese, Thai, Lao, Khmer (in South-East Asia), and others in Central Asia
- The **Neo-Brāhmī** group is so named to cover all such scripts used today and which are based on Brāhmī

# Features of Brāhmī based scripts

- Brāhmī is written from **left to right**
- It has an angular shape. As it evolved this angular feature was gradually replaced by rounded shapes in cultures where palm leaves where used as a medium of written communication.
- The main feature of Brāhmī is the written syllable or **akshara** a concept admitting a **full Consonant or Vowel as a node**
- Vowels admit Vowel Modifiers such as nasals or vowel lengtheners
- **Consonants** are at times **modified by** a combining mark functioning as "vowel-killer" (termed **Halanta**), truncating the following vowel and thereby constituting a conjunct
- In turn these can be modified by vowel signs: Matras and further by Nasals or Vowel lengtheners
- The **adjuncts** to the Vowel or Consonant nodes are **appended in a strict rule-order**
- **This feature has been remarkably stable** over the evolution of Brāhmī and has been followed by all the later Indic and Southeast scripts derived from the script.

# Principal Neo-Brāhmī Languages South Asian Scripts

- **Devanāgarī**: 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. 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.
- **Gujarati**: Used for writing Gujarati and Kacchi, Gujarati is extensively spoken in large parts of Africa, Madagascar, UK and the USA as well as by emigrant communities around the world.
- Gurumukhi which evolved separately in the Northern family is used to write the Punjabi language in the Indian state of Punjab and elsewhere in India. Gurmukhi stabilised around the 16th century when it was used to transcribe the holy Granth Sahib.
- **Bengali**: Often termed as Bangla by linguists and grammarians is historically related and similar in design to the Devanāgarī script and with one or two exceptions has the same consonant and vowel set.  Bengali is used to transcribe quite a few languages of which the most prominent are Assamese and Manipuri. The former differs from Bengali in a few consonant characters. The same is the case with Manipuri which today is also written in Meetei Mayek.

# Principal Neo-Brāhmī Languages South Asian Scripts

- **Oriya [Odia]** can be traced back to the Ashokan inscriptions: 3rd century B.C.E. Because of the prevalence of a large number of tribal languages belonging to the Munda and Dravidian families in the state of Odisha (Orissa), the Oriya script is used in writing these languages.
- Sinhala used for writing Sinhala language and at times also Pali, is derived from Brāhmī as early as the third-second century B.C.E. Although it belongs historically to the Northern family, it has been considerably influenced by the early Grantha script of South India.
- **Tamil** (also spelt as "Tamizh") More than any other script derived from Brāhmī, it is highly alphabetical in nature and admits no ligatures with the exception of two consonant conjuncts. Apart from being the official language of Tamil Nadu, Tamil is also an official and national language of Sri Lanka and one of the official languages of Singapore. Tamil is also spoken by significant minorities in Malaysia, England, Mauritius, Canada, South Africa, Fiji, Indonesia, as well as emigrant communities around the world.
- **Kannada and Telugu** are closely related scripts used to write two Dravidian languages: Kannada in the state of Karnataka, and Telugu in Telangana and Andhra. Over the centuries, Brāhmī evolved with marked characteristics in the south. Around the tenth century, these crystallised into the Old Kannada script, used where both Kannada and Telugu are now spoken. By around 1500, this script divided into Kannada and Telugu. As a result, there are very few differences between these two scripts.
- **Malayalam**: Subject to reforms, modern Malayalam has introduced alphabetic writing into the script, although the main structure of Malayalam still adheres to the akṣara.

# Current status of Neo-Brahmi GP Work

- Devanagari LGR is almost ready and with some final touches, can be sent to the Integration Panel
- We can discuss the same at length in this F2F meeting as well
- Similar exercise can be immediately undertaken for other scripts

Before starting with the

Devanagari LGR

let us take a look at

Akshar Formalism that binds brahmi based scripts

# Character classification

## Components of the Syllable

- **Consonants(C) :**
  - क ख ग घ ङ च छ ज झ ञ ट ठ ड ढ ण त थ द ध न ऩ य य़ र ऱ ल ळ ऴ व श ष स ह ॻ ॼ ॾ ॿ
- **Vowels (V) :**
  - ऄ अ आ इ ई उ ऊ ऋ ऍ ऎ ए ऐ ऑ ऒ ओ औ ॲ ॳ ॴ ॵ ॶ ॷ
- **Vowel Signs / Matras (M) :**
  - ◌ॆ ◌ा ◌ि ◌ी ◌ु ◌ू ◌ृ ◌े ◌ै ◌ॉ ◌ॊ ◌ो ◌ौ ◌ॅ ॕ ऻ ॏ ॖ ॗ
- **Vowel modifiers (D) :** ◌ँ ◌ं ◌ः
- **Halant (H) :** ◌्
- **Nukta (N) :** ◌़

# Formalism at a glance …

# Formalism Illustrated…

- **Variables :**

| | | |
|---|---|---|
| Dash → | Hyphen - | |
| Digit | → | Indo-Arabic digits [0-9] |
| C | → | Consonant |
| V | → | Vowel |
| M | → | Matra |
| D | → | Anusvara/Bindi/Tippi/Sunna |
| B | → | Chandrabindu/Anunasika/Arasunna |
| X | → | Visarga/Aytham |
| H | → | Halant/Chandrakala/Virama |
| A | → | Addak |
| N | → | Nukta |
| Y | → | Avagraha/Praslesham |
| L | → | Chillu |
| Z | → | Khanda Ta |
| k | → | Number of possible Consonant Halanta Sequences |

# Formalism Illustrated…

- Formalism Operators :

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

# Formalism Illustrated...

**The Formalism:**

```
Consonant-Syllable →

        *k(C[N]H) C[N] [H|D|B|X|BD|BX|M[D|B|X|BD|BX]]
                       | [CH]Z
                       | L[HC[D|H|M[D]]]
                       | AC[D|X|M[D|X]]

Vowel-Syllable     →   V[D|B|X|BD|BX]

Syllable           →   Consonant-Syllable [Y] | Vowel-Syllable[Y]

IDN-Label      →   (Syllable | digit)*([dash](Syllable | digit))
```

# Formalism Illustrated..

Consonant-Syllable :

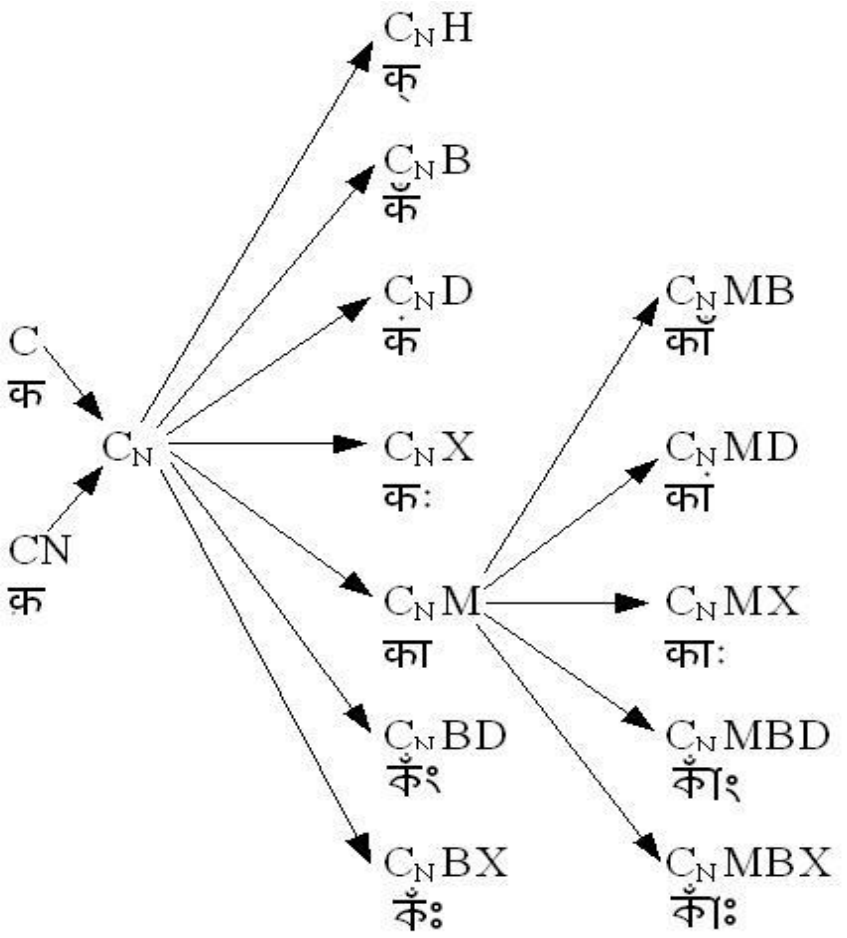
```
*k(C[N]H) C[N] [ H|D|B|X|BD|BX|M[D|B|X|BD|BX] ]
        |       [CH] Z
        |       L[ HC [ D | H | M[D] ] ]
        |        AC[ D | X | M[D|X] ]
```

# ABNF Illustrated..

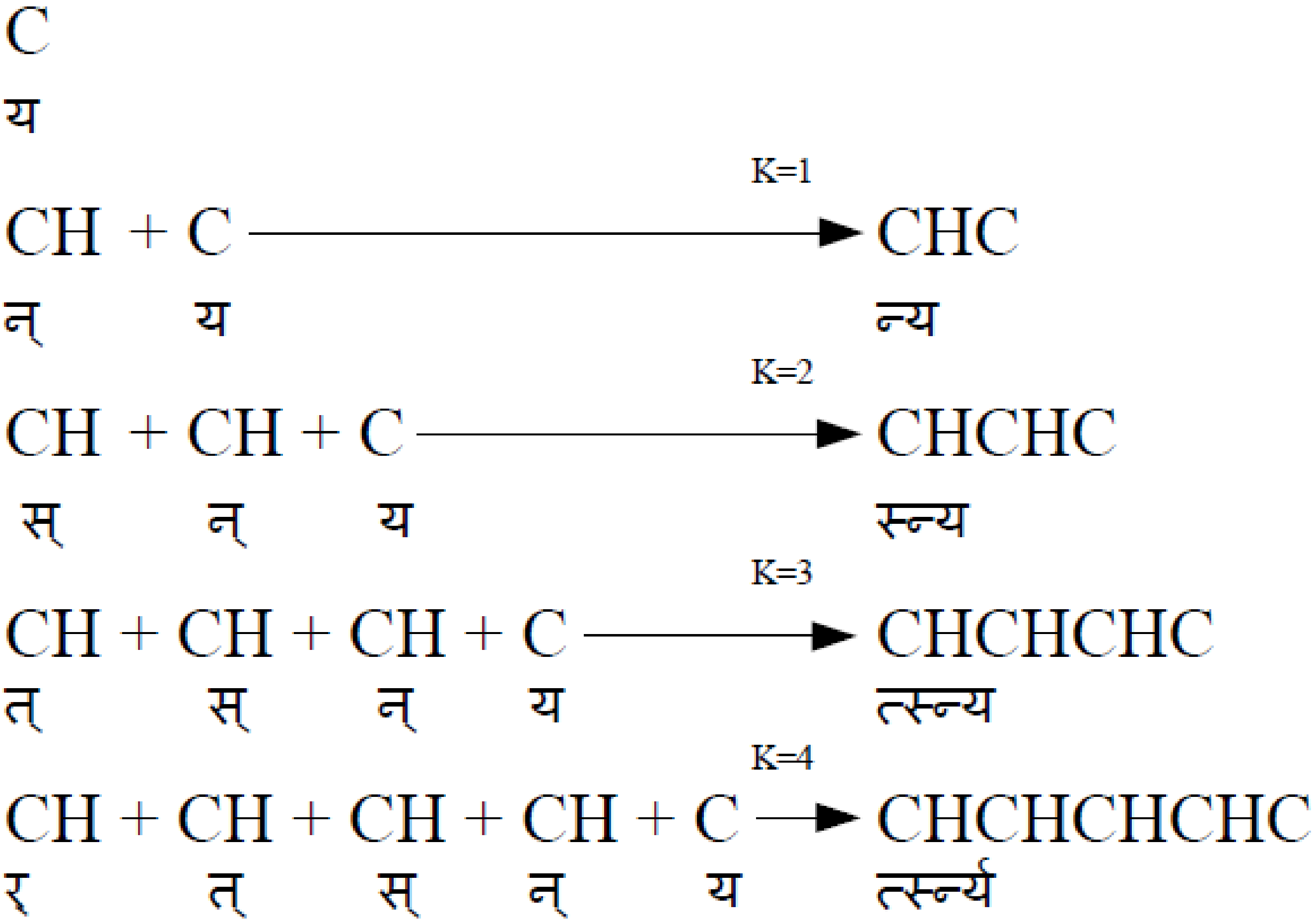
Consonant-Syllable :

```
   *k(C[N]H) C[N] [
H|D|B|X|BD|BX|M[D|B|X|BD|BX] ]
   |   [CH] Z
   |   L[ HC [ D | H | M[D] ] ]
   |    AC[ D | X | M[D|X] ]
```

*k(C[N]H) C[N] [ H|D|B|X|BD|BX|M[D|B|X|BD|BX] ]



*k(C[N]H) C[N] [ H|D|B|X|BD|BX|M[D|B|X|BD|BX] ]



$C$

य

$CH + C \xrightarrow{K=1} CHC$

न्  य  न्य

$CH + CH + C \xrightarrow{K=2} CHCHC$

स्  न्  य  स्न्य

$CH + CH + CH + C \xrightarrow{K=3} CHCHCHC$

त्  स्  न्  य  त्स्न्य

$CH + CH + CH + CH + C \xrightarrow{K=4} CHCHCHCHC$

र  त्  स्  न्  य  त्स्न्र्य

# ABNF Illustrated...

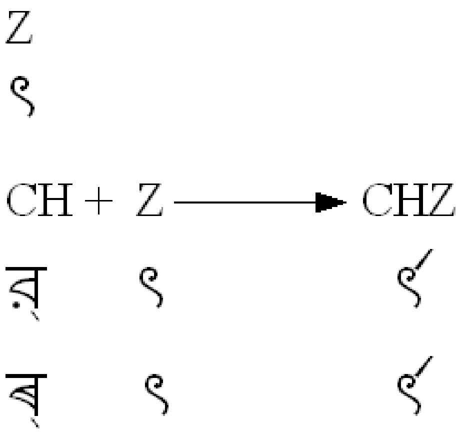
Consonant-Syllable :

```
   *k(C[N]H) C[N] [
H|D|B|X|BD|BX|M[D|B|X|BD|BX] ]
   |   [CH] Z
   |   L[ HC [ D | H | M[D] ] ]
   |    AC[ D | X | M[D|X] ]
```

# Consonant Syllable continues...

Syllable with Khanda Ta only exists in Bangla and Assamese language.

[CH] Z



# ABNF Illustrated…

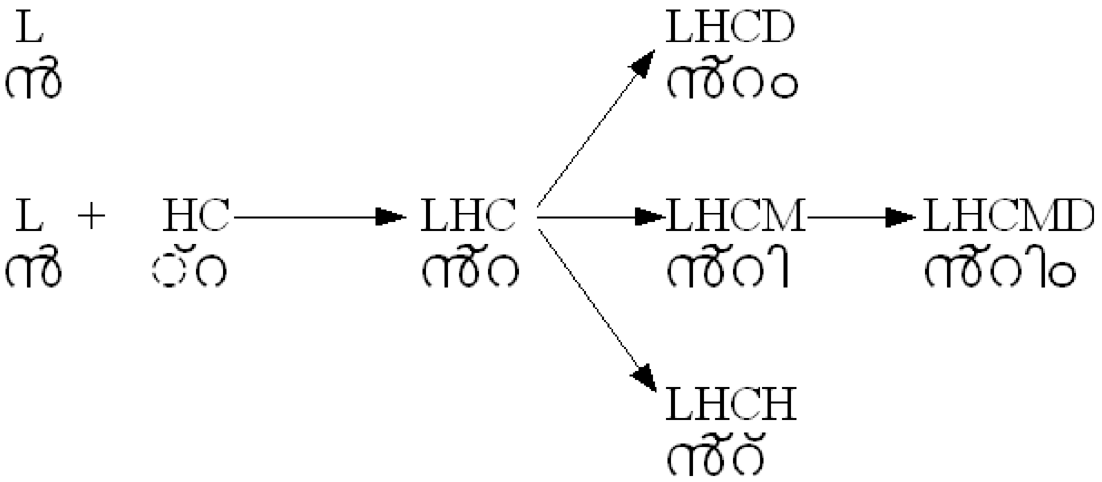
Consonant-Syllable :

```
   *k(C[N]H) C[N] [
H|D|B|X|BD|BX|M[D|B|X|BD|BX] ]
   |   [CH] Z
   |   L[ HC [ D | H | M[D] ] ]
   |    AC[ D | X | M[D|X] ]
```

# Consonant Syllable continues…

Syllable with Chillu characters only exists in Malayalam language.

L[ HC [ D | H | M[D] ] ]



# ABNF Illustrated...

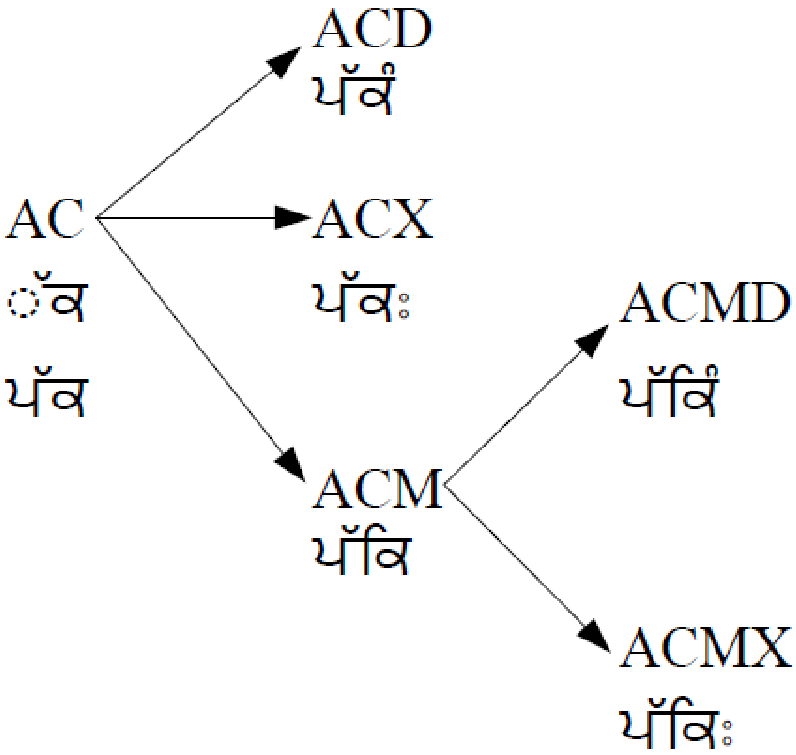
Consonant-Syllable :

```
   *k(C[N]H) C[N] [
H|D|B|X|BD|BX|M[D|B|X|BD|BX] ]
   |   [CH] Z
   |   L[ HC [ D | H | M[D] ] ]
   |    AC[ D | X | M[D|X] ]
```

# Consonant Syllable continues...

Syllable with Addak only exists in Punjabi language.

AC[ D | X | M[D|X] ]



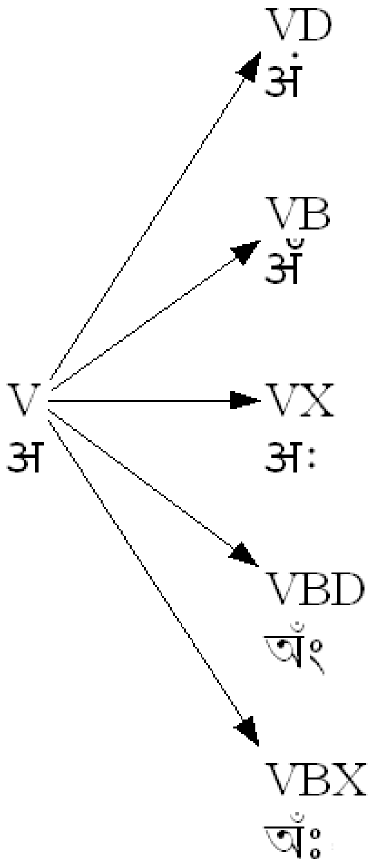
# ABNF Illustrated..

Vowel-Syllable :

V [ D | B | X | BD | BX]

# Vowel Syllable continues...

V [ D | B | X | BD | BX]



# ABNF Illustrated...

Syllable :

Consonant-Syllable | Vowel-Syllable

-where

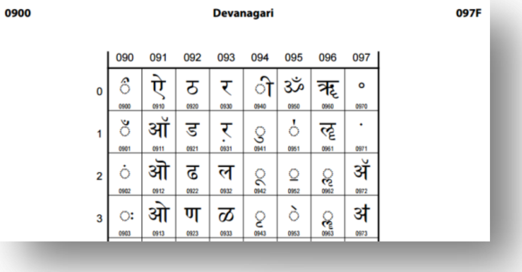
```
Consonant-Syllable →
                *k(C[N]H) C[N] [H|D|B|X|BD|BX|M[D|B|X|BD|BX]]
                | [CH]Z
                | L[HC[D|H|M[D]]]
                | AC[D|X|M[D|X]]

Vowel-Syllable    →    V[D|B|X|BD|BX]
```

# Devanagari LGR

# Devanagari LGR

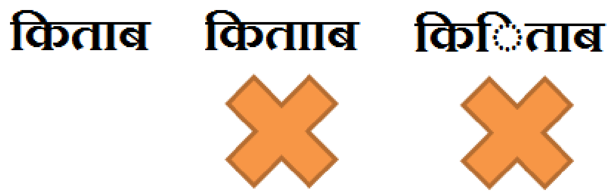
- **Fundamental Blocks:**
  - Code point repertoire



0900 **Devanagari** 097F

|   | 090 | 091 | 092 | 093 | 094 | 095 | 096 | 097 |
|---|---|---|---|---|---|---|---|---|
| 0 | ◌ऀ<br>0900 | ऐ<br>0910 | ठ<br>0920 | र<br>0930 | ◌ी<br>0940 | ॐ<br>0950 | ॠ<br>0960 | ॰<br>0970 |
| 1 | ◌ँ<br>0901 | ऑ<br>0911 | ड<br>0921 | ऱ<br>0931 | ◌ु<br>0941 | ◌॑<br>0951 | ॡ<br>0961 | ॱ<br>0971 |
| 2 | ◌ं<br>0902 | ऒ<br>0912 | ढ<br>0922 | ल<br>0932 | ◌ू<br>0942 | ◌॒<br>0952 | ◌ॢ<br>0962 | ॲ<br>0972 |
| 3 | ◌ः<br>0903 | ओ<br>0913 | ण<br>0923 | ळ<br>0933 | ◌ृ<br>0943 | ◌॓<br>0953 | ◌ॣ<br>0963 | ॳ<br>0973 |

  - Whole Label Evaluation rules



  - Variant Rules

अद्रक
अद्रक
अद्रक

# Root LGR procedure

- Binding principles:
  - LONGEVITY PRINCIPLE
  - LEAST ASTONISHMENT PRINCIPLE
  - INCLUSION PRINCIPLE
  - SIMPLICITY PRINCIPLE
  - PREDICTABILITY PRINCIPLE
  - STABILITY PRINCIPLE
  - LETTER PRINCIPLE

# Root LGR procedure

- **LONGEVITY PRINCIPLE**
  - The panels are supposed to begin using the latest version of Unicode, but also to take into consideration the stability of Unicode character properties.
  - If the panels both fail to behave in this way, then there is a risk either that code points will be permitted for allocation in the root zone that do not work with multiple versions of Unicode, or that code point substitution rules will be adopted that work well in peculiar contexts, but that will work poorly in other (perhaps future) contexts.

# Root LGR procedure

- **LEAST ASTONISHMENT PRINCIPLE**
  - The Least Astonishment Principle aims at ensuring that the allocated code points included in the zone repertoire are useful as elements in unique identifiers. To the extent that a code point is confusing to the user population or can be used in surprising ways –whether to members of the original linguistic target community or, in the case of the root, to members of other linguistic communities – use of the code point fails to adhere to the Least Astonishment Principle in that context.
  - The integration panel, especially, is responsible to ensure adherence to the Least Astonishment Principle. Because the Root Zone is a shared resource, the Integration panel is explicitly charged with considering the entire user population, which is everyone on the Internet.

# Root LGR procedure

- **INCLUSION PRINCIPLE**
  - The procedure is an example of the Inclusion Principle in action, since every rule or code point is excluded until reviewed and then explicitly included.

# Root LGR procedure

- **SIMPLICITY PRINCIPLE**
  - Part of the point of having the integration panel is that it performs a check of the Simplicity Principle. The integration panel cannot possibly include experts in every language and script, but the members must have general knowledge of Unicode, IDNA, DNS, or all of the above. If any member of the integration panel cannot understand the rationale for inclusion of some rule, then that member will not support the rule, and it will not proceed. This is the purpose of the unanimity requirement for the integration panel.

# Root LGR procedure

- **PREDICTABILITY PRINCIPLE**
  - The proposal follows the Predictability Principle in much the same way it follows the Simplicity Principle: if the integration panel does not immediately agree with the recommendations of the generation panel, or if members of the integration panel disagree with each other, that is a good reason to suppose that the rule in question is not really predictable.

# Root LGR procedure

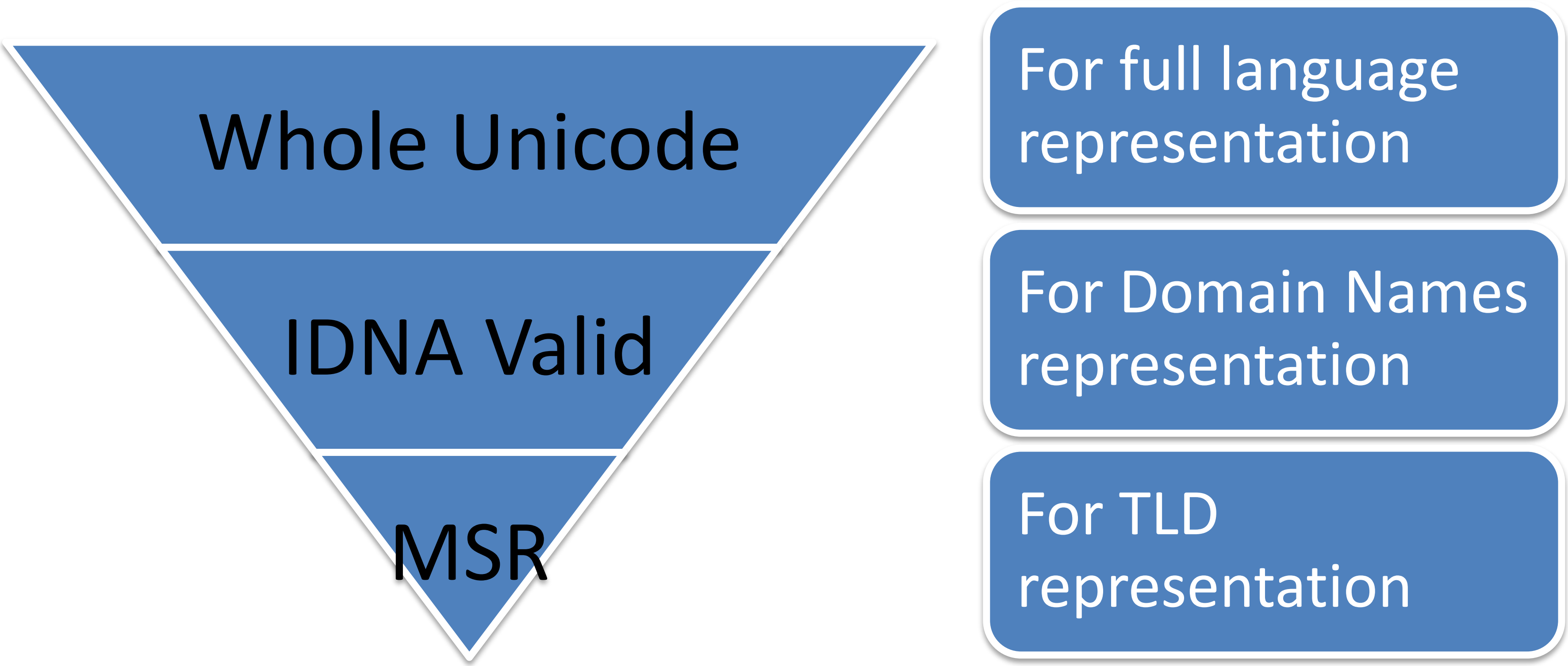
- **STABILITY PRINCIPLE**
  - Especially in the case of the root zone, the Stability Principle is less a matter of guidance and more a statement of fact. The proposed procedure attempts to minimize the possibility that any label generation rule will be permitted for the root zone without that rule having been considered as carefully as possible for any negative consequences. If there is a failure such that the integration panel determines that a previously active rule needs to be removed, this procedure requires that the procedure itself be subject to review.

# Root LGR procedure

- **CONSERVATISM PRINCIPLE**
  - The proposal is consistent with the Conservatism Principle in two ways. First and most important, because the integration panel is supposed to reject anything it does not positively think is safe, the Conservatism Principle is built in to the integration panel's criteria. Second, in the event of disagreement between the generation and integration panels, the proposed rule that is the subject of the disagreement is automatically excluded from the root label generation rules.

# Root LGR procedure

- Starting point:
  - Maximal Starting Repertoire:



Whole Unicode

IDNA Valid

MSR

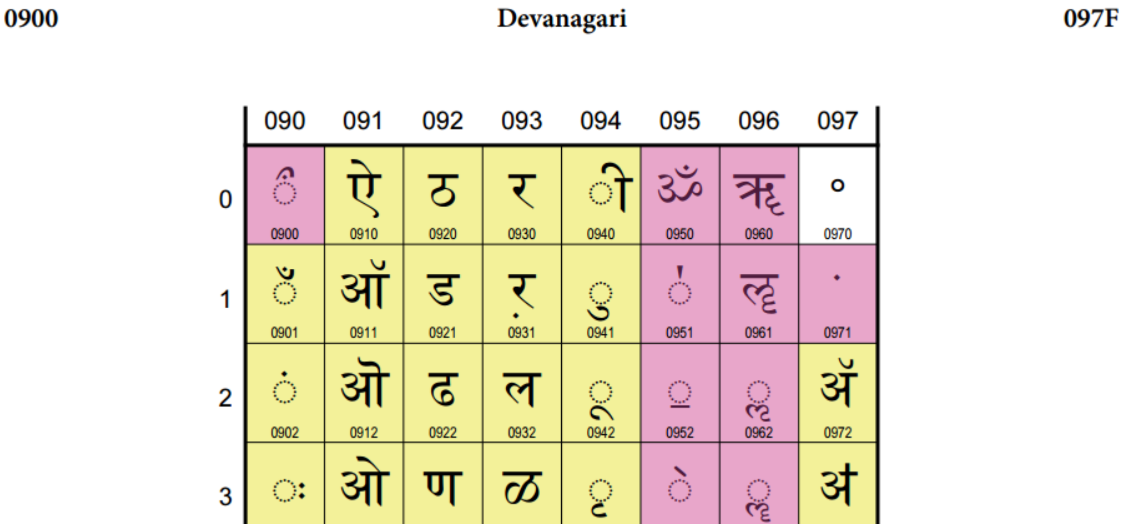
For full language representation

For Domain Names representation

For TLD representation

# Root LGR procedure

- Maximal Starting Repertoire:
  - MSR-1 released by the Integration Panel on 20th Jan. ’14
  - MSR-2 released on 27th Apr. 2015



0900 Devanagari 097F

| | 090 | 091 | 092 | 093 | 094 | 095 | 096 | 097 |
|---|---|---|---|---|---|---|---|---|
| 0 | ◌ऀ<br>0900 | ऐ<br>0910 | ठ<br>0920 | र<br>0930 | ◌ी<br>0940 | ॐ<br>0950 | ॠ<br>0960 | ॰<br>0970 |
| 1 | ◌ँ<br>0901 | ऑ<br>0911 | ड<br>0921 | ऱ<br>0931 | ◌ु<br>0941 | ◌॑<br>0951 | ॡ<br>0961 | ॱ<br>0971 |
| 2 | ◌ं<br>0902 | ऒ<br>0912 | ढ<br>0922 | ल<br>0932 | ◌ू<br>0942 | ◌॒<br>0952 | ◌ॢ<br>0962 | ॲ<br>0972 |
| 3 | ◌ः | ओ | ण | ळ | ◌ृ | ◌॓ | ◌ॣ | ॳ |

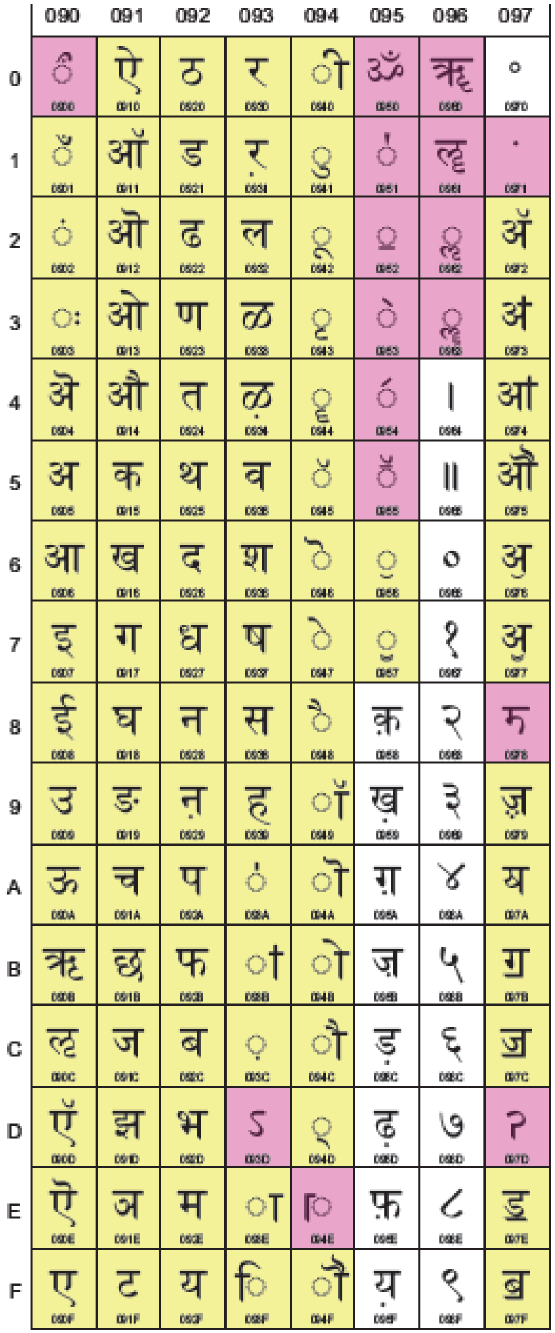
# Devanagari Code Block
# - MSR

**Convention:**

White: Blocked by IDNA Protocol

Pink: Blocked by MSR

Yellow: Permitted by MSR for final decision by the GPs.



| | 090 | 091 | 092 | 093 | 094 | 095 | 096 | 097 |
|---|---|---|---|---|---|---|---|---|
| 0 | ◌ऀ<br>0900 | ऐ<br>0910 | ठ<br>0920 | र<br>0930 | ◌ी<br>0940 | ॐ<br>0950 | ॠ<br>0960 | ॰<br>0970 |
| 1 | ◌ँ<br>0901 | ऑ<br>0911 | ड<br>0921 | ऱ<br>0931 | ◌ु<br>0941 | ◌॑<br>0951 | ॡ<br>0961 | ॱ<br>0971 |
| 2 | ◌ं<br>0902 | ऒ<br>0912 | ढ<br>0922 | ल<br>0932 | ◌ू<br>0942 | ◌॒<br>0952 | ◌ॢ<br>0962 | ॲ<br>0972 |
| 3 | ◌ः<br>0903 | ओ<br>0913 | ण<br>0923 | ळ<br>0933 | ◌ृ<br>0943 | ◌॓<br>0953 | ◌ॣ<br>0963 | ॳ<br>0973 |
| 4 | ऄ<br>0904 | औ<br>0914 | त<br>0924 | ऴ<br>0934 | ◌ॄ<br>0944 | ◌॔<br>0954 | ।<br>0964 | ॴ<br>0974 |
| 5 | अ<br>0905 | क<br>0915 | थ<br>0925 | व<br>0935 | ◌ॅ<br>0945 | ◌ॕ<br>0955 | ॥<br>0965 | ॵ<br>0975 |
| 6 | आ<br>0906 | ख<br>0916 | द<br>0926 | श<br>0936 | ◌ॆ<br>0946 | ◌ॖ<br>0956 | ०<br>0966 | ॶ<br>0976 |
| 7 | इ<br>0907 | ग<br>0917 | ध<br>0927 | ष<br>0937 | ◌े<br>0947 | ◌ॗ<br>0957 | १<br>0967 | ॷ<br>0977 |
| 8 | ई<br>0908 | घ<br>0918 | न<br>0928 | स<br>0938 | ◌ै<br>0948 | क़<br>0958 | २<br>0968 | ॸ<br>0978 |
| 9 | उ<br>0909 | ङ<br>0919 | ऩ<br>0929 | ह<br>0939 | ◌ॉ<br>0949 | ख़<br>0959 | ३<br>0969 | ॹ<br>0979 |
| A | ऊ<br>090A | च<br>091A | प<br>092A | ◌ऺ<br>093A | ◌ॊ<br>094A | ग़<br>095A | ४<br>096A | ॺ<br>097A |
| B | ऋ<br>090B | छ<br>091B | फ<br>092B | ◌ऻ<br>093B | ◌ो<br>094B | ज़<br>095B | ५<br>096B | ॻ<br>097B |
| C | ऌ<br>090C | ज<br>091C | ब<br>092C | ◌़<br>093C | ◌ौ<br>094C | ड़<br>095C | ६<br>096C | ॼ<br>097C |
| D | ऍ<br>090D | झ<br>091D | भ<br>092D | ऽ<br>093D | ◌्<br>094D | ढ़<br>095D | ७<br>096D | ॽ<br>097D |
| E | ऎ<br>090E | ञ<br>091E | म<br>092E | ◌ा<br>093E | ◌ॎ<br>094E | फ़<br>095E | ८<br>096E | ॾ<br>097E |
| F | ए<br>090F | ट<br>091F | य<br>092F | ◌ि<br>093F | ◌ॏ<br>094F | य़<br>095F | ९<br>096F | ॿ<br>097F |

# Proposed Devanagari LGR Code point repertoire

- Over and above the MSR recommendations, further suggests barring following additional characters:

- ऄ ऌ ऎ ऒ ऩ ऴ ◌ॆ ◌ॊ

# Proposed Devanagari LGR – Additional rule for ऱ

| Sr. No. | Unicode Code Points | Sequence | Character Names | Unicode General Category (gc) | Reference |
|---|---|---|---|---|---|
| **1.** | 0931<br>094D<br>092F | ऱ्य | DEVANAGARI LETTER RRA<br>DEVANAGARI SIGN VIRAMA<br>DEVANAGARI LETTER YA | Lo<br>Mn<br>Lo | [INSCRIPT] |
| **2.** | 0931<br>094D<br>0939 | ऱ्ह | DEVANAGARI LETTER RRA<br>DEVANAGARI SIGN VIRAMA<br>DEVANAGARI LETTER HA | Lo<br>Mn<br>Lo | [INSCRIPT] |

# Proposed Devanagari LGR

## - WLE Rules -

| **C** | **Consonant** |
|---|---|
| **M** | Matra |
| **V** | Vowel |
| **D** | Anusvara / Chandrabindu |
| **X** | Visarga |
| **H** | Halant / Virama |
| **N** | Nukta |
| **S** | Eyelash Reph (C1HC2) where<br>C1 is 0931 (ऱ - DEVANAGARI LETTER RRA)<br>H is 094D (◌् - DEVANAGARI SIGN VIRAMA)<br>C2 is either - 092F (य - DEVANAGARI LETTER YA) or 0939 (ह - DEVANAGARI LETTER HA) |

- N: must be preceded only by either of specific set of Cs viz.
  - क (U+0915)
  - ख (U+0916)
  - ग (U+0917)
  - ज (U+091C)
  - ड (U+0921)
  - ढ (U+0922)
  - फ (U+092B)
- H: must be preceded by C or N
- X: must be preceded by either of V, C, N or M
- D: must be preceded by either of V, C, N or M (Can be combined with rule for X)
- M: must be preceded either by C or N
- V: Can **NOT** be preceded by H

# Proposed Devanagari LGR

## - Variants -

- Currently None.
- There are various factors involved
- ICANN ideally does not want homographic variants to be part of this variant set
- Most of the **prominent** variants in Brahmi Based scripts/languages **are Homographic**

# Possible Variant Cases - Devanagari

- Confusingly similar – Single characters

| | |
|---|---|
| घ<br>U+0918 | ध<br>U+0927 |
| भ<br>U+092D | म<br>U+092E |

# Possible Variant Cases - Devanagari

- Confusingly similar – Composite characters

| | | |
|---|---|---|
| द्ग<br>U+0926 U+094D U+O917 | द्र<br>U+0926 U+094D U+0930 | द्न<br>U+0926 U+094D U+0928 |
| द्ध<br>U+0926 U+094D U+0927 | द्घ<br>U+0926 U+094D U+0918 | |
| ष्ट<br>U+0937 U+094D U+091F | ष्ठ<br>U+0937 U+094D U+0920 | |
| द्व<br>U+0926 U+094D U+ 0935 | द्ब<br>U+0926 U+094D U+092C | |

# Possible Variant Cases - Devanagari

- Confusingly similar – Cross script

| DEVANAGARI SCRIPT | COGNATE SCRIPT | CODEPOINT IN COGNATE SCRIPT |
|---|---|---|
| **VOWELS** | | |
| उ 0909 | Bangla | ও 0993 |
| उ 0909 | Gurmukhi | ਤ 0A24 |
| ऋ 090B | Gujarati | ૠ 0AE0 |
| **CONSONANTS** | | |
| क 0915 | Bangla | ক 0995 |
| ग 0917 | Gujarati | ગ 0A97 |
| ग 0917 | Gurmukhi | ਗ 0A17 |
| घ 0918 | Gurmukhi | ਬ 0A2C |
| घ 0918 | Gujarati | ઘ 0A98 |
| ङ 0919 | Gujarati | ઙ 0A99 |
| छ 091B | Gujarati | છ 0A9B |
| ञ 091E | Gujarati | ઞ 0A9E |

| DEVANAGARI SCRIPT | COGNATE SCRIPT | CODEPOINT IN COGNATE SCRIPT |
|---|---|---|
| ट 091F | Gurmukhi | ਟ 0A17 |
| ठ 0920 | Gujarati | ઠ 0AA0 |
| ठ 0920 | Gurmukhi | ਠ 0A20 |
| ड 0921 | Gujarati | ડ 0AA1 |
| ढ 0922 | Gurmukhi | ਫ 0A2B |
| त 0924 | Gujarati | ત 0AA4 |
| ध 0927 | Gujarati | ધ 0AA7 |
| न 0928 | Gujarati | ન 0AA8 |
| न 0928 | Bangla | ন 09A8 |
| न 0928 | Bangla | ণ 09A3 |
| प 092A | Gujarati | પ 0AAA |
| प 092A | Gurmukhi | ਧ 0A17 |
| प 092A | Gurmukhi | ਪ 0A2A |

| DEVANAGARI SCRIPT | COGNATE SCRIPT | CODEPOINT IN COGNATE SCRIPT |
|---|---|---|
| प 092A | Gurmukhi | ੫ 0A6B |
| म 092E | Gurmukhi | ਸ 0A38 |
| म 092E | Gujarati | મ 0AAE |
| य 092F | Gujarati | ચ 0A9A |
| र 0930 | Gujarati | ૨ 0AAE |
| र 0930 | Gurmukhi | ਕ 0A15 |
| ल 0932 | Bangla | ল 09B2 |
| व 0935 | Gujarati | વ 0AB5 |
| श 0936 | Gujarati | શ 0AB6 |
| श्<br>0936+094D | Bangla | ঽ 09BD |
| ष 0937 | Gujarati | ષ 0AB7 |
| स 0938 | Gujarati | સ 0AB8 |
| ह 0939 | Gujarati | હ 0AB9 |

| DEVANAGARI SCRIPT | COGNATE SCRIPT | CODEPOINT IN COGNATE SCRIPT |
|---|---|---|
| **Nukta characters** | | |
| ग़ 095A<br>Or<br>0917+094D | Gurmukhi | ਗ਼ 0A5A |
| ढ़ 095D<br>Or<br>0922+094D | Gurmukhi | ਫ਼ 0A5E |

धन्यवाद   !