

Sinhala Generation Panel Update

Harsha Wijayawardhana
Sinhala GP Co-Chair

Agenda Overview

1

Scripts Covered and
Where They are
Used

2

Members of the GP

3

Code Points
Repertoire

4

Variant Analysis

5

Whole Label
Evaluation Rules

6

Next Steps

Sinhala GP – Languages Using Sinhala Script

- ⦿ Sinhala Script is primarily used in Sri Lanka to write Sinhala language which belongs to Indo-European Language family and Indo-Arya Sub family
- ⦿ The Script is Abugida Script which sprang from family of Southern Brahmi Script to which Telugu, Malayalam and Tamil belong to
- ⦿ Languages covered by the script:
 - Sinhala
 - Pali
 - Sanskrit

Sinhala GP – Members

- ⊙ Dr. Ruvan Weerasinghe
- ⊙ Mr. Harsha Wijayawardhana
- ⊙ Mr. Chamila Liyanage
- ⊙ Mr. Pathum Egodawatha
- ⊙ Mr. Viraj Welgama
- ⊙ Ms. Aruni Goonathilake
- ⊙ Mr. Chamara Dissanayake
- ⊙ Ms. Sagarika Wickramasekara
- ⊙ Prof. J.B.Dissanayake
- ⊙ Mr. Champika Wijayathunga
- ⊙ Mr. Rajeewa Abeygoonaratne
- ⊙ Rev. Mettavihari
- ⊙ Ms. Nimasha Dilshani

Code Point Repertoire

- Starting from MSR-3, the repertoire includes:
 - 72 code points
 - 4 sequences
- The repertoire excludes:

| # | Unicode Code Point | Glyph | Character Name | Reason for exclusion |
|---|--------------------|-------|--|--|
| 1 | 0D8E | ස෦෦ | SINHALA LETTER IRUUYANNA | Usage unknown |
| 2 | 0D8F | ඌ | SINHALA LETTER ILUYANNA | Usage unknown |
| 3 | 0D90 | ඌ෦෦ | SINHALA LETTER ILUUYANNA | Usage unknown |
| 4 | 0D9E | ඞ | SINHALA LETTER KANTAJA | Not in modern usage |
| 5 | 0DA6 | ඌ෦෦ | SINHALA LETTER SANYAKA | Only used in the word ‘ඌ෦෦’ (this word is used to call dogs) |
| 6 | 0DDF | ඌ෦෦ | SINHALA VOWEL SIGN GAYANUKITTA | Usage unknown |
| 7 | 0DF3 | ඌ෦෦ | SINHALA VOWEL SIGN DIGA GAYANUKITTA | Usage unknown |

In-Script Variant Analysis

- Sinhala GP decided the following are in-script variant code points due to similar shapes and characters which could be used interchangeably
 - ස (U+0DC3) and ස (U+0D9D)
 - බ (U+0DB6) and බ (U+0D9B)
 - භ (U+0DC4) and භ (U+0DB7)
 - ච (U+0DA0) and ච (U+0DC0)
 - ඔ (U+0D94) and ඔ (U+0DB9)
 - ඵ (U+0D91) and ඵ (U+0DB5)
 - සා (U+0D8D) and සා (U+0DC3 U+0DD8)
 - ඔඵ (U+0D93) and ඔඵ (U+0DB5 U+0DD9)
 - ඵ් (U+0D92) and ඵ් (U+0DB5 U+0DCA)
 - ඔ් (U+0D95) and ඔ් (U+0DB9 U+0DCA)

Cross-Script Variant Analysis

- Sinhala GP concluded there is no cross-script variant rules
- Following are confusable cases
 - U+0D82 (SINHALA SIGN ANUSVARAYA, ീ)

| Sinhala | Telugu | Kannada | Malayalam |
|---------------|----------------|----------------|----------------|
| ീ (U+0D82) | ీీ (U+0C02) | ೀೀ (U+0C82) | ഀഀ (U+0D02) |

- U+0D83 (SINHALA SIGN VISARGAYA, ു)

| Sinhala | Devanagari | Gujarati | Telugu | Kannada | Malayalam |
|---------------|---------------|---------------|---------------|---------------|---------------|
| ു (U+0D83) | ः (U+0903) | ઃ (U+0A83) | ః (U+0C03) | ಃ (U+0C83) | ഃ (U+0D03) |

Cross-Script Variant Analysis

- Following are confusable cases (cont.)

- Sinhala and Malayalam

| Sinhala | Malayalam |
|---------------|---------------|
| ඉ (U+0D9C) | ඹ (U+0D17) |
| ඊ (U+0DC1) | ඹ (U+0D36) |
| උ (U+0DCF) | ඹ (U+0D3E) |

- Sinhala and Myanmar

| Sinhala | Myanmar |
|---------------|---------------|
| ඉ (U+0D9C) | ၀ (U+1010) |
| ඊ (U+0DC1) | ၀ (U+107B) |

Whole Label Evaluation Rules

- ⊙ Code point category
 - C → Consonant
 - V → Vowel
 - M → Matras / Vowel Signs
 - B → Anusvara (Bindu)
 - X → Visarga
 - H → Halanta / Virama
 - J → Sannjakas

- ⊙ Whole Label Evaluation Rules
 - H: must be preceded by C
 - M: must be preceded by C or J
 - X: must be preceded by either V, C, or M
 - B: must be preceded by either V, C, J or M

Timeline and Next Step

