

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

---

LGR Version: 3.0

Date: 2018-10-06

Document version: 1.1

Authors: Myanmar Generation Panel

## 1. General Information/ Overview/ Abstract

This document lays down the Label Generation Ruleset for Myanmar script. Three main components of the Myanmar Script LGR i.e. Code point repertoire, Variants and Whole Label Evaluation Rules have been described in detail here having given the historical background of the Script under Section 3.

## 2. Script for which the LGR is proposed

ISO 15924 Code: Mymr

ISO 15924 Key N°: 350

ISO 15924 English Name: Myanmar (Burmese)

Latin transliteration of native script name: Myanmar

Native name of the script: မြန်မာ

Maximal Starting Repertoire [MSR] version: 3

## 3. Background

### 3.1 Background on Script and Principal Languages Using It

Burmese Language is the national and official language of Myanmar and is a member of the Tibeto-Burman Language family which is a sub-family of the Sino-Tibetan family languages. The letters of the alphabet used in Myanmar script are said to have been derived from the Brahmi script which has flourished in the Indian subcontinent between 5th Century B.C and 3rd Century A.D. Existing records of original Myanmar script from votive tablets, terracotta plaques, stone inscriptions, painted glosses on murals, palm leaf inscriptions, bell inscriptions and inscriptions made on writing tablets made of handmade paper, gold or copper foil, show that earliest surviving records can be dated as far back as 5th Century of the Myanmar Era (12th Century AD).

Myanmar (Burmese) is written from left to right and requires no spaces between words, except to make phrase or sentence segments. The Myanmar script is used for a number of languages. This means that when considering the script as a whole, care must be taken not to over specify constraints on what character sequences should be considered valid or in error. The temptation is to use script level sequence constraints as a form of spell checking. But spell checking is inherently language specific. The result is that script constraints need to be the lowest common denominator of all the orthographies supported by the script. The orthography list is not closed: we have not described all the existing orthographies yet; languages change and develop and their orthographies with them. As a result, script constraints cannot simply be the intersection of all known writing system constraints, but must take a more intentional approach. The basic principle used here is not to try to constrain what users can generate, but only to ensure that there are no two different valid sequences that look the same, within a writing system. We do this by specifying a valid string as being a sequence of slots. Each slot may be empty or contain a character (or sequence as specified by the slot). Implementations may well add further, language specific, constraints to help their users.

A further concern when reading a developing document such as this is the stability criteria. What can we be sure about going into the future? The approach taken in this document follows the core principle of stability in Unicode: Any valid data today will always remain valid. This requires that any changes to the sequence order, for example, will always be to loosen it. Thus more sequences will be allowed rather than less. This means that invalid data today may not always remain invalid in future versions of this document. It should also be born in mind that while the unity of the script as a whole may well be affected by the addition or changes in a single language, each language stands alone in its encoding and needs its own consistency. Care is taken that any changes that a difference in language may cause on the script as a whole (adding more legal sequences), do not cause any changes in other language

encodings. This may result in some decisions made for a particular language, looking different from those for another language and the temptation to try to over unify languages should be avoided.

The languages that currently use Myanmar scripts which have EGIDS scale from 1 to 4 and EGIDS 5 languages which have more than 500,000 users are included in the analysis of the Generation Panel. These languages are listed in Table 1 – List of languages in consideration.

Language	ISO 639-3 Code(s)	Countries	Local Name of the Script	EGIDS Scale	Total Users in All Countries
Burmese	[mya]	Myanmar	မြန်မာ	1	42,906,490
Shan	[shn]	Myanmar, China, Thailand	လိၵ်ႈတႆး	3	3,295,000
Rakhine	[rki]	Myanmar	ရခိုင်	3	2,020,000
Karen, Sgaw	[ksw]	Myanmar, Thailand	ꨀꨣ	3	1,560,000
Mon	[mnw]	Myanmar, Thailand	မန်	5	851,000
Pa'O Karen	[blk]	Myanmar	ပအိုဝ်း	5	560,740

Table 1: List of languages in consideration.

Three additional language are added to this proposal.

Rakhine language[rki] is one variation of Burmese language[mya]. Rakhine script form is 100% the same as Burmese Script. All work for Myanmar language will cover Rakhine language as well. This language has EDIS scale 3 with 2,020,000 total speakers.

Mon language[mnw] is used in some parts of southern Myanmar and along the Thai-Myanmar border area. This language has EDIS scale 5 with 851,000 total speakers.



### Shan Consonants

There are 24 consonants in Shan. They are treated the same way as Burmese consonants, despite having different sets of Medials, Vowels and Tones.

ဂ	ခ	စ	င	လ	သ	ဃ	ဆ
u1075	u1076	u1077	u1004	u1078	u10u1E	u107A	u1079
တ	ထ	တံ	ဆဲ	ပ	ဖ	ဗ	စ
u1010	u1011	u107B	u107C	u1015	u107D	u107E	u107F
မ	ယ	ရ	လ	ဝ	ဆ	ဂ	က
u1019	u101A	u101B	u101C	u101D	u1080	u1081	u1022

### S'gaw Karen Consonants

က	ခ	ဂ	ဃ	င	စ	ဆ	ရ	
u1000	u1001	u1002	u1003	u1004	u1005	u1006	u1061	
ည	တ	ထ	ဒ	န	ပ	ဖ	ဘ	
u100A	u1010	u1011	u1012	u1014	u1015	u1016	u1018	
မ	ယ	ရ	လ	ဝ	သ	ဟ	အ	ဧ
u1019	u101A	u101B	u101C	u101D	u101E	u101F	u1021	u1027

### Mon Consonants

က	ခ	ဂ	ဃ	င	စ	ဆ	ဇ	ဈ	ည	ဋ
u1000	u1001	u1002	u1003	u105A	u1005	u1006	u1007	u105B	u1009	u100A
ဋ	ဌ	ဍ	ဎ	ဏ	တ	ထ	ဒ	ဓ	န	ပ
u100B	u100C	u100D	u100E	u100F	u1010	u1011	u1012	u1013	u1014	u1015
ဖ	ဗ	ဘ	မ	ယ	ရ	လ	ဝ	သ	ဟ	ဌ
u1016	u1017	u1018	u1019	u101A	u101B	u101C	u101D	u101E	u101F	u1020
အ	ဓ	ဍ								
u1021	u105C	u105E								

## Pa'O Consonants

All the same as Myanmar Consonants.

က	ခ	ဂ	ဃ	င	စ	ဆ	ဇ	ဈ	ည	ည
u1000	u1001	u1002	u1003	u1004	u1005	u1006	u1007	u1008	u1009	u100A
ဋ	ဌ	ဍ	ဎ	ဏ	တ	ထ	ဒ	ဓ	န	ပ
u100B	u100C	u100D	u100E	u100F	u1010	u1011	u1012	u1013	u1014	u1015
ဖ	ဗ	ဘ	မ	ယ	ရ	လ	ဝ	သ	ဟ	ဌ
u1016	u1017	u1018	u1019	u101A	u101B	u101C	u101D	u101E	u101F	u1020
အ	ဿ									
u1021	u103F									

### 3.3.2 The Vowels and vowel diacritics

Myanmar

Myanmar: Independent Vowels

အ	ဤ	ဥ	ဦ	ဧ	ဩ	ဪ
u1023	u1024	u1025	u1026	u1027	u1029	u102A

Myanmar: Dependent Vowels

ိ	ု	ိ	ိ	ိ	ိ	ေ	ဲ
u102B	u102C	u102D	u102E	u102F	u1030	u1031	u1032

Myanmar: Dependent Vowel Sequences

ဝေါ	ော	ေါ်	ော်	ို	ိ	ိ
u1031	u1031	u1031	u1031	u102D	u102D	u102F
u102B	u102B	u103A	u103A	u102F	u1036	u1036

Shan

Shan: Dependent Vowels

ာ	ိ	ိ	ော	ော	ို	ိ	ိ
u1083	u102D	u102E	u1031	u1084	u102F	u1030	u1086

Shan: Dependent Vowel Sequences:

ော	ေါ်	ို	ိ	ာ
u1031	u1031	u102D	u102D	u1062
u1083	u103A	u102F	u1030	u1086

Sgaw Karen

Sgaw Karen: Dependent Vowels

ါ	ိ	ာ	ို	ိ	ိ	ဲ	ိ	ိ
102B	1036	1062	102F	1030	1037	1032	102D	102E

Mon

Mon: Independent Vowels

ိ	ဲ	ဲ	ြ	ြ
u1023	u1025	u1028	u1029	u102A

Mon: Independent Vowel Sequences

ꨀ	ꨁ
u1023	u1025
u1033	u102F

Mon: Dependent Vowels

ꨀ	ꨁ	ꨂ	ꨃ	ꨄ	ꨅ	ꨆ	ꨇ	ꨈ	ꨉ	ꨊ
u102B	u102C	u102D	u1033	u102F	u1030	u1031	u1032	u1034	u1036	u1038

Mon: Dependent Vowel Sequences:

ꨀꨁ	ꨀꨂ	ꨂꨃ	ꨃꨄ
u1031	u1031	u102D	u1035
u102B	u102C	u102F	u102F

Pa'O Karen

Pa'O Karen: Dependent Vowels

ꨀ	ꨁ	ꨂ	ꨃ	ꨄ	ꨅ	ꨆ	ꨇ
u102B	u102C	u102D	u102E	u102F	u1030	u1031	u1032

Pa'O Karen: Dependent Vowel Sequences

ꨀꨁ	ꨀꨂ	ꨂꨃ	ꨃꨄ	ꨂꨃꨄ	ꨃꨄꨅ	ꨃꨄꨅꨆ
u1031	u1031	u1031	u1031	u102D	u102F	u102F
u102B	u102B	u102B	u102C	u102F	u1032	u1036

### 3.3.3 The other type of character (Medials and Tones)

Myanmar

Myanmar Medials

ꨀ	ꨁ	ꨂ	ꨃ
u103B	u103C	u103D	u103E

Myanmar Medial Sequences:

ꨀꨆ	ꨁꨆ	ꨀꨇ	ꨁꨇ	ꨂꨇ	ꨁꨆꨇ
u103B	u103C	u103B	u103C	u103D	u103C
u103D	u103D	u103E	u103E	u103E	u103D
					u103E

Myanmar Tones:

ꨆ	ꨆꨄ
u1037	u1038

Myanmar Other Diacritics:

ꨆ	ꨆꨅ	ꨆꨃ	ꨆꨄꨃ
u1036	u1039	u103A	u1004
			u103A
			u1039

Shan

Shan Medials:

ꨀ	ꨁ	ꨃ
u103B	u103C	u1082

Shan Tones:

ꨆꨄ	ꨆꨄꨃ	ꨆꨄꨃꨄ	ꨆꨄꨃꨄꨃ	ꨆꨄꨃꨄꨃꨄꨃ
u1087	u1088	u1038	u1089	u108A

Shan Other Diacritics:

◌ၢ	◌ၣ	◌ၤ	◌ၥ
u1062	u1035	u1085	u103A

Sgaw Karen

Sgaw Karen: Medials

◌ၦ	◌ၧ	◌ၨ	◌ၩ	◌ၪ
u103E	u1060	u103B	u103C	u103D

Sgaw Karen: Tones

◌ၫ	◌ၬ
u1038	u1064

Sgaw Karen: Tone sequences

ၢၣ	ၢၤ	ၢၥ
1062	102C	1063
103A	103A	103A

Sgaw Karen: Other Diacritic

◌ၭ
u103A

Mon

Mon: Medials

◌ၮ	◌ၯ	◌ၰ	◌ၱ	◌ၲ	◌ၳ	◌ၴ
u105E	u105F	u103B	u103C	u1060	u103D	u103E

Mon: Other Diacritics

◌ +	◌ ζ	◌ ε
u1039	u103A	u105A u103A u1039

### Pa'O Karen

#### Pa'O Karen: Medials

◌ ꠆	◌ ꠇ	◌ ◌
u103B	u103C	u103D

#### Pa'O Karen: Tones

◌ ◻	◌ ◌	◌ ◌
uAA7B	u1038	u108F

#### Pa'O Karen: Other Diacritics

◌ ◌	◌ ◌	◌ ◌
u1036	u1037	u103A

### 3.3.4 Structural Formation of Myanmar

One consonant can have up to 7 diacritics which can be combination of medial marks, vowel diacritics and tone marks. All diacritics are encoded behind the main character, which can be Consonant or some Independent Vowel.

Certain Independent Vowels can be followed by another Dependent Vowel or a Tone Mark. The examples are :

◌ ◌	◌ ◌	◌ ◌
u1026 u1038	u1025 u102F u1036	u1029 u1031 u102C

## 4. Overall Development Process and Methodology

Myanmar Generation Panel has been formed by community members having experience in Natural Language Processing (NLP), Digital linguistic, software engineers and especially Unicode experts.

The Myanmar community members are having online meetings and face to face meetings. Where needed, the panel members also discussed linguistic matters with external experts from Department of Myanmar and Language Education (Formerly known as Department of Myanmar Language Commission) under Ministry of Education, which is an authoritative organization for languages and scripts of Myanmar as well as ethnic group.

The panel also got feedback of the interim work from the Integration Panel at ICANN to ensure that any concerns were addressed during the development of the proposal.

Finally, upon completion of the draft proposal, a public workshop was organized on in Yangon where all the members of panel as well as other participants from private sector, public sector especially from Posts and Telecommunications Department under Ministry of Transport and Communications, ISPs, Universities and relevant civil society were invited in order to elicit the feedback and receive recommendation on the proposal before it was finalized.

## 5. Repertoire

Myanmar code point repertoire is discussed in this section.

### 5.1 Myanmar section of Maximal Starting Repertoire [MSR] Version 3

Myanmar Unicode chart is given below and has mentioned the characters included and excluded in the Myanmar script root zone.

	100	101	102	103	104	105	106	107	108	109
0	က	ခ	ဂ	ဃ	င	စ	ဆ	ဗ	ဆ	ဝ
1	ခ	ထ	အ	ေ	ဘ	ဖ	ရှ	ိ	ှ	်
2	ဂ	ဒ	ဒ	ဲ	၂	ဇ	ာ	ိ	့	း
3	ယ	စ	ဆ	ိ	၃	ဗ	ာ	ိ	ါ	ှ
4	င	န	ဗြ	်	၄	ဇ	ာ	ိ	ေ	ှ
5	စ	င	ဇ	ိ	၅	ဆ	ဘ	ဂ	ိ	ှ
6	ဆ	ဖ	ဇ	ိ	၆	ာ	ှ	စ	ိ	ှ
7	ဇ	ဇ	ဇ	ိ	၇	ာ	ှ	ဂ	ိ	ှ
8	ဇ	ဒ	ဇ	ိ	၈	ာ	ှ	ဂ	ိ	ှ
9	ဇ	ဖ	ဇ	ိ	၉	ာ	ှ	ဂ	ိ	ှ
A	ဇ	ယ	ဇ	ိ	၁၀	ာ	ှ	ဂ	ိ	ှ
B	ဇ	ရ	ိ	၁၁	၂	ာ	ှ	ဂ	ိ	ှ
C	ဇ	လ	ိ	၁၂	၃	ာ	ှ	ဂ	ိ	ှ
D	ဇ	ဝ	ိ	၁၃	၄	ာ	ှ	ဂ	ိ	ှ
E	ဇ	သ	ိ	၁၄	၅	ာ	ှ	ဂ	ိ	ှ
F	ဇ	ဟ	ိ	၁၅	၆	ာ	ှ	ဂ	ိ	ှ

Color convention<sup>1</sup>:

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

Pinkish background - Characters that are excluded from the [MSR]

White - Characters that are disallow by IDNA2008

Figure 1 : Myanmar Table in MSR-3

## 5.2 Code Point Repertoire

This section provides the code point repertoire that Myanmar GP proposes to be included in the Myanmar LGR.

<sup>1</sup> Recommend to print this page of the document in color to identify the color convention used.

#	Unicode Code Point	Glyph	Character Name	Category	Language(s), with EGIDS Value	Reference
1	U+1000	က	MYANMAR LETTER KA	Consonant, CS1, CS2, CS3, CS5, CS7	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
2	U+1001	ခ	MYANMAR LETTER KHA	Consonant, CS1, CS2, CS3, CS5, CS7	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
3	U+1002	ဂ	MYANMAR LETTER GA	Consonant, CS1, CS2, CS3, CS5, CS7	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
4	U+1003	ဃ	MYANMAR LETTER GHA	Consonant, CS1, CS2, CS3, CS4, CS8, CS9	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
5	U+1004	င	MYANMAR LETTER NGA	Consonant	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Pa'O Karen	[101], [105]
6	U+1005	စ	MYANMAR LETTER CA	Consonant	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
7	U+1006	ဆ	MYANMAR LETTER CHA	Consonant	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
8	U+1007	ဇ	MYANMAR LETTER JA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103]
9	U+1008	ဈ	MYANMAR LETTER JHA	Consonant	1 Burmese, 5 Pa'O Karen	[101]
10	U+1009	ည	MYANMAR LETTER NYA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103]
11	U+100A	ဋ	MYANMAR LETTER NNYA	Consonant, CS3, CS4, CS9	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]

#	Unicode Code Point	Glyph	Character Name	Category	Language(s), with EGIDS Value	Reference
12	U+100B	ꠇ	MYANMAR LETTER TTA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103]
13	U+100C	ꠈ	MYANMAR LETTER TTHA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103]
14	U+100D	ꠉ	MYANMAR LETTER DDA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103]
15	U+100E	ꠊ	MYANMAR LETTER DDHA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103]
16	U+100F	ꠋ	MYANMAR LETTER NNA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103]
17	U+1010	ꠌ	MYANMAR LETTER TA	Consonant, CS1, CS2, CS3	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
18	U+1011	ꠍ	MYANMAR LETTER THA	Consonant, CS1, CS2, CS3	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
19	U+1012	ꠎ	MYANMAR LETTER DA	Consonant, CS1, CS2, CS3	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
20	U+1013	ꠏ	MYANMAR LETTER DHA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103]
21	U+1014	ꠐ	MYANMAR LETTER NA	Consonant, CS1, CS3, CS4, CS9	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]

#	Unicode Code Point	Glyph	Character Name	Category	Language(s), with EGIDS Value	Reference
22	U+1015	၀	MYANMAR LETTER PA	Consonant, CS1, CS2, CS3, CS5, CS7	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
23	U+1016	၁	MYANMAR LETTER PHA	Consonant, CS1, CS2, CS3, CS7	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
24	U+1017	၂	MYANMAR LETTER BA	Consonant, CS1, CS2, CS3	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103]
25	U+1018	၃	MYANMAR LETTER BHA	Consonant, CS1, CS2, CS3	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
26	U+1019	၄	MYANMAR LETTER MA	Consonant, CS1, CS2, CS3, CS4, CS5, CS6, CS7, CS8, CS9, CS10	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
27	U+101A	၅	MYANMAR LETTER YA	Consonant, CS1, CS3, CS4	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
28	U+101B	၆	MYANMAR LETTER RA	Consonant, CS3, CS4, CS9	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
29	U+101C	၇	MYANMAR LETTER LA	Consonant, CS1, CS3, CS4, CS6, CS9	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
30	U+101D	၈	MYANMAR LETTER WA	Consonant	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]

#	Unicode Code Point	Glyph	Character Name	Category	Language(s), with EGIDS Value	Reference
31	U+101E	၁	MYANMAR LETTER SA	Consonant, CS3, CS6	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
32	U+101F	၂	MYANMAR LETTER HA	Consonant	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
33	U+1020	၃	MYANMAR LETTER LLA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103]
34	U+1021	၄	MYANMAR LETTER A	Consonant	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
35	U+1022	၅	MYANMAR LETTER SHAN A	Consonant	3 Shan	[101], [105]
36	U+1023	၆	MYANMAR LETTER I	Independent Vowel	1 Burmese, 5 Mon	[101], [103]
37	U+1024	၇	MYANMAR LETTER II	Independent Vowel	1 Burmese,	[101]
38	U+1025	၈	MYANMAR LETTER U	Independent Vowel	1 Burmese, 5 Mon	[101], [103]
39	U+1026	၉	MYANMAR LETTER UU	Independent Vowel	1 Burmese,	[101], [103]
40	U+1027	၁၀	MYANMAR LETTER E	Independent Vowel	1 Burmese, 3 S'gaw Karen	[101]
41	U+1029	၁၁	MYANMAR LETTER O	Independent Vowel	1 Burmese, 5 Mon	[101], [103]
42	U+102A	၁၂	MYANMAR LETTER AU	Independent Vowel	1 Burmese, 5 Mon	[101], [103]

#	Unicode Code Point	Glyph	Character Name	Category	Language(s), with EGIDS Value	Reference
43	U+102B	ၵ	MYANMAR VOWEL SIGN TALL AA	Dependent Vowel Sign LV	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
44	U+102C	ၵ	MYANMAR VOWEL SIGN AA	Dependent Vowel Sign LV	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103]
45	U+102D	ၵ	MYANMAR VOWEL SIGN I	Dependent Vowel Sign SV	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
46	U+102E	ၵ	MYANMAR VOWEL SIGN II	Dependent Vowel Sign LV	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Pa'O Karen	[101], [105]
47	U+102F	ၵ	MYANMAR VOWEL SIGN U	Dependent Vowel Sign SV	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
48	U+1030	ၵ	MYANMAR VOWEL SIGN UU	Dependent Vowel Sign LV	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
49	U+1031	ၵ	MYANMAR VOWEL SIGN E	Dependent Vowel Sign LV	1 Burmese, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
50	U+1032	ၵ	MYANMAR VOWEL SIGN AI	Dependent Vowel Sign LV	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
51	U+1035	ၵ	MYANMAR VOWEL SIGN E ABOVE	Dependent Vowel Sign LV	3 Shan	[101], [105]

#	Unicode Code Point	Glyph	Character Name	Category	Language(s), with EGIDS Value	Reference
52	U+1036	◌◌	MYANMAR SIGN ANUSVARA	ANUSVARA	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
53	U+1037	◌◌◌	MYANMAR SIGN DOT BELOW	Tone-Short	1 Burmese, 3 S'gaw Karen, 5 Pa'O Karen	[101], [103]
54	U+1038	◌◌◌◌	MYANMAR SIGN VISARGA	Tone-Long	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
55	U+1039	◌◌+	MYANMAR SIGN VIRAMA	VIRAMA	1 Burmese, 5 Mon	[101], [103]
56	U+103A	◌◌◌	MYANMAR SIGN ASAT	KILLER	1 Burmese, 3 Shan, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [105], [103]
57	U+103B	◌◌◌◌	MYANMAR CONSONANT SIGN MEDIAL YA	Dependent Consonant Sign	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
58	U+103C	◌◌◌◌	MYANMAR CONSONANT SIGN MEDIAL RA	Dependent Consonant Sign	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
59	U+103D	◌◌◌◌	MYANMAR CONSONANT SIGN MEDIAL WA	Dependent Consonant Sign	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103]
60	U+103E	◌◌◌◌	MYANMAR CONSONANT SIGN MEDIAL HA	Dependent Consonant Sign	1 Burmese, 3 S'gaw Karen, 5 Mon	[101], [103]
61	U+103F	◌◌◌◌	MYANMAR LETTER GREAT SA	Consonant C1	1 Burmese , 5 Pa'O Karen	[101]

#	Unicode Code Point	Glyph	Character Name	Category	Language(s), with EGIDS Value	Reference
62	U+1061	၂	MYANMAR LETTER SGAW KAREN SHA	Consonant	3 S'gaw Karen	[101]
63	U+1062	၂	MYANMAR VOWEL SIGN SGAW KAREN EU	Dependent Vowel Sign	3 S'gaw Karen, 3 Shan	[101], [105]
64	U+1063	၂	MYANMAR TONE MARK SGAW KAREN HATHI	Tone	3 S'gaw Karen	[101]
65	U+1064	၂	MYANMAR TONE MARK SGAW KAREN KE PHO	Tone	3 S'gaw Karen	[101]
66	U+1075	၂	MYANMAR LETTER SHAN KA	Consonant	3 Shan	[101], [105]
67	U+1076	၂	MYANMAR LETTER SHAN KHA	Consonant	3 Shan	[101], [105]
68	U+1077	၂	MYANMAR LETTER SHAN GA	Consonant	3 Shan	[101], [105]
69	U+1078	၂	MYANMAR LETTER SHAN CA	Consonant	3 Shan	[101], [105]
70	U+1079	၂	MYANMAR LETTER SHAN ZA	Consonant	3 Shan	[101], [105]
71	U+107A	၂	MYANMAR LETTER SHAN NYA	Consonant	3 Shan	[101], [105]
72	U+107B	၂	MYANMAR LETTER SHAN DA	Consonant	3 Shan	[101], [105]

#	Unicode Code Point	Glyph	Character Name	Category	Language(s), with EGIDS Value	Reference
73	U+107C	ခေ	MYANMAR LETTER SHAN NA	Consonant	3 Shan	[101], [105]
74	U+107D	ဃေ	MYANMAR LETTER SHAN PHA	Consonant	3 Shan	[101], [105]
75	U+107E	ဃှ်	MYANMAR LETTER SHAN FA	Consonant	3 Shan	[101], [105]
76	U+107F	ဃ	MYANMAR LETTER SHAN BA	Consonant	3 Shan	[101], [105]
77	U+1080	ဆ	MYANMAR LETTER SHAN THA	Consonant	3 Shan	[101], [105]
78	U+1081	ဟ	MYANMAR LETTER SHAN HA	Consonant	3 Shan	[101], [105]
79	U+1082	့	MYANMAR CONSONANT SIGN SHAN MEDIAL WA	Dependent Consonant Sign	3 Shan	[101], [105]
80	U+1083	့ၵ	MYANMAR VOWEL SIGN SHAN AA	Dependent Vowel Sign	3 Shan	[101], [105]
81	U+1084	့ၶ	MYANMAR VOWEL SIGN SHAN E	Dependent Vowel Sign	3 Shan	[101], [105]
82	U+1085	့ၷ	MYANMAR VOWEL SIGN SHAN E ABOVE	Dependent Vowel Sign	3 Shan	[101], [105]
83	U+1086	့ၸ	MYANMAR VOWEL SIGN SHAN FINAL Y	Dependent Vowel Sign	3 Shan	[101], [105]
84	U+1087	့ၹ	MYANMAR SIGN SHAN TONE-2	Tone	3 Shan	[101], [105]

#	Unicode Code Point	Glyph	Character Name	Category	Language(s), with EGIDS Value	Reference
85	U+1088	◌း	MYANMAR SIGN SHAN TONE-3	Tone	3 Shan	[101], [105]
86	U+1089	◌့	MYANMAR SIGN SHAN TONE-5	Tone	3 Shan	[101], [105]
87	U+108A	◌့း	MYANMAR SIGN SHAN TONE-6	Tone	3 Shan	[101], [105]

Table 2: Code point repertoire

### 5.3 Code point sequence

#	Unicode Code Point Sequence	Glyph	Note
S5	U+103B U+103D	ꨀꨂ	For WLE Rule #16, See Appendix B
S6	U+103B U+103E	ꨀꨃ	For WLE Rule #17, See Appendix B
S7	U+103C U+103D	ꨂꨂ	For WLE Rule #18, See Appendix B
S8	U+103C U+103E	ꨂꨃ	For WLE Rule #19, See Appendix B
S9	U+103D U+103E	ꨂꨃ	For WLE Rule #20, See Appendix B
S10	U+103C U+103D U+103E	ꨂꨂꨃ	For WLE Rule #21, See Appendix B
S11	U+1004 U+103A + U+1039	ꨀ	For WLE Rule #8, See Appendix B
S12	U+102D U+102F	ꨀꨃꨂ	Long vowel sequence, For WLE Rule #10, #11
S13	U+102F U+1036	ꨀꨃꨂ	long vowel sequence, For WLE Rule #10, #11
S14	U+1031 U+102B	ꨀꨂꨂ	long vowel sequence, For WLE Rule #7, #11
S15	U+1031 U+102C	ꨀꨂꨂ	long vowel sequence, For WLE Rule #7, #11
S16	U+1037 U+ 103A	ꨀꨂ	Short Vowel sequence, , For WLE Rule #7
SV1	U+1000 U+1039 U+1000	ꨀꨂꨀ	See 6.1

#	Unicode Code Point Sequence	Glyph	Note
SV2	U+1009 U+1037 U+103A	၌	See 6.1
SV3	U+1009 U+103A	၌	See 6.1
SV4	U+101B U+103E	၍	See 6.1
SV5	U+101E U+103C	၊	See 6.1
SV6	U+101E U+103C U+1031 U+102C U+103A	၊	See 6.1
SV7	U+1025 U+1037 U+103A	၌	See 6.1
SV8	U+1025 U+103A	၌	See 6.1
SV9	U+1029 U+1031 U+102C U+103A	၊	See 6.1
SV10	U+107D U+103E	၍	See 6.1
SS1	U+1025 U+102F U+1036	၊	1025 cannot join with other diacritics
SS2	U+1025 U+102F U+1038	၊	1025 cannot join with other diacritics
SS3	U+1026 U+1038	၊	1026 cannot join with other diacritics
SS4	U+1005 U+103D	၊	See Appendix B
SS5	U+1006 U+103D	၊	See Appendix B
SS6	U+1007 U+103D	၊	See Appendix B
SS7	U+1009 U+103E	၊	See Appendix B
SS8	U+100F U+103E	၊	See Appendix B

#	Unicode Code Point Sequence	Glyph	Note
SS9	U+1013 U+103D	၈	See Appendix B
SS10	U+101D U+103E	၉	See Appendix B
SS11	U+101F U+103D	၉	See Appendix B
SS12	U+1020 U+103E	၉	See Appendix B
SS13	U+1021 U+103D	၉	See Appendix B

Table 3: Code point sequences

## 5.4 Code point not included:

Following code points have not been included in the repertoire.

#	Unicode Code Point	Glyph	Character Name	Reason for exclusion
1	108B	◌◌	Myanmar sign Shan Council Tone-2	Never used in colloquial Shan
2	108C	◌◌◌	Myanmar sign Shan Council Tone-3	Never used in colloquial Shan
3	108D	◌◌◌◌	Myanmar sign Shan Council Emphatic Tone	Never used in colloquial Shan

Table 4: Code points not included

## 6. Variants

### 6.1. In-Script Variants

Myanmar script has in-script variants visually.

Set#	Unicode Code Point	Glyph	Unicode Code Point	Glyph
1	u1023	၉	u1000 + u1039 + u1000	၉

2	u1029	ၩ	u101E + u103C	ၩ
3	u102A	ၩ	u1029 + u1031+u102C+u103A	ၩ
4	u102A	ၩ	u101E+ u103C + u1031+ u 102C+ u103A	ၩ
5	u1061	ၩ	u101B + u103E	ၩ
6	u1009 + u103A	ၩ	u1025 + u103A	ၩ
7	u1009 + u1037 + u103A	ၩ	u1025 + u1037 + u103A	ၩ
8	u107E	ၩ	u107D + u103E	ၩ

Table 5: In-script Variants

## 6.2. Cross-Script Variants

Some characters of Myanmar look the same as some characters only in Malayalam.

### 6.2.1 Cross-Script Variants Analysis for Myanmar and Malayalam

Myanmar has the following variant code points with Malayalam script.

The Malayalam and Myanmar code point sets in Table(6) are cross-script variant code points.

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Malayalam Character Name
1	◌ <sup>◌</sup>	U+1002	MYANMAR LETTER GA	◌ <sup>◌</sup>	U+0D31	MALAYALAM LETTER RRA
2	◌ <sup>◌</sup>	U+101D	MYANMAR LETTER WA	◌ <sup>◌</sup>	U+0D20	MALAYALAM LETTER TTHA

Table 6: Myanmar and Malayalam code point and homoglyph analysis

### 6.2.2 Cross-Script Variants Analysis for Myanmar and Oriya

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Oriya Character Name
1	◌ <sup>◌</sup>	U+101D	MYANMAR LETTER WA	◌ <sup>◌</sup>	U+0B20	ORIYA LETTER TTHA
2	◌ <sup>◌◌</sup>	U+1031	MYANMAR VOWEL SIGN E	◌ <sup>◌</sup>	U+0B47	ORIYA VOWEL SIGN E

Table 7: Myanmar and Oriya code point and homoglyph analysis

## 7. Whole Label Evaluation Rules (WLE)

This section provides the WLE's that are required by all the languages mentioned in section 3 when written in Myanmar 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:

### Classification

C → Consonant

IV → Independent Vowel

DVS → Dependent Vowel Sign

ANUSVARA → 1036 \$◌̣

T\_SHORT → 1037 \$◌̣

T\_LONG → 1038 \$◌̣:

K → Killer: 103A \$◌̣

VIRAMA → 1039

M → Dependent Consonant Sign (Medial)

C1 → 103F ◌̣◌̣

LV → Long Vowel: 102B, 102C, 102E, 1030, 1031, 1032, 1036,

LVS → Long Vowel Sequence: 102D+102F , 1031+102B , 1031+102C ,  
102F+1036

SV → Short Vowel: 102D, 102F

Below are the specific WLE rules:

1. [DVS or ANUSVARA] must follow C or M
2. M must follow C
3. (C+ K) or (C +1037+K) must follow C or [C+M]
4. C1 must follow C

5. T\_SHORT must not follow [C or M or 102D or 102F or 103A or 1039]
6. T\_LONG must not follow [C or M or 102D or 102F or 1039]
7. S16 can only follow these consonants ( င, ဥ, ည, က, န, မ, ဝ ) or S14 or S15  
// this whole thing cannot be at the beginning of the string
8. S11 must be followed by C  
// this whole thing cannot be at the beginning of the string
9. VIRAMA must be between two consonants [ C Virama C ]  
// this whole thing cannot be at the beginning of the string
10. T\_LONG must follow [ LV or S12 or S13 or K]
11. T\_SHORT must follow [LV or S12 or S13 or S14 or S15 or T\_LONG]
12. U+103B must follow CS1
13. U+103C must follow CS2
14. U+103D must follow CS3
15. U+103E must follow CS4
16. S5 must follow CS5
17. S6 must follow CS6
18. S7 must follow CS7
19. S8 must follow CS8
20. S9 must follow CS
21. S10 must follow CS10

## 8. Contributors

Myanmar Generation Panel

## 9. References

- [101] Unicode Technical Note (11), Myanmar ( Range: 1000–109F ), The Unicode Standard 11.0, Copyright © 1991-2018 Unicode.
- [102] Myanmar Language - <https://www.ethnologue.com/country/MM>
- [103] Mon News - <http://mon.monnews.org/>
- [104] <http://www.omniglot.com/writing/burmese.htm>
- [105] <https://shanneews.org/>
- [106] [https://en.wikipedia.org/wiki/Pa%270\\_people](https://en.wikipedia.org/wiki/Pa%270_people)
- [107] <https://www.omniglot.com/writing/karen.htm>

[108] <http://guides.mmunicode.org/index.php/>

[109] <http://www.unicode.today/paoh-keyboard-2016-v2/>

## Appendix A - Confusable Code Points

Some code points were discussed and the conclusion is that these are not variant code points. However they are similar and confusable. This table should be used by other panel working on the string similarity cases.

Set#	Unicode Code Point	Glyph	Unicode Code Point	Glyph	Note
1	u1008	၈	u1005 + u103B	၈	The sequence U+1005 U+103B is invalid
2	u1009 + u102C	၉	u1025+ u102C	၉	The sequence u1025+ u102C is invalid

Table A-1: In-script confusable code points

## Confusable Characters between Myanmar Script and Malayalam

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Malayalam Character Name
1	က	U+1000	MYANMAR LETTER KA	ന	U+0D28	MALAYALAM LETTER NA
2	ယ	U+101A	MYANMAR LETTER YA	ധ	U+0D27	MALAYALAM LETTER DHA
3	ခ	U+1000, U+102C	MYANMAR LETTER KA , MYANMAR VOWEL SIGN AA	ക്ഷ	U+0D28, U+0D4D, U+0D28	MALAYALAM LETTER NA, MALAYALAM SIGN VIRAMA, MALAYALAM LETTER NA

Table A-2: Myanmar-Malayalam confusable code points

## Confusable Characters between Myanmar and Sinhala

Following are the comparison of Myanmar characters and Sinhala characters. Sinhala Diacritics have similarities with Myanmar diacritic characters and vowel sequence, however, there are no Sinhala consonant which look like Myanmar to be paired with these similar diacritics.

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Sinhala Character Name
1	꠵	U+1025	MYANMAR LETTER U	ඊ	U+0D8B	SINHALA LETTER UYANNA
2	꠶	U+102C	MYANMAR VOWEL SIGN AA	ඈ	U+0DCF	SINHALA VOWEL SIGN AELA-PILLA
3	꠷	U+1031	MYANMAR VOWEL SIGN E	ඉ	U+0DD9	SINHALA VOWEL SIGN KOMBUVA
4	꠸	U+1031, U102C	MYANMAR VOWEL SIGN AA, MYANMAR VOWEL SIGN E	ඊ	U+0DDC	SINHALA VOWEL SIGN KOMBUVA HAA AELA-PILLA

Table A-3: Myanmar-Sinhala confusable code points

### Confusable Characters between Myanmar and Telugu

There are two Telegu digits which look like variants of Myanmar Consonants. The GP notes that digits are out of scope of Root Zone LGR, however the table might be useful for the second level labels.

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Telugu Character Name
1	꠵	U+101D	MYANMAR LETTER WA	౦	U+0C66	TELUGU DIGIT ZERO
2	꠶	U+1002	MYANMAR LETTER GA	౧	U+0C67	TELUGU DIGIT ONE

Table A-4: Myanmar-Telugu confusable code points



	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
<b>C</b> <b>[u1000</b> <b>to</b> <b>u1021]</b>	ꠘ u103B	ꠙ u103C	ꠚ u103D	ꠛ u103E	ꠜ u103B u103D	ꠝ u103B u103E	ꠞ u103C u103D	ꠟ u103C u103E	ꠠ u103D u103E	ꠡ u103C u103D u103E
ꠣ U+100D										
ꠤ U+100E										
ꠥ U+100F				ꠦ						
ꠧ U+1010	꠨	꠩	꠪							
꠨ U+1011	꠬	꠭	꠮							
꠯ U+1012	꠰	꠱	꠲							
꠳ U+1013			꠴							
꠵ U+1014	꠶		꠷	꠸					꠹	
꠺ U+1015	꠻	꠼	꠽		꠾		꠿			
꠿ U+1016	ꡀ	ꡁ	ꡂ				ꡃ			
ꡄ U+1017	ꡅ	ꡆ	ꡇ							
ꡈ U+1018	ꡉ	ꡊ	ꡋ							
ꡌ U+1019	ꡍ	ꡎ	ꡏ	ꡐ	ꡑ	ꡒ	ꡓ	ꡔ	ꡕ	ꡖ
ꡗ U+101A	ꡘ		ꡙ	ꡚ						

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
<b>C</b> [u1000 to u1021]	ချ u103B	င u103C	ဝ u103D	ှ u103E	ပ u103B u103D	ဖ u103B u103E	ဇ u103C u103D	ဇ u103C u103E	ဝ u103D u103E	င u103C u103D u103E
ရ U+101B			ရ u101B	ရ u101B					ရ u101B	
လ U+101C	လ u101C		လ u101C	လ u101C		လ u101C			လ u101C	
ဝ U+101D				ဝ u101D						
သ U+101E			သ u101E			သ u101E				
ဟ U+101F			ဟ u101F							
ဇ U+1020				ဇ u1020						
အ U+1021			အ u1021							

Table B-1: The combinations of Consonant Medial Sequences

## Appendix C

Thee GP plans to propose the code points to b included in the next version of MSR as follows:

#	Unicode Code Point	Glyph	Character Name	Category	Language (s), with EGIDS Value	Reference
1	U+1028	ꠅ	MYANMAR LETTER MON E	Independent Vowel	5 Mon	[101], [102], [103]
2	U+1033	꠆	MYANMAR VOWEL SIGN MON II	Dependent Vowel Sign	5 Mon	[101], [102], [103]
3	U+1034	ꠇ	MYANMAR VOWEL SIGN MON O	Dependent Vowel Sign	5 Mon	[101], [102], [103]
4	U+105A	ꠊ	MYANMAR LETTER MON NGA	Consonant	5 Mon	[101], [102], [103]
5	U+105B	ꠋ	MYANMAR LETTER MON JHA	Consonant	5 Mon	[101], [102], [103]
6	U+105C	ꠌ	MYANMAR LETTER MON BBA	Consonant	5 Mon	[101], [102], [103]
7	U+105D	ꠍ	MYANMAR LETTER MON BBE	Consonant	5 Mon	[101], [102], [103]
8	U+105E	ꠎ	MYANMAR CONSONANT SIGN MON MEDIAL NA	Medial	5 Mon	[101], [102], [103]
9	U+105F	ꠏ	MYANMAR CONSONANT SIGN MON MEDIAL MA	Medial	5 Mon	[101], [102], [103]
10	U+1060	ꠐ	MYANMAR CONSONANT SIGN MON MEDIAL LA	Dependent Consonant Sign	3 S'gaw Karen, 5 Mon	[101], [102], [103]
11	108F	ꠑ	MYANMAR SIGN RUMAI PALAUNG TONE-5	Tone mark	5 Pa'O Karen	[101], [106], [108]
12	AA7B	ꠒ	MYANMAR SIGN PAO KAREN TONE	Tone mark	5 Pa'O Karen	[101], [106], [108]

Table C-1: Candidate code points for MSR-4