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

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 stated in detail here.

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: 5

3.　　Background

3.1.　　History of the Myanmar Script

The Burmese Language (endonym- Myanmar) is the official national language of Myanmar and is a member of the Tibeto-Burman Language family which is a sub-family of the Sino-Tibetan. To avoid confusion with the script family name "Myanmar", the language name will be used as the xenonym "Burmese" in this proposal.

The letters of the alphabet used in the 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 the earliest surviving records can be dated as far as the 5th Century of the Myanmar Era (12th Century AD).

The rounded appearance of letters is a result of the use of palm leaves as the traditional writing material. Straight lines would have torn the leaves. The Burmese name for the script is "စာလုံး" /sɑ-loʊː/, which means 'round script'.

Burmese is a tonal language with three main tones (high, low and creaky) and two other tones (stopped and reduced). The tones are indicated in writing using diacritics or special letters. [104]

The old Burmese script was derived from the Mon-Khmer script and adapted to the languages spoken in the Irrawaddy-Chindwin region such as Pyu, Thet (Sak) and Kanyan (Kamyam). In the Bagan era (AD 11-12), the first spelling normalization was carried out by inviting monks from Sri Lanka Island and Mon Kingdom (also known as Suvarnabhumi Kingdom). In the Inwa (Ava) era (AD 15-16), tone marks are introduced to differentiate the sound of the words and spelling was normalised. Modern standard Burmese language is developed based on Inwa (Ava) literature. In the early 20th Century, scholars of Yangon University did research on colloquial realization of Burmese language. The Government of the Union of Myanmar declared Burmese (endonym- Myanmar) as the official language of the country, and standardized the language curricula, textbooks and dictionaries with the ability of *Myanmar Language Commission* [100] from 1963 to date.

3.2. Principal Languages Using the Myanmar Script

The languages that currently use the Myanmar script, and which have EGIDS scale from 1 to 4 as well as 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.

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 /mja–ma/	1	42,906,490
Shan	[shn]	Myanmar, China, Thailand	လိၵ်ႈတႆး /lĩk tái/	3	3,295,000
Rakhine	[rki]	Myanmar	ရက္ခိုင် /rə.khái/	3	2,020,000
Sgaw Karen	[ksw]	Myanmar, Thailand	စှီၢ် /sɣɔʔ/	3	1,560,000

Mon	[mnw]	Myanmar, Thailand	မန် /mun/	5	851,000
Pa'O Karen	[blk]	Myanmar	ပအိုဝ်း /pə.ʔəʊ:/	5	560,740

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

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

Pa'O Karen[blk] is one variation of the Karen languages, written using almost the same letters as Burmese, with two additional tone marks (U+108F and U+AA7B), only excluding one medial mark (U+103E). This language has EGIDS scale 5 with 560,740 total speakers.

The word "Hello" written in listed languages:

Burmese: "မင်္ဂလာပါ" /mɪŋ-gə-la-ba/

Shan: "မ်ႂ်ႇသုင်ၶႃႈ" /mɪ-suŋ-kha̰/

Rakhine: "သာလီစွပါ" /θa-li-zwa̰-ba/

Sgaw Karen: "နံၤလၤအဂ့ၤ" /niː-lḛə-ə-ɣeː/

Mon: "မ္ၚိရအဴ" /mŋeə-rə-aʊ/

Pa'O Karen: "မင်္ဂလားဒျားၩ" /mɪŋ-gə-la̰-drḛ-ɔː]

3.3. Features and Categories of Characters

The Myanmar script is written from left to right and requires no spaces between words, except to make phrase or sentence segments.

The type of writing system is syllabic alphabet - each letter has an inherent vowel [a]. Other vowel sounds are indicated using separate letters or diacritics which appear above, below, in front of, after or around the consonant.[104]

The categories of Myanmar characters are:

1. Consonant (C), see 3.3.1
2. Independent Vowel (IV), see 3.3.2
3. Diacritic
    3.1. Dependent Vowel (DV), see 3.3.3
    3.2. Medial or Dependent Consonant (M), see 3.3.4
    3.3. Tone Mark (T), see 3.3.5
    3.4. Other Signs(O), see 3.3.5

There is usually a Consonant(C) at the head of each syllable. If not, the syllable is headed by an Independent Vowel(IV). A Consonant can be stand-alone or followed by Medials (M), Dependent vowels (DV), Other Signs (O), or Tone marks (T).

Example word:

ေပြး (ပ+ ြ◌ + ေ◌ + ◌း)   [transcription 'pje:'],
                                      meaning is "Run" and it is a sequence of (C+M+DV+T)

/ ပ / ြ◌ / ေ◌ / ◌း /

/ U+1015 / U+103C / U+1031 / U+1038 /

Figure 1: Composing the syllable "ေပြး" according to canonical order

The sound of Consonant Pa U+1015 is fused with Medial Ya U+103C, and then Dependent vowel and Tone mark are followed orderly. Vowel cannot come between Consonant and Medial, thus, WLE rule #1 restricts the order.

A single consonant can have up to 7 diacritics, which can be a combination of medial marks, vowel diacritics and tone marks. All diacritics are encoded following 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.

Example words with Independent vowel :

ဦး (U+1026 U+1038)   [transcription 'u:'],
                                      meaning is "head" (also used as Prefix of Men's name),
                                      it is a sequence of IV + T

ဧသာ (U+1029 U+1031 U+102C)   [transcription 'ɔ:'],
meaning is "alas" and it is a sequence of IV + DV + DV

ဥုံ (U+1025 U+102F U+1036)   [transcription 'oʊɴ'],
meaning is "holy", it is a sequence of
IV + DV_Sequence(S13)

Dependent Vowels, Diacritics, Medials, and Tone marks cannot occur at the beginning of the syllable. They must follow a consonant or an independent vowel.

Some consonant-diacritic combinations are impossible according to each ethnic language. These will cause strange combinations and rendering issues. Thus, the characters should be treated according to respective language categories in the WLE rules (part 7).

Some characters can be one category in one language and be a different category in another language. For the consonant and independent vowel there is only one case: U+1027 ဧ MYANMAR LETTER E. It is a S'gaw Karen consonant as well as as an Independent Vowel in Burmese. There are more of these cases for Dependent Vowels, Diacritics, Medials, and Tone marks. See Table 2, 3, 4, 5 and 6 for mapping matrices.

Table 7 lists all the code points included in this LGR. The sequences in Table 8 are to handle rendering and rule exceptions. They are there to make sure the unicode sequence is in proper order to create a well-formed label. These are not spelling or grammar rules.

3.3.1. The Consonants

Consonants in the Myanmar script relating to the languages are shown in Table 2: *List of Consonants*. Consonants usually stay at the head of each syllable. As described earlier, a consonant can be stand-alone or be followed by Medials(M), Dependent Vowels (DV), Other Signs (O), or Tone Mark (T). When a consonant is followed by Asat or Asat sequences (S16, S17) , it becomes a follower.
Different set of Myanmar script consonants are used in five languages covered in this proposal. Table 2 maps consonants and their usages.

Table 2: List of Consonants

Code Point	Glyph	Unicode Name	Myanmar	Shan	S'gaw Karen	Mon	Pa'O
U+1000	က	MYANMAR LETTER KA	✓		✓	✓	✓
U+1001	ခ	MYANMAR LETTER KHA	✓		✓	✓	✓
U+1002	ဂ	MYANMAR LETTER GA	✓		✓	✓	✓
U+1003	ဃ	MYANMAR LETTER GHA	✓		✓	✓	✓
U+1004	င	MYANMAR LETTER NGA	✓	✓	✓		✓

Code	Glyph	Name					
U+1005	စ	MYANMAR LETTER CA**	✓		✓	✓	✓
U+1006	ဆ	MYANMAR LETTER CHA	✓		✓	✓	✓
U+1007	ဇ	MYANMAR LETTER JA	✓			✓	✓
U+1008	ဈ	MYANMAR LETTER JHA	✓				✓
U+1009	ဉ	MYANMAR LETTER NYA	✓			✓	✓
U+100A	ည	MYANMAR LETTER NNYA	✓		✓	✓	✓
U+100B	ဋ	MYANMAR LETTER TTA	✓			✓	✓
U+100C	ဌ	MYANMAR LETTER TTHA	✓			✓	✓
U+100D	ဍ	MYANMAR LETTER DDA	✓			✓	✓
U+100E	ဎ	MYANMAR LETTER DDHA	✓			✓	✓
U+100F	ဏ	MYANMAR LETTER NNA	✓			✓	✓
U+1010	တ	MYANMAR LETTER TA	✓	✓	✓	✓	✓
U+1011	ထ	MYANMAR LETTER THA	✓	✓	✓	✓	✓
U+1012	ဒ	MYANMAR LETTER DA	✓		✓	✓	✓
U+1013	ဓ	MYANMAR LETTER DHA	✓			✓	✓
U+1014	န	MYANMAR LETTER NA	✓		✓	✓	✓
U+1015	ပ	MYANMAR LETTER PA	✓	✓	✓	✓	✓
U+1016	ဖ	MYANMAR LETTER PHA	✓		✓	✓	✓
U+1017	ဗ	MYANMAR LETTER BA	✓			✓	✓
U+1018	ဘ	MYANMAR LETTER BHA	✓		✓	✓	✓
U+1019	မ	MYANMAR LETTER MA	✓	✓	✓	✓	✓
U+101A	ယ	MYANMAR LETTER YA	✓	✓	✓	✓	✓
U+101B	ရ	MYANMAR LETTER RA	✓	✓	✓	✓	✓
U+101C	လ	MYANMAR LETTER LA	✓	✓	✓	✓	✓
U+101D	ဝ	MYANMAR LETTER WA	✓	✓	✓	✓	✓
U+101E	သ	MYANMAR LETTER SA**	✓	✓	✓	✓	✓
U+101F	ဟ	MYANMAR LETTER HA	✓		✓	✓	✓
U+1020	ဠ	MYANMAR LETTER LLA	✓			✓	✓
U+1021	အ	MYANMAR LETTER A	✓		✓	✓	✓
U+1022	ဢ	MYANMAR LETTER SHAN A		✓			
U+1027	ဧ	MYANMAR LETTER E	*		✓		
U+103F	ဿ	MYANMAR LETTER GREAT SA**	✓				✓

U+105A	ç	MYANMAR LETTER MON NGA				✓	
U+105B	ၛ	MYANMAR LETTER MON JHA				✓	
U+105C	ၜ	MYANMAR LETTER MON BBA				✓	
U+105D	ၝ	MYANMAR LETTER MON BBE				✓	
U+1061	ၡ	MYANMAR LETTER SGAW KAREN SHA			✓		
U+1075	ၵ	MYANMAR LETTER SHAN KA		✓			
U+1076	ၶ	MYANMAR LETTER SHAN KHA		✓			
U+1077	ၷ	MYANMAR LETTER SHAN GA		✓			
U+1078	ၸ	MYANMAR LETTER SHAN CA		✓			
U+1079	ၹ	MYANMAR LETTER SHAN ZA		✓			
U+107A	ၺ	MYANMAR LETTER SHAN NYA		✓			
U+107B	ၻ	MYANMAR LETTER SHAN DA		✓			
U+107C	ၼ	MYANMAR LETTER SHAN NA		✓			
U+107D	ၽ	MYANMAR LETTER SHAN PHA		✓			
U+107E	ၾ	MYANMAR LETTER SHAN FA		✓			
U+107F	ၿ	MYANMAR LETTER SHAN BA		✓			
U+1080	ႀ	MYANMAR LETTER SHAN THA		✓			
U+1081	ႁ	MYANMAR LETTER SHAN HA		✓			

*S'gaw Karen consonant "ဧ" U+1027 is also used as an Independent Vowel in Burmese.

** The letters SA and GREAT SA are transcribed as [θ], the letter CA, is transcribed as [s] according to native pronunciation.

3.3.1.1.   Burmese Consonants

There are 35 consonants in Burmese. The consonants can be standalone as they possess inherent vowel "-a" sound without any vowel diacritics.

Example words with standalone consonants:
-    ပထမ (U+1015 U+1011 U+1019)
   -    [transcription 'pəˈ-təˈ-ma̰'], meaning is "first"
-    မဒမ (U+1019 U+1013 U+1019)
   -    [transcription 'məˈ-dəˈ-ma̰'], meaning is "pole" (as vertical roof support in buildings)

6

- တသသ (U+1010 U+101E U+101E) .
  - [transcription 'tə'-θa̰-θa̰'],
    meaning is "reminiscence"

More than three instances of the same consonant do not occur consecutively. The same consonant can repeat up to 3 times only, however, GP will not restrict to allow creative naming.

Example words with 3 consecutive consonants:
- စစစ (U+1005 U+1005 U+1005)
  - [transcription 'sa̰-sa̰-sa̰'],
    meaning is "acronym for detective department"
- ကကကြည်း (U+1000 U+1000 U+1000 U+103C U+100A U+103A U+1038)
  - [transcription 'Ka̰-Ka̰-tʃi:'],
    meaning is "Army specific"
    See explanation of 'K' to 'tʃ' transformation in section 3.3.4 Diacritic - Medials.

The special consonant "သသ" (U+103F) /θa̰/ can only follow other consonants (with or without medial, vowel combinations) or "ဣ" (U+1023) /ḭ/ or "ဥ"(U+1025) /ṵ/ .

3.3.1.2.   Shan Consonants

There are 24 consonants in Shan. They are treated the same way as Burmese consonants, despite having different sets of Vowels (3.3.2) and Diacritics(3.3.3).
14 Out of 24 Shan consonants are non-Burmese, and Shan has own code points for those.

3.3.1.3.   S'gaw Karen Consonants

There are 25 consonants in S'gaw Karen. They are treated the same way as Burmese consonants, despite having different sets of Vowels (3.3.2) and Diacritics(3.3.3). In S'gaw Karen scripts, Myanmar Letter E ("ဧ" U+1027) is regarded as a consonant. The spellings with character "ဧ" as consonant are defined in Table 8.

3.3.1.4.   Mon Consonants

There are 36 consonants in Mon. Most Mon consonants are the same as Burmese Consonants except the following: "ၚ" (U+105A) , "ၛ" (U+105B) , "ၜ" (U+105C) and "ၝ" (U+105D) . Mon does not use Burmese consonants "င" (U+1004) and "ဈ" (U+1008) and "သသ" (U+103F).

3.3.1.5.   Pa'O Consonants

All the same as Burmese Consonants.

3.3.2. The Independent Vowels

Only Burmese, Mon and Pa'O languages use Independent Vowels. An Independent Vowel can be at the beginning of a syllable. Most of these Independent Vowels can be stand-alone. The exception sequences are defined in Table 8 (see 3.3.2.5 and 5.3, Table 8, SS1 and SS2).

Table 3: List of Independent Vowels

Code Point	Glyph	Unicode Name	Myanmar	S'gaw Karen	Mon	Pa'O
U+1023	အ	MYANMAR LETTER I	✓		✓	✓
U+1024	ဤ	MYANMAR LETTER II	✓			✓
U+1025	ဥ	MYANMAR LETTER U	✓		✓	✓
U+1026	ဦ	MYANMAR LETTER UU	✓			✓
U+1027	ဧ	MYANMAR LETTER E	✓	*		✓
U+1028	ဨ	MYANMAR LETTER MON E			✓	
U+1029	ဩ	MYANMAR LETTER O	✓		✓	✓
U+102A	ဪ	MYANMAR LETTER AU	✓		✓	✓

*Myanmar Independent vowel "ဧ" U+1027 is used as a consonant in S'gaw Karen language.

3.3.2.1. Burmese Independent Vowels

Burmese Independent vowels can be stand-alone and also can be the head of the string.

Some of these vowels can be spelled with diacritics or Devowelizers (being followed by (Consonant+Virama +Consonant ...) or special consonant ("သ" U+1036). [101] Virama is a Diacritic.

Example with diacritics :

- ဥုံ (U+1025 U+102F U+1036)   [transcription 'oʊ'],
  meaning is "holy", it is a sequence of
  IV + DV_Sequence(S13)
- ဥူး (U+1025 U+102F U+1038)   [transcription 'uː'],
  meaning is "Prefix of Monk's name",
  it is a sequence of IV + DV + T

Example with Devowelizer (Consonant+Virama+Consonant …) :

- ဣ္တ္တိယ (U+1023 U+1010 U+1039 U+1011 U+102D U+101A)
    - [transcription ' e̤ɪʔ-thḭ-ja̰'],
      meaning is "women"

- ဥက္ကလာ (U+1025 U+1000 U+1039 U+1000 U+101C U+102C)
    - [transcription 'oʊʔ-kə-la'],
      meaning is "Head-related"

Example with special consonant:

- ဣဿရမ္ဴလီ (U+1023 U+103F U+101B U+1019 U+1030 U+101C U+102E)
    - [transcription 'eɪʔ-θə-ra̰-mu-li'],
      meaning is "spiral" (as in plant)

- ဥဿဖရား (U+1025 U+103F U+1016 U+101B U+102C U+1038)
    - [transcription 'oʊʔ-θə-phə-yɑː'],
      meaning is "zircon"

3.3.2.2.   Mon Independent Vowels

There are five Mon Independent Vowels in Table 3 and there are two sequences of Independent Vowels. No other diacritics or devowelisers need to follow.

Mon Independent Vowel Sequences

1. ဣ္ဳ U+1023 U+1033 (See Table 8, S_Mon1)

    This is equivalent to Myanmar vowel "ဤ" U+1024. Mon uses the sequence "ဣ္ဳ" for the sound of character "ဤ" [transcription 'i']

2. ဥု U+1025 U+102F (See Table 8, S_Mon2)

    This is equivalent to Myanmar vowel "ဦ" U+1026. Mon uses the sequence "ဥု" for the sound of character "ဦ" [transcription 'u']

3.3.3. Diacritic - Dependent Vowels

Dependent vowel signs add vowel property to consonants. These signs appear in top/below/left/right positions of a center consonant or consonant+medial. The dotted-circle indicates where the centre character would be.

Dependent vowel signs cannot be repeated and Dependent vowels cannot be adjacent to each other unless within sequences defined in Table 8. Dependent vowels also cannot be followed by Asat (U+103A) unless within sequences defined in Table 8.

Table 4: List of Dependent Vowels

| Code Point	Glyph	Unicode Name	Myanmar	Shan	S'gaw Karen	Mon	Pa'O
U+102B	ါ	MYANMAR VOWEL SIGN TALL AA	✓		✓	✓	✓
U+102C	ာ	MYANMAR VOWEL SIGN AA	✓		✓	✓	✓
U+102D	ိ	MYANMAR VOWEL SIGN I	✓	✓	✓	✓	✓
U+102E	ီ	MYANMAR VOWEL SIGN II	✓	✓	✓		✓
U+102F	ု	MYANMAR VOWEL SIGN U	✓	✓	✓	✓	✓
U+1030	ူ	MYANMAR VOWEL SIGN UU	✓	✓	✓	✓	✓
U+1031	ေ	MYANMAR VOWEL SIGN E	✓	✓		✓	✓
U+1032	ဲ	MYANMAR VOWEL SIGN AI	✓		✓	✓	✓
U+1033	ဳ	MYANMAR VOWEL SIGN MON II				✓	
U+1034	ဴ	MYANMAR VOWEL SIGN MON O				✓	
U+1035	ဵ	MYANMAR VOWEL SIGN E ABOVE		✓		✓	
U+1036	ံ	MYANMAR SIGN ANUSVARA	✓		✓	✓	✓
U+1037	့	MYANMAR SIGN DOT BELOW	✓		✓		✓
U+1038	း	MYANMAR SIGN VISARGA	✓	✓	✓	✓	✓
U+1062	ၢ	MYANMAR VOWEL SIGN SGAW KAREN EU			✓		
U+1083	ႃ	MYANMAR VOWEL SIGN SHAN AA		✓			
U+1084	ႄ	MYANMAR VOWEL SIGN SHAN E		✓			
U+1086	ႆ	MYANMAR VOWEL SIGN SHAN FINAL Y		✓			

Myanmar Vowel Sign AA "ာ" U+102C is a dependent vowel for the Burmese, Mon and Pa'O languages. In Sgaw Karen, it is used as Tone sequence "ာ်" (U+102C U+103A) (see 3.3.5.3 S'gaw Karen Tones; see also ST2 in Table 8).

Myanmar Vowel Sign E Above " ဵ " U+1035 is only used in the Dependent Vowel Sequence " ဵ " (U+1035 U+102F) in Mon (See 3.3.3.4). In Shan, " ဵ " U+1035 is used as Other Sign (See 3.3.5.2 Shan Other Signs).

Myanmar Sign Anusvara "ံ" U+1036 is considered a Dependent Vowel in S'gaw Karen and Mon, but it is part of the Other Sign category for Burmese and Pa'O Karen. In addition, both Burmese and Pa'O Karen use U+1036 as part of a vowel sequence "ုံ" (U+102F U+1036) (See 3.3.3.1, 3.3.3.5).

Myanmar Sign Dot Below "့" U+1037 is a Dependent Vowel for the S'gaw Karen language, however, it is a Short Tone Sign for Burmese and Pa'O Karen (see 3.3.5.1 ,3.3.5.5).

Myanmar Sign Visarga "း" U+1038 is a vowel for Mon, however, it is a Tone Sign for Burmese, Shan, Sgaw Karen and Pa'O Karen (see 3.3.3.1, 3.3.3.5).

Myanmar Vowel Sign Sgaw Karen "ၢ" U+1062 is a Dependent Vowel when used alone, as well can be used as part of a tone sequence "ၢ်" (U+1062 U+103A) (see 3.3.5.3 S'gaw Karen Tones).

3.3.3.1. Burmese Dependent Vowels

Due to the overlapping issue shown in Table 4, the dependent vowels for each language are respectively listed below for each language for clarification.

Burmese Dependent Vowels:
1. U+102B   ါ   MYANMAR VOWEL SIGN TALL AA
2. U+102C   ာ   MYANMAR VOWEL SIGN AA
3. U+102D   ိ   MYANMAR VOWEL SIGN I
4. U+102E   ီ   MYANMAR VOWEL SIGN II
5. U+102F   ု   MYANMAR VOWEL SIGN U
6. U+1030   ူ   MYANMAR VOWEL SIGN UU
7. U+1031   ေ   MYANMAR VOWEL SIGN E
8. U+1032   ဲ   MYANMAR VOWEL SIGN AI

Dependent vowels require a consonant or consonant+medial or special consonant ("ဿ" U+103F) to lead. If the dependent vowel follows the special consonant "ဿ" U+103F, there should be some character before it.

Example :

- ဘူ (U+1018 U+1030)

- [transcription: 'bu']

- ဇွာ (U+1007 U+103D U+102C)
  - [transcription: 'zwa']

- ဥဿာ (U+1025 U+103F U+102C)

  - [transcription: 'ʊʔ-θa']

Burmese: Dependent Vowels: Short Vowels

The dependent vowels that originally had sharp sound are "ိ" (U+102D) and "ု" (U+102F). These vowel signs are not followed by tone marks "့" (U+1037) and "း" (U+1038).

Burmese: Dependent Vowels: Long Vowels

The dependent vowels that originally have normal tones, also known as long sound are "ါ" (U+102B) , "ာ" (U+102C) , "ီ" (U+102E) , "ူ" (U+1030) , "ေ" (U+1031) , "ဲ" (U+1032) . All of these vowels can be followed by tone marks "့" (U+1037) or "း" (U+1038).

The short tone mark "့" (U+1037) shortens the vowel sound while the long tone mark "း" (U+1038) creates a vowel sound with the higher tone.

Burmese: Dependent Vowel Sequences

1.	ေါ	U+1031 U+102B	Defined as S14 in Table 8
2.	ော	U+1031 U+102C	Defined as S15 in Table 8
3.	ေါ်	U+1031 U+102B U+103A	Defined as S19 in Table 8
4.	ော်	U+1031 U+102C U+103A	Defined as S20 in Table 8
5.	ို	U+102D U+102F	Defined as S12 in Table 8
6.	ိံ	U+102D U+1036	Defined as S18 in Table 8
7.	ုံ	U+102F U+1036	Defined as S13 in Table 8

Some dependent vowels can be paired as the sequences above. These vowel sequences also must follow consonants or consonants+medial or special consonant ("ဿ" U+103F). All the vowel sequences are "Long vowel sequences".

As the short tone sign "့" (U+1037) must be always in front of "်" (U+103A) , the long vowel sequence "ေါ်" (U+1031 U+102B U+103A) becomes "ေါ့်" (U+1031 U+102B U+1037 U+103A). and sequence "ော်" (U+1031 U+102C U+103A) becomes "ော့်" (U+1031 U+102C U+1037 U+103A). For the rest of long vowel sequences, the short tone "့" (U+1037) comes last.

The long tone sign "ေ:" (U+1038) can follow two of the long vowel sequences. These are "ို" (U+102D U+102F) and "ုံ" (U+102F U+1036) .

Exception cases where the vowel sequence follows an Independent Vowel are documented. (See part 5.3, Table 8, SS1 and SS2.

Example:

- ကို (U+1000 U+102D U+102F)
    - [transcription: 'ko']

- လျှော် (U+101C U+103B U+103E U+1031 U+102C U+103A)
    - [transcription: 'ʃɔ']

- ဥဿုံ (U+1025 U+103F U+102F U+1036)

    - [transcription: 'oʊʔ-θoʊ']

3.3.3.2. Shan Dependent Vowels

Shan Dependent Vowels:
1. U+102D   ိ   MYANMAR VOWEL SIGN I
2. U+102E   ီ   MYANMAR VOWEL SIGN II
3. U+102F   ု   MYANMAR VOWEL SIGN U
4. U+1030   ူ   MYANMAR VOWEL SIGN UU
5. U+1031   ေ   MYANMAR VOWEL SIGN E
6. U+1083   ႃ   MYANMAR VOWEL SIGN SHAN AA
7. U+1084   ႄ   MYANMAR VOWEL SIGN SHAN E
8. U+1086   ႆ   MYANMAR VOWEL SIGN SHAN FINAL Y

Shan dependent vowels require Shan consonants or Shan consonants + Shan medial to lead. In Shan writing, there are only dependent vowels and no stand-alone Independent Vowels.

Shan: Dependent Vowel Sequences

1. ေႃ   U+1031 U+1083   Defined as S_Sh1 in Table 8
2. ေႃ်   U+1031 U+1083 U+103A   Defined as S_Sh2 in Table 8
3. ို   U+102D U+102F   Defined as S12 in Table 8
4. ိူ   U+102D U+1030   Defined as S_Sh3 in Table 8
5. ႂႆ   U+1062 U+1086   Defined as S_Sh4 in Table 8
6. ႂ်   U+1082 U+103A   Defined as S_Sh5 in Table 8

Some dependent vowels can be paired as above. Shan Vowel sequences can only follow Shan consonants or Shan medial. Only the defined vowel sequences are used. A dependent vowel cannot be repeated or randomly follow another one.

3.3.3.3.   Sgaw Karen Dependent Vowels

Sgaw Karen Dependent Vowels:
1. U+102B   ◌ါ   MYANMAR VOWEL SIGN TALL AA
2. U+102D   ◌ိ   MYANMAR VOWEL SIGN I
3. U+102E   ◌ီ   MYANMAR VOWEL SIGN II
4. U+102F   ◌ု   MYANMAR VOWEL SIGN U
5. U+1030   ◌ူ   MYANMAR VOWEL SIGN UU
6. U+1032   ◌ဲ   MYANMAR VOWEL SIGN AI
7. U+1036   ◌ံ   MYANMAR SIGN ANUSVARA
8. U+1037   ◌့   MYANMAR SIGN DOT BELOW
9. U+1062   ◌ၢ   MYANMAR VOWEL SIGN SGAW KAREN EU

S'gaw Karen dependent vowels require a S'gaw Karen Consonant or S'gaw Karen Consonant + S'gaw Karen Medial to lead. A dependent vowel cannot be repeated or randomly follow another one.

3.3.3.4.   Mon Dependent Vowels

Mon Dependent Vowels:
1. U+102B   ◌ါ   MYANMAR VOWEL SIGN TALL AA
2. U+102C   ◌ာ   MYANMAR VOWEL SIGN AA
3. U+102D   ◌ိ   MYANMAR VOWEL SIGN I
4. U+102F   ◌ု   MYANMAR VOWEL SIGN U
5. U+1030   ◌ူ   MYANMAR VOWEL SIGN UU
6. U+1031   ေ◌   MYANMAR VOWEL SIGN E
7. U+1032   ◌ဲ   MYANMAR VOWEL SIGN A
8. U+1033   ◌ၳ   MYANMAR VOWEL SIGN MON II
9. U+1034   ◌ၴ   MYANMAR VOWEL SIGN MON O
10. U+1036   ◌ံ   MYANMAR SIGN ANUSVARA
11. U+1038   ◌း   MYANMAR SIGN VISARGA

Mon dependent vowels require Mon consonants or Mon consonant + Mon medial to lead.

Mon: Dependent Vowel Sequences

1. ေါ   U+1031 U+102B   Defined as S14 in Table 8
2. ော   U+1031 U+102C   Defined as S15 in Table 8
3. ◌ိ   U+102D U+102F   Defined as S12 in Table 8
4. ◌ၱ   U+1035 U+102F   Defined as S_Mon3 in Table 8

5. ◌ါံ   U+102B U+1036   Defined as S_Mon5 in Table 8

6. �ော် U+102C U+1036　　　Defined as S_Mon6 in Table 8

7. ို U+102D U+102F U+1032　　　Defined as S_Mon7 in Table 8

8. ေါ် U+1031 U+102B U+1036　　　Defined as S_Mon8 in Table 8

9. ော် U+1031 U+102C U+1036　　　Defined as S_Mon9 in Table 8

10. ေဲ U+1031 U+1032　　　Defined as S_Mon10 in Table 8

11. ဲာ U+1032 U+102B　　　Defined as S_Mon11 in Table 8

12. ဲာ U+1032 U+102C　　　Defined as S_Mon12 in Table 8

13. ျ် U+103E U+103A　　　　　Defined as S_Mon13 in Table 8

14. ေျ် U+103E U+103A U+1031　　　　　Defined as S_Mon14 in Table 8

15. ေျာ် U+103E U+103A U+1031 U+102C　　　Defined as S_Mon15 in Table 8

16. ောျ် U+1031 U+102C U+103E U+103A　　　Defined as S_Mon16 in Table 8

Mon dependent vowels can be used in the above sequences. These sequences come after a Mon consonant standing alone or a Mon consonant + Mon medial. Only the defined vowel sequences are used. A dependent vowel cannot be repeated or randomly follow another one.

3.3.3.5.　　　Pa'O Karen Dependent Vowels

Pa'O dependent vowels are all the same as Burmese dependent vowels. The only one difference is that in Pa'O Karen writing, there are only dependent vowels and no stand-alone Independent Vowels.

Pa'O Karen: Dependent Vowel Sequences

1. ေါ U+1031 U+102B　　　　　Defined as S14 in Table 8
2. ော U+1031 U+102C　　　　　Defined as S15 in Table 8
3. ေါ် U+1031 U+102B U+103A　　　Defined as S19 in Table 8
4. ော် U+1031 U+102C U+103A　　　Defined as S20 in Table 8
5. ို U+102D U+102F　　　　　Defined as S12 in Table 8
6. ုဲ U+102F U+1032　　　　　Defined as S_Pao in Table 8
7. ုံ U+102F U+1036　　　　　Defined as S13 in Table 8

Pa'O Karen dependent vowels can be paired as in the above sequences. These sequences follow Pa'O Karen Consonant or Pa'O Karen Medial. Only the defined vowel sequences are used. A dependent vowel cannot be repeated or randomly follow another one.

3.3.4. Diacritic - Medials

Medials are used to enhance the sound of Consonants. They are also noted as Dependent Consonants as they need a leading Consonant to attach to. All five languages in this proposal use Medials.

Medial is not an alternative consonant. One consonant being followed by Medial_Ra is different from being followed by Consonant_Ra.

Example 1: "ကျ" [U+1000 U+103C] vs "ကရ" [U+1000 U+101B]

The pronunciation of Medial add Letter_Ka becomes /tʃə̰/ , and the latter is /kə-jə̰/.

Example 2: "ပြ" [U+1015 U+103C] vs "ပရ" [U+1015 U+101B]

The pronunciation of Medial added Letter_Pa becomes /pjə̰/ , and the latter is /pə-jə̰/.

"ပြ" [U+1015 U+103C] vs "ပရ" [U+1015 U+101B]
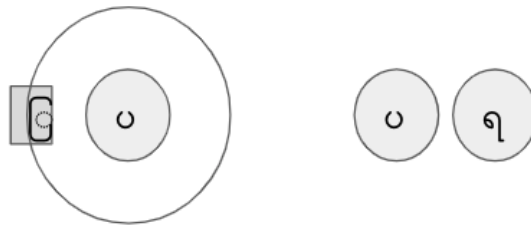
Figure 2: Composing "ပြ" [U+1015 U+103C] vs "ပရ" [U+1015 U+101B]

Example 3: Combining Consonant La "လ" and Medial Ha "ှ" gives result to a nasal consonant sound, which has grapheme "လှ" , with transcription "/l̥ə̰/". Adding medial 'Ha' to the consonant 'La' does not make two sounds added together. Instead, they become a new phoneme which is a fusion of the two.

Table5: List of Medials

Code Point	Glyph	Unicode Name	Myanmar	Shan	S'gaw Karen	Mon	Pa'O
U+103B	ျ	MYANMAR CONSONANT SIGN MEDIAL YA	✓	✓	✓	✓	✓
U+103C	ြ	MYANMAR CONSONANT SIGN MEDIAL RA	✓	✓	✓	✓	✓
U+103D	ွ	MYANMAR CONSONANT SIGN MEDIAL WA	✓		✓	✓	✓

U+103E	◌	MYANMAR CONSONANT SIGN MEDIAL HA	✓		✓	✓	
U+105E	◌	MYANMAR CONSONANT SIGN MON MEDIAL NA				✓	
U+105F	◌	MYANMAR CONSONANT SIGN MON MEDIAL MA				✓	
U+1060	◌	MYANMAR CONSONANT SIGN MON MEDIAL LA			✓	✓	
U+1082	◌	MYANMAR CONSONANT SIGN SHAN MEDIAL WA		✓			

3.3.4.1. Burmese Medials

Burmese Medials

There are four Burmese Medial signs as listed in Table 5, they are led by Burmese Consonants, some Medials can follow another combination of certain Consonant-Medial combinations and can be followed by Burmese Dependent Vowels, Burmese Vowel Sequences or Anusavra (U+1036). The same medial cannot be repeated and it cannot be followed by Asat U+103A. Medials cannot randomly follow each other except for the defined sequences in Table 8 .

For example:

- ချ (U+1001 U+103B)
    - [transcription: 'tʃʰa̰']

- မြဲ (U+1019 U+103C U+1032)
    - [transcription: 'mja̰e']
- ရှာ (U+101B U+103E U+102C)

    - [transcription: 'ʃa']

Burmese Medial Sequences

When more than one Medial are following a consonant, the medials need to be arranged in definite order. The leading consonant acts as a nucleus in the middle and the medials can be on its left, right and below of the consonant. If the Medials happened to be in arbitrary order, they could be overlapped or become ill-formed graphemes. To prevent unexpected renderings, the medial sequences are listed in Table-8 and used with WLE rules.

1. ◌ျ   U+103B U+103D   MM1 in Table-8   See Appendix B

2.	◌ျ	U+103B U+103E	C_MM2 in Table-8	See Appendix B
3.	◌ြ	U+103C U+103D	MM3 in Table-8	See Appendix B
4.	◌ြ	U+103C U+103E	C_MM4 in Table-8	See Appendix B
5.	◌	U+103D U+103E	MM5 in Table-8	See Appendix B
6.	◌ြ	U+103C U+103D U+103E	CMMM in Table-8	See Appendix B

Medial sets 1, 3 and 5 can follow the consonant groups CMM1, CMM3, CMM5 (see Table-7) respectively. (see WLE rule #3)

No.2: This medial set only occurs with three nasal consonants, thus, three sequences for this combination C_MM2. (see Table-8)

No.4: This medial set only occurs with three nasal consonants, thus, three sequences for this combination C_MM4. (see Table-8)

No.6: There is only one consonant which can be combined with these three Medials. Thus, one sequence for this Consonant-Medial-Medial-Medial combination is listed. (see Table-8)

For example:

- ကြ (U+1000 U+103C U+103D)
    - [transcription: 'tjwa̰']

- မျ (U+1019 U+103B U+103E)
    - [transcription: 'm̥ja̰']

- လ (U+101C U+103D U+103E)

    - [transcription: 'l̥wa̰']

3.3.4.2.   Shan Medials

Shan Medials

There are three Shan Medials as listed in Table 5. Shan Medials must follow Shan consonants. Only Shan vowels or Shan vowel sequences or Shan other signs can follow Shan Medials . The same medial cannot be repeated and it cannot be followed by Asat U+103A.

3.3.4.3.　　Sgaw Karen Medials

There are five Sgaw Karen Medials as listed in Table 5. S'gaw Karen Medials must follow S'gaw Karen consonants. The same medial cannot be repeated and it cannot be followed by Asat U+103A.

3.3.4.4.　　Mon Medials

There are seven Mon Medials as listed in Table 5. Mon Medials must follow Mon consonants. The same medial cannot be repeated and it cannot be followed by Asat U+103A.

3.3.4.5.　　Pa'O Karen Medials

There are three Pa'O Karen Medials as listed in Table 5. Pa'O Karen Medials must follow Pa'O Karen consonants. The same medial cannot be repeated and it cannot be followed by Asat U+103A.

3.3.5. Diacritic - Tone Marks and Other Signs

Table 6: List of Tone Marks and Other Signs
(Legend: T- Tone, V- Vowel, O- Others)

Code Point	Glyph	Unicode Name	Myanmar	Shan	S'gaw Karen	Mon	Pa'O
U+102C	ာ	MYANMAR VOWEL SIGN AA	V		T	V	V
U+1035	�e	MYANMAR VOWEL SIGN E ABOVE		O			
U+1036	ံ	MYANMAR SIGN ANUSVARA	O		V	V	O
U+1037	့	MYANMAR SIGN DOT BELOW	T		V		O
U+1038	း	MYANMAR SIGN VISARGA	T	T	T	V	T
U+1039	္	MYANMAR SIGN VIRAMA	O			O	
U+103A	်	MYANMAR SIGN ASAT	O	O	O	O	O
U+1062	ၢ	MYANMAR VOWEL SIGN SGAW KAREN EU		O	V/T		
U+1063	ၣ	MYANMAR TONE MARK SGAW KAREN HATHI			T*		

U+1064	�ၤ	MYANMAR TONE MARK SGAW KAREN KE PHO			T		
U+1085	၅	MYANMAR VOWEL SIGN SHAN E ABOVE		O			
U+1087	၇	MYANMAR SIGN SHAN TONE-2		T			
U+1088	၈	MYANMAR SIGN SHAN TONE-3		T			
U+1089	၉	MYANMAR SIGN SHAN TONE-5		T			
U+108A	၊	MYANMAR SIGN SHAN TONE-6		T			
U+108F	၏	MYANMAR SIGN RUMAI PALAUNG TONE-5					T
U+AA7B	ꩻ	MYANMAR SIGN PAO KAREN TONE					T

Myanmar Vowel Sign AA "ာ" U+102C is a dependent vowel for Burmese, Mon and Pa'O. In Sgaw Karen, it is used as Tone sequence "ာ်" (U+102C U+103A) (see 3.3.3).

Myanmar Vowel Sign Sgaw Karen "ၢ" U+1062 is a dependent vowel, as well, used as a tone sequence " ၢ် " (U+1062 U+103A) (see 3.3.3.3 S'gaw Karen Dependent Vowels).

Myanmar Tone Mark Sgaw Karen Hathi "ၣ" U+1063 is used only in the following tone sequence: "ၣ်" (U+1063 U+103A)(See 3.3.5.3).

3.3.5.1.   Burmese Tone Marks and Other Signs

Burmese Tone Marks

Tone marks complement the long vowel. The Short-tone sign (U+1037 MYANMAR SIGN DOT BELOW) and Long-tone sign (U+1038 MYANMAR SIGN VISARGA) are used to enhance the sound of Long Vowels.

The Short tone sign " ့ " (U1037 MYANMAR SIGN DOT BELOW) must follow a Long vowel or the Anusvara "ံ" (U+1036), to add creakiness to a syllable.

When U+1036 is in the sequence [Short tone+Asat] (see S16 of Table-8), it can follow the consonants: င (U+1004), ဉ (U+1009), ည (U+100A), ဏ (U+100F), န (U+1014), မ (U+1019), ယ (U+101A). These consonants are marked as C2 in Table-7.

Example:

- ေတ္ြ့ (U+1010 U+103D U+1031 U+1037)
    - [transcription: 'twɛɪ']
- မြို့ (U+1019 U+103C U+102D U+102F U+1037)
    - [transcription: 'mjo̰']
- ပျံ့ (U+1015 U+103B U+1036 U+1037)
    - [transcription: 'pja̰ɴ']

Long tone sign "◌း" (U1038 MYANMAR SIGN VISARGA) appears at the end of the syllable.

Example:

- ေတ္ြး (U+1010 U+103D U+1031 U+1038)
    - [transcription: 'twɛɪ:']
- မျိုး (U+1019 U+103B U+102D U+102F U+1038)
    - [transcription: 'mjo:']
- လုံး (U+101C U+102F U+1036 U+1038)
    - [transcription: 'loʊɴ:']

Burmese Other Signs

There are four other signs and sequence in Burmese:

1. U+1036 ◌ံ MYANMAR SIGN ANUSVARA
2. U+1039 ◌္ MYANMAR SIGN VIRAMA
3. U+103A ◌် MYANMAR SIGN ASAT
4. U+1004 U+103A U+1039 ◌ၤ KINZI

MYANMAR SIGN ANUSVARA - U1036
Linguistically, Anusvara is considered a shortened form of Consonant Ma and Asat. However, it sometimes has other properties, such as joining with other vowel signs to create Diphthong vowels. The Short tone sign "◌့" (U+1037) can follow Anusvara.

Example words: သံ , တိံ , ကြိုံ

သံ - U+101E **U+1036** [transcription: /θaɴ/]

This is a homophone of other spellings "သန်" and "သမ်" . Different spellings are used for different meanings. This discussed spelling "သံ" means 'iron', and as a suffix , it means 'voice'.

တိံ - U+1010 U+102D **U+1036** [transcription: /teɪɴ/]

The sanskrit-derived word means "of the sky" or "something from the sky". Pronounced in similar way as "တိမ်". Although "တိံ" and "တိမ်" both come from the same origin, the word

"တိမ်" is used as specific meaning "cloud". The historical spelling "တိ" is still well known and used in names.

> ကြို - U+1000 U+103C U+102F **U+1036** [transcription: /tʃoʊɴ/]

There are homophone spellings "ကြုန်" or "ကြုမ်" , however, only the spelling "ကြို" has meaning "meet", the rest are just syllabic pronunciations. These may be used to spell foreign names.

MYANMAR SIGN VIRAMA - U1039

Virama has two properties, as killer (devoweliser) and joiner of syllable chaining. This virama brings the consonant after it to be rendered below the consonant before.

Pattern of syllable chaining : Consonant + Virama + Consonant (Covered in WLE rules)

Example words: ကဏ္ဍ , ဣတ္ထီ , လိမ္မော်

ကဏ္ဍ - There are two syllables in the word, ကဏ် and ဍ.

(The word as a whole means "category")

- - [transcription: 'gaɴ-da̰']

    Syllabic spelling : U+1000 U+100F U+**103A** U+100D

    Standard spelling : U+1000 U+100F U+**1039** U+100D

ဣတ္ထီ - There are two syllables in the word, ဣတ် and ထီ.

(The word as a whole means "woman")

- - [transcription: 'eḭʔ-thḭ']

    Syllabic spelling: U+1023 U+1010 U+**103A** U+1011 U+102D

    Standard spelling: U+1023 U+1010 U+**1039** U+1011 U+102D

လိမ္မော် - There are two syllables in the word, လိမ် and မော်.

(The word as a whole means "orange".)

- - [transcription: 'leɪɴ-mɔ']

    Syllabic spelling: U+101C U+102D U+1019 **U+103A** U+1019 U+1031 U+102C U+103A

    Standard spelling: U+101C U+102D U+1019 **U+1039** U+1019 U+1031 U+102C + 103A

The words which are usually written with Virama (U+1039) will look strange if they are spelled with Asat (U+103A) . These two are not interchangeable with each other although they may have the same 'killer' property.

MYANMAR SIGN ASAT (KILLER) - U103A

This sign is used to remove the consonant sound of a letter and take only the vowel property to create more vowel sounds out of consonants.

Example words: ဝန် , လျှပ် , စွမ်း , မှုန့်

ဝန် - Consonant Wa , Consonant Na, **Asat**

- [transcription: 'wuɴ']

Although there are two consonants, there is only one syllable [W ʊn] . When the second consonant combines with Asat, it becomes a vowel to compliment the first consonant.

လျှပ် - Consonant La , Medial Ya, Medial Ha , Consonant Pa, **Asat**

- [transcription: 'l̥ja̰ʔ']

Consonant Asat combination can also complement Consonants with Dependant signs (either medials or vowels) . It is not limited to Consonants only.

စွမ်း - Consonant Ca , Medial Wa, Consonant Ma, **Asat**, Visarga

- [transcription: 'swaɴˑ']

Tone sign Visarga can follow Asat sign to enhance the tone to the highest.

မှုန့် - Consonant Ma, Medial Ha, Vowel U , Consonant Na, Dot Below, **Asat**

- [transcription: 'm̥o̰ʊɴ']

Combination of Dot Below (Short-Tone) and Asat can make the vowel to have Creaky(short) Tone.

The following is the graphical explanation of how Asat is considered. When there is a Consonant being followed by Asat, it must follow the preceding Consonant or Consonant+Medial or Consonant+Dependent Vowel or Consonant+Medial+Dependent Vowel to form one complete syllable. In this graphic example, the leading combination is Consonant+Medial.
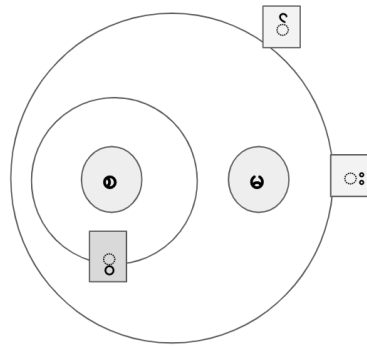
/ စ / ◌ွ / မ / ◌် / ◌း /

/ U+1005 / U+103D / U+1019 / U+103A / U+1038 /

Figure 3: Composing of a one-syllable word "စွမ်း" which has Consonant Asat combination

Another way of using Asat is that it can be tone modifier of the vowel sequences "ေ◌ါ" and "ေ◌ာ" , which is mentioned in 3.3.2 .

Example word: ေဝ္ှာ်

ေဝ္ှာ် - U+101D U+103E U+1031 U+102C U+**103A**

- [transcription: 'l̥ɔ']
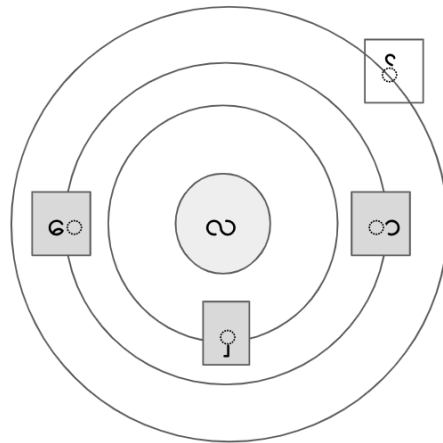
/ လ / ◌ှ / ေ◌ / ◌ာ / ◌် /

/ U+101C / U+103E / U+1031 / U+102C / U+103A /

Figure 4: Composing of syllable with U+103A as vowel modifier

24

Except the defined sequences in Table 8 (S16, S17, S19, S20, ST1, ST2, ST3, S_Sh2, S_Sh5), Asat cannot follow other Diacritics - Medial, Medial sequence, Vowel, vowel sequence, Tone, Tone sequence or Other Signs.

KINZI - U+1004 U+103A U+1039

This special rendering is used as Orthographical variant form of Consonant Nga and Asat (င်္ [u1004+u103A]). Kinzi was believed to be invented during the stone-script era to save more space. Kinzi is used in Sanskrit-derived vocabulary, historical words and people/location name spellings.

Example words: မင်္လာ , စင်္ပူ , သိင်္တ္တရ

မင်္လာ

  - [transcription: mıŋ-gə-la]
  - Sanskrit derived word of "blessing" , where the syllables are read as မင်-ဂ-လာ . This format is standard spelling and widely used.

စင်္ပူ

  - [transcription: sıŋ-ka-pu]

Location name spelling of "Singapore", where the syllables are read as စင်-ကာ-ပူ . This is standard spelling.

သိင်္တ္တရ

  - [transcription: θeɪɴ-goʊ-tə-ɹa̰]

Location name spelling of a Pagoda in Yangon, which is believed to be in existence since the 6th and 10th centuries AD. Although it is read as "သိင်-ဂုတ်-တ-ရ", the historical spelling "သိင်္တ္တရ" is reserved as standard spelling and used until today.

3.3.5.2.   Shan Tone Marks and Other Signs

Shan Tone Marks

There are five Shan Tone Marks. They are diacritics which come as the last character of a syllable.

Shan Other Signs

There are four Other Signs in Shan

1. U+1035   ိ   MYANMAR VOWEL SIGN E ABOVE
2. U+103A   ်   MYANMAR SIGN ASAT
3. U+1062   ၢ   MYANMAR VOWEL SIGN SGAW KAREN EU
4. U+1085   ႅ   MYANMAR VOWEL SIGN SHAN E ABOVE

These diacritics are used to spell closed syllables, which means the sequence must be as follow:

Shan Consonant, (U+1062 or U+1035 or U+1085), Shan Consonant, Asat (U+103A).

Diacritic U+1062 also can appear as vowel sequence with U+1086 (See 3.3.3.2)

3.3.5.3.   Sgaw Karen Tone Marks and Other Signs

Sgaw Karen Tone Marks

Due to remarks in Table 6 (in section 3.3.5) all code points are listed here for clarification.

1. U+1038   း   MYANMAR SIGN VISARGA
2. U+1064   ၤ   MYANMAR TONE MARK SGAW KAREN KE PHO

In Sgaw Karen writing, Tone indication can be either by one character or a sequence of characters. A Tone sign can appear as the last character of a syllable.

Sgaw Karen Tone Sequences

1. ၢ်   U+1062 U+103A   Defined as ST1 in Table 8
2. ာ်   U+102C U+103A   Defined as ST2 in Table 8
3. ၣ်   U+1063 U+103A   Defined as ST3 in Table 8

These are Sgaw Karen tone sequences which can appear at the end of a syllable.

Sgaw Karen Other Signs

1. U+103A   ်   MYANMAR SIGN ASAT

This sign is used in the pairing of tone sequences. It can also be used as devowelizer when following a consonant.

3.3.5.4.   Mon Tone Marks and Other Signs

Tone Marks are not used in Mon language.

Mon Other Signs

There are three other signs used in Mon.

1. U+1039   ္   MYANMAR SIGN VIRAMA
2. U+103A   ်   MYANMAR SIGN ASAT

3. U+105A U+103A U+1039   ၚ်   MON KINZI

The Mon Kinzi is encoded using Mon Consonant Nga (U+105A) which is the equivalent of Myanmar Consonant Nga (U+1004). The usage of these diacritics - *Virama* (U+1039) , *Asat* (*Killer* U+103A) and *Kinzi* are the same as in Burmese.

3.3.5.5.   Pa'O Karen Tone Marks and Other Signs

Pa'O Karen Tone Marks

There are three Pa'O Karen tone marks as listed in Table 6. In Pa'O Karen writing, Tone marks are the diacritics which come as the last character of a syllable.

The code points 108F and AA7B are exclusively tone marks for Pa'O Karen. These cannot follow other tone marks as it can confuse other combinations with dots, therefore, these singleton 108F or AA7B cannot follow 1037 or 1038.

Pa'O Karen Tone Sequences

- ◌ၟ   U+1037 U+AA7B   Defined as ST4 in Table 8

This is Pa'O Karen tone sequence which can appear at the end of a syllable.

Pa'O Karen Other Signs

There are three Pa'O Karen other signs as listed in Table 6.These diacritics are used to devowelize the consonants.

4.   Overall Development Process and Methodology

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

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

The panel also received 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 in Yangon where all the panel members as well as other participants from the private sector, public sector especially from Posts and Telecommunications Department under Ministry of Transport and Communications, ISPs, Universities and relevant civil societies were invited in order to elicit the feedback and receive recommendations on the proposal before it was finalized.

Considerations based on history of Myanmar script encoding

The consideration of variants and the WLE rules are based on Unicode compliant fonts such as Padauk font family, Pyidaungsu font and Noto San Myanmar font. Non-unicode compliant fonts and ASCII fonts are not considered. The nature of Myanmar script requires different shapes for some of the characters. For example, Consonant_Na "န" would have to shorten the leg when it is paired with Vowel_U "ု" and the combination becomes "နု". Before unicode font rendering was supported, there used to be multiple codepoints for some characters, for example, two different code points were used for Consonant_Na with normal leg and shortened leg. There were many multiple-code-point issues for certain Consonants which confuse non-unicode font users and developers. Unicode has solved these issues and simplified the requirements for Myanmar script.

However, there are other fonts which are made to imply with unicode standards and backward compatible in both ways. The special font called 'Zawdecode' is used on 50% of mobile devices in Myanmar. This special font decides which is the most likely form of grapheme and displays accordingly.

Using the backward compatible font can cause displaying the text in unexpected form.

Example: "ကြို" U+1020 U+103C U+101D

The medial U+103C in the middle is supposed to render on the leading consonant U+1020, however, using Zawdecode font, U+103C could be displayed attached to the following consonant instead of the leading one in some devices.

The form which is supposed to be displayed: ကြို

The form displayed on device: ကြ

Moreover, The different sequence of code points could be rendered into the same grapheme.

Example: "ျ" U+1005 U+103C (This combination is unlikely) is rendered to the same grapheme as "ွ" U+1005 U+103D .

Code points	On devices with Unicode-compliant fonts	On devices with backward compatible font
U+1005 U+103C (ၐ)	ၐ	ၐ
U+1005 U+103D (ၵ)	ၵ	ၵ

Comparison of Displaying text on different devices or fonts
Generated using the tool at https://mmunicode.org/converter/zawd.html.

To prevent unexpected rendering issues, the GP has chosen to allow only the expected combinations of character combinations to be allowed. (See more in session 3.3.4.1, and Appendix B)

5.   Repertoire

5.1.   Myanmar section of Maximal Starting Repertoire [MSR] Version 5

The Myanmar Unicode chart is given below and has mentioned the characters included and excluded in the Myanmar script root zone based on MSR-5.

Figure 5: Myanmar Table in MSR-5

Color convention[1]:

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

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

White - Characters that are disallowed by IDNA2008

5.2. Code Point Repertoire

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

Table 7: Code point repertoire

#	Unicode Code Point	Glyph	Character Name	Category	Language(s), with EGIDS Value	Refer-e nce
1	U+1000	က	MYANMAR LETTER KA	Consonant, CMM1, CMM3, C_mon, C_103B, C_103C	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
2	U+1001	ခ	MYANMAR LETTER KHA	Consonant, CMM1, CMM3, C_mon, C_103B, C_103C	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
3	U+1002	ဂ	MYANMAR LETTER GA	Consonant, CMM1, CMM3, C_mon, C_103B, C_103C	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
4	U+1003	ဃ	MYANMAR LETTER GHA	Consonant, C_mon, C_103B, C_103C	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
5	U+1004	င	MYANMAR LETTER NGA	Consonant, C2, C3, CMM3, CMM5, C_shan, C_103B, C_103C, C_103E	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Pa'O Karen	[101], [105], [108], [110]
6	U+1005	စ	MYANMAR LETTER CA	Consonant, C_mon, C_103E	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
7	U+1006	ဆ	MYANMAR LETTER CHA	Consonant, C_mon, C_103B, C_103C	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
8	U+1007	ဇ	MYANMAR LETTER JA	Consonant, C_mon,	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103],

					[108], [110]
9	U+1008	ဈ	MYANMAR LETTER JHA	Consonant, C_n103D	1 Burmese, 5 Pa'O Karen	[101], [108]
10	U+1009	ဉ	MYANMAR LETTER NYA	Consonant, C2, C3, CMM5, C_mon, C_103E	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103], [108]
11	U+100A	ည	MYANMAR LETTER NNYA	Consonant, C2, C3, CMM5, C_mon, C_103E	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
12	U+100B	ဋ	MYANMAR LETTER TTA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103], [108]
13	U+100C	ဌ	MYANMAR LETTER TTHA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103], [108]
14	U+100D	ဍ	MYANMAR LETTER DDA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103], [108]
15	U+100E	ဎ	MYANMAR LETTER DDHA	Consonant	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103], [108]
16	U+100F	ဏ	MYANMAR LETTER NNA	Consonant, C2, C3, CMM5, C_mon, C_103B, C_103C, C_103E	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103], [108]
17	U+1010	တ	MYANMAR LETTER TA	Consonant, CMM1, CMM3, C_mon, C_shan, C_103B, C_103C	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]
18	U+1011	ထ	MYANMAR LETTER THA	Consonant, CMM1, CMM3, C_mon, C_shan, C_103B, C_103C	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]

19	U+1012	ဒ	MYANMAR LETTER DA	Consonant, CMM1, CMM3, C_mon, C_103B, C_103C	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
20	U+1013	ဓ	MYANMAR LETTER DHA	Consonant, CMM1, CMM3, C_mon, C_103B, C_103C	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103], [108]
21	U+1014	န	MYANMAR LETTER NA	Consonant, C2, C3, CMM1, CMM3, CMM5, C_mon, C_103B, C_103C, C_103E	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
22	U+1015	ပ	MYANMAR LETTER PA	Consonant, CMM1, CMM3, C_mon, C_shan, C_103B, C_103C	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]
23	U+1016	ဖ	MYANMAR LETTER PHA	Consonant, CMM1, CMM3, C_mon, C_103B, C_103C	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
24	U+1017	ဗ	MYANMAR LETTER BA	Consonant, CMM1, CMM3, C_mon, C_103B, C_103C	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103], [108]
25	U+1018	ဘ	MYANMAR LETTER BHA	Consonant, CMM1, CMM3, C_mon, C_103B, C_103C	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
26	U+1019	မ	MYANMAR LETTER MA	Consonant, C2, C3, CMM1, CMM3, CMM5, C_mon, C_shan, C_103B, C_103C, C_103E	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]
27	U+101A	ယ	MYANMAR LETTER YA	Consonant, C2 , CMM1, CMM5, C_mon, C_shan, C_103B, C_103E	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]

28	U+101B	ရ	MYANMAR LETTER RA	Consonant, CMM5, C_mon, C_shan, C_103E	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]
29	U+101C	လ	MYANMAR LETTER LA	Consonant, CMM1, CMM5, C_mon, C_shan, C_103B, C_103E	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]
30	U+101D	ဝ	MYANMAR LETTER WA	Consonant, C3, C_mon, C_shan, C_103B, C_103E	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]
31	U+101E	သ	MYANMAR LETTER SA	Consonant, C_mon, C_shan, C_103E	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]
32	U+101F	ဟ	MYANMAR LETTER HA	Consonant, CMM1, CMM3, C_mon, C_103B, C_103C, C_103E	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
33	U+1020	ဠ	MYANMAR LETTER LLA	Consonant, C_mon, C_103E	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103], [108]
34	U+1021	အ	MYANMAR LETTER A	Consonant, C_mon	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
35	U+1022	ဢ	MYANMAR LETTER SHAN A	Consonant, CMM1, C_shan, C_103C	3 Shan	[105]
36	U+1023	ဣ	MYANMAR LETTER I	Independent Vowel	1 Burmese, 5 Mon	[101], [103]
37	U+1024	ဤ	MYANMAR LETTER II	Independent Vowel	1 Burmese	[101]

38	U+1025	ဥ	MYANMAR LETTER U	Independent Vowel	1 Burmese, 5 Mon	[101], [103]
39	U+1026	ဦ	MYANMAR LETTER UU	Independent Vowel	1 Burmese	[101]
40	U+1027	ဧ	MYANMAR LETTER E	Independent Vowel, Consonant	1 Burmese, 3 S'gaw Karen	[101], [110]
41	U+1028	ဨ	MYANMAR LETTER MON E	Independent Vowel	5 Mon	[103]
42	U+1029	ဩ	MYANMAR LETTER O	Independent Vowel	1 Burmese, 5 Mon	[101], [103]
43	U+102A	ဪ	MYANMAR LETTER AU	Independent Vowel	1 Burmese, 5 Mon	[101], [103]
44	U+102B	ါ	MYANMAR VOWEL SIGN TALL AA	Dependent Vowel, lv	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
45	U+102C	ာ	MYANMAR VOWEL SIGN AA	Dependent Vowel, lv	1 Burmese, 5 Mon, 5 Pa'O Karen	[101], [103], [108]
46	U+102D	ိ	MYANMAR VOWEL SIGN I	Dependent Vowel, sh_vowel	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]
47	U+102E	ီ	MYANMAR VOWEL SIGN II	Dependent Vowel, lv, sh_vowel	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Pa'O Karen	[101], [105], [108], [110]
48	U+102F	ု	MYANMAR VOWEL SIGN U	Dependent Vowel, sh_vowel	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]

49	U+1030	◌ု	MYANMAR VOWEL SIGN UU	Dependent Vowel, lv, sh_vowel	1 Burmese, 3 S'gaw Karen, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]
50	U+1031	ေ◌	MYANMAR VOWEL SIGN E	Dependent Vowel, lv, sh_vowel	1 Burmese, 3 Shan, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108]
51	U+1032	◌ဲ	MYANMAR VOWEL SIGN AI	Dependent Vowel, lv	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
52	U+1033	◌ဳ	MYANMAR VOWEL SIGN MON II	Dependent Vowel	5 Mon	[103]
53	U+1034	◌ဴ	MYANMAR VOWEL SIGN MON O	Dependent Vowel	5 Mon	[103]
54	U+1035	◌ဵ	MYANMAR VOWEL SIGN E ABOVE	Other Various Sign	3 Shan	[105]
55	U+1036	◌ံ	MYANMAR SIGN ANUSVARA	Dependent Vowel, Anusvara, lv	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
56	U+1037	◌့	MYANMAR SIGN DOT BELOW	T_Short, Dependent Vowel	1 Burmese, 3 S'gaw Karen, 5 Pa'O Karen	[101], [108], [110]
57	U+1038	◌း	MYANMAR SIGN VISARGA	T_Long, Dependent Vowel	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	101], [103], [108], [110]
58	U+1039	◌္	MYANMAR SIGN VIRAMA	Virama	1 Burmese, 5 Mon	[101], [103]

59	U+103A	်	MYANMAR SIGN ASAT	Killer	1 Burmese, 3 Shan, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [105], [108], [110]
60	U+103B	ျ	MYANMAR CONSONANT SIGN MEDIAL YA	Medial, C_n103D	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
61	U+103C	ြ	MYANMAR CONSONANT SIGN MEDIAL RA	Medial, C_n103D	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
62	U+103D	ွ	MYANMAR CONSONANT SIGN MEDIAL WA	Medial	1 Burmese, 3 S'gaw Karen, 5 Mon, 5 Pa'O Karen	[101], [103], [108], [110]
63	U+103E	ှ	MYANMAR CONSONANT SIGN MEDIAL HA	Medial	1 Burmese, 3 S'gaw Karen, 5 Mon	[101], [103], [110]
64	U+103F	ဿ	MYANMAR LETTER GREAT SA	Consonant, C1	1 Burmese , 5 Pa'O Karen	[101], [108]
65	U+105A	ၚ	MYANMAR LETTER MON NGA	Consonant, C_mon, C_n103D	5 Mon	[103]
66	U+105B	ၛ	MYANMAR LETTER MON JHA	Consonant, C_mon, C_n103D	5 Mon	[103]
67	U+105C	ၜ	MYANMAR LETTER MON BBA	Consonant, C_mon, C_103B, C_103C, C_103E	5 Mon	[103]

68	U+105D	ဍ	MYANMAR LETTER MON BBE	Consonant, C_mon, C_n103D	5 Mon	[103]
69	U+105E	ဍ	MYANMAR CONSONANT SIGN MON MEDIAL NA	Medial, M_mon	5 Mon	[103]
70	U+105F	ဍ	MYANMAR CONSONANT SIGN MON MEDIAL MA	Medial, M_mon	5 Mon	[103]
71	U+1060	ဍ	MYANMAR CONSONANT SIGN MON MEDIAL LA	Medial, M_mon	3 S'gaw Karen, 5 Mon	[103], [110]
72	U+1061	ရ	MYANMAR LETTER SGAW KAREN SHA	Consonant	3 S'gaw Karen	[110]
73	U+1062	ဍ	MYANMAR VOWEL SIGN SGAW KAREN EU	Dependent Vowel, Other Various Sign	3 S'gaw Karen, 3 Shan	[105], [110]
74	U+1064	ဍ	MYANMAR TONE MARK SGAW KAREN KE PHO	Sgaw_Tone	3 S'gaw Karen	[110]
75	U+1075	ဂ	MYANMAR LETTER SHAN KA	Consonant, C3, CMM1, C_shan, C_103B, C_103C	3 Shan	[105]
76	U+1076	ခ	MYANMAR LETTER SHAN KHA	Consonant, CMM1, C_shan, C_103B, C_103C	3 Shan	[105]

77	U+1077	ဟ	MYANMAR LETTER SHAN GA	Consonant, CMM1, C_shan, C_103B, C_103C	3 Shan	[105]
78	U+1078	ၸ	MYANMAR LETTER SHAN CA	Consonant, CMM1, C_shan, C_103B, C_103C	3 Shan	[105]
79	U+1079	ၹ	MYANMAR LETTER SHAN ZA	Consonant, CMM1, C_shan, C_103B, C_103C	3 Shan	[105]
80	U+107A	ၺ	MYANMAR LETTER SHAN NYA	Consonant, C3, C_shan	3 Shan	[105]
81	U+107B	ၻ	MYANMAR LETTER SHAN DA	Consonant, CMM1, C_shan, C_103B, C_103C	3 Shan	[105]
82	U+107C	ၼ	MYANMAR LETTER SHAN NA	Consonant, C3, CMM1, C_shan, C_103B, C_103C	3 Shan	[105]
83	U+107D	ၽ	MYANMAR LETTER SHAN PHA	Consonant, CMM1, C_shan, C_103B, C_103C	3 Shan	[105]
84	U+107E	ၾ	MYANMAR LETTER SHAN FA	Consonant, C_shan, C_n103D	3 Shan	[105]
85	U+107F	ၿ	MYANMAR LETTER SHAN BA	Consonant, CMM1, C_shan, C_103C	3 Shan	[105]
86	U+1080	ႀ	MYANMAR LETTER SHAN THA	Consonant, CMM1, C_shan, C_103C	3 Shan	[105]
87	U+1081	ႁ	MYANMAR LETTER SHAN HA	Consonant, C_shan	3 Shan	[105]

88	U+1082	◌	MYANMAR CONSONANT SIGN SHAN MEDIAL WA	Medial, M_shan	3 Shan	[105]
89	U+1083	◌	MYANMAR VOWEL SIGN SHAN AA	Dependent Vowel, sh_vowel	3 Shan	[105]
90	U+1084	◌	MYANMAR VOWEL SIGN SHAN E	Dependent Vowel, sh_vowel	3 Shan	[105]
91	U+1085	◌	MYANMAR VOWEL SIGN SHAN E ABOVE	Other Various Sign	3 Shan	[105]
92	U+1086	◌	MYANMAR VOWEL SIGN SHAN FINAL Y	Dependent Vowel, sh_vowel	3 Shan	[105]
93	U+1087	◌	MYANMAR SIGN SHAN TONE-2	Sh_Tone	3 Shan	[105]
94	U+1088	◌	MYANMAR SIGN SHAN TONE-3	Sh_Tone	3 Shan	[105]
95	U+1089	◌	MYANMAR SIGN SHAN TONE-5	Sh_Tone	3 Shan	[105]
96	U+108A	◌	MYANMAR SIGN SHAN TONE-6	Sh_Tone	3 Shan	[105]
97	U+108F	◌	MYANMAR SIGN RUMAI PALAUNG TONE-5	Pao_Tone	5 Pa'O Karen	[108]

98	U+AA7B	◌ၞ	MYANMAR SIGN PAO KAREN TONE Pao_Tone	5 Pa'O Karen	[108]

<p style="text-align:center">Table 7: Code point repertoire</p>

5.3. Code Point Sequences

The following code point sequences are also included. Table 8 lists the sequences arranged by the Unicode code point value.

Myanmar script consonants can have up to six dependent signs. The signs will add up to one nucleus character, thus the correct order of the dependent sign is required.

<p style="text-align:center">Table 8: Code point sequences by Unicode codepoint value</p>

Seq. #	Unicode Code Point	Glyph	Name	Note
1	U+1000 U+1039 U+1000	က္က	SV1	WLE rule #7. See 6.1 for variant requirements.
2	U+1000 U+103B U+103D U+1014 U+103A U+102F U+1015 U+103A	ကျွန်ုပ်	SS06	Exception to WLE rule #1
3	U+1004 U+103A	င်	-	See Table 10
4	U+1004 U+103A U+1039	င်္	S11	WLE rule #6
5	U+1004 U+103C U+103E	ၚ	C_MM4	See Appendix B
6	U+1014 U+103B U+103E	နျ	C_MM2	See Appendix B
7	U+1014 U+103C U+103E	ၞ	C_MM4	See Appendix B
8	U+1015 U+102C U+103A	ပာ်	-	See Table 10
9	U+1015 U+1039 U+1015 U+102C	ပ္ပာ	SV2	Exception of WLE rule #14. WLE rule #7.
10	U+1019 U+103B U+103E	မျ	C_MM2	See Appendix B
11	U+1019 U+103C U+103D U+103E	ၿ	CMMM	See Appendix B
12	U+1019 U+103C U+103E	ၿ	C_MM4	See Appendix B

13	U+101A U+1031 U+102C U+1000 U+103A U+103B U+102C U+1038	ယောက်ျား	SS07	Exception to WLE rule #3 Asat U+103A is followed by Medial U+103B
14	U+101A U+103B U+103E	ယ္ဠ	C_MM2	See Appendix B
15	U+101B U+103E	ရှ	SV3	Section 6.1
16	U+101C U+1000 U+103A U+103B U+102C	လက်ျာ	SS08	Exception to WLE rule #3 Asat U+103A is followed by Medial U+103B
17	U+101C U+103B U+103E	လ္ဠ	C_MM2	See Appendix B
18	U+101E U+103B U+103E	သ္ဠ	C_MM2	See Appendix B
19	U+101E U+103C U+1014 U+103A	သြန်	SS10	[103] The only Mon word with U+101E U+103C
20	U+101E U+103C U+1083 U+1087	သြၴ	SS09	[105] The only Shan word with U+101E U+103C
21	U+101F U+103A	ဟ်	-	See Table 10
22	U+1023 U+1033	ဣ	S_Mon1	This is Independent vowel sequence for Mon Language (See 3.3.2.2)
23	U+1025 U+102F	ဥ	S_Mon2	This is Independent vowel sequence for Mon Language (See 3.3.2.2)
24	U+1025 U+102F U+1036	ဥံ	SS01	1025 cannot join with other signs
25	U+1025 U+102F U+1038	ဥး	SS02	1025 cannot join with other signs
26	U+1026 U+1038	ဦး	SS03	1026 can only join with 1038, not with other signs

27	U+102C U+103A	ာ်	ST2	This is used as Tone Sequence in Sgaw Karen language. (See 3.3.5.3)
28	U+102D U+102F	ို	S12	Long vowel sequence, For WLE rule #11
29	U+102D U+1030	ိူ	S_Sh3	Dependent Vowel Sequence for Shan Language. (See 3.3.3.2)
30	U+102D U+1036	ိံ	S18	Long vowel sequence, For WLE rule #1
31	U+102F U+1032	ုဲ	S_Pao	Dependent Vowel Sequence for Pa'O Language. This must follow c_pao or m_pao (See 3.3.3.5)
32	U+102F U+1036	ုံ	S13	long vowel sequence, For WLE rule #1
33	U+1031 U+102B	ေါ	S14	long vowel sequence, For WLE rule #1, #11
34	U+1031 U+102B U+1037 U+103A	ေါ့်	S21	S14 + S16
35	U+1031 U+102B U+103A	ေါ်	S19	Long vowel sequence
36	U+1031 U+102C	ော	S15	long vowel sequence, For WLE rule #1, #11
37	U+1031 U+102C U+1037 U+103A	ော့်	S22	S15 + S16
38	U+1031 U+102C U+103A	ော်	S20	Long vowel sequence
39	U+1031 U+1083	ေ	S_Sh1	Dependent Vowel Sequence for Shan Language.(See 3.3.3.2)
40	U+1031 U+1083 U+103A	ေ	S_Sh2	Dependent Vowel Sequence for Shan Language. (See 3.3.3.2)

41	U+1035 U+102F	ိ◌ု	S_Mon3	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)
42	U+1037 U+ 103A	◌့်	S16	Short Vowel sequence, For WLE rule #5
43	U+1037 U+AA7B	◌ႛ	ST4	This is used as Tone Sequence in Pa'O Karen language. (See 3.3.5.5)
44	U+103A U+1038	◌်း	S17	Long tone sequence, For WLE rule #5
45	U+103B U+103D	◌ျွ	MM1	See Appendix B
46	U+103C U+103D	◌ြွ	MM3	See Appendix B
47	U+103D U+103E	◌ွှ	MM5	See Appendix B
48	U+105A U+103A	◌ၚ်	-	See Table 10
49	U+105A U+103A U+1039	◌ၚ်္	S_Mon4	Mon Kinzi (See 3.3.5.4)
50	U+1062 U+103A	◌ၢ်	ST1	This is used as Tone Sequence in Sgaw Karen language. (See 3.3.5.3)
51	U+1062 U+1086	◌ၢႆ	S_Sh4	Dependent Vowel Sequence for Shan Language. (See 3.3.3.2)
52	U+1063 U+103A	◌ၣ်	ST3	This is used as Tone Sequence in the Sgaw Karen language. (See 3.3.5.3)
53	U+1082 U+103A	◌ႂ်	S_Sh5	Dependent Vowel Sequence for Shan Language. (See 3.3.3.2)
54	U+102B U+1036	◌ါံ	S_Mon5	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)
55	U+102C U+1036	◌ာံ	S_Mon6	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)

56	U+102D U+102F U+1032	ိုဲ	S_Mon7	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)
57	U+1031 U+102B U+1036	ေါံ	S_Mon8	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)
58	U+1031 U+102C U+1036	ောံ	S_Mon9	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)
59	U+1031 U+1032	ေဲ	S_Mon10	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)
60	U+1032 U+102B	ဲါ	S_Mon11	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)
61	U+1032 U+102C	ဲာ	S_Mon12	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)
62	U+103E U+103A	ှ်	S_Mon13	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)
63	U+103E U+103A U+1031	ေှ်	S_Mon14	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)
64	U+103E U+103A U+1031 U+102C	ေှ်ာ	S_Mon15	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)
65	U+1031 U+102C U+103E U+103A	ောှ်	S_Mon16	Dependent Vowel Sequence for Mon Language. (See 3.3.3.4)

The sequences SS07 "ယောက်ျား" ([transcription: /jɔʔ-tʃaː/], meaning "men"), and SS08 "လက်ျာ" ([transcription: /læʔ-ya/], meaning "right"), are exceptions to WLE rule #3, Medial U+103B cannot follow Asat U+103A. Both these sequences have the common part [U+1000 U+1038 U+103B U+102C] "က်ျာ", which is neither meaningful nor pronounceable. This combination only occurs in these two words (SS07 and SS08) in Burmese language.

If only the part [U+1000 U+1038 U+103B U+102C] "က်ျာ" is defined as one sequence, WLE rule for Tone Mark U+1038 to follow this particular sequence is to be considered. The sequence also becomes confusable with the possible combination [U+1000 U+103B U+102C U+103A]

"ကျစ်" [transcription: /tʃaɪt/], meaning "respect", which is already aligned with WLE rules. To prevent further complications, the only exact words to be affected are defined as two sequences, SS07 and SS08 separately.

5.4. Code Point Included via Sequence

Sgaw Karen tone mark U+1063 is excluded from repertoire since it will never be used alone. It will be used as sequence ST3 (Table 8).

Table 8a: Code point Included via Sequence

#	Unicode Code Point	Glyph	Character Name	Note
1	U+1063	◌ၣ	Myanmar Tone Mark Sgaw Karen Hathi	Never used alone. Only used together with U+103A as in Tone sequence: ST3 in Table 8

5.5. Code Points Not Included

Code points not used at all

Following three code points have not been included in the repertoire.

Table 9: Code points not included

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

6. Variants

6.1. In-script Variants

The Myanmar script has the following in-script variants.

Set #	Unicode Code Point	Glyph	Unicode Code Point	Glyph	Disposition	Detail
1	U+1023 MYANMAR LETTER I	အ	U+1000 U+1039 U+1000 SV1	အ	blocked	Homoglyph
2	U+101F U+103A	ဟ်	U+1015 U+102C U+103A (confusable with U+1015 ST1)	ဟ်	blocked	Homoglyph
3	U+1061	ျ	U+101B U+103E	ျ	blocked	Homoglyph
4	U+1004	င	U+105A	ၚ	allocatable with condition	Becomes homoglyph when diacritics added
5	U+1004 U+103A	င်	U+1004 U+103A U+1039	င်	allocatable with condition	Semantic, see 3.3.5.1 (Kinzi)
6	U+1008	ဈ	U+105B	ၛ	allocatable with condition	Semantic, Homoglyph
7	U+1001	ခ	U+1076	ၶ	allocatable with condition	Homoglyph (for most of the fonts)
8	U+102E	ီ	U+1033	ု	allocatable with condition	Semantic
9	U+102B	ါ	U+102C	ာ	allocatable with condition	Semantic
10	U+105A U+103A	ၚ်	U+105A U+103A U+1039	ၚ်	allocatable with condition	Semantic, see 3.3.5.4 (Kinzi)

Table 10: In script variants

In order to limit the number of allocatable variant labels, allocatable variant code points are grouped in two grapheme sets (set1 which is the first column of code points and set2 which is the second column of code points). The code points on the left side (set1) are either the simpler forms (shorter sequence) or the lower Unicode code point values.

No.4: Myanmar Letter Nga "င" (U+1004) and Myanmar Letter Mon Nga "ၚ" (U+105A) are shown side-by-side. They look identical when followed by Asat (U+103A). Both of them have the same semantic meaning and property and only one of them should be in a single given label.

47

U+1004 U+103A " င်" = **U+105A** U103A "ၚ်"

Kinzi, **U+1004** U+103A U+1039 " င်္" = **U+105A** U103A U+1039 "ၚ်္"

It is suggested that characters that have the same meaning, pronunciation and property should not appear both in the same string. Based on the language, only one type should be chosen and written consistently.

No.6: Myanmar Letter Jha "ဈ" and Myanmar Letter Mon Jha "ၛ". They are read as the same pronunciation [za]. Since the characters are closely related, visual similarity increases when vowels or other signs are added to form more meaningful words.

U+1008 U+1031 U+1038 "ဈေး" = U+105B U+1031 U+1038 "ၛေး"

Homophone [zeɪː] Meaning: 'market' in Burmese.

No.7: Myanmar Letter Kha "ခ" and Myanmar Letter Shan Kha "�х". They are read as the same pronunciation [kha]. Since the characters are closely related, visual similarity increases when vowels or other signs are added.

U+1001 U+102D U+102F "ခို" = U+1076 U+102D U+102F "ၶို"

Homophone [kho]. Meaning: 'pigeon' in Burmese.

No.8: Diacritics, which have the same property, and are used in different languages are shown side-by-side. One type of diacritic to be chosen for one string at a time.

For example, the word kiwi fruit [ki.wi] could be written as

"ကီဝီ" U+1000 **U+102E** U+101D **U+102E** or "ကဝ" U+1000 **U+1033** U+101D **U+1033**

No.9: Dependent vowel sign [-a] has two forms and these are interchangeable. Most standard spelling use "ာ" U+102C with a certain set of exception which uses the taller form "ါ" U+102B, however, the taller form can be used with all the consonants.

"လာပါ" U+101C U+102C U+1015 U+102B "လါပါ" U+101B U+102C U+1015 U+102B

These are homophones [la.pa] and synonyms, meaning 'come', simultaneously.

There are other semantically similar codepoints listed in the Appendix, Table A-1: In-script confusable code points. Having less visual similarity, they are not considered as actual grapheme-variants but listed for reference.

Some of these pairs cannot occur in the same labels. Please see Section 7, WLE rules.

In addition, to limit the number of allocatable variant labels, additional rules are included to allow only Grapheme-set-1 (left column in Table 10) or Grapheme-set-2 (right column in Table 10) in an allocatable variant labels.

Each of grapheme-set-1 code points has been given the reflexive variant type "r-set1" and each of grapheme-set-2 code points has been given the reflexive variant of type "r-set2". (By

convention, the prefix "r-" marks a type used in a reflexive variant mapping, that is, it represents an instance of the original code point at that location in a variant label, see Section 5.3.4 in [RFC 7940].)

The variant mapping from each of grapheme-set-1 code point to grapheme-set-2 code point is of type "set1-to-set2", while the variant type for the mapping from grapheme-set-2 code point to grapheme-set-1 code point is of type "set2-to-set1".

Special <action> elements defined for this LGR use these types to ensure the following restrictions.

- **blocked** — a variant label containing a blocked variant will receive a disposition of "blocked".

- **r-set1 r-set2** — a label containing one or more of these reflexive variant types and no others represents an original label and receives a disposition of "allocatable"
- **r-set1 set2-to-set1** — a label containing one or more of these variant types and no others receives a disposition of "allocatable"

- **r-set2 set1-to-set2** — a label containing one or more of these variant types and no others receives a disposition of "allocatable"

- **set1-to-set2 set2-to-set1** — a label containing mixed of these variant types receives a disposition of "blocked"

6.2.   Cross-script Variants

Some Myanmar characters look the same as characters in Malayalam, Oriya and Georgian scripts.

6.2.1. Myanmar and Malayalam

The Malayalam and Myanmar code point sets in Table 11 are cross-script variant code points.

Table 11: Myanmar and Malayalam code point and homoglyph analysis

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Malayalam Character Name
1	ၐ	U+101D	MYANMAR LETTER WA	ഠ	U+0D20	MALAYALAM LETTER TTHA

2	ဂ	U+1002	MYANMAR LETTER GA	ഠ	U+0D31	MALAYALAM LETTER RRA

6.2.2. Myanmar and Oriya

The Oriya and Myanmar code point sets in Table 12 are cross-script variant code points.

Table-12: Myanmar and Oriya code point and homoglyph analysis

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Oriya Character Name
1	ဝ	U+101D	MYANMAR LETTER WA	ଠ	U+0B20	ORIYA LETTER TTHA
2	ေဝ	U+1031	MYANMAR VOWEL SIGN E	େ	U+0B47	ORIYA VOWEL SIGN E

Myanmar Letter Wa U+101D is also variant of other characters. (See Table-14)

6.2.3. Myanmar and Georgian

The Georgian and Myanmar code point sets in Table 13 are cross-script variant code points. Georgian word "ითი" (meaning: "digit") can be composed using Myanmar Consonants "တ" and "ဂ".

Table 13: Myanmar and Georgian code point and homoglyph analysis

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	GEORGIAN Character Name
1	ဂ	U+1002	MYANMAR LETTER GA	ი	U+10D8	GEORGIAN LETTER IN
2	တ	U+1010	MYANMAR LETTER TA	თ	U+10D7	GEORGIAN LETTER TAN

6.2.4. Myanmar and Others

Myanmar Character Wa U+101D is a full circle shape and could be mistaken with other cross-script letters which are listed in the table.

Table 14: Myanmar and Other codepoints which have full circle shapes

No.	Glyph	Code Point	Character Name
1	ဝ	U+101D	MYANMAR LETTER WA

2	o	U+006F	LATIN SMALL LETTER O
3	ο	U+03BF	GREEK SMALL LETTER OMICRON
4	о	U+043E	CYRILLIC SMALLER LETTER O
5	օ	U+0585	ARMENIAN SMALL LETTER OH

This LGR inherits additional cross-script variants by integration; they may not be listed here unless they result in in-script variants.

7.   Whole Label Evaluation Rules (WLE)

This section provides the WLE rules 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.

The Whole Label Evaluation Rules (WLE) rely on a syllable based operation for which the identification of syllable boundaries. The basic principle rule in Myanmar Script as syllable boundary as for WLE rule is Consonant > Medials > Vowels > Finals > Tone

Below are the symbols used in the WLE rules.

7.1.   Classification

Repertoire Partition:

C   → Consonant (Table 2)
IV   → Independent Vowel (Section 3.3.2)
                 1023,1024, 1025, 1026, 1027, 1028, 1029, 102A
DV   → Dependent Vowel Sign (Section 3.3.3)
                 102B, 102C, 102D, 102E, 102F, 1030, 1031, 1032, 1033, 1034, 1036,
                 1037, 1038, 1062, 1083, 1084, 1086
K   → Killer or Asat: 103A $ͣ
VIRAMA   → 1039 "ͦ" (It is always in between two consonants and invisible, "ကျ")
M   → Dependent Consonant Sign (Medial)
Sh_Tone   → 1087, 1088, 1089, 108A
Pao_Tone   → 108F, AA7B
Sgaw_Tone   → 1064
OV   → Other Various Sign (Section 3.3.3)
                 1035, 1062, 1085

Sub-classes of Consonants (C)

C1   → 103F သ (GREAT SA, theoretical combination of two Myanmar Letter Sa)

C2   → င , ဉ , ည , ဏ , န , မ , ယ
                  (U+1004, U+1009, U+100A, U+100F, U+1014, U+1019, U+101A)

C3   → င , ဉ , ည , ဏ , န , မ , ဝ , � , �, ႜ
                  (U+1004, U+1009, U+100A, U+100F, U+1014, U+1019, U+101D,
                  U+1075, U+107A, U+107C)

Subclasses of Dependent Vowels (DV)

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

SV   → Short Vowel: 102D, 102F

Sh_Vowel → U+102D, U+102E, U+102F, U+1030, U+1031, U+1083, U+1084, U+1086

ANUSVARA → 1036 $ံ

T_SHORT   → 1037 $့

T_LONG   → 1038 $း
                 103B, 103C, 103D, 103E, 105E, 105F, 1060, 1082

7.2.   Whole Label Evaluation rules:

1. DV must follow C or M   (Section 3.3.3)
   - LV must follow C or M
   - SV must follow C or M
   - S12, S13, S14, S15, S18, S19, S20, S_Mon3, S_Mon5, S_Mon6, S_Mon7, S_Mon8, S_Mon9, S_Mon10, S_Mon11, S_Mon12, S_Mon16, S_Sh1, S_Sh2, S_Sh3, S_Sh4, S_Sh5, S_Pao must follow C or M
   - OV must follow C or M
2. ANUSVARA must follow C or M   (Section 3.3.5)
3. Rules of Medials combining with Consonants   (Section 3.3.4)
   - Rules for Single Medial
     - 103B must follow consonant C_103B
     - 103C must follow consonant C_103C
     - 103E must follow consonant C_103E
     - 103D must follow consonant but not C_n103D
     - M_mon must follow C_Mon
     - 1082 must follow C_shan
   - Rules for Combined Medial
     - MM1 must follow CMM1
     - MM3 must follow CMM3
     - MM5 must follow CMM5
     - S_Mon13 must follow consonant C_103E
     - S_Mon14 must follow consonant C_103E

- S_Mon15 must follow consonant C_103E

Note: This rule set is to prevent arbitrary Consonant-Medial(s) combinations which can cause chaotic rendering.

4. The Myanmar Great Saa (C1 or U+103F) must follow C or M or DV or 1023 or 1025

(Section 3.3.1)

5. (C+K) or (C2 + S16) or (C3+S17) must follow
   - C or M or DV or OV   (Section 3.3.5.1)
     Note: this means that K must follow C, S16 must follow C2, and S17 must follow C3, and any of these three combinations then must follow C or M or DV or OV. This is equivalent, for example, to K following C+C, M+C, DV+C or OV+C, and so on. Because of limitations of the LGR syntax, the XML adopts the second formulation.

6. S11 must follow C or M or DV and another C must follow S11   (Section 3.3.5.1)

7. VIRAMA must be in between two C (C+VIRAMA+C)   (Section 3.3.5.1)
   - C cannot be in between VIRAMAs to prevent virama-c-virama-c
   - The sequence containing a VIRAMA (U+1039) must not follow or precede a VIRAMA

(Table 8)

8. T_LONG or T_SHORT must follow C or M or LV or S12   (Section 3.3.5)

Note: The tone mark T_SHORT (dot below) was treated the same way as T_Long while considering the underlined-hyperlink is not applicable for Myanmar script.

9. Sh_Tone must follow Sh_Vowel or (C+K) or S_Sh2 or S_Sh5   (Section 3.3.3)
      Note: here (C + K) means any sequence where a consonant precedes a K.

10. Pao_Tone must follow M or K or DV except 1037, 1038   (Section 3.3.5)

11. ST4 must follow U+1031, U+1032, S12, S14, or S15   (Section 3.3.5)

12. ST1, ST2, ST3 or Sgaw_Tone must follow C or M or DV   (Section 3.3.5)

13. S_Mon4 must follow C or M or DV
    and another C must follow S_Mon4   (Section 3.3.5.1)

14. U+102C cannot follow any of the three consonants U+1002, U+1015, U+101D

(Table 10)

According to Table 10, there are variants which cannot occur both in the same string. The characters written in comparison from row 1-4, 6, 7, and 8 are equivalent consonants or diacritics for ethnic languages which have equal pronunciation function. If the left column of the pair (grapheme-set-1) occurs in a string, the other equivalent on the right column

(grapheme-set-2) should not be allowed in the same string and vice versa. For example, the word "ခင်မင်", transcription [kɪŋ.mɪŋ], should be either the sequence of [U+1001 **U+1004** U+103A U+1019 **U+1004** U+103A] or [U+1001 **U+105A** U+103A U+1019 **U+105A** U+103A] only.

7.3. Rules regarding the variants

To limit the number of allocatable variant labels, additional dispositions are applied on variant mapping row 4-10 in Table 10 and the corresponding additional action rules are included. These to ensure that, besides the original label, the total number of allocatable variant is limited to two; (1) the label comprised of original code point and grapheme-set-1, or the label comprised of original code point and grapheme-set-2.

The code point U+1004 and U+105A are confusable when they are followed by diacritics. Therefore, in order to limit the cross-script variant impact of U+1004 and U+105A, the contextual rule "not-when(followed-by-c-end) is applied in variant mapping. Such context would give limit a cross-script variant label containing U+1004 to be alternations of Myanmar script.

8.   Contributors

Myanmar Generation Panel

Advisor	Dr. Myint Myint Than	
	Dr. Khin Aye	Mr. Zaw Htut
	Mr. Ngwe Tun	Mr. Maung Sun
Chair	Ms. Thin Zar Phyo	
Co-Chair	Ms. Yin May Oo	
Member	Mr. Naing Win Oo	Mr. Sai Zin Di Di Zone
	Mr. Ye Zarni Aung	Ms. Nancy Aye
	Mr. San Lin Naing	Mr. Thura Soe
	Mr. Min Paing Khant Oo	Mr. Nai Saik Chan
	Mr. Kaung Khant Zaw	Mr. Khun Maung Maung

9.   References

[0]   The Unicode Standard 1.1

[3]   The Unicode Standard 3.0

[9]   The Unicode Standard 5.1

[10]   The Unicode Standard 5.2

[100]   Myanmar Language Commission
         http://www.myanmarlanguagecommission.myn.asia/en/

[101]   Section "Burmese", p. 21   ff in "Representing Myanmar in Unicode", UTN#11,
         Details and Examples, Version 4,
          https://www.unicode.org/notes/tn11/UTN11_4.pdf

[103]   Section "Mon", p. 31   ff in "Representing Myanmar in Unicode", UTN#11, Details and Examples, Version 4
https://www.unicode.org/notes/tn11/UTN11_4.pdf

[104]   Omniglot, Burmese, (Accessed 6 October 2019),
http://www.omniglot.com/writing/burmese.htm,

[105]   Section "Shan", p. 41   ff in "Representing Myanmar in Unicode", UTN#11, Details and Examples, Version 4,
https://www.unicode.org/notes/tn11/UTN11_4.pdf

[107]   Omniglot, Karen (ကညီကျိ်ာ), (Accessed 6 October 2019),
https://www.omniglot.com/writing/karen.htm,

[108]   Section "Pa'o Karen", p. 37ff in "Representing Myanmar in Unicode", UTN#11, Details and Examples, Version 4,
https://www.unicode.org/notes/tn11/UTN11_4.pdf

[109]   Unicode Today, Pa'O Karen language keyboard, (Accessed 6 October 2019),
http://unicodetoday.org/paoh-keyboard-2016-v2/

[110]   Section "Sgaw Karen", p. 33ff in "Representing Myanmar in Unicode", UTN#11, Details and Examples, Version 4,
https://www.unicode.org/notes/tn11/UTN11_4.pdf

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. On the right side of the table are the combinations which are blocked by the rules.

Table A-1: In-script confusable code points

Set#	Unicode Code Point	Glyph	Unicode Code Point	Glyph	Note
1	U+1008	စျ	U+1005 U+103B	စျ	The sequence U+1005 U+103B is invalid
2	U+1009 U+102C	ဉာ	U+1025 U+102C	ဉာ	The sequence U+1025 U+102C is invalid
3	U+105B	ၛ	U+1007 U+103B U+103E	ၛ	The sequence U+1007 U+103B is invalid
4	U+1070	ၰ	U+1003 U+103E	ၰ	The sequence U+1003 U+103E is invalid
5	U+107E	ၾ	U+107D U+103E	ၾ	The sequence U+107D U+103E is invalid
6	U+1009 U+103A	ဉ်	U+1025 U+103A	ဉ်	The sequence U+1025 U+103A is invalid
7	U+1009 U+1037 U+103A	ဉ့်	U+1025 U+1037 U+103A	ဉ့်	The sequence U+1025 U+1037 U+103A is invalid
8	U+1029	ဩ	U+101E U+103C	ဩ	The sequence U+101E U+103C is invalid
9	U+102A	ဪ	U+101E U+103C U+1031 U+102C U+103A	ဪ	The sequence U+101E U+103C U+1031 U+102C U+103A is invalid

10	U+102A	ဪ	U+1029 U+1031 U+102C U+103A	ဩော်	The sequence U+1029 U+1031 U+102C U+103A is invalid
11	U+1000	က	U+1075	ၵ	Semantic similarity
12	U+1002	ဂ	U+1077	ၷ	Semantic similarity
13	U+1005	စ	U+1078	ၸ	Semantic similarity
14	U+1007	ဇ	U+1079	ၹ	Semantic similarity
15	U+100A	ည	U+107A	ၺ	Semantic similarity
16	U+100F	ဏ	U+107C	ၼ	Semantic similarity
17	U+1014	န	U+107C	ၼ	Semantic similarity
18	U+1016	ဖ	U+107D	ၽ	Semantic similarity
19	U+1017	ဗ	U+107F	ၿ	Semantic similarity
20	U+1021	အ	U+1022	ၢ	Semantic similarity
21	U+102B	ါ	U+1083	ၣ	Semantic similarity
22	U+102C	ာ	U+1083	ၣ	Semantic similarity

From No.11 to No.22, the code points are used by different languages although they originally came from the same Nagri characters and still hold the same phonetic values. Since they are visually not similar enough and no simultaneous synonym and homophones exist, they are listed as confusables.

Confusable Characters between Myanmar Script and Malayalam

Table A-2: Myanmar-Malayalam confusable code points

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 | ကာ | U+1000, U+102C | MYANMAR LETTER KA , MYANMAR VOWEL SIGN AA | ന | U+0D28, U+0D4D, U+0D28 | MALAYALAM LETTER NA, MALAYALAM SIGN VIRAMA, MALAYALAM LETTER NA |

Confusable Characters between Myanmar and Sinhala

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

Table A-3: Myanmar-Sinhala confusable code points

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Sinhala Character Name
1	ာ	U+102C	MYANMAR VOWEL SIGN AA	ා	U+0DCF	SINHALA VOWEL SIGN AELA-PILLA
2	ေ	U+1031	MYANMAR VOWEL SIGN E	ෙ	U+0DD9	SINHALA VOWEL SIGN KOMBUVA
3	ော	U+1031, U+102C	MYANMAR VOWEL SIGN AA, MYANMAR VOWEL SIGN E	ො	U+0DDC	SINHALA VOWEL SIGN KOMBUVA HAA AELA-PILLA

Confusable Characters between Myanmar and Telugu

There are two Telugu 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. In the second level, numbers may be included, therefore, it is good to consider Myanmar number zero "၀" U+1040 which also has the exactly round circle shape.

Table A-4: Myanmar-Telugu confusable code points

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Telugu Character Name
1	ဝ	U+101D	MYANMAR LETTER WA	౦	U+0C66	TELUGU DIGIT ZERO
2	၀	U+1040	MYANMAR DIGIT ZERO	౦	U+0C66	TELUGU DIGIT ZERO
3	ဂ	U+1002	MYANMAR LETTER GA	౧	U+0C67	TELUGU DIGIT ONE

Appendix B. Consonant Combinations

The combinations of Consonant Medial Sequences.

Table B-1: The combination of Consonant Medial Sequences

	M	M	M	M	MM1	MM2	MM3	MM4	MM5	MMM	M_mon	M_mon	M_mon	M_shan
Consonants (1000.. 1080)	U+103B ျ	U+103C ြ	U+103D ွ	U+103E ှ	U+103B U+103D ျွ	U+103B U+103E ျှ	U+103C U+103D ြွ	U+103C U+103E ြှ	U+103D U+103E ွှ	U+103C U+103D U+103E ြွှ	U+105E ၞ	U+105F ၟ	U+1060 ၠ	U+1082 ႂ
က U+1000	ကျ	ကြ	ကွ	ကှ	ကျွ	ကျှ	ကြွ	ကြှ	ကွှ	ကြွှ	ကၞ	ကၟ	ကၠ	ကႂ
ခ U+1001	ချ	ခြ	ခွ	ခှ	ချွ	ချှ	ခြွ	ခြှ	ခွှ	ခြွှ	ခၞ	ခၟ	ခၠ	ခႂ
ဂ U+1002	ဂျ	ဂြ	ဂွ	ဂှ	ဂျွ	ဂျှ	ဂြွ	ဂြှ	ဂွှ	ဂြွှ	ဂၞ	ဂၟ	ဂၠ	ဂႂ
ဃ U+1003	ဃျ	ဃြ	ဃွ	ဃှ	ဃျွ	ဃျှ	ဃြွ	ဃြှ	ဃွှ	ဃြွှ	ဃၞ	ဃၟ	ဃၠ	ဃႂ
င U+1004	ငျ	ငြ	ငွ	ငှ	ငျွ	ငျှ	ငြွ	ငြှ	ငွှ	ငြွှ	ငၞ	ငၟ	ငၠ	ငႂ
စ U+1005	စျ	စြ	စွ	စှ	စျွ	စျှ	စြွ	စြှ	စွှ	စြွှ	စၞ	စၟ	စၠ	စႂ
ဆ U+1006	ဆျ	ဆြ	ဆွ	ဆှ	ဆျွ	ဆျှ	ဆြွ	ဆြှ	ဆွှ	ဆြွှ	ဆၞ	ဆၟ	ဆၠ	ဆႂ
ဇ U+1007	ဇျ	ဇြ	ဇွ	ဇှ	ဇျွ	ဇျှ	ဇြွ	ဇြှ	ဇွှ	ဇြွှ	ဇၞ	ဇၟ	ဇၠ	ဇႂ
ဈ U+1008	ဈျ	ဈြ	ဈွ	ဈှ	ဈျွ	ဈျှ	ဈြွ	ဈြှ	ဈွှ	ဈြွှ	ဈၞ	ဈၟ	ဈၠ	ဈႂ
ဉ U+1009	ဉျ	ဉြ	ဉွ	ဉှ	ဉျွ	ဉျှ	ဉြွ	ဉြှ	ဉွှ	ဉြွှ	ဉၞ	ဉၟ	ဉၠ	ဉႂ
ည U+100A	ညျ	ညြ	ညွ	ညှ	ညျွ	ညျှ	ညြွ	ညြှ	ညွှ	ညြွှ	ညၞ	ညၟ	ညၠ	ညႂ
ဋ U+100B	ဋျ	ဋြ	ဋွ	ဋှ	ဋျွ	ဋျှ	ဋြွ	ဋြှ	ဋွှ	ဋြွှ	ဋၞ	ဋၟ	ဋၠ	ဋႂ
ဌ U+100C	ဌျ	ဌြ	ဌွ	ဌှ	ဌျွ	ဌျှ	ဌြွ	ဌြှ	ဌွှ	ဌြွှ	ဌၞ	ဌၟ	ဌၠ	ဌႂ
ဍ U+100D	ဍျ	ဍြ	ဍွ	ဍှ	ဍျွ	ဍျှ	ဍြွ	ဍြှ	ဍွှ	ဍြွှ	ဍၞ	ဍၟ	ဍၠ	ဍႂ
ဎ U+100E	ဎျ	ဎြ	ဎွ	ဎှ	ဎျွ	ဎျှ	ဎြွ	ဎြှ	ဎွှ	ဎြွှ	ဎၞ	ဎၟ	ဎၠ	ဎႂ

ဏ U+100F	ဏာ	ဩ	ဏ္	ဏ်	ဏ္တ	ဏ္တ	ဩ္တ	ဩ္တ	ဏ္တ	ဩ္တ	ဏ္တ	ဏ္တ	ဏ္တ	ဏ္တ
တ U+1010	တာ	ြတ	တ္	တ်	တ္တ	တ္တ	ြတ္	ြတ္	တ္	ြတ္	တ္	တ္	တ္	တ္
ထ U+1011	ထာ	ြထ	ထ္	ထ်	ထ္	ထ္	ြထ္	ြထ္	ထ္	ြထ္	ထ္	ထ္	ထ္	ထ္
ဒ U+1012	ဒာ	ြဒ	ဒ္	ဒ်	ဒ္	ဒ္	ြဒ္	ြဒ္	ဒ္	ြဒ္	ဒ္	ဒ္	ဒ္	ဒ္
ဓ U+1013	ဓာ	ြဓ	ဓ္	ဓ်	ဓ္	ဓ္	ြဓ္	ြဓ္	ဓ္	ြဓ္	ဓ္	ဓ္	ဓ္	ဓ္
န U+1014	နာ	ြန	န္	န်	န္	န္	ြန္	ြန္	န္	ြန္	န္	န္	န္	န္
ပ U+1015	ပါ	ြပ	ပ္	ပ်	ပ္	ပ္	ြပ္	ြပ္	ပ္	ြပ္	ပ္	ပ္	ပ္	ပ္
ဖ U+1016	ဖာ	ြဖ	ဖ္	ဖ်	ဖ္	ဖ္	ြဖ္	ြဖ္	ဖ္	ြဖ္	ဖ္	ဖ္	ဖ္	ဖ္
ဗ U+1017	ဗာ	ြဗ	ဗ္	ဗ်	ဗ္	ဗ္	ြဗ္	ြဗ္	ဗ္	ြဗ္	ဗ္	ဗ္	ဗ္	ဗ္
ဘ U+1018	ဘာ	ြဘ	ဘ္	ဘ်	ဘ္	ဘ္	ြဘ္	ြဘ္	ဘ္	ြဘ္	ဘ္	ဘ္	ဘ္	ဘ္
မ U+1019	မာ	ြမ	မ္	မ်	မ္	မ္	ြမ္	ြမ္	မ္	ြမ္	မ္	မ္	မ္	မ္
ယ U+101A	ယာ	ြယ	ယ္	ယ်	ယ္	ယ္	ြယ္	ြယ္	ယ္	ြယ္	ယ္	ယ္	ယ္	ယ္
ရ U+101B	ရာ	ြရ	ရ္	ရ်	ရ္	ရ္	ြရ္	ြရ္	ရ္	ြရ္	ရ္	ရ္	ရ္	ရ္
လ U+101C	လာ	ြလ	လ္	လ်	လ္	လ္	ြလ္	ြလ္	လ္	ြလ္	လ္	လ္	လ္	လ္
ဝ U+101D	ဝါ	ြဝ	ဝ္	ဝ်	ဝ္	ဝ္	ြဝ္	ြဝ္	ဝ္	ြဝ္	ဝ္	ဝ္	ဝ္	ဝ္
သ U+101E	သာ	ြသ	သ္	သ်	သ္	သ္	ြသ္	ြသ္	သ္	ြသ္	သ္	သ္	သ္	သ္
ဟ U+101F	ဟာ	ြဟ	ဟ္	ဟ်	ဟ္	ဟ္	ြဟ္	ြဟ္	ဟ္	ြဟ္	ဟ္	ဟ္	ဟ္	ဟ္
ဠ U+1020	ဠာ	ြဠ	ဠ္	ဠ်	ဠ္	ဠ္	ြဠ္	ြဠ္	ဠ္	ြဠ္	ဠ္	ဠ္	ဠ္	ဠ္

U+1021													
U+1022													
U+105A													
U+105B													
U+105C													
U+105D													
U+1075													
U+1076													
U+1077													
U+1078													
U+1079													
U+107A													
U+107B													
U+107C													
U+107D													
U+107E													
U+107F													
U+1080													

The table shows the combinations of all Consonants and Medial possibilities. The white color boxes are the combinations which are being used in generic words. The green boxes are possible combinations which could be spelled in creative names in future, and have no complicated rendering issue. The orange code means these graphemes are neutral and less probable to form the combination. The red are the combines which are unpronounceable, not handled by all fonts uniformly, or may cause unexpected rendering issues.

Since each Medial has their own set of Consonants to be combined with, the combination rules are tailored accordingly under WLE rule #3. The Medial sets, column MM2 and MM4 have very few consonants which can actually be combined, therefore, all combinations relating to MM2 and MM4 are defined as sequences in Table-8.

Of all consonant+medial combinations, "သျ" [101E+103C] appears to be visually almost identical to Independent vowel "ဩ" [U+1029]. There are only two valid words spelled with "သျ" [101E+103C], therefore, the words are defined as sequences SS09 "သျ၊" [U+101E U+103C U+1083 U+1087] and SS10 "သျန်" [U+101E U+103C U+1014 U+103A] in session 5.3, Code Point Sequences, Table 8.