

Date: 14 May 1998

Dear Michael,

After studying your letter, and the document# N1729, I've got a few questions and things to write to you about a few misinterpretations I found.

QUESTIONS

1) As far as I've seen there is no letter 'ဝ်' (WiteCha). I've seen -ါ (MoukCha) only.

2) I see no difference between

U+1737 င် and U+1738 င်

3) I'm curious where all other Burmese glyphs not listed in the UCS table are located. Please refer me for more readings if available on-line. I need more education on this Unicode's policy:

The UCS specifies that only characters should be encoded for data interchange and storage, and that the specific glyph representations should be handled at another level of abstraction, in the font.

4) I urge you to educate me more on this and what it means too:

Burmese. *The basic proposal was agreed to easily; it is a virama model. The main items of discussion were character ordering, initial vowels, and subscript YA, RA, VA, HA letters. We decided to encode characters in natural Burmese order. Lee Collins had proposed to delete eight vowels from N1523, but after discussion, we decided to retain the six vowels used in Pali and Sanskrit. The four subscript letters proposed by Lee Collins were not required when using the virama model, so they were not accepted. The consensus of the group was to progress both proposals as PDAMS.*

5) Where can I get a copy of document# N1523 ?

(From now on, please note that I'm going to describe a Burmese blank character as '–' (hyphen), instead of the dim 'O'.)

THE ISSUE OF 'RAYIT' AND 'YAPINT'

From what I've learnt, there is a slight misinterpretation in Prof. John Okell's papers. In your letter, you mentioned:

I note that your code table is both character-based and glyph-based. That is, it contains unique characters (like ဝ် SA (THA)) as well as the varying glyph representations of characters (like ရှ င် င် င် င် all of which mean RA KOK (YA GAU)). (Note that I am using Professor John Okell's scholarly transliteration as well as a popular transliteration. I hope I make no errors.)

... In the case of Burmese, for instance, the initial vowel *O* could be composed (in UCS terms) of SA (THA) + VIRAMA + RA ROK (YA GAU): $\text{ဆ} + \overset{\text{c}}{\text{}} + \text{ရ} = \text{ဆ}$.

It is NOT...

$\text{က} = \text{က} + \overset{\text{c}}{\text{}} + \text{ရ} = \text{ကရ်} \quad (\text{Ka+Virama+RaKok})$

$\text{ကျ} = \text{က} + \overset{\text{c}}{\text{}} + \text{ယ} = \text{ကယ်} \quad (\text{Ka+Virama+YaPakLek})$

If these are decoded this way, they will pronounce 'Kar' and 'Kay'.

it is ...

$\text{က} = \text{က} + \overset{\text{c}}{\text{}} + \text{ရ} + \text{ရ} = \text{ကရ်ရ} \quad (\text{Ka+Virama+RaKok+RaKok})$

$\text{ကျ} = \text{က} + \overset{\text{c}}{\text{}} + \text{ယ} + \text{ယ} = \text{ကယ်ယ} \quad (\text{Ka+Virama+YaPakLek + YaPakLek})$

Only when it is correctly decoded, it will be 'Kra (KaRa)' and 'Kya (KaYa)'.

All of $\text{က} \text{ကျ} \text{က}$ developed from the abbreviated forms of certain combinations. (I can find reference info if you need.)

$\text{က} = \text{က} = \text{က} + \overset{\text{c}}{\text{}} + \text{ရ} + \text{ရ}$

$\text{ကျ} = \text{ကျ} = \text{က} + \overset{\text{c}}{\text{}} + \text{ယ} + \text{ယ}$

$\text{က} = \text{က} = \text{က} + \overset{\text{c}}{\text{}} + \text{ဟ} + \text{ဟ}$

Therefore, my point is $\text{က} \text{ကျ} \text{က}$ are no more glyph-representations. They are the outcome of a development, evolved over centuries. Many Burmese of today do not even know that, according to history က is equivalent to မရ်ရ . Because it also has developed into a different speaking nature: က is no more pronounced 'MaRa'; It is 'Mia' today. If you say 'MaRa' these days, you'll be understood as မရ .

PAT-SINT and KINZI

I still can't figure out how all the whole bunch of PatSints, which are not yet in the UCS table yet, will work in a font. It must be because of the policy that "only characters should be encoded for data interchange and storage, and that the specific glyph representations should be handled at another level of abstraction, in the font."

However PatSints are predicted to be obsolete in a few decades, they are still very well in use today. **My question is where all these PatSints are going to be assigned. Are they, including KinZi, going to use up the space in the reserved area, between U+175A to U+177F?** If they are not predefined, the character assignments might not be stable in the industry.

SPECIAL CHARACTERS

It is true that the vowel letter သ cannot be decomposed.

In the case of Burmese, for instance, the initial vowel O could be composed (in UCS terms) of SA (THA) + VIRAMA + RA ROK (YA GAU): သ + [◌] + ရ = သ.

I certainly agree with this decision to have non-decomposable characters, such as သ. It has to be noted that the letter သ is not considered as a vowel, but အော (its identical pronunciation) is. သ is considered as a special character.

Please note that သ (pronounced 'O') definitely is NOT သ+ရံ+ရ , if it is THA+RAYIT it will pronounce 'ThRa'. And this is the reason why သ+၂ (ThRa) does not exist in our language. So it is a good idea to keep it non-decomposable.

Take a note that you've mentioned:

Burmese initial vowels, some of which (အ က ဤ ဥ ဇ) cannot be decomposed, and some of which (အ ဉီ အဲ သြ သြော်) can be decomposed,

The following is the BASIC set of Burmese vowels: (row after row)

အ	အာ	အာ့
အိ	အီ	အိ့
အု	အူ	အု့
အေ	အော့	အေ့
အဲ	အဲ့	
အော	အော့	အော်
အံ	အံ့	
အို	အို့	အို့

These do pronounce as same as one of those in this basic set of vowels, but these are considered special characters (some are decomposable):

က ဤ ဥ ဦ ဦး ဥး ဇ ဩ ဩော် ဣ

And these are other special characters which do not sound like any of the vowels:

၍ ဌ ၎င်း

Therefore, I believe encoding the following decomposable vowels in the UCS is not appropriate:

U+1722 အာ

U+1728 အဲ

I believe these decomposable special characters, are also not appropriate:

U+1726 ဦ

U+172A ဩော်

I can understand that these were encoded for the sake of multi-platform, multi-lingual applications and data-conversion. I'll go further on this under next topic.

ECONOMIC CODING MODEL

In your letter, you mentioned:

... In the context of international and intra-script data interchange, it was decided for both Sinhala and Burmese that the Sanskrit and Pali vowels be represented in the Standard as they are for many Brahmi-derived scripts...

The convenience of using this model for encoding cannot be understated. The Burmese script behaves much like all Brahmic scripts: if the Burmese coding conventions in the UCS are similar to those for the other Brahmic scripts, software for Burmese will be easier and cheaper – and therefore more attractive – to implement. Conversion software, for converting a Pali text from Burmese to Sinhala or Devanagari script, will be much easier to write.

But such applications which convert Roman-Pali text to Burmese or Devanagari are already invented. They've been used in the Buddhist Bible CD published by Indian Buddhist Mr.Goenka's Family. In this CD the whole Buddhist Bible is browsable and searchable in Burmese, Roman-Pali, or Devanagari.

Like I've said, the Burmese encoding on that CD is almost identical from WinMyanmar Systems' encoding, except the fact that they have precomposed ၁၁ and ၁၂ . I agree with them too.

TECHNICAL LIMITATIONS

I'm not sure whether I've told you that I've designed several dozens of fonts from scratch. You can take a look at them at

<http://www.geocities.com/Tokyo/Pagoda/8160/wms.htm>

To my experience, today's typographic technology will not support to adjust space for these characters. They will not be compatible when used together with other glyphs and letters.

U+1732 ၁၁

U+1733 ၁၂

Kerning won't help those characters from loose/tight problems, when we type these:

မော ကော ကြောင်း ကျောင်း

HARMONIZATION versus ESTABLISHED TYPING METHOD

Harmonization is a good principle. However, since economic encoding in Burmese character set is unlikely to work out, harmonizing to it will be far more complicated.

It is very likely to conflict with the well-established Burmese typing method. This conflict will be similar to that of "Dvorak v.s. QWERTY", even though it will be better, it can be hardly accepted.

The draw-backs I can see now is:-

- there will be less data storage, but more keys to be remembered;
- or if an IME (Input Method Editor) is used instead, this will not be compatible with all applications in the market.

Well, that's all for now. I hope you'll have time to study and get back to me soon.

Sincerely,

Zaw Htut

stevehtut@usa.net