KOBE – ICANN GDD: Myanmar GP Meeting with Integration Panel
Tuesday, March 12, 2019 – 17:00 to 18:30 JST
ICANN64 | Kobe, Japan

UNIDENTIFIED FEMALE: Okay. Thank you all for joining the ICANN GDD-Myanmar GP meeting with the Integration Panel on the 12ᵗʰ of March. I think we can start by introducing ourselves around the table because this is being recorded. So maybe I can start with the [front]? Please go ahead.

MICHAEL BAULAND: Hi. My name is Michael Bauland. I'm a member of the Latin Generation Panel.

BILL JOURIS: My name is Bill Jouris and I'm also a member of the Latin Generation Panel.

JOYCE CHEN: Hello, everyone. I'm Joyce Chen from the ICANN APAC regional office based in Singapore.

PITINAN KOOARMORNPATANA: Hi, I'm Pitinan Kooarmornpatana, ICANN staff.

Note: The following is the output resulting from transcribing an audio file into a word/text document. Although the transcription is largely accurate, in some cases may be incomplete or inaccurate due to inaudible passages and grammatical corrections. It is posted as an aid to the original audio file, but should not be treated as an authoritative record.

MICHEL SUIGNARD:	Michel Suignard, Integration Panel.
WIL TAN:	Wil Tan, Integration Panel.
YIN MAY OO:	Yin May Oo. I'm Co-Chair of Myanmar GP. Thanks.
SARMAD HUSSAIN:	Sarmad Hussain, ICANN staff.
UNIDENTIFIED FEMALE:	Thank you, all. Then I will handle over to Yin May to present. Thank you.
YIN MAY OO:	Okay. So this is the update about our GP work, second round. Do the slide, please.

It's okay.

So forgive me for not checking all the renderings in here, but it's the list of the languages that the Myanmar script has covered for this round. So for Myanmar, Burmese is the official language for the whole country, but we also have Shan *or Chin?*, [Rakine,

Garin] Mon, and [Bolkerin.] So basically some of the languages share all the scripts, the same as Burmese. They are only pronounced differently and they are spelled differently. But for certain cases, like Shan and [Korean] and Mon, they also have their own set of characters, and they have their own ways to spell the things.

Next slide, please. Our generation panel have included a lot more people, also from ethnic groups, who have learned their native scripts since they were young. So this is the list of the people that we have included in our panel.

So to begin, this is the list of the – excuse me for a minute. Is there a way that we can project the card rendering?

UNIDENTIFIED FEMALE: [inaudible]

YIN MAY OO: Yeah. So it looks like – all right. Yeah, that would be great.

So Myanmar script rendering is so unique that sometimes Adobe doesn't get it.

EN

UNIDENTIFIED FEMALE:	Oh, right. So I think I'll post the link in the chat, so please follow that as well because, in the Adobe system, we can only display the PDF.

YIN MAY OO:	Okay. So … okay. It's correct. Yeah.

So [I think 200]. Yes, that's fine. Okay.

So what we have here is the [graphic] of the variant. From the left side is the correct character, and on the right side is other combinations that we can create to look like the left one. So some of these combinations are allowed. In Burmese, we don't write the way on the right, but in Mon or Shan they have this combination. So we must be aware of what other situations we are dealing with.

So we have a bunch of in-script variant lists. These are the ones with Myanmar characters – different combinations. Then the last three are also consonants: from [inaudible] Karen, a consonant from Shan, and a consonant from Mon.

UNIDENTIFIED MALE:	[inaudible]

YIN MAY OO:	Yes.

[SARMAD HUSSAIN]:   Just a question about the last variant. That looks visibly different. So is there a non-visual motivation for that variant?

YIN MAY OO: Yes. So Myanmar has the consonant [na] and Mon also has the consonant [na] on the right. So when they pronounce [na] they want to spell with the right-side character. So it's the same pronunciation, some consonant, but we have different writing.

Also, when we put some diacritic on the right one, then it looks visibly similar with the left.

Next slide, please.

[SARMAD HUSSAIN: I think it may be useful to add that explanation, maybe even possibly with an example of combination in the proposal to make it clearer for somebody who does not understand that background, I guess.

YIN MAY OO: Yeah, sure. We should explain by showing in which role what is the motivation of these variants.

So for the next one, this is also the same case for Burmese [Fazez] Mon. Burmese has our own consonant [za] and then Mon has their own consonant [za]. They won't happen together in the same string.

So the second one is independent vowels, vowel sign A for Burmese. Then on the right is the independent vowel sign A also for Mon. Also, that one, the fourth one, the fifth one, are from the Shan family. Shan also – although we may share the same diacritics, we different consonants to write their own names, their own scripts. So these are the variants based on the meaning of consonants and pronunciation.

So the list goes on. These are all consonants.

Next slide, please. So Burmese comes from [Brahmese] script, so we have two [pa], two [na]. For deep [na] and soft [na], we have different consonant characters. But for the Shan language they have only one [na]. They have simplified. So they will not use Burmese [na] if they are [right] spelling their own names. So the rest of the consonants are also the same thing.

For Rule #21, 22, and 23, they are vowel diacritics, so they also have the same property to change the sound of the consonant, but they won't appear to together in the same place.

EN

So these are all in-script. Now we move to cross-script. So for cross-script, we have a few characters with [malialan]. Only two consonant characters are similar. [Orea] also we have one consonant and one diacritic, which looks similar.

For the next slide, an interesting [lay]. Georgian has something very close to the Burmese consonant. So we have consonant [ka] and consonant [da] which coincides with Georgian letter [da] and Georgian vowel [inaudible]. So the square box is supposed to be a character that looks like [da]. So these are all cross-scripts.

Okay. So Slide 10. Confusable. Here was we have are the characters or the combinations that we have in Myanmar script and different ethnic scripts. So the combinations on the right are the ones that can be confused with the actual character on the left. So we have defined the ones on the right as invalid strings. So this is just for recording that we have these combinations that can be written and form the same visual thing.

Okay. So updates on the rules. The definitions Which ones are called consonants [inaudible] vowels. They are all the same. And we have updated the rules.

Next slide, please. So some of the rules which are contracted with each other are removed. Now we have simplified what other things to look for, like some of the characters are not supposed to be at the head of the string. Some of the combinations are

supposed to be not at the head of the string. They must be at the second place. So those things are defined.

Rule 8 is about the [celeba] chain character that we use to stack consonants together. We also updated so that we cannot stack too many consonants together, like stack only once.

These are for Burmese. Some certain combinations only come with a certain set of consonants so that we don't confuse them with very random kinds of combinations to form very strange [graphics] which cannot be pronounced. So these are all the updates on our whole label rules.

So we got the review for our second proposal, and we would like to submit it by this month. Then, after that, we would like to have a public workshop in Yangon and we would like to open for public comment. We'll gather the information and then we'll finalize by this coming July.

So these are all the updates for the Myanmar GP work. Thank you.

UNIDENTIFIED FEMALE: Thank you. Then we open the floor for questions.

SARMAD HUSSAIN: Oh, you wanted to go? Please.

UNIDENTIFIED MALE: No.

SARMAD HUSSAIN: Oh. So thank you for a very comprehensive update on the Myanmar GP. It's excellent work, so thank you for that as well.

Just a question on this slide. When are you considering to finalize this draft you are thinking of sharing with the IP? Or do we have currently shared feedback? I just want to clarify the status of where we are.

YIN MAY OO: We have received the review for the second round, and we have finalized the rules and the variants. So we have to finalize the second draft by a few days. Then we will submit again.

BILL JOURIS: We did a little work at the request of the IP on some generic symbols – a straight line, a circle, a crescent – and we think we found in Myanmar script: Myanmar letter [wa] – I'm sure I'm pronouncing that wrong – code point 101D, and also code point 0D20. I don't know if you could consider those close enough to be variants of a Latin letter O or not, but you might want to look at that.

Likewise, the crescent code point 1004 and the Latin small letter c were just a couple that we came across. Thank you.

YIN MAY OO: That's great information. We also noticed it looks a little bit alike. So you are saying Latin letter small c and Latin letter small o look alike with 1004 and 101D?

BILL JOURIS: Yes.

YIN MAY OO: So only these two?

BILL JOURIS: The only other one is there is a vowel sign 1062, which there is a code point in Latin called a dotless letter i. I don't know if that Myanmar vowel sign can be combined with the other two or not. There might be some potential there as well. I can send you the little spreadsheet we made of what the various code points are if that would be helpful.

YIN MAY OO: Oh, yeah. That would be great. And please share it with me. Thank you so much.

BILL JOURIS: [Will do].

MICHEL SUIGNARD: We in fact already spotted that c and the o. To some degree we don't necessarily expect all the GPs to find these kinds of out-of-the ordinary variants. So we in fact, when we do integration, do look at those cases. We had spotted the o and the c already as possible variants. In fact, they do show up in my presentation tomorrow as already spotted. So it's something we see even if the GP doesn't find them. We tend to pay attention to those cases.

Also, we have to be careful that, if you have combining characters that are confusable, you need to have a base to go with it. That's kind of the rule. What we've seen in this case – I've not done the work yet – we see you always have to have, in addition to the combining marks, then a possible base that goes with it. If you have a combining on its own, that is confusable. It's not enough. We have a lot of those cases, in fact, in the situation where you have, not a lot of combining marks that could be confusable but there's no base to go with. So then we are not making them confusable.

BILL JOURIS: Right. That was why I said I don't know if that combining mark can be done with the circle and the crescent.

MICHEL SUIGNARD: [Could be].

BILL JOURIS: If it can't, then it's not a problem, but if it can be, then it needs to be looked at.

YIN MAY OO: Yeah. The diacritic 1062 can be appear next to 1004 or 101D. It is possible.

Any more questions?

PITINAN KOOARMORNPATANA: I have a question maybe for follow-up with Michel. Does it mean that the GP doesn't have to include all the [c and c] roles? That's the IP's role to work on it?

MICHEL SUIGNARD: I would say that is optional. To find them is even better. We will basically had them later as part of the integration. We're not forcing every GP to analyze every other script. It's just too much sometimes. On some GP, basically [inaudible] drown in the

ICANN 64 COMMUNITY FORUM
KOBE
9–14 March 2019

exercise. We don't want them to spend too much time on that because it's not really their duty to look. Otherwise, then you could say, well, maybe you also have to look at [CGAK]. I mean, there's no limit to where you can go there.

BILL JOURIS: Latin GP is very familiar with drowning on this.

YIN MAY OO: Okay. So can I have the slide next to this?

So the Georgian script, I would like to clarify because there's one word that can use Myanmar characters, too [phon]. So this kind of case, we need to include, right?

MIGUE SUGINARD: Again, we have been debating this kind of case for a while now because we basically allow cross-script when there's very simple shapes, like the o on the vertical [bow] or the c or the partial circle, basically, because those are so simple. Our first position was in fact not to include them as variants because they were just two scripts not related. We didn't see how you could create a label. In those it would be confusing as a script. That was our initial position.

EN

Then we discovered to our great horror that triple-o had been delegated – the o repeated three times. So that meant that in fact we had to be prepared to look at that. [Some cases] were in fact very simple shapes that could be confused among many, many, many scripts [that] could be in fact allocated.

So that's why we added in fact the o now, even those they are not related script. They're not even Latin. They're nothing like Myanmar, as a good example.

I think when you go on the case like the [ta] on the Georgian, that's pushing too much because, again, variants live as labels. They don't live as code points. So you have to be able to produce a multiple of them [inaudible] makes sense of it.

So you're relying here on having the Myanmar repeated multiple times or the Georgian character being repeated multiple times [to] make sense, at least in one the scripts. If it doesn't make any sense in any script, that's pushing a bit too far in my opinion. But this is a judgement call here.

SARMARD HUSSAIN: I'm just noticing that this is a capital letter. I think we should probably verify also whether this [b] valid or not. If it's not [b] valid, then this can be taken out.

MICHEL SUIGNARD:	And this is in the extended block, not in the regular Georgian block.
	So a question would be whether uppercase letters would be in scope for this and [our] letters are not.
	It is not. So, yeah. That's the thing. When you look at [inaudible], please limit yourself to MSR. Otherwise, you'd be finding too many characters that are out of scope. 1C97 is in fact out of a scope. That's a good point. The same for the one above, 1c98. That's also out of scope.
YIN MAY OO:	Oh, I see.
MICHEL SUIGNARD:	[inaudible]
YIN MAY OO:	Oh, all right. So this can be regarded as …
MICHEL SUIGNARD:	This is deleted.
YIN MAY OO:	Deleted? Okay, thanks.

EN

MICHEL SUIGNARD: That is another Georgian letter. That is 10D8. That one is …

SARMAD HUSSAIN: So an alternative is actually 10D7 and 10D8.

MICHEL SUIGNARD: Yeah.

SARMAD HUSSAIN: Those are the ones you should use here [inaudible].

MICHEL SUIGNARD: Yes. Yeah, Georgian went through a revision recently. [inaudible] was added [as] a case pair to the great unhappiness of other people. But that happened.

                         But still, for an IDN point of view, those don't really]exist. Only the origin and characters from the 1000-block exist. So, yeah.

YIN MAY OO: So thanks for clarifying about the Georgian letters. Any more questions, please?

EN

SARMAD HUSSAIN: Since there are many rules that are required by these languages beyond Burmese, I think one of the things we requested from the Myanmar GP was to perhaps to get a corpus or word list from these languages so that we can test the rules. Is that something you've been able to compile, or is that something which is still being worked on?

YIN MAY OO: We have a word list for Burmese given, and we have the list for Shun and Mun. So we are still working on a [Skal] Karen list and [Bolkin] list. So we have two more to go.

SARMAD HUSSAIN: One thing you should probably consider is that the data you provide when you run the rules on the data should be sufficient that each rule is at least run one or two times so that we can verify that rules are actually consistent with what the expectation is.

PITINAN KOOARMORNPATANA: Can I have the last question? So far, the latest IP feedback is requesting to maybe simplify the rules and to make it script-based, not language-based. I know you've been working on it. So can you help us a little bit on how it goes?

YIN MAY OO:

Yes. So we have excluded some of the rules which are repeating. The thing is, each ethnic group has their own characters. Just now, I showed it in the in-script variants. So their have their own way of writing the same consonant. So if we can simplify it, Burmese consonant [ka] won't happen the same time as the Shun consonant [ka]. Something like that. Then we can just simplify it very much already. For the diacritics that can be used as either the vowel or final tone we can just be flexible.

So there are some combinations which will only occur with certain types of consonants, like certain [medials] for Mun will only pair with a certain set of consonants that we can also define so the arbitrary combinations will not happen.

MICHEL SUIGNARD:

Yeah. I think the point here is that as long as the rules – even if they're language-specific, expressed in code points – that's fine. We don't really care. If, for example, some code points are only using some languages, then they're expressed as rules on code points. Then there's no issue for us.

But we see we cannot have a language [identifier]. We don't know what language this thing is about. So we have no way to know the language from an LGR point of view, so everything has to be expressed within code points or a group of code points or classes.

From what I could see in the presentation, there seems to be quite a few less rules. Still a lot of them. So we see that's a bit of a concern because it looks like a complicated system. But we see if we have to see an actual LGR proposal, and then we will basically look at it and do a new review. It looks complicated, but that's, I guess, part of the system.

But for sure if we could find any simplification, we will again suggest anything we could find to simplify because, overall, the goal of the LGRs for the root we want to be as simple as possible because the more complicated they are, the more the chance we mimic a mistake or miss something.

Anyway, we have to wait for evidence at this point. I don't want to prejudge what the LGR will be. We'll wait. But at least we could see by the presentation again that there's a bit less rules than before.

YIN MAY OO: One more thing to add about the rules is that we have also some kind of writings adapted from old Burmese with contradict with the rules. But those are a limited number of words, like three or four words, which are written not according to these rules. So we can just define them as exceptions instead of just bending the rules.

UNIDENTIFIED MALE: Can you clarify? When you say that there is some old Burmese writing that contradicts with the rules, does that mean the rules, as you have currently defined, will not allow those old Burmese writings? Is that right?

YIN MAY OO: Yes. We have certain rules regarding [U]103A, that it should not go next to [medial], and there are a few Burmese words that are of shortened kind of spellings adapted from traditional old Burmese. So there are only, like, three words which we have found. When we checked with the [SMO] rules for the first time with the word list, these three words came up failed. So also we have noted down what the words are that we still use but that failed with the rules. So we can just put them as exceptions.

MICHEL SUIGNARD: There's a solution for that. In some cases, you can always express sequences as part of the LGR repertoire. That way you escape the rule because then you pre-define the sequence that doesn't have to obey the rules. So if there's only a few words that are exceptions, you can in fact encode them as sequences on the LGR.

UNIDENTIFIED MALE: And I guess it's always worth asking the question of, are we creating spelling rules versus writing a system—

MICHEL SUIGNARD: Yeah. Well-formed.

UNIDENTIFIED MALE: Yeah. Well-formed. That the goal is to just make sure that … I'm not familiar with the Myanmar writing system, but when you mention that certain rules are there to prevent spellings that cannot be pronounced, that could indicate to me that maybe we're concentrating too much on pronunciation. But it also could be very valid from a writing system that's well-formed point of view.

YIN MAY OO: Yeah. So we just want to prevent the arbitrary combinations of certain cases. So Burmese has a lot of types of diacritics, like which type goes closer to the consonant, which type goes farthest from the consonant – those orders – so we have to make sure that the input string is in the right sequence so that we don't confuse.

MICHEL SUIGNARD: I think, to [put it into] perspective, we are not necessarily concerned about the correct [autography]. Our main concern is

in fact what we call well-formed. Basically, well-formed is another way of saying that most of the rendering engines that do render Burmese will run correctly that sequence, no matter what's [spinning] or [wrong-spinning] this thing is.

So we see that a lot in the context where, if you enter a wrong sequence, basically your rendering engine waves a hand on the [inaudible] start to show a dotted circle. That's a good sign that something is really wrong: when the sequence doesn't make any sense, no matter what, because the combination is totally against the well-formed rules of the writing system.

So we see a bit of a gray area because to some degree there's some consensus to be reached among the various rendering engines to display, for example, Myanmar.

The good news is in fact, among vendors at this point, the number of rendering engines that do display many of those writing systems is in fact going down. There's less and less because there tends to be some sort of harmonization of rendering engines across the industry at this point. So it's in fact less challenging than it used to be.

But that's the point. Basically, if you see that a sequence is never rendered or is rendered in a very inconsistent way by most of the platform, that's a very good sign that this sequence should not be allowed because users will not see a consistent rendering of the

sequence. So that's a major security list. So that goes beyond [spinning]. It's nothing to do with [spinning]. It's more to do with, can you render in your consistent way that sequence in all platforms that are used commonly in not just your area but at least in most of the rendering systems?

[BILL JOURIS]: Just to expand a little on what Michel said, I don't know whether it works in Burmese, but in the Latin script area, we have acronyms which are not pronounceable but used all the time, like BMW, the automobile company. You can't pronounce that, but it's a perfectly normal set of characters to have in a label.

YIN MAY OO: For us, acronyms are all consonants. We don't have diacritics most of the time. And every consonant is pronounceable alone.

[BILL JOURIS]: Then you don't have that problem.

UNIDENTIFIED FEMALE: Okay. So is there any questions? Comments? We don't have anything online, so does anybody want to have the last word?

Okay. Then thank you, everyone. We close the meeting. Thank you.

YIN MAY OO: Okay. Thank you, everyone.

[END OF TRANSCRIPTION]

ICANN
COMMUNITY FORUM
64
KOBE
9–14 March 2019