

Unicode Formation & Rules & Problems with Unicode Rules

submitted by Omi Azad

L2/03-233

1. Following problem is not well understood by Unicode.

"A consonant following (preceding) a consonant is represented by a modifier called consonant modifier if the shape of other character to which it is attached, remains unaltered. Among consonant modifier 𑄎 (ya-phala) is the most frequent one. This character is a presentation form of the character য় (Ya, U+09AF). It takes this shape depending upon the context. It could be applied to nominal form of any consonant or conjunct or even to the vowel.

Application of this character may either bring in repetitive sound or 'ya'-ness of the character it is applied to. Though it is a presentation form of an existing character, it should be treated like a separate character because of the two reasons. This character may appear along with a vowel where the conjunct formation rules cannot be applied. As for example:

অ + 𑄎 + ৳ = অ্যা

Apart from that, ৳ (Ra, U+09B0) and য় (Ya, U+09AF) combination has two different conjunction forms:

৳ + য = য়

৳ + য = ৳য়

Using the same conjunct rule both of them cannot be formed. If it is treated as a combination character, rather than a original form of Ya, both of these issues could be solved."

"In The following, technical issues related to few characters (including some special characters) are the separately elaborated:

Khanda-ta (঳)

It is a combining character, which is widely used in Bangla script. For some transliterated words it might appear at the beginning. Examples: হৎকম্পন, মহৎ, ঞ্জিলাড and নাৎসী"

2. Half form of each characters which the table shown below.

মূল বর্ণ	Pre Base	Above Base	Post Base	Below Base
ক	x	x		x
খ			x	

Unicode Formation & Rules & Problems with Unicode Rules

গ	x			
ঘ				
ঙ	x			
চ	x			
ছ				
জ	x			
ঝ				
ঞ				
ট				
ঠ				
ড				
ঢ				
ণ	x			x
ত	x			x
থ				✗
দ	x			✗
ধ				x
ন	x	x		x
প	x			
ফ			✗	
ব			x	x
ভ				x
ম	x	x	x	
য			x	
র	x	x	x	x
ল	x	x		x
শ	x	✗		
ষ				
স	x	x		
হ				
ড়				
ঢ়				

Unicode Formation & Rules & Problems with Unicode Rules

য				
---	--	--	--	--

3. Rule from Indic FAQ where did Unicode mistake with (ja-phala)?

Q: What are the Bengali characters used to transcribe the sound "a" (as in English "bat") in Unicode?

A: Ya-phalaa (pronounced "jo-phola" in Bengali) is a presentation form of U+09AF য় BENGALI LETTER YA. Represented by the sequence <U+09CD U+09AF> (য়) ya-phalaa has a special form য়̃. When combined with U+09BE া BENGALI VOWEL SIGN AA it is used for transcribing the English vowel "a" as in "bat". This ya-phalaa can be applied to initial vowels as well:

অ্য̃ = <0985, 09CD, 09AF, 09BE> (a- hasant ya -aa)

এ্য̃ = <098F, 09CD, 09AF, 09BE> (e- hasant ya -aa)

Note: Japhala character may appear along with a vowel where the conjunct formation rules cannot be applied.

If a candrabindu or other combining mark needs to be added in the sequence it comes at the end of the sequence. For example:

অ্য̣̃ = <0985, 09CD, 09AF, 09BE, 0981> (a- hasant ya -aa candrabindu)

Further examples:


Unicode Formation & Rules & Problems with Unicode Rules

অ + ্ + য + া = অ্যা

এ + ্ + য + া = এ্যা

ত + ্ + য + া = ত্যা

র + ্ + য + া = র্যা [ræ]

র + ্ +  + য + া = র্যা [rja:]

Note: Japhala character may appear along with a vowel where the conjunct formation rules cannot be applied.

Note: If Microsoft follows the same rule here then we wouldn't have the problem with Reph nor Japhala as single sequence.

4. One had a problem and cause trouble to understand linguistic and typographic.

I took some snapshot for a dictionary not from Bangladesh from Indian Dictionary where it also uses in Bangladesh. So, it solves for the problem that one has still doubt about forming linguistic, is important to have in the font.

The Author this Dictionary is A.T. Dev.

Following Images will have to understand why is needed in USP or in the font.

ক্রীড়াঙ্কল, *n.* Pretence or feigning of play or sport.
 ক্রীড়াঙ্কলে, *ad.* Playfully; sportingly; out of fun; in jest; as a joke.
 ক্রীড়াভূমি, ক্রীড়াঙ্গল, *n.* Playground.
 ক্রীড়াময়, *a.* Playful; sportive.
 ক্রীড়ারথ, *n.* Child's playing car.
 ক্রীড়াশিক্ষক, *n.* Games Master.
 ক্রীড়াশীল, ক্রীড়াসক্ত, *a.* Playful; sportive; fond of play or sport.
 ক্রীড়াস্থান—Same as ক্রীড়াভূমি।
 ক্রীড়োদ্যান, *n.* Pleasure-garden; pleasure-grove.
 ক্রীত, *a.* Bought; purchased.
 ক্রীতদাস, *n.* Slave; serf; bondsman.
 ক্রীতদাসী, *n.* Slave-woman; slave girl.
 ক্রীশ্চান, *n. & a.* Christian.
 ক্রুজার, *n.* Cruiser.
 ক্রুদ্ধ, *a.* Angry; enraged; wrathful; resentful; indignant.
 ক্রুশ, ক্রুস, *n.* cross. ক্রুশে বিম্ব করা—Crucify.
 ক্রুশ-কাঠি, ক্রুশ-কাঠি, *n.* Crochet.
 ক্রুশবিম্ব, *a.* Crucified. ক্রুশবিম্ব করা—Crucify.
 ক্রুশাকার, *a.* Cruciform.
 ক্রুর, *a.* Cruel; unfeeling; pitiless; hard-hearted; hard; harsh; mischievous; crooked-minded; severe pitiless; cruel; fierce.
 ক্রুরকর্মা, *a.* Relentless; unrelenting.
 ক্রুরতা, *n.* Cruelty; brutality.
 ক্রুরমতি, ক্রুরমনা, *a.* Evil-minded; malicious; cruel; mischievous.
 ক্রেডিট, *n.* Credit.
 ক্রেতব্য, *a.* Purchasable; worth buying or purchaser; buyable.
 ক্রেতা, *n.* Purchaser; buyer; vendee; customer. to be bought.
 ক্রেয়, *a.* Purchasable; worth buying.
 ক্রোক, *n.* Distraint of goods; attachment (মাল ক্রোক—attachment of movables). ক্রোক করা—Seize; distraint; attach.

Unicode Formation & Rules & Problems with Unicode Rules

প্রীতিপ্রফুল্ল, a. Beaming with delight.
প্রীতিভরে, ad. Very affectionately;
out of affection.
প্রীতিভাজন-Same as প্রীতিপাত্র।
প্রীতিভোজ, n. Dinner-party; tea-
party; picnic.
প্রীতিসন্তাষণ, n. Friendly or corial
greeting.
প্রীয়মাণ, a. Satisfied; gratified.
প্রুফ-পেষকার, n. Proof Pressman.
প্রুফ-ফর্মাধাবক, n. Forme Washer.
প্রুফ-শোধক, n. Reader (W. B. G. Pr-
ess).
প্রুফ-সংশোধন, n. Proof-reading; pro-
of-correction.
প্রুষ্ট, a. Burnt.
প্রেক, পেরেক, n. Nail; spike; pin.
প্রেক্ষক, n. Observer; spectator

শুভকর, শুভংকর, শুভঙ্কর, শুভদ, শুভদায়ক,
a. Auspicious; beneficial; favou-
rable; lucky; good.
শুভকর্ম, শুভকার্য, n. Auspicious action;
religious action; auspicious cere-
mony.

Unicode Formation & Rules & Problems with Unicode Rules

This images will show that we need both of them the linguistic and typographical. I will suggest to Paul that not to change this leave it they way it is in the USP beta version.

5. To solve problem of the following rule.

I mention Paul before and I am mentioning again to him to follow the below.

ঝ + ◌ + anything = ঞ

Ra + Hasanta + Anything = Reph

anything + ◌ + ঞ = ঞ

Anything + Hasanta + Ya = Yaphala

ঝ + ◌ + anything = ঞ

Ba + Hasanta + Anything = Baphala

anything + ◌ + ঞ = ঞ

Anything + Hasanta + Ra = Raphala

Why we need this?

If one wants to write book and demonstrate this formation how it works so that reason, it must support it.

6. Why not support in Bengali Script in the USP other featuring?

Unicode Formation & Rules & Problems with Unicode Rules

I have found in different document have different type of glyph what I mean by one glyph might have different ornamental way to represent. So, I went to Microsoft Typography where I have read about the feature of ornament and so and others which will help to make the font more productive and efficiently for user and whoever makes books for people to learn language.

I want to tell Paul & others about this document as consideration if there is any mistake any kind of problem which is unclear to anyone inform me by E-mail solaiman_k@hotmail.com.