

Minutes of the meeting

NBGP Bangla meeting

IHC, New Delhi

December 21-22, 2018

Names of participants (in meeting room):

1. Prof Udaya Narayana Singh
2. Mr Samiran Gupta
3. Mr Mamum Or Rashid
4. Mr Haseeb
5. Mr Abhijit Dutta
6. Dr Atiur Rahman Khan
7. Mr Akshat Joshi

Distant participants (online):

1. Dr Ajay Data
2. Dr Sarmad
3. Ms Pitinan K
4. Mr Rajib Chakraborty
5. Mr Rajiv Gupta
6. Ms Neha Gupta

Day 1

Introduction: Co-chair Prof. Udaya Narayana Singh inaugurated the day's session with his introductory speech describing the NBGP Bangla team's activities and responsibilities with respect to the Bangla LGR and the discussions in line with the issues that were raised in the last meeting in Dhaka and the subsequent online and email meetings.

Proceedings:

Dr. Atiur Rahman Khan, C-DAC Pune listed the 11 points accumulated from previous meetings with regard to Bangla LGR. Every point was taken one by one subsequently answered and discussed by other members present physically in the meeting room and also in remote attendance.

Mr. Haseeb from Bangladesh added another point of *ajji* and Bengali letter *Li* (U+098C) bringing the total number to twelve.

1. Point 1: The correct name for the script in question (Bengali vs. Bangla).

Prof. Singh suggested to the Unicode consortium in changing the name of the script in question to Bangla. However, the work of LGR should not be delayed or restrained due to the changing process.

NBGP Bangla may also establish contact with Sierra Leone (West Africa) in terms of LGR document as Bangla enjoys the status of official language in this African nation. Mr. Haseeb

added that Bangladesh may officially take this responsibility of contacting Sierra Leone in this regard so that their say in the LGR is also taken.

2. Technical and other information with regard to census statistics, relation with Brahmi-Devanagari script etc.

Latest census documents and other relevant authorized sources to be referred to for such details and information.

3. Issues regarding term/ alternative names for vowel markings on consonant ('-kar' vs. 'Matra'); 'hasanta' vs. the term 'halanta'

It was decided that the Bangla or native terms such as kar and hashanta be used in brackets wherever the technical terms have been used in the LGR document.

4. Incorporation of activities and contribution of Dhaka Bangla Academy.

Mr. Mamun and Mr. Hasib were asked to provide information to be added in the LGR with respect to various activities in connection with Bangla and Bangla script in particular.

5. Nukta character in the Bangla repertoire.

The panel members unanimously agreed that the term and the concept of *nukta* is borrowed in Bangla. Dr. Atiur noted that can *nukta* be used for demonstrating the characteristics sounds of loan words from Persian and Arabic and for that matter other foreign languages. There could be cases of demand to write Muslim names, Urdu poetic words and loan words with *nukta* under ক, খ, গ, জ and ঝ only for the sake of correct pronunciation and maintaining the sanctity of the loan word. I was however, decided to do away with the *nukta* in Bangla. Mr. Rajib pointed out that ড, ঢ and ঝ are different from ক, খ, গ, জ and ঝ. Mr. Haseeb was of the opinion that there is no need of *nukta* in Bangla. Mr. Abhijit proposed not to remove *nukta* if absolutely necessary to do so. Mr. Akshat pointed out the possibility what if singularly disown *nukta* character as atomic characters (ড, ঢ and ঝ) are not normalized in Unicode. There could be problem in domain names and that they are broken in Unicode. Mr. Akshat further stated that ড, ঢ and ঝ could be proposed as atomic characters to the Unicode. The panel agreed to this concept referring to the fact that ঝ has originally been incorporated as a single character without *nukta* and never as ঝ+*nukta*. Hence, the case of making ড, ঢ and ঝ as atomic characters in the repertoire could be easily justified. The co-chair, Prof. U. N. Singh suggested two ways that either the *nukta* may not be used at all or put up formal note to Unicode and see it is also normalized in a way they look like ঝ (atomic).

Dr Sarmad, through online medium, made the panel aware that the LGR procedure suggests that if there is no consensus in the community, the default is to leave the relevant code points out, until a time when there is a unanimous agreement. He further added that it is not easy to get the recommendations through Unicode.

Prof. Singh finally stated that the *nukta* may not be allowed whatever be the consequences.

There were a couple of comments from the Bangladesh minister of Information Technology which he conveyed via Mr. Haseeb that *nukta* characters such as ক্, ঞ্, ঙ্, ঙ্ and ঞ্ may not be allowed. The three essential characters (ঊ, ঋ and ঌ) will not be used in the Top level domains (TLDs). This decision anticipates normalization which when done the NBGP team will reconvene and include them back into the Bangla repertoire.

6. Additions of numerals in Bangla repertoire:

As a standard practice and norms of the MSR-3 no numerals are to be used for any language in the TLDs due to cyber security. Therefore, it may not be allowed for Bangla as well.

7. Use of native terms for technical terminologies.

It was discussed and decided that Bangla names such as *onushshar* for anusvara, *hashanta* for halanta etc. may be used in the document with IPA equivalents within brackets.

8. Use of Avagraha in the LGR.

The NBGP Bangla group recollected Dr Swarochis Sarkar's comment on the retention of Avagraha which was discussed in the Dhaka meet in June 2018. However, Avagraha is blocked in TLDs but used in Sanskrit, Pali and Maithili texts. Perhaps can be used in the second level domain.

9. In-script and cross-script variants.

Discussed in Day 2 proceedings.

10. Relevance of Unicode character code U+9C4

The Bengali vowel sign Vocalic RR was unanimously decided to be not relevant in the context of modern Bangla writing system. The Indian side as well as Mr. Mamun were of the opinion that it may be excluded.

11. *Anji* and *Li* (U+098C) characters in the LGR.

Unicode character U+0983 and *Anji* symbol be blocked. Mr. Mamun suggested to exclude the character *Li* from the Bangla repertoire in the LGR.

Day-2:

Introduction: Co-chair Prof. Udaya Narayana Singh opened the day's session by recapitulating the previous day's discussions on the 11 major points gathered from last meetings.

The second day of the NBGP Bangla was mostly dedicated to the discussions on *nukta* and in-script and cross-script variants with regard to the perspectives and opinions of Mr. Mamun Or Rashid and Mr. Haseeb from Bangladesh. Other points discussed were *Visarga* and *nukta*.

1. Comparison of Devanagari and Bangla *Visarga*

Both are different and distinct as far as the shape of the character is concerned. The Bangla *Visarga* (*Bishargo*) is more like a circled character unlike the two dots of Devanagari *Visarga*.

2. In-script variants

The following were the issues pertaining to in-script variants pointed out by Mr Mamun:

চ্ছ চ্ছব চ্ছর,	দধ দ্ব	ট্ট ট্র	মত্ৰ মত্রি
----------------	--------	---------	------------

As far as মত্ৰ and মত্রি is concerned the pair could be kept as confusables. The same is with the pair ট্ট and ট্র.

These two characters may be added as confusables only in the LGR document.

3. Cross-script Variants

The following were the issues pertaining to in-script variants pointed out by Mr Mamun:

Devanagari Vs Bangla	Odia Vs Bangla	Gurumukhi Vs Bangla	Gujarati Vs bangla
ज (ङ) Vs अ न (त) Vs न ल (ल) Vs ल २ (२) Vs २	ଏ (এ) Vs এ ଓ (ও) Vs ও ସ (স) Vs স	घ (ब), घ (थ), घ (थ) Vs घ १ (१) Vs १ ४ (४) Vs ४	૧ (૧) Vs ૧ ૨ (૨) Vs ૨

Devanagari घ [gha] and Bangla ঘ [gha] be added as confusables.

With respect to **Odia** script the [ସ sa] and [ঘ gha] and Odia [ଏ e] and Bangla [এ e] may be ruled out as variants. Odia [ଓ o] and Bangla [ও o] may be treated as confusables. Mr. Mamun concluded that it is very important that the issue of variant meticulously and take people's perception which is very important. Moreover, with respect to other scripts Mr. Mamun suggested that we must have people's opinion on this issue.

With respect to comparison with **Gurmukhi** and **Gujarati** scripts the cases pointed out by Mr Mamun were discussed and decided. The numbers are already disallowed by MSR hence there is not question to allow them in the repertoire.

4. Nukta

Mr Mamun Or Rashid suggested the exclusion of *nukta* from the Bangla repertoire in the LGR document. He elaborated that when single stroke typing is already provided by the Unicode then there why the use of ଢ † and ଢ †. Mr. Akshat elaborately described how *nukta* is relevant to the Bangla repertoire keeping in mind the limitations given by MSR. Pink colour debar the respective characters from being used in the TDLs. There is only yellow code available for us to use. The IDN protocol says that it cannot have these as atomic forms. What we have to do

is to change the recommendation for normalization at the Unicode and say to keep the atomic forms and then change. ৩ and ৪ are in the white code point hence, not usable. So in TDL these two characters are not allowed. Moreover, he concluded by saying that the problem is that our three characters have not come from the Unicode to IDNA. So the problem lies with the Unicode level.

Mr Mamun further said that the NBGP can propose from this forum to add ৩, ৪ and ঞ as atomic characters and do away with the issue of typing these with dot (*nukta*).

Prof. U.N.Singh reiterated that the panel along with its arguments may convince the Integration Panel (IP). The IP take sit up with IDNA and the Unicode. This could be a smooth, but long, way of solving the problem. Prof. Singh said blocking will result in disenfranchising the Bangla community to appear on the top level domains. Therefore, holding LGR activity was not a wise effort and such an important decision may not be taken at once. The co-chair recommended gathering opinion with special respect to having feedback from Bangladesh in particular.

Dr Ajay Data emphasized that the Bangla LGR must go and there was no need or question of holding or delaying the task. Efforts may be made to make a note of this disagreements and attach with the proposal.

In short, all of the panel supported the deletion of nukta. All agreed on approach the MSR. The panel also reiterated that the inclusion of ৩, ৪ and ঞ in the primary body of the consonants and not perceive as additional consonants. At the same time this issue could be floated on the public forum for further steps. The public comments, as a consequence, will create pressure and thus endeavour to facilitate the task of change at the IDNA and Unicode level.

Action Item:

The Bangladeshi experts agreed to get back with their comment on these points in two weeks, preferably by Jan 15 2019.