

Subject: Malayalam Feedback from the Online Reporting Form

Date: January 29, 2008

The following feedback on specific issues of Malayalam for Unicode 5.1.0 was received on the reporting form Jan 27 - 28, 2008. I have separated this from the other public feedback on various issues because of its total length and to assist in the South Asian Subcommittee discussions.

Rick

Date/Time: Sun Jan 27 19:33:47 CST 2008
Contact: pravi.a@gmail.com
Name: Praveen A
Report Type: Feedback on an Encoding Proposal
Opt Subject: Code points for chillus is unacceptable.

The Atomic chillu's are unacceptable because it destroys the link of a chillu with its base character.

1. The examples used to justify semantic difference between words only separated by zwj are non-existent in dictionary or are grammatically wrong or meaningless without proper context.

a) വൻയവനിക/വന്യവനിക (vanYavanika/vanyavanika), കൺവലയം/കണ്ഡലയം (kanvalayam/kanualayam) ... contrived examples not found in dictionary

b) ആ മനുഷ്യൻ കൊടുക്കുന്നു (that man is giving)
ആ മനുഷ്യന് കൊടുക്കുന്നു (giving to the man)

as per malayalam linguistic rules the sentence is a mistake.
it will be completed if and only if you need to write it as following.

Structure:

ആ മനുഷ്യൻ <to whom & what he is gives> കൊടുക്കുന്നു
ആ മനുഷ്യന് <who is giving & what is giving> കൊടുക്കുന്നു.

Example:

ആ മനുഷ്യൻ (man) പൂച്ചക്ക് (to cat) പാൽ(milk) കൊടുക്കുന്നു (That man is giving milk to cat)
ആ മനുഷ്യന് (to man) പൂച്ച (cat) പാൽ (milk) കൊടുക്കുന്നു. (That cat is giving milk to man) :-)

Fundamental problem lies here in the unicode's way of treating only representational forms without checking linguistic correctness. Most of the indic languages are unlike latin and collations are based on linguistic base. If you are not considering it, it will become a play yard of people with vested interests

2. All these arguments were once considered and rejected by UTC and the only new argument in support of atomic chillus is the issue of missing domain names in IDN. The examples given in 1) can't be considered real as these are contrived just to make a case for atomic chillus. Even if were real it is similar to case folding in Latin (You can't register two sites PenIsland.com and PenisLand.com). How can already rejected proposal be accepted when the new arguments in supports is not only proved to be real, but creates a lot of new chaos and security problems.

3. This will create dual encoding and makes URL spoofing very easy.

റാൽമീനോവ്.blogspot.com (using chillu joiner sequence)

റാ® മീനോവ്.blogspot.com (using atomic chillu)

because both of these have different punicode. The existing chillu encoding with joiners is best solution because all of the combinations of joiners and non-joiners give exactly same punicode.

4. Since the joiners has to be supported for backward compatibility it creates unnecessary complexity to all text processing application (sorting, searching) and it makes atomic chillus redundant and useless.

5. Why isn't a canonical equivalence to old sequences not provided?

6. Even after atomic chillus are made part of the standard many words cannot be written without joiners and it would be increasing the chaos.

കൊയ്‌രാള (koirala), സദ്‌വാരം (sadvaram)

7. Using virama with chillus is linguistically incorrect (function of virama is to create vowel-less and you can't use it with a chillu or pure consonant because these are already vowel-less forms of the underlying consonants)

I strongly oppose including this characters in the standard as it not only fail to solve all the problems with joiner it creates lots of new problems and the need for providing backward compatibility will produce more chaos in encoding chillus.

Praveen A
Swathanthara Malayalam Computing
www.smc.org.in

(End of Report)

Date/Time: Mon Jan 28 05:40:32 CST 2008
Contact: jinesh.k@gmail.com
Name: Jinesh K J
Report Type: Feedback on an Encoding Proposal
Opt Subject: reg:[the issues related to the encoding of chillus and making conjuncts using chillus]

The proposals to give separate code points to the 'chillu' characters in Malayalam is supposed to tackle some problems like

1. Problems in the change of meaning in text just due to presence of ignorable ZWJ in the present accepted and standard sequence.

-> After encoding chillu as a separate entity, this problem is not at all solved(just think of the sequences which are produced with ZWNJ in it!). A solution which never solves a problem and arises a lot of other problem like, continuing support to currently accepted sequences just gives a chance for dual encoding and more complexity to all process related to language processing(using grammar rules its comparatively easy to identify sequence of words process them, but when there is chance for same letter can be represented in different ways, even able to form conjuncts which is always against the rules of language, the language is changed for sake of computation).

2. Problems in using some Malayalam sequences registering domain names for websites.

-> Adding atomic chillu never solves the problem, but increases by giving a change for spoofing of the websites. Currently when we use a chillu, it is represented by ZWJ and this sequence will be supported for ever as a part of backward compatibility so, one who tries to register a website using the new chillu should also register the old sequence as the display with ZWJ and with atomic chillu will be same but they will go to different sites.

3. The specification of standards for making conjuncts in a language.

-> As far as i know, this is the first time unicode consortium is taking the role of specifying how to create different conjuncts and sequences. This it self shows the confusion that will arise with the inclusion of the chillu in the base character set. Also, in the tables provided, the existing encoding and sequence formation is not at all considered. Then I don't know how the data written in current encoding is gonna be interpreted by system which follows new standard(losing backward compatibility). Or if they are ready to accept those sequences also, then what is the need for such new sequence specification(giving some specific job for newly encoded chillus?). I understands the problem is a single ZWJ makes the spelling difference(Also these documents talks nothing about the problems created by missing a single ZWNJ).

4. Conjunct making chillus!

-> Chillus are half forms of a consonant(actually the half sound of a consonant). By the language definition it will not make any conjunct. There is no validation for the argument and specification to use chillus to form a dot reph. This proposal should be reconsidered.

Finally, what i have to say is, to use the current accepted standard(some says there is no such spec, first i don't know who is supposed to specify all spec on all world languages UNICODE? I think he was joking). It is not creating any problem so far and just invalidating that sequence simply and thinking of encoding all characters making problem is not the right way i think. Also some solutions which actually doesn't solve the problem is also something funny. Then finally, making some new rules for a language, which was not there for the last many many years of existence of language(even when malayalam lipi was reformed to fit in type writers, i think they consulted people in all fronts of the society and summarized their opinions, here the sequence of conjunct forming chillus is just coming after a discussion in unicode list, where i haven't seen many linguistic specialists from Kerala). As Malayalam is my mother tongue, i have concerns about the encoding and the consultations the consortium made before trying to specify some clear cut specifications for the representation of characters in a language.

(End of Report)

Date/Time: Sun Jan 27 13:17:17 CST 2008
Contact: mangalat@yahoo.com
Name: Dr.Mahesh Mangalat
Report Type: Feedback on an Encoding Proposal
Opt Subject: Faulty Encoding Proposal for Malayalam

The proposed additional code points for Malayalam have been a point of debate in Indic list and many other forums. It has been argued that the proposals would result in dual encoding and linguistically incorrect encoding. More over it will isolate Malayalam from other Indic Languages, which had enjoyed comfortable position with regard to transliteration to other Indic Languages. ISCII encoding for Indic languages has been the foundation for all Indic languages in Unicode. Malayalam will be isolated form the rubric of Indic languages due to the proposed code points.

The debates, so far taken place in any one of the forums, have never proved the merits of atomic encoding of Chillu characters. In spite of that, a lobby has been arguing for this, for reasons best known to them. I write this to bring to your notice, as an academic from the field of Malayalam, who has been closely monitoring all these debates.

I do not hesitate to attest that those who argue for a deviation from the present encoding for Malayalam has not been there, when the code points were initially worked out, based on the input from the learned sources of ISCII. Those who come forward with the arguments have proved their incompetence in language and technology, and have failed to support their arguments with any valid rationale. Please refer to the comments of James Kass in the Indic list to see how absurd their arguments are.

The governmental committees, who come in disguise to support this argument, consist of people who never use either Computers or not Malayalam on computers. They are people who have been there by virtue of their position that they hold in Universities or in some other organisations. This will be a matter that the Unicode authorities would not be able to understand in their Euro-American setting. These specialists in government committees do not have a competence to differentiate between Encoding and Keyboard settings. Please see that the governmental proposals are from Encoding and Keyboard Standardisation Committee of Kerala Government.

Let me inform you about a movement in Malayalam Computing, named as Rachana Aksharavedi that came forward to save Malayalam computing from the idiosyncrasies of this group promoted by the establishment. Rachana Aksharavedi was responsible for solely restoring the original character set of Malayalam language in computing, fighting with this coterie. The lobby which works against this movement, happens to be a governmental committee, which has always been proposing ridiculous arguments which are linguistically and technically wrong. If you go to the genesis of the Chillu encoding argumet, you would see that, it was complimentary to a proposal, which argued for removal of ഋ along with some other characters from the code page for Malayalam that was summarily rejected by UTC.

Let me conclude this submission with this request to summarily dismiss the proposal for atomic encoding of Chillu characters. Had that been ever encoded,Unicode Consortium will loose its credibility in Malayalam user community which is growing in a fast phase, by the usage of their mother tongue on computers, as a part their school education. This is an ever enlarging community.

Finally let me also remind you that the so called specialists, who have a proven record of their incompetence with regard to language and technology, who spoke in support of Chillu encoding has been citing words that has never been present in any of the dictionaries or any other printed materials. The discussions so far taken place in Indic list are not known to the user community. I am sure that once they happen to know about it, they would respond it in a befitting manner. I hope that the Unicode Consortium would take a decision in this matter in a responsible manner.

I am only happy to expand any one of the points that I put forward here.

(End of Report)

Date/Time: Sat Jan 26 15:33:06 CST 2008
Contact: naa.ganesan@gmail.com
Name: Dr. Naga Ganesan
Report Type: Other Question, Problem, or Feedback
Opt Subject: Table 2 Malayalam chillu sequences in Unicode 5.1

Table 2 in
www.unicode.org/versions/Unicode5.1.0/#Significant_Character_Additions

Compare Tables 1 and 2. Obviously when chillus occur we need to use the chillu sequence from Table 1 in Table 2. <0D28, 0D4D> should change to <0D28, 0D4D, 200D> zwj (U+200D) is missing in Rows 1, 2, 4a, 5, 7a and 8a in Table 2..

So, please add a zwj whenever chillu n sequences are used in Table 2 as this will be consistent with chillu n sequence in Table 1.

Major fonts in use by Malayalee community such as Microsoft Kartika, and Rachana fonts have chillu sequences <cons, virama, zwj> throughout. Earlier bug in Rachana font creating chillus without zwj has been corrected by the developer (see Note).

N. Ganesan

Note:

Cibu Johny wrote on 26/January/2008 in Indic list:
>It is a fact that 5.0 is not at all clear about chillus.
>We could say it is hardly defined. Because of that many
>implentations took different routes. For example, the chillus
>appearing in the middle of a word is not defined. So in
>GPLed Rachana font, if there is no conjunct for
><consonant1, virama, consonant2> and consonant1 has a chillu
>form, it takes the chillu form. However, Kartika consonant1
>will form chillu only if zwj is specified as
><consonant1, virama, zwj, consonant2>.

Praveen wrote a reply:

"Rachana fonts use zwj for chillus. An earlier version did not use it, it was a bug (the developer never wanted it that way) and it was fixed long ago."

This is very important information. Now that we know Rachana fonts use ZWJ wherever Chillu sequences are needed. So, Cibu Johny's documents of the last 6 months or so, showing "Rachana interpretation" of Chillus as <cons, virama> only need to be updated as <cons, virama, zwj>, just as Microsoft Kartika font.

Table 2 in

www.unicode.org/versions/Unicode5.1.0/#Significant_Character_Additions

Are all these are attested? or, many are just *constructed* with no attestations?

(End of Report)

Date/Time: Mon Jan 28 02:01:31 CST 2008
Contact: hiran.v@gmail.com
Name: Hiran Venugopalan
Report Type: Feedback on an Encoding Proposal
Opt Subject: Malayalam Unicode Issues

1.Introduction of atomic glyphs will surely make spoofing. The detailed description are on
pravi.livejournal.com/19722.html
santhoshspeaking.blogspot.com/2008/01/blog-post_24.html

2.In the [www.unicode.org/versions/Unicode5.1.0/#Significant_Character_Additions] Unicode 5.1.0 description Table 2. NTA Cluster the first two rows defines the combination of 0D28 + 0D4D + 0D31 in two ways. Its actually the same character which

the language allows two ways for its writing and hence it is the font designer whom decided the usage. Personally I prefer the first as there is a chance for visual misunderstanding as in the 4th row. The example given in the same table row 8.b is incorrect. In the Table 3. RRA - RRA Clusters the same issue exists.

3.The Table 4a. Dot Reph RA + YA, suggests a virama between the CHILLU RR and YA. for usage of dot reph RA. The fact is that the row 1 and 3 are the same. Usage of two encoding for the same character will make spoofing and also will make issues on sorting. Same for the row 1 and 4 of Table 4b. Dot Reph RA+VA. Table 4c. Dot Reph - Rest of the Consonants suggestion will also introduce spoofing.

(End of Report)

Date/Time: Mon Jan 28 03:01:22 CST 2008
Contact: aashiks@gmail.com
Name: Ashik Salahudeen
Report Type: Feedback on an Encoding Proposal
Opt Subject: About inclusion of atomic chillu codepoints for Malayalam in Unicode.

Hello

I am a Malayalee by birth (one whose native place is Kerala and mother tongue is Malayalam) and software engineer by profession. I have been speaking , reading and writing Malyalam all my life.

I recently came to know about the proposal to include separate code points for chillu's in Unicode Standards. I became a part of a group of people, called Swathanthra Malayalam Computing . I read about the current state of unicode proposals for malayalam and read about the decision to give separate codepoints for atomic chillu in Unicode.

I don't see any compelling need for including these characters in Unicode. They are different forms of characters that are already present in Unicode. I cannot comprehend the issues and problems that this particular decision (inclusion of atomic chillus) will solve. I think that instead of solving anything , it creates more confusion and chaos, primary concerns being things like URL spoofing and dual encoding which have been discussed on the Indic lists . The sheer volume of existing contents alone proves that there is no need at all for adding anything to the Unicode Standards .

The arguments that are introduced in favor of atomic chillus are plain idiotic - there is not even a single grammatically correct usage in them. And I have never seen/heard any of those in actual usage. I have talked to people who has studied malayalam about this , i

talked to people who regularly writes articles about this. None of them sees any logic in the arguments put forth in favor of atomic chillus.

I have read the indic@unicode.org mailing lists and i am convinced that the move to include atomic chillu characters would be a bad decision. Our language has been torn apart once with the advent of typewriters. We are gradually recovering from it with the help of computers. If this decision goes through , it would be a blow to a movement that is currently on the right track. And it will create a lot of unnecessary confusion.

What surprises me is that even as the Unicode standards are being formulated , there has been no public announcement of so serious an event - Noted academicians and writers from our language has not been consulted about this. They do not know what unicode is - much less the issue of atomic chillus. What we need is time , to make our language experts understand what this is all about , and then draft further proposals with their help . This hasn't been done yet.

In spite of all this , if the Unicode committee decides to go ahead with the proposal , it would be willful murdering of a language which has produced outstanding literature and cinema in India. (As a side note , India has been put on the map of world cinema several times by Malayalam films , notably those by Adoor Gopalakrishnan).

Thank you

(End of Report)

Date/Time: Mon Jan 28 03:10:52 CST 2008
Contact: santhosh.thottingal@gmail.com
Name: Santhosh Thottingal
Report Type: Feedback on an Encoding Proposal
Opt Subject: New Malayalam Characters in Unicode 5.1

Hi,

I would like to comment on the new chillu characters in unicode 5.1

First of all let me question why one encoding proposal which got rejected once is again taken into consideration, even though the subject remains controversial and there was no consensus in Indic mailing list. If anybody from UTC read the discussion archive read the discussion will came to understand that encoding new codes does not have any advantage other than breaking the existing logical stability of Malayalam language.

Currently there are 5000+ articles in Malayalam wikipedia and thousands of Malayalam blogs. There are many Operating system interface localization and there are many linguistic softwares in malayalam. All of them are working fine without any problem with chillus.

All of the arguments presented in the chillu encoding were proved false in the discussion.

The 5.1 document does not say anything about the backward compatibility or about supporting the existing sequences. If we want to support the existing sequence as per the stability policy, then the canonical equivalence has to be defined. That did not happen. So what will happen to all of the Malayalam Unicode text created so far? All these writers are fools?

Everybody knows that there was no consensus reached in the discussion in indic@unicode.org mailing list and still the problem is controversial. Another thing is even though the new changes will have a major impact on the language technology, the linguistics and language experts in Malayalam is not at all aware of the facts. We doubt that language experts/authorities accepted by the public were given an explanation of what Unicode is and what the atomic chillu proposal is about. Only some ivory-tower discussions among some academicians were carried out and even those have reached the conclusions that there is no particular necessity for atomic chillus. Even among the IT literate Malayalees (people who use Malayalam on a regular basis) only a handful know the Unicode representation of Malayalam and issues surrounding it. So any hurry on adding new code points will, in our opinion, be ill-informed and will have a bad impact on the future of the language.

Unlike Latin languages the language and its letters are highly complex and it is related to linguistics. A mere consideration of letters will not give any picture of the language and if we make any decision based on that, it is going to destroy the language's root.

So, please don't hurry to encode the chillus, until there is a consensus or at least majority of people supports it.

Please don't kill a language again which has undergone many murder attempts

(End of Report)

Date/Time: Mon Jan 28 06:26:43 CST 2008
Contact: sulochana@cdactvm.in
Name: K.G.Sulochana
Report Type: Public Review Issue
Opt Subject: Malayalam text in Unicode 5.1.0

I have specific comments on the item numbers 4 and 5 in tables 4a and 4b (dot rep of RA+YA and RA+VA) in the Malayalam related text in Unicode 5.1.0

The dot reph forms for Malayalam need not be listed in the text of Unicode 5.1.0.

Reasons:

1. The authors themselves admit that there are problems with both the proposals.
2. We have alternate unambiguous forms to represent these sequences
3. These forms are obsolete and nobody uses it now a days.
4. The script reform committee of the Govt. of Kerala have recomended discontinuing the use of dot reph in Malayalam. Please see the attached Govt. order. The recommendation part is in Malayalam. Page 3, top sentence says that the committee recommends discontinuing the use of dot reph.

(End of Report)

Date/Time: Mon Jan 28 20:45:12 CST 2008
Contact: naa.ganesan@gmail.com
Name: Naga Ganesan
Report Type: Other Question, Problem, or Feedback
Opt Subject: Comments on Unicode 5.1 New Sequences for Malayalam Reph

Comments on Unicode 5.1 New Sequences for Malayalam Reph
(Tables 4-a, 4-b and 4-c)

>The Unicode 5.1 (draft) web page has recently been updated with new text
>on Malayalam additions. The most important new characters for Malayalam are
>the six new chillu characters, U+0D7A - U+0D7F. There are a number of
>important issues in this new text, including some where feedback is
>requested.

>Please see:
>www.unicode.org/versions/Unicode5.1.0/#Significant_Character_Additions

Nowadays, reph forms are obsolete, only in old-style orthography
Malayalam script needs the rephs.

Even in the old books printed prior to 1950 or so,
the dot-reph over just YA or just VA is extremely rare.
For those dot-reph over YA or VA can be represented
as <RA, VIRAMA, ZWJ, YA> and <RA, VIRAMA, ZWJ, VA>
respectively. The fall-back in fonts with no dot-reph

possibility in the dot-reph over YA and VA is that <RA, VIRAMA, ZWJ> will be simply Chillu RR/R. This is admissible by the Kerala Government Order available at www.malayalamresourcecentre.org/Mrc/order.pdf which states that Chillu RR/R can indeed replace dot-reph. Of course, the current sequence for Chillu RR/R is only <RA, VIRAMA, ZWJ> and will be available for backward compatibility uses.

So, in fonts having reph capability (now, Malayalam reph is obsolete) <ra, virama, zwj> can show the reph. In fonts without reph capability, the fall back has to happen. reph will be replaced with a chillu R which is OK as per Kerala G.O. on how to treat reph in Reformed script.

Pl. see:
indology2.googlepages.com/Malayalam_reph_with_zwj.pdf

N. Ganesan

PS:
For any reason, this sequence is a problem, an alternate is <RA, ZWJ, VIRAMA, YA> and <RA, ZWJ, VIRAMA, VA> for the extremely rare dot-reph ya or va.

(End of Report)