

Сэдэв: Монголын авиа зүйн загварыг сайжруулах санал ба саналын хураангуй

Title: Summary of suggestions and proposals for improvements to the Mongolian phonetic mode

Зохиогч: Рүүзбэх Парнадер (WhatsApp/Facebook)

Author: Roozbeh Pournader (WhatsApp/Facebook)

Огноо: 2018 оны 4 сарын 5 өдөр

Date: 2018-04-05

Энэхүү баримт бичигт АНУ-н Калифорни муж Сан Хосе хотноо 2018 оны 4-р сард Монгол бичгийн ажлын хэсгийн 2-р хуралдаан дээр гарсан авиа зүйн загвартай холбоотой бүхий л саналыг товчоор хавсаргав. Оролцогчид, мэргэжилтэн, шинжээчдийн хувьд хамгийн чухал хэлэгдсэн зүйлсийг тодорхойлж, тодруулсан болно.

This document briefly lists all the proposals related to the phonetic model raised at the Mongolian Working Group 2 meeting in April 2018 in San José, California. Higher priorities for the participating experts are specified and highlighted.

С.Бадралын зөвлөгөө ба санал хүсэлт (Үндсэн баримт бичиг MWG / 2-N3R).

Proposals and suggestions from Badral Sanlig (main document MWG/2-N3R)

- МОНГОЛ ҮСЭГ ХА ба МОНГОЛ ҮСЭГ ГА-ийн эм хэлбэрүүд нь тус тусдаа нэг үсэг болгож, тусдаа тэмдэглэгээтэй байх ёстой, эсвэл FVS1-тэй тогтмол тохируулгатай нэг хэлбэр байх ёстой. (Онцлох зүйл 1.)

The feminine forms of MONGOLIAN LETTER QA and MONGOLIAN LETTER GA either should be separately encoded as separate characters, or as fixed variants with FVS1. (Priority 1.)

- Бид NNBSР тэмдэгтийг аль болох хурдан засах эсвэл монголын шинэ дагавар холбогчоор кодчилох хэрэгтэй. (Онцлох зүйл 2.)

We should fix the NNBSР character as soon as possible or encode a new suffix connector for Mongolian. (Priority 2.)

- Монгол блок дахь загварлаг болон түүхэн тэмдэгтүүд, хувилбарын дараалалуудыг цэвэрлэж, Монголын блок дахь Али Галын хэсэгт дахин зохион байгуулж болно. (Онцлох зүйл 3.)

Stylistic and historical characters and variation sequences in the Mongolian block need to be cleaned up and potentially reorganized into the Ali Gali section of the Mongolian block. (Priority 3.)

- Монголын үсгийн фонтуудын дүрмийг (OTF) хялбаршуулан стандартчилан олон нийтэд ил тод болгон мэдээлэх хэрэгтэй.

Font rules for Mongolian fonts (OTF) need to become simplified and standardized and made public.

- Монголын скриптүүдийн олон янзын боловсоруулсан шийдлүүд нь нэгдэн нийлүүлэх буюу нэгтгэх шаардлагатай байна.

Different implementations of Mongolian script need to converge and get unified.

- Текст рэндэрлэх хөдөлгүүрүүдэд монгол хэл дээр рэндэрлэх дүрмийг жишиг болгон оруулах хэрэгтэй.

Rendering rules for Mongolian in text rendering engines should be standardized.

- Байрлалын нийцэмжгүй байдлууд нь Монгол хэлний хувилбарын дараалал болон Юникод анхдагч хувилбар болон тэдгээрийн бодит товчилсон нийлбэр хувилбаруудыг хооронд нь холбох хэрэгтэй.

Positional mismatches, between the name of Mongolian variation sequences and default forms in Unicode and their actual cursive joining positional variants, need to be fixed.

- Чөлөөт хувилбарыг сонгох товчлуурын тоог дор хаяж нэг болгож ядаж оруулгад оруулах.
- Хэрэглэгчдийг чөлөөт хувилбар сонгох товчлуурынбурүү хэрэглэхээс сэргийлж үсгийн механизм дотор яг шивж бичих мэт шил дараалалсан хэлбэрээр нэмж өгөх шаардлагатай.

The number of Free Variation Selectors should be reduced to just one, at least in input.

A font mechanism may need to be added to make the users aware of incorrect use of Free Variations Selectors, like typing more than one of them consecutively.

- Шинэ тэмдэгтүүд нь хашилтын тэмдэглэгээг кодчилох хэрэгтэй байж магадгүй. Тэмдэглэл: Санал асуулгад оролцсон бусад оролцогчид U + 00AB ЗҮҮН ТАЛЫГ-2 ДАВХАР ХАШИЛТЫН ТЭМДЭГ ЛЭЛГЭЭ ба U + 00BB БАРУУН ТАЛЫН- 2 ДАВХАР ХАШИЛТЫН ТЭМДЭГЛЭЛГЭЭ тэмдэглэж оруулах.

New characters may need to be encoded for quotation marks. Note: Other participants at the meeting suggested using U+00AB LEFT-POINTING DOUBLE ANGLE QUOTATION MARK and U+00BB RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK.

- Нэг цэгтэй товчилсон үгийн тэмдэглэгээг нэмж оруулсанаар Монгол хэлээр овог, нэрийг товчлоход тусалдаг байх шаардлагатай. Тэмдэглэл: Уулзалтанд оролцсон бусад хүмүүс шинэ тэмдэгтийг кодлохын оронд энэ асуудлыг фонтын дотор зүүн-хэсэг болон баруун-хэсгийн хажууг U + 1802 МОНГОЛ ТАСЛАЛ (босоогоор бичдэг шиг дээд хэсгээс доош чиглэсэн) цөөлж, эсвэл бусад цэг таслалыг U-00B7 ДУНД ОРОХ ЦЭГ гэж тохируулан оруулах боломжтой гэж үзэж байгаа ба шаардлагатай бол хэвийн бүрэн зогсолтоос өмнө буюу дараа нэмэлт зайг ашиглана.

A single-dot abbreviation sign may need to be added for Mongolian to help with name and surname abbreviations. Note: other participants at the meeting suggested that instead of encoding a new character, we may be able to solve this in fonts, by adjusting and reducing the left-side and right-side bearings of U+1802 MONGOLIAN COMMA (top-side and bottom-side in vertical writing) or other punctuation such as U+00B7 MIDDLE DOT, and using an additional space character before or after the normal full stop, as needed.

- Нийлмэл үг нийлүүлэгчийн тэмдэгтийг нийлмэл үг бичихэд нэмэх шаардлагатай байж болно. Энэ тэмдэглэгээ нь өмнөх болон дараахь үе болно эгшгийн зохицон нийцсэн байдлыг нүдэнд харагдахуйцаар дүрслэн харуулдаг.

A compound word joiner character may need to be added for writing compound words. This character would have a visual effect on the previous and following syllables by modifying the shaping and vowel harmony context.

- Худам, Али Гали, Тодын алга болсон хувилбаруудыг шинэ тэмдэгт болгон эсвэл хувилбаруудын дараалал хэлбэрээр кодчлогдох шаардлагатай болж магадгүй юм.
There are missing forms in Hudum, Ali Gali, and Todo that may need to be encoded as new characters or variations sequences.
- Монгол бичгийн скрипт нь одоо кодчилогдсоноор аюулгүй байдлын үүднээс шалтгаалж олон улсын домэйн нэрийг гаргахад тохиромжгүй байж болох юм.
The Mongolian script, as presently encoded, may be unsuitable for internationalized domain names due to security issues.

Жиримүтү, Сэчин, Бао Хай Шан, Буригуду болон Мөнхзаяа нарын санал хүсэлт ба зөвлөмжүүд (Үндсэн баримт бичиг MWG / 2-N1)

Proposals and suggestions from Jirimutu, Siqin, Bao Haishan, Burigudu, and Menghejiya (main document MWG/2-N1)

- NNBSPP нь Монгол хэлний хувьд Монгол бичгийн маш чухал хэлбэрийг өөрчилдөг, удирдах товчлуурын нэг ба текст рэндэрлэхэд маш эмзэг байдаг, ялангуяа үсгийн фонт унах үед, скриптийг үелэх болон хэрхэн текстийг скрипт болгох талаар хоёрдмол утгатай байдаг.
Because NNBSPP, as a crucial shaping control character for Mongolian, is fragile in text rendering, especially when there is font fallback and there's ambiguity in how text should be segmented into script runs and font runs:
 - NNBSPP-ийн зан төлөвийг өөрчилсөн үсэг дээр ямар нэгэн хам сэдвийн нөлөө үзүүлэхгүй. Үүний оронд дагаварын онцгой хэлбэрийг хянахын тулд Чөлөөт Хувилбар Сонгогчыг ашиглана.
Change behavior of NNBSPP to not have any contextual effect on succeeding letters. Instead, use Free Variation Selectors to control the special shaping of suffixes.
 - Компьютерийн гарын зайн товчлуурыг дагавар холбогчоор(өнөөгийн NNBSPP-ийг хэрэглэж байх үед) үг хооронд зай гаргахын тулд ашиглахыг зөвшөөрөх.
Allow the use of normal space as a suffix joiner (when NNBSPP is used today) for cases where it's allowable to break the line at the position.

Мөнхзаяагын санал хүсэлт ба зөвлөмжүүд (Үндсэн баримт бичгийн дугаар MWG / 2-N7)

Proposals and suggestions from Menghejiya (main document MWG/2-N7)

- Хэв маягын хувилбар нь хувилбарын дарааллаар бус фонтыг ашиглан хэрэгжих ёстой. Ийм загварын хувилбаруудыг үзүүлэхийг хүссэн бичгүүд нь ижил кодчилолтой байх боловч орчин үеийн эсвэл түүхэн фонтыг хүссэн байдлаас хамааран харуулах ёстой.

Style variants should be implemented using fonts, not variations sequences. Texts that want to display such style variants would have the same encoding, but would use a modern or historic font depending on the desired display.

- Өөр өөр загвар, гүйцэтгэлүүдээс бүх төрлийн хувилбаруудыг нэгтгэх арга.

Unify all variation sequences across different models and implementations:

- Цөөн хэдэн байрлалын хувилбаруудыг (хэв маягын хувилбараас ялгаатай бөгөөд ижил хэв маягт тохиолдож болох) нэмэх шаардлагатай байх.

A few positional variants (which are different than style variants and can happen in the same style) may need to be added.

- Төрөл бүрийн дэд бүтэцтэй болон хэрэгжүүлэлтийн олон янзын ялгааг засаж нэгтгэх ёстой. ТАЙЛБАР: Дэд бүтэцтэй гэдэг нь авиан загварыг хэрэгжүүлэхэд хэрэглэгдэх өөр өөр загваруудыг хэлнэ.

Various differences in submodels and implementations should be fixed and the implementations should be unified and use the same submodel. Note: submodel here means different models used for implementing the phonetic model by different implementations.

- Байршлын бүх хувилбарууд (ялангуяа тусгаарлагдсан хувилбарууд) нь стандартад тодорхойлогдож, тодорхойлогдсон байх ёстой.

All positional variants (especially isolated variants) need to be defined, specified, and shown in the standard.

- Чөлөөт хувилбар сонгогчыг ашиглах үедээ, чөлөөт хувилбар сонгогчын нөлөөлөл нь суурь үсгийг сонгохдоо хам сэдэв дээр тулгуурлах ёсгүй. Монгол хувилбарын дарааллын хэлбэр нь түүний үндсэн үсэг дээр өмнөх болон дараагийн тэмдэгтүүд нь нэгдэн нийлсэн эсвэл нэгтгэгдээгүй тохиолдолд хамаарна.

When Free Variation Selectors are used, the effect of the Free Variation Selectors on a base letter should not depend on the context. The shape of a Mongolian variation sequence may only depend on its base letter, its variation selector, and if the preceding and succeeding characters are joining or non-joining.

- Үсгийн бүх байрлалын хувилбарууд, түүний үндсэн хэлбэр нь хувилбаруудын дараалалтай тохирч байх ёстой. (Одоогийн байдлаар, үсгийн үндсэн хувилбар нь түүнд тохирсон хувилбаруудын дараалал байх ёсгүй). (Онцлох зүйл 1.)

All positional variants of a letter, including its default form, should have a corresponding variations sequence. (Presently, the default form of a letter in a position does not have a variations sequence specified for it.) (Priority 1.)

- NNBS-тэй холбоотой олон асуудал байна. Бид (Онцлох зүйл 2):

There are various problems with the NNBS. We should either (Priority 2):

- A. Шинэ Монгол дагавар холбогч үсгийг Монгол блок дээр нэмэх, эсвэл:

Add a new Mongolian Suffix Connector character in the Mongolian block; or

- B. NNBS нь Монголын бичгийг хэлбэржүүлэхэд нөлөөлөхгүйгээр тодорхойлох бөгөөд байрлалын хооронд үг хооронд зайгүй байлгах ёстой юм.

Specify that NNBSF does not affect Mongolian shaping, and its only role is to make sure there is no word break at that position.

- Дөрөвдэх Чөлөөт хувилбарын дараалалд үгийн дунд ордог цэггүй ГА үсгийг оруулах болон бусад үсгүүдтэй холбоотой бусад хувилбарууд, үүнд шинэ Чөлөөт Хувилбар сонгогч 4 мөн багтаасан байх хэрэгтэй. (Онцлох зүйл 1.1, мөн Онцлох зүйл 1-н үр дүн.)

A fourth Free Variation Sequence needs to be added for a dotless medial form of GA and potentially other positional variants for other letters, including encoding a new Free Variation Selector 4. (Priority 1.1, a corollary to Priority 1.)

- Бидэнд удирдах товчлуур зайлшгүй хэрэгтэй(3-р Priority). Үүнд: 1. Үгийн үндэс хоёр ба түүнээс дээш эр, эм гэсэн хоёр эгшгээс гаралтай байдаг ба эр эсвэл эм эгшгийн үр нөлөөг хязгаарлах. 2. Автомат холбогчыг тусргаалах.

We need a control character for (Priority 3): 1. For limiting the effect of feminine or masculine vowels in a word when the word has two or more roots of different genders. 2. For breaking automatic ligatures.

- U + 180E МОНГОЛЫН ЭГШИГ ТУСГААРЛАГЧ-ын зан төлөв, шинж чанарыг тодруулж, түүний хэрэглээ болон монгол бичгийн скриптэд тохируулан өөрчилж болох юм.

The behavior and properties of U+180E MONGOLIAN VOWEL SEPARATOR need to be clarified and potentially modified to match its usage and effects in the Mongolian script.

Баатарын.Энхдалайн санал хүсэлт ба зөвлөмж (Үндсэн баримт бичгийн дугаар MWG / 2-N9) Suggestions and proposals by Enkhdalai Baatar (main document MWG/2-N9)

- ХА ба ГА үсэг одоогийн кодчилсон систем дотор нэгтгэгдээгүй байгаа ба ХА ба ГА үсгийн эм хувилбарын хувьд шинэ тэмдэгт болгон нэмэх шаардлагатай. Энэ нь Монголчууд уг үсгийг тодорхойлоход шаардлагатай дүрмүүдийн тоог багасгахад тустай болно.(Онцлох зүйл 1.)

A new character should be added for the feminine form of QA and GA, disunified from the currently encoded QA and GA. This would help reduce the number of rules needed for shaping Mongolian. (Priority 1.)

- Хэв маягын болон түүхэн хувилбарууд нь үсгийн дарааллаар бус форматтай дарааллаар хэрэглэгдэх ёстой. Юникод кодчилол нь загварын хувилбаруудыг оруулах ёсгүй бөгөөд Юникодын дүрмийг хялбаршуулж, богиносгох хэрэгтэй. (Онцлох зүйл 2.)

Stylistic and historical variants should be implemented using fonts, not variations sequences. Unicode encoding should not include style variants, they should be removed so the rules in Unicode would be simplified and shortened. (Priority 2.)

- MVS болон гурван чөлөөт хувилбар сонгогч товчлуур зөвхөн нэг чөлөөт хувилбар сонгохын оронд (хоёр ба түүнээс дээш дараалан ашигладаг) солигдох ёстой. (Онцлох зүйл 3.)

MVS and the three Free Variation Selectors should be replaced by just one Free Variation Selector (with potentially two or more of it used in a row). (Priority 3.)

- Монгол Юникод скрипт блокуыг Юникод дээр ашиглах нь эцсийн хэрэглэгчид хялбар болгох бөгөөд фонт, програм хөгжүүлэгчдэд хялбар байх болно.

Using the Mongolian script block in Unicode should become easier for end users and supporting it should become easier for font and application developers.

- Фонтын алдааг хянахын тулд, О/У/Ө/Ү гэсэн эгшигүүдийн авиа, дуудлагыг ялгахын тулд дуудлага ялгах тэмдэгийг оруулах хэрэгтэй. Жнь: Хэвтээ буюу босоо барилтай эсвэл үгийн дунд ордог дэгээ хэлбэрийн өмнө цэг бичих гэх мэт.

For proofreading fonts, diacritic changes could be made to the o/u/oe/ue vowels to differentiate their pronunciation, for example, a horizontal or vertical stroke, or a dot in the middle of their loop.

- Хуудас гүйлгэгч болон <textarea> ба <textview> гүйлгэгч тус тусдаа болон автоматаар зохицуулагдах тул хэвтээ ба босоо текстээс өөр байж болно.

Page scrolling and <textarea> and <textview> scrolling should be handled separately and automatically, so they could be different for horizontal and vertical text.

- Монгол босоо текстыг баруунаас зүүн тийш нь дээрээс доош, доороос дараагийн мөрөнд шилжихдээ дээрээр эхлэх хэлбэрээр оруулах ёстой.

When right-to-left text is inserted in Mongolian vertical text, the right-to-left text should go top-to-bottom, and not bottom-to-top.

Оролцогчид болон мэргэжилтэн, шинжээчдийн хэлэлцүүлэгт суурилсан бусад шинэ мэдээлэл.

Other new information based on discussions by participating experts.

- U + 1806 МОНГОЛ ТОД БИЧГИЙН ЗӨӨЛӨН ҮЕ ТАСЛАГЧ

U+1806 MONGOLIAN TODO SOFT HYPHEN:

- Энэ тэмдэгтийн хэрэглээ нь зөвхөн МОНГОЛ ТОД скриптээр хязгаарлагддаггүй. Өөрөөр хэлбэл энэ тэмдэглэгээ нь Монгол Худам бичигт нийлмэл үгсийг салгахад ашигладаг.

The usage of this character is not limited to the Todo script. It is also used for Hudum in Mongolia to separate compound words.

- U + 202F НАРИЙН ЗАЙГҮЙ ОРОН ЗАЙН ТОВЧЛУУР (дагавар холбогч гэж дууддаг):

U+202F NARROW NO-BREAK SPACE (also known as suffix connector):

- Тэмдэгтийн өргөн нь ердийн үсгийн хэвийн зайнаас багагүй байдаг. Өргөн нь загвар болон хувилбарын байдлаас шалтгаалж хэвлэгдэн гардаг юм.

The width of the character is not always less than a normal space. The width is a typographical choice depending on the style.

- Монголд хэлний зарим тохиолдолд Монгол бичгийнхэл дээр, номын нэр, дэлгүүрийн нэр гэх мэт зүйл дээр тус тэмдэгт мөрийг таслахад хүчинтэй байдаг. Хэрвээ дээрх тохиолдол гарсан бол дагавар холбогч барууны орнуудын бичгийн хэв маягтай адил аргаар эвдэрч, тасархай шугам тасарсан тохиолдолд орон зай байхгүй болно.

In some use cases in Mongolian, including in book titles, store signs, etc, it is valid to break the line at this character. When and if the break happens, the suffix connector disappears at the line break the same way that in Western typography, a space disappears if a line is broken at it.

- Зөвшөөрөгдөх шалтгаанаас болж NNBSpace нь шугаман дахь зай нь сунгагдах болон (одоогийн хэрэглэгдэж байгаа өргөн нь ихэнхи хэрэгжүүлэлтүүдэд засагдсан байгаа) өргөтгөх шаардлагатай болж магадгүй юм.

Because of justification, NNBSpace may also need to get expanded if the spaces in the line are stretched (currently its width is fixed in most implementations).

- Зарим хэрэгжүүлэлтүүд нь NNBSpace-г ердийн зайгаар сольдог ба буруу байдлаар зохицуулагддаг. Заримдаа NNBSpace-ийг Монгол текстэнд буруугаар зохицуулж оруулсан байдаг бөгөөд үүнийг засахад хялбар байдаг. Гэхдээ заримдаа NNBSpace-ийн нэг тал нь цифр эсвэл монгол бус тексттэй байдаг бөгөөд энэ нь засварлахад хэцүү байдаг.

Some implementations continue to either replace NNBSpace with a normal space, or handle it incorrectly. Sometimes the mishandling happens when NNBSpace is both preceded and followed by Mongolian text, which is easier to fix. But sometimes on one side of NNBSpace there are digits or non-Mongolian text, which is harder to fix.

- **Зэрэгцээ загвар.**

Parallel models:

- Лианг Жинбао, Жиримүтэ, Лианг Хай болон бусад хүмүүсийн танилцуулга, тайлбар дээр үндэслэн зарим хэрэглээний тохиолдолд авиан загвар нь график загвараас илүү сайн байдаг бол өөр нэг тохиолдолд график загвар нь авиан загвараас илүү сайн байдаг тохиолдол ч мөн байна. Хэдийгээр Солонгост Хангул скриптийг хэрхэн яаж кодчилсонтой адилаар Юникод дотор буй Монгол скриптийг хоёр зэрэгцээ загварыг мөн ашиглахыг хичээх хэрэгтэй. Ингэснээр өөр өөр програмууд өөрсдийн хэрэглээнд нийцсэн, тохирсон хувилбарыг сонгох ба бусад хэрэглэгчид ч мөн өөрсдөд нийцсэн хамгийн сайн хувилбарыг сонгох боломжтой юм. Хангултай төстэй байдлаар, хоёр загварыг хооронд нь хөрвүүлэх алгоритмуудыг хэрэгжүүлэлтийн хооронд хоёрдмол утга болон нийцэж таарахгүй байгаа хэрэгжүүлэлтийг багасгахын тулд тодорхойлж болно. Бичиг үсгийн загварыг монгол бичгээр бичсэн бусад хэлийг авч үзсэний дараа тусдаа блокод кодлож болно. Юникод стандартын авиан загварыг сайжруулах нь нэн тэргүүний асуудал хэвээр байх ёстой.

Based on presentations and comments by Liang Jinbao, Jirimutu, Liang Hai, and others, it is clear that the phonetic model is better than the graphetic model for some use cases, while the graphetic model is better than the phonetic model for some other use cases. We should consider pursuing two parallel models for Mongolian script in Unicode, similar to how Korean in Hangul script is encoded, so that different applications could choose the best model for their use case or their users. Similar to Hangul, algorithms for converting between the two models could be specified to reduce ambiguity and disparity between implementations. The graphetic model could be encoded in a separate block, after considering other languages written in the Mongolian script. Fixing the phonetic model in the Unicode Standard should remain a high priority.