音码：Phonetic code
形码：Graphic code
义码：Semantic code

Revised Proposal for minimizing the current phonetic code

Meng he ji ya
March 23, 2018

Introduction

GB 25914-2010 Information Technology—Traditional Mongolian Nominal Characters, Presentation Characters and Use Rules of Controlling Characters (shortened form *Transformational Rule*) was published seven years ago. Research and development based on Mongolian standard code achieved plenty of goals during these seven years. However, because of existing problems in the current transformational rule, and differences on understanding the *Transformational Rule* and Mongolian grammar, the transformation between Mongolian nominal characters and presentation characters is not completely unified.

Based on the most urgent problem, *Transformational Rule* and Unicode 10 which solved the editing error and position mismatch, a new revised proposal to minimize revise contents is brought forward.

2 Revise Content

2.1 Making a uniform variation sequence (Perfect sequence of current character variants)

In order to unify differences between ligature positional model and in-word positional model, this proposal will directly adopt classified method of ligature position model which unifying their differences through assigning the same FVS. Therefore, the differences would only exist in the textual level and customers wouldn’t feel the differences. This is the only way to unify their differences and also doable as proved.

There are five aspects as followed in order to perfect sequence of current character variants:

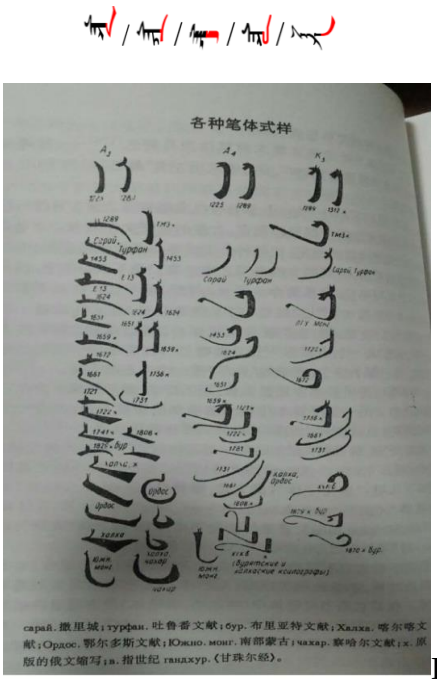
(1) Supplement of positional variants which were treated as ancient variants but not exactly

When it comes to add modern positional variant to HA (182C) and GA (182D), some scholar would bind them to digitalization of ancient documents. Only few words could be defined as positional variants in ancient documents until now. It means that if code positions of these positional variants were set up, the problem of code positions of ancient documents digitalization would basically solved and making different fonts of ancient documents would be the only thing left.

Style variant:

Style variant means different ways to writing. In Orthography, Style variant is only different writing styles in the same alphabet of the same word.

It could transformed between modern and ancient by transforming between different fonts (ancient and modern for instance) in the same coding sequence. As for the time span between ancient and modern, it just means different writing styles because of different writers, different writing time and different writing materials and they have no linguistic meanings. E.g.:

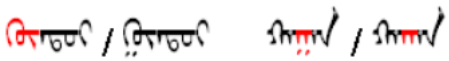


Different writing styles

Positional variant:

In Orthography, positional variant means varying positions in the same character. As in the time span between ancient and modern, positional variant is isolated

variant that completely mutually contradictory and complementation. E.g.:



Though the description is not very precise, however we still could decide whether it is style variant[①] by using ways as followed:

- Whether a reasonable explanation exist in linguistics[②]. Every variant should have a reasonable explanation in Mongolian Orthography.
- Whether a continuous transition exist between these two variants. If it has a continuous transition and infinite variant possibilities, then we could sure that it is style variant. It could not transform fonts by using FVS and VS, instead it has to switch among different writing style to achieve the purpose of display.

According to the above ways, few positional variants should be add on.

E.g.:

	No.	Form	25914	U10	diff	FVS
ISOL	1			+	☆	
	2		+	+	☆	1
	3			+	☆	2
	4			+	☆	3
INIT	1					4
	2			+	●	1
	3		+		●	2
	4		+	+	●	3
MEDI	1					4
	2			+	●	1
	3		+		●	2
	4		+	+	●	3
	5			-	★	MVS
	6					1+MVS
FINA	1		+		★	
	2		+	+	●	1

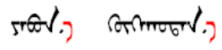
(2) Unifying divergences caused by difference between ligature position model and in-word positional model

Divergences caused by differences between ligature position model and in-word

① In this Chinese paper, the author used the expression “style variant”, I was wondering whether the author made a mistake or not. I guess she want to write “positional variant”. There are two reasons: first, she already mentioned style variant in the above paragraph, why she describe style variant in the “positional variant” part? Second, this section is about positional variant, why she want to discuss style variant in the positional variant section?

② As we mentioned in the above, we know that style variant is only different writing styles and it has nothing to do with Linguistic. So if it has a reasonable explanation in Linguistic. It should not be a style variant. However in the second sentence of the first way, it says that every variant should have a reasonable explanation in Mongolian Orthography. I don’t quite understand what she was trying to say.

positional model only theoretically or textually exists. We could avoid these kinds of divergences. It has been proved technically feasible. It is lucid to the user and they wouldn't see the theoretical understanding divergence between these two theories. E.g.:



	No.	Form	25914	U10	diff	FVS
ISOL	1					3
	2					1
	3	[3]	+	[3]	★	2
INIT	1					2
	2		+		★	1
MEDI	1					
	2					1
		-		-	★	
FINA	1	[1]	[1]	[1]		1
	2	[2]	[2]	+	★	MVS

(3) Unifying divergences caused by positional differences of initials of Mongolian case supplementary element

Which position is the initials of Mongolian case supplementary element, word-initial, word-medial, word-final or isolating? These are divergences caused by positional differences. Problems wouldn't exist while unifying these divergences as long as unifying them according to one theory or viewpoint. E.g.:

	No.	Form	25914	U10	diff	FVS
ISOL	1					3
	2					1
	3	[3]	+	[3]	★	2
INIT	1					2
	2		+		★	1
MEDI	1					
	2					1
		-		-	★	
FINA	1	[1]	[1]	[1]		1
	2	[2]	[2]	+	★	MVS

(4) Agreement on a way forward of whether isolated variants of characters should be set up

Agreement on a way forward of whether isolated variants of consonants should be set up and I think isolated variants of consonants should be set up from the application point of view. As for the question of whether isolated variants should be set up or switched by other variants in Unicode, the form should be displayed isolated variants because isolated variants is not about showing variants of word-medial (the situation of ZWJ doesn't exist in the context). E.g.:

	No.	Form	25914	U10	diff	FVS
ISOL	1			+	☆	
INIT	1					
MEDI	1					
FINA	1					

(5) FVS setting is non-context-related

Except varying according to grammar context, A FVS should be specifically allotted to every variant. It could transform any variant application of every word. The current Phonetic code program sets up concept of default font while designing but FVS doesn't allocated. This program just think about positional default font without considering font of grammar context, which situation of unsettled FVS would appeared when the requirement of asking to be the first variant while it was a default grammar context variant (it wasn't the first variant before). E.g.:

Default:  ->1834+1822+182E+1821+1833+1823+1833+183C+1820+1837

Request:  ->1834+1822+182E+1821+1833+FVS2+1823+FVS1+1833+183C+1820+1837

1833 MONGOLIAN LETTER DA

	No.	Form	25914	U10	diff	FVS
ISOL	1			+	☆	
	2		+	+	☆	1
INIT	1					2
	2					1
MEDI	1					2
	2					1
FINA	1					
	2					1

In conclusion, when we identify variant, we should allocate specific FVS to every variant. The first variant which is regarded as positional default variant was included. Every variant should be allocated a specific FVS.

2.2. Adding connectives of Mongolian supplementary element

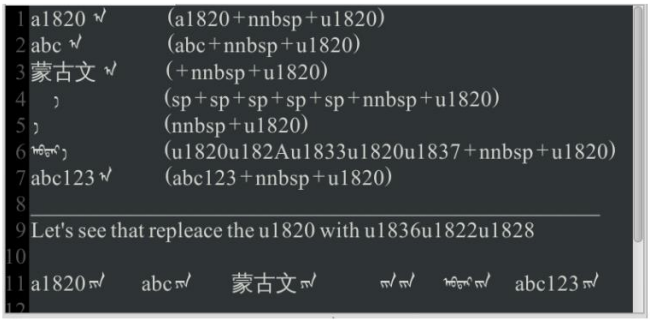
According to Unicode standard, NNBSP (0x202F) is a universal character, pertain to a common language not a specific language. NNBSP is treated as a special operational character of Mongolian languages that include classic Mongolian, Todo Mongolian, Xibe language, Manchu,Galik and great Mongolian with digital punctuations, which means it controls the variant of characters before and after. However, in some special circumstances, identification of NNBSP's languages may cause some problem based on the situation of many kinds of languages mix together.

For instance, English abc+nnbsp + Mongolian a(0x1820) combines to be one string, which shows on windows as followed:

Explanation: English "abc" belongs to English language while nnbsp is a common language. Mongolian "a" belongs to Mongolian language. Because of nnbsp, the English character varied and the Mongolian "a" was asked to be de word-ending postfix form, which showed on windows and it in line with Unicode norms.

In Linux system, Pango would separate the string into different fragments according to their languages and each fragment could be handled by its own engine. If there is a common language, its characters could be districted into the language before it. In this case, Mongolian "a" was dealt with as an isolated language. After that, the Mongolian engine could only feel one isolated Mongolian "a" and it would handle it as an isolate. It is not line with Unicode norms.

The following picture shows variants under all kinds of situations. In all those situations, number 4,5 and 6 were dealt with the Mongolian "a" together because of no other language character at the beginning, which are correct ones. Others are all error situations.



In the current phonetic code program, we need to set up an operational character which has functions as followed. We know that the recent NNBSP (202F) also satisfies part of the needs but doesn't satisfy perfectly and doesn't support well. We doesn't use it well, so I was thinking we should add a new specific Mongolian use MSC which belongs to 1800 sections. The problem is we settled too many extra and unnecessary functions for NNBSP. We should set a code position isolate in 1800 section, such as 0x180F and etc.

There are two ways to solve this problem:

Solution A: Adding an operational character of Mongolian supplementary element with connective function in Mongolian coding section (0x18xx) such as Monglian[③] Suffix Connector and we settle functions like controlling variant of the word-initial and word-suffix.

Solution B: Be explicitly stipulated that current NNBSP must not attend in Mongolian variant rules and its only function is doesn't tokenize words.

2.3. Adding selector of free variant

While designing the current coding system, we only thought about positional default variant and we didn't expect default variant according to the context. As a result, when the variant number are over three (for instance four, H, G), the computer couldn't solve the problem of requesting to be the first positional variant from another positional variant.

E.g.: If there is no red FVS4, we couldn't input the following Request font.

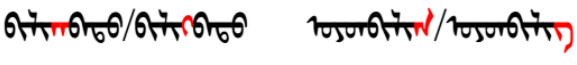
Default:  -> HAGAN (182C+1820+182D+1820+1828)
Request:  -> HAGAN (182C+1820+182D+FVS4+1820+1828)
Default:  -> MOGAI (182E+1823+182D+1820+1822)
Request:  -> MOGAI (182E+1823+182D+FVS4+1820+1822)

2.4. Adding root delimiter for Mongolian multi-root words

Though problems could be solved without this delimiter, it would be much more complicated. If we set up this delimiter, the whole coding system would be much better.

Request example:

(1) Diagnosing positive or negative of multi-root words for it is general to mix positive and negative together.

For instance: 

In this case, we should make sure the font of g according to positive or negative of the word. In default situation, the g in words of biligbatv and wyvnbilig is supposed to

[③] I think this word is supposed to be Mongolian, I didn't change it because the author write down this word in English and it was an operational character.

be positive for these two words are positive. However, g in these two words is negative for these two words are combined by two other words. If we write these two words as bilig/batv and wyvn/bilig (/ means root delimiter), we don't need to diagnose word class by ourselves, the root delimiter could help us.

(2) Stopping automatic fit (even new graphic program requires function of stopping automatic fit)

Chinese Pinyin for instance: xian =xi' an= xi an. It is classic alphabetic writing. We could use space key to write Chinese Pinyin xi an. However when ligature requires, the computer couldn't distribute the difference between xi' an and xian. Though Mongolian is also an alphabetic writing like Chinese Pinyin, still it is a alphabetic writing that requires characters varied in different positions of words.

Mongolian example: hexig' undur

Mongolian is not like Latin, it needs this function but punctuations couldn't be used, and the normal FVS couldn't solve this issue. Even though we could use special ways like zwj(u200D), nirugu(u180A) and etc. to solve this problem, yet we have a perfect and precise way to use. Why not use it. Isn't 1800 sections that enough? ZWJ is the last way to use while input Mongolian names.

2.5. Editing MVS property

Unicode identifies MVS as a spacebar with space property as the following picture shows. However, it doesn't have space property from display of the font companies, instead it plays the role of an invisible operational character.

Table 6-2. Unicode Space Characters

Code	Name
U+0020	SPACE
U+00A0	NO-BREAK SPACE
U+1680	OGHAM SPACE MARK
U+180E	MONGOLIAN VOWEL SEPARATOR
U+2000	EN QUAD
U+2001	EM QUAD
U+2002	EN SPACE
U+2003	EM SPACE
U+2004	THREE-PER-EM SPACE
U+2005	FOUR-PER-EM SPACE
U+2006	SIX-PER-EM SPACE
U+2007	FIGURE SPACE
U+2008	PUNCTUATION SPACE
U+2009	THIN SPACE
U+200A	HAIR SPACE
U+202F	NARROW NO-BREAK SPACE
U+205F	MEDIUM MATHEMATICAL SPACE
U+3000	IDEOGRAPHIC SPACE

Actually, considering the original intention of design, MVS should contain properties as followed:

(1) Blank space, MVS is not a spacebar, it only has 1/6 of blank space. It doesn't have

the function of punctuation, neither does it is the end of a word.

(2) An ordinary Mongolian alphabet

MVS is supposed to be an ordinary Mongolian alphabet and it shouldn’t have the property of operational character.

3. Conclusion

Compared Transformation Rules and U10 and issues that appeared while display and testing, we would cost less in time, money and tech to realize Unicode if we could solve and unify all those issues of these five aspects.