

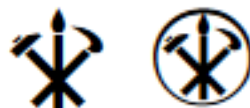
Proposal to reconsider compatibility symbols and punctuation used in the DPRK

Eduardo Marin Silva

03/01/2018

Introduction. In the mailing list (<http://www.unicode.org/mail-arch/unicode-ml/y2015-m05/0276.html>) it was brought to my attention that many symbols used in the KPS 9566-97 standard, were not encoded into the Unicode standard. Since not even looking at the UTC minutes helps disambiguate the discussion that was had that day, we can only approximate the objections raised against many of these characters based on personal experience. In this document, I propose the inclusion of some symbols and presentation forms, either through new characters, named character sequences or variation sequences.

Since I have no direct access to North Korean experts or even just experts in that standard, my main source of information will be the proposal document L2/01-349 (<http://unicode.org/L2/L2001/01349-N2374-DPRK-AddSymbols.pdf>) as well as the feedback by Asmus Freytay in L2/02-102 (<http://www.unicode.org/L2/L2002/02102-n2417-dprk.pdf>).



Hammer And Sickle And Brush. This symbol, along with its circled counterpart, were probably excluded due to its use in the logo of a political party, however this has not stopped the consortium in the past, we have the FLEUR-DE-LIS ♁ the OUTLINED WHITE STAR ☆, the MAPLE LEAF 🍁 and even the HAMMER AND SICKLE ⚒ itself. North Korean users may feel like that it isn't fair that their logo is excluded for apparently arbitrary reasons. This is not similar to emoji, where one has to prove multiple possible usages; the consortium is dedicated to the best compatibility with the most standards, and this is part of a legacy standard that probably still has encoded documents with that symbol on.

One possible argument against it, would be that it would hurt the sensibilities of the enemies of North Korea, like South Korea or Japan or even North Korean defectors. We could think of it as a similar situation as if one were to propose a NAZI CIRCLED SVASTIKA (which is different in shape and meaning to the already encoded svastikas) to transcribe WW2 era German documents; the controversy may be too much for the consortium to bear, and it may hamper all other proposals.

One thing to note, is that it would be unacceptable to just encode one sign and not the other.

Map symbols. These symbols are battleground markers on maps, they would be similar to the already encoded symbols that are used in the ARIB standard in Japan. Here I list them, with a proposed Unicode name. It must be noted that names that refer to their semantics would be more confusing than names that describe their glyphs.

Glyph	KPS code	Proposed name
	A2F1	UP POINTING BLACK TRIANGLE ENCLOSING BLACK TRIANGLE
	A2F2	UP POINTING BLACK TRIANGLE WITH HORIZONTAL FILL
	A2F3	UP POINTING BLACK TRIANGLE WITH UPPER LEFT TO LOWER RIGHT FILL
	A2F4	UP POINTING BLACK TRIANGLE WITH UPPER RIGHT TO LOWER LEFT FILL

The names proposed here are different than the ones proposed by the DPRK, in order to make them more like other already encoded characters. The adjective “BLACK” is added because that appears to be the dominant color; if one were trying to describe a glyph with equal parts black and white, one could drop such adjective from the name of the last three characters.

Freytag did not bring any objections to their encodings, but rather he brought attention to other unencoded symbols. Such symbols need their own separate proposal.

A good place to encode them would be the Geometric Shapes block.

Leftwards Scissors. These were not encoded, because as Freytag already pointed out, these are not used to contrast with other scissors symbols and so it can be unified with BLACK SCISSORS with a font change. However, it must be pointed out, that a good case can be made for scissor symbols pointing in different directions with independent merits. I may soon write a proposal for such symbols.



Circled White Up Pointing Index. This character has the KPS code ACD5 and the original name proposed by the DPRK was CIRCLED UPWARD INDICATION, but I propose the name: CIRCLED WITH UP POINTING INDEX, which is more similar to already encoded characters. One could argue that it could be unified with WHITE UP POINTING INDEX 🇮🇸, but if one looks at the original code chart included in the proposal, the glyph is in contrast with a WHITE RIGHT POINTING INDEX 🇮🇸 that is not circled. And since the consortium has ruled many times that legacy standards with circled versions of other characters are not unifiable, the same may be argued here.

This would be encoded in the Supplemental Symbols and Pictograms, preferably at U+1F90C with an annotation stating that it is not an emoji.

Vertical presentation forms of punctuation. Of the proposed characters only, a few have been separately encoded, and those were:

- PRESENTATION FORM FOR VERTICAL COMMA
- PRESENTATION FORM FOR VERTICAL IDEOGRAPHIC FULL STOP
- PRESENTATION FORM FOR VERTICAL IDEOGRAPHIC COMMA

The ones that are left are to be encoded have a similar function to the rest of the vertical presentation forms, which means only a minority of users dealing with legacy data from North Korea would need them, since otherwise the glyph positioning is better handled by fonts. Nevertheless, a rationale for its exclusion must be explicitly provided or their exclusion seems arbitrary, (especially in comparison with the GB 18030 standard) so they are proposed here for further discussion.

Glyph	KPS code	Proposed name	Decomposition
·	A1C0	PRESENTATION FORM FOR VERTICAL FULL STOP	<vertical> 002E .
§	A1C5	PRESENTATION FORM FOR VERTICAL TILDE	<vertical> 007E ~
,	A1DE	PRESENTATION FORM FOR VERTICAL LEFT SINGLE QUOTATION MARK	<vertical> 2018 ‘
‘	A1DF	PRESENTATION FORM FOR VERTICAL RIGHT SINGLE QUOTATION MARK	<vertical> 2019 ’

”	A1E0	PRESENTATION FORM FOR VERTICAL LEFT DOUBLE QUOTATION MARK	<vertical> 201C “
“	A1E1	PRESENTATION FORM FOR VERTICAL RIGHT DOUBLE QUOTATION MARK	<vertical> 201D ”

Assuming one is working in an environment that does not allow to change glyphs of characters in a vertical context, all of these characters are of great use. Of course they are mainly for compatibility, so the consortium doesn't need to worry about encoding other marks.

There is just enough space in the Vertical Forms block to accommodate all six characters.

Double punctuation. Two punctuation marks are proposed for handling the mono-width nature of fonts, when a FULL STOP precedes either a RIGHT PARENTHESIS or a RIGHT DOUBLE ANGLE BRACKET. They are also convenient for vertical text, when one wants both characters to occupy the same cell.

It is unclear how these should be handled in Unicode, either they are encoded as separate characters in the Supplemental Punctuation block, with their corresponding decompositions, or they can be handled as named character sequences. The latter is simpler to implement but it would break precedent of other similar marks, like the ellipsis or the double question mark. Either way, names must be proposed for either situation.

Glyph	KPS code	Proposed name	Decomposition or named sequence
.)	A1DC	FULL STOP RIGHT PARENTHESIS	≈ 002E . 0029)
.>>	A1DD	FULL STOP RIGHT DOUBLE ANGLE BRACKET	≈ 002E . 300B >>



Enclosed postal mark symbol. This is no different than the already encoded variants of the postal mark, so I have no clue why it wasn't included.

The name could be POSTAL MARK ENCLOSED IN DOWN POINTING TRIANGLE, it could be encoded in the Ideographic Symbols and Punctuation block and the references would include: POSTAL MARK 𠄎, POSTAL MARK FACE 𠄏 and CIRCLED POSTAL MARK 𠄐.

Vulgar fractions. The DPRK implies that both forms of some vulgar fractions (diagonal and horizontal bar) are both encoded in their standard. Whatever use distinguishing them may have is unclear, and we don't have a print of their codechart where the vulgar fractions appear to confirm the claim, unlike the other characters.

Assuming however that it is true, encoding the horizontal version of the fractions separately would not have been ideal, since they would have been confusable with the already existing fraction. Nevertheless, it is still possible to encode variation sequences for them, for the purposes of compatibility.

Entries:

- 00BC FE00; horizontal stroke form; # VULGAR FRACTION ONE QUARTER
- 00BD FE00; horizontal stroke form; # VULGAR FRACTION ONE HALF
- 00BE FE00; horizontal stroke form; # VULGAR FRACTION THREE QUARTERS

- 2153 FE00; horizontal stroke form; # VULGAR FRACTION ONE THIRD
- 2153 FE00; horizontal stroke form; # VULGAR FRACTION TWO THIRDS

Emphasized Hangul syllables. The DPRK has dedicated some codepoints to write the name of two of their presidents in a different font for emphasis. Not only is the effect easily done using markup, but the problems associated with spoofing would be increased if the same were to be done in Unicode, particularly since they do not reuse the same emphasized syllables in both names.

Nonetheless, we have precedent in other scripts of additions for compatibility that are religiously motivated, such as many arabic presentation forms, used explicitly for Islam.

As a compromise, I originally planned proposing implementing them as variation sequences of five Hangul syllables. This made sure that spoofing would no longer be an issue and it even avoided hurting the sensibilities of others; because all the enemies of the DPRK would have to do, is to ignore the variation selector whenever it appears for any Hangul syllable.

This may have lead to abuse, where they keep on adding special emphasized glyphs as different characters of their standard, whenever a new president takes office (which would set the precedent for other nations to do the same). To avoid this, the consortium would have explicitly stated that no new emphasized syllables will be encoded, if they were defined past a certain date.

The DPRK may request a different SVS for each of the three names, making it a total of nine, however it is obvious that they use the same font for all emphasized names and because variation sequences only are supposed to capture glyphic variation, these entries would have been enough.

Entries:

- AE40 FE00; emphasized; # HANGUL SYLLABLE GIM
- C131 FE00; emphasized; # HANGUL SYLLABLE SEONG
- C740 FE00; emphasized; # HANGUL SYLLABLE EUN
- C77C FE00; emphasized; # HANGUL SYLLABLE IL
- C815 FE00; emphasized; # HANGUL SYLLABLE JEONG

Unfortunaetly Hangul syllables have canonical decompositions with their respective jamos and so are not eligible for SVS's.

As yet another workaround, we could introduce three named character sequences, corresponding to the three names of the leaders of North Korea, so that software engineers can implement the automatic emphasis more easily. This can also be easily ignored by everybody else, so there is no possible controversy.

Named sequence	Sequence
NAME OF KIM IL SUNG	AE40 C77C C131
NAME OF KIM JONG IL	AE40 C815 C77C
NAME OF KIM JONG UN	AE40 C815 C740