Author: Kent Karlsson
Status: Expert Contribution
Date: 2000-03-02

Comments on DPRK New Work Item proposal on Korean characters.

These are some a bit more detailed comments supplementing the Swedish NB NO vote the New Work Item proposal from DPR of Korea, details that for brevity are not included in the vote itself.

On point 1 (and general) of JTC1/N5999:

The Hangul characters (syllables, conjoining, and compatibility letters) have assigned code points already, and these code point assignments are being used in more and more implementations. They cannot be changed now, since any change would cause major disruptions with existing and now produced applications.

The Hangul character assignments were changed by amendment 5. However, changing code points a few years ago was highly unfortunate even then, even if the result was more consistent than the original, due to that it forced incompatible changes to implementations. Now, when there are many more implementations, any change other than additions of characters would be very problematic. However, the change done by amendment 5 caused simplifications to implementations, in particular that the (Unicode) canonical decomposition of Hangul syllables became easy to express algorithmically.

But doing any changes now to already assigned code points is very far from being acceptable.

On point 2:

Please see Unicode technical report 10 and ISO/IEC DIS 14651, for methods for tailoring of the ordering of Hangul Jamo and Hangul syllables well as compatibility forms (nominal, half-width, circled, and parenthesised Hangul letters). Note how various compatibility forms of the same Hangul Jamo letter are ordered "close" to each other, despite them having quite "far distant" code points.

Nearly none of the scripts covered by the UCS can be ordered culturally correct just by ordering by code point value. For CJK Unified Ideographs and for Hangul Syllables, code point order can be used. But tailoring of the ordering can be applied also to CJK ideographs and Hangul. A tailoring, or delta, for *ordering* Hangul (for comparison, and ultimately sorting) in the way the New Work Item proposal suggests can be given without changing any code point to character allocations.

By using the methods of UTR 10 and ISO/IEC 14651, one can get any ordering desired for the Hangul Jamo, implying a corresponding ordering of the Hangul

syllables. The effort of reordering Hangul characters should be directed at creating one or several tailoring(s) of the 14651 Common Template Table, tailorings that are suitable for Hangul.

Note that such tailorings are needed e.g. also for the Latin script (several different tailorings), and for the Japanese scripts (to take the length mark into account).

On point 3:

See text regarding Annex B below.

On point 4:

Adding cross-references to KPS 9566-97 in 10646 itself is not needed. However, North Korea could provide a machine-readable mapping table (like those found on the Unicode web site) for mapping KPS 9566-97 encoded data into 10646 encoded data (and back). There are many encodings for which 10646 itself does not give any cross mapping. Even the cross-mappings given in 10646 for CJK ideographs are somewhat misplaced, and should ideally have been given elsewhere.

A machine-readable mapping table is much more helpful than having the mapping information printed in the 10646 document (or any other document which is not easily machine readable for data extraction).

On point 5:

Character name changes for already assigned characters cannot be accepted. It would cause problems for numerous implementations. And, as far as we know, the term Hangul is the accepted name of the script, in both North and South Korea as well as in the rest of the world. Furthermore, it is not true that the name of "national" characters bears the name of "their" nation. For example "Katakana", "Hebrew", "Yi". These are used mostly, though like for Hangul not exclusively, in only one nation. But they do not bear the name of that nation.

On Annex A:

(K+ below refers to the North Korean proposal.)

Initial consonants:

1. Two existing characters appear not to be covered by the proposal:
   U+1134 and U+1146.

2. 6 "new" characters are proposed K+1113, K+1114, K+1117, K+1120,

K+1128, and K+112A. 4 of these APPEAR to be glyph
variants of existing characters, while one is brand new (K+1117),
and one is a new combination (K+1128).

3. It is not clear if K+1157 is the same as U+1145 or not.
   Given the new names as well as the glyph change, a
   difference may be intended.

4. Some of the glyphs used in the North Korean proposal appear
   to be more in the Hangul typographic style, than those in the
   Unicode charts (in particular those for U+113C-U+1140, U+1146,
   and U+114E-U+1151; but also for U+114C, which currently has
   too little distinction from U+110B in our eyes). There are
   some other glyph differences as well (like for U+1110),
   but can only guess that that is only glyph differences.

Middle vowels:

1. Three existing characters appear not to be covered by the North Korean
   proposal: U+118B, U+1192, and U+1197.

2. 7 new characters appear to be proposed: K+1180, K+1184,
   K+1186, K+118B, K+118D, K+118E, and K+1193.

3. Again there are some glyph differences, especially
   related to how sub-letters join.

Final consonants:

1. Here there appears to be only rearrangement and some example
   glyph differences (some of which should be considered for
   the "typical" glyphs in the Unicode/10646 charts).

The rearrangement and renaming of Hangul Jamo characters is of course
unacceptable at this time. Many implementations depend on the existing names and
allocations for Hangul Jamo characters.

To exclude already assigned characters is also unacceptable.

The possible glyph improvements should be considered further by someone familiar
with Hangul. North Korea could provide specifics about which "typical" glyphs should
be changed, and why.

Apparently, 7 new initial consonants are proposed, as well as 7 new middle vowels.
These need to be considered further by someone familiar with Hangul. Since these

apparently are historic, not modern, Hangul characters, historic documents need be produced that show the use of each proposed new Hangul Jamo.

On Annex B:

1. Part of this appears to suggest a rearrangement of the
   Hangul Compatibility Jamo (U+3130-U+318F). Collation order
   tailoring techniques should be used to get the ordering desired,
   not code point changes.

2. Another bit suggests other characters, some of which
   would have compatibility decompositions. Some of
   the suggested characters already exist in the UCS.

3. Part of this even suggest having special characters for
   some (2) names of North Korean political significance,
   plus the logo of a party (twice). Logos and personal names
   are not acceptable as characters.

A small number of the characters suggested in Annex B *may* merit further consideration. But not those that would have compatibility decompositions (in Unicode), and/or are names/logos, or those that already are in the UCS (like quotation marks, the vulgar fractions, some of the arrows, and the compatibility Jamo) should not be included into the UCS.

On Annex C:

The character name changes are unacceptable. The names are normative and must not be changed (even if wrong). However, these names can, non-normatively, be translated into other languages, and errors may then be corrected.

The rearrangement of code point/character assignments is unacceptable. However, tailoring the collation ordering of the Hangul Jamo character should imply a corresponding collation order reordering of the syllable characters.

Furthermore, the suggested syllables rearrangement looses simple algorithmic decomposition even in reference to the New Work Item proposal's proposal for Hangul Jamo.

Summary:

1. The glyph changes, for the example glyphs, for some (historic) Hangul Jamo characters need be considered by people familiar with the Hangul script, especially historical Hangul. No new work item is needed for this.

2. Some of the "new" Hangul Jamo characters suggested need be considered further by people familiar with the Hangul script, especially historical Hangul. Examples from historic documents need to be provided if any of these characters are to be accepted. No new work item is needed for this.

3. Some of the "miscellaneous" characters suggested in annex B may be considered further. Some of them are already under consideration as math symbols. No new work item is needed for this.

4. Cross references between 10646 and KPS 9566-97 (except for things that are not accepted into 10646) should be provided separately in a machine readable file. No new work item is needed for this.

5. A tailoring of the 14651 CTT to suit the desired ordering of Hangul characters could be provided by DPRK (as a national Standard, if desired). However, such a tailoring is out of scope for 10646 itself.

------------------------------------**end of N2167**--------------------------------