

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Document type: Working group document

Title: Proposal to add nine multi-letter Jamo characters

Source: Kent Karlsson

Status: Expert contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2008-04-21

Hangul Compatibility Jamo (HANGUL LETTERs)

The Hangul script is an alphabetic script, originally with 28 letters (a small number of letters were since added, but have fallen out of use). The alphabetic nature of the script has not changed in any way since the script was designed in the 1440-ties. What is a little bit unusual about Hangul compared to other alphabetic scripts is that letters are grouped in syllable clusters. In ISO/IEC 10646 this division into syllable clusters is codified by using a double set of consonant letters, lead and trail. Furthermore, in the Jamo system, partial syllables, like a single consonant or a single vowel, are represented by a formal syllable, that contains a “filler” letter. So a single isolated consonant, KIYEOK, for instance, is represented by the consonant Jamo followed by a filler vowel Jamo, <lead KIYEOK, vowel filler> (<HANGUL CHOSEONG KIYEOK, HANGUL JUNGSEONG FILLER>, <U+1100, U+1160>).

Ideally, one would have used the same approach for this alphabetic script as has been used for all other alphabetic scripts (with some minor exceptions), one character per letter in the script. However, this is not how Hangul has been handled. There are many multi-letter sequences that have been coded as single characters. In this document we concentrate on the HANGUL LETTER ones.

For reasons of compatibility with pre-existing coded representations for Hangul, some partial syllables are also represented as single characters. For instance, KIYEOK, as a partial syllable, is also represented as HANGUL LETTER KIYEOK, U+3131. Similarly, <U+115F, U+1161> (<HANGUL CHOSEONG FILLER, HANGUL JUNGSEONG A>) is the Jamo representation for an isolated Hangul A vowel, also represented by U+314F, HANGUL LETTER A.

However, there are also representations of sequences of consonants and sequences of vowels as single characters. For instance, HANGUL LETTER SSANGKIYEOK, U+3132, represents a double Hangul KIYEOK as a partial syllable, in the Jamo system represented as <HANGUL CHOSEONG KIYEOK, HANGUL CHOSEONG KIYEOK, HANGUL JUNGSEONG FILLER>. But in addition, the sequence <HANGUL CHOSEONG KIYEOK, HANGUL CHOSEONG KIYEOK> is also represented as HANGUL CHOSEONG SSANGKIYEOK,

U+1101. So <HANGUL CHOSEONG SSANGKIYEOK, HANGUL JUNGSEONG FILLER> is another representation for a double KIYEOK as a partial Hangul syllable.

Taking the letters in this example sequence one by one, and replacing by a more composed representation as one goes, gives:

<U+1100, U+1100> → U+1101

<U+1101, U+1160> → U+3132

Continuing this analysis for all of the HANGUL LETTER characters gives a table where most of the HANGUL LETTER characters are composed most simply from two other characters. The composition can be applied stepwise, as exemplified above, using also similar compositions for multiletter Hangul Jamo characters.

However, and this is the topic of this proposal, some of the HANGUL LETTER characters are composed most simply (using only characters already allocated, or already in the process of being allocated) composed of three characters. For instance, U+3133, HANGUL LETTER KIYEOK-SIOS, is most simply composed from a three-character sequence, <U+1100, U+1109, U+1160> (using only already allocated characters, or characters already proposed, see N3242 which includes proposed code point allocation). This is unfortunate, since it makes it harder to convert between the different representations of the same Hangul letter sequences, requiring exceptional handling in nine cases. In order not to need exceptional handling in these nine cases for mapping between the different representations of the same letter sequences, nine new Hangul characters are needed. In the U+3133 case, a new HANGUL CHOSEONG KIYEOK-SIOS is proposed to enable its simplest mapping to be a two-character one. These nine characters are proposed in this document (see below).

Mappings for HANGUL LETTERS

The full list of mappings between HANGUL LETTER characters and Hangul Jamo characters is as follows in the list below. Note that some of the mappings here refer to already proposed characters (see N3242), and nine of them refer to characters otherwise not elsewhere proposed, but are proposed in this document, nine new characters. These mappings, together with similar mappings for multi-letter Jamos, can be used by processes where one wishes to map between different representations for the same Hangul text.

With a certain few Hangul characters already proposed (see N3242) all of the multi-letter Jamos can get stepwise mappings to single Hangul Jamo letters (these mappings are not listed in this document) using mappings to two other Hangul Jamo for multiletter Jamo characters.

NOTE, HOWEVER, THAT THE MAPPINGS CONSIDERED HERE ARE NEVER TO BE USED IN UNICODE NORMALISATIONS.

```
3131; <compound> 1100 1160 # HANGUL LETTER KIYEOK
3132; <compound> 1101 1160 # 1100 1100 1160, HANGUL LETTER SSANGKIYEOK
3133; <compound> A97D 1160 # 1100 1109 1160, HANGUL LETTER KIYEOK-SIOS [the mapping
      uses a NEW proposed character]
3134; <compound> 1102 1160 # HANGUL LETTER NIEUN
3135; <compound> 115C 1160 # 1102 110C 1160, HANGUL LETTER NIEUN-CIEUC [the mapping
      uses a N3242 character]
3136; <compound> 115D 1160 # 1102 1112 1160, HANGUL LETTER NIEUN-HIEUH [the mapping
      uses a N3242 character]
3137; <compound> 1103 1160 # HANGUL LETTER TIKEUT
3138; <compound> 1104 1160 # 1103 1103 1160, HANGUL LETTER SSANGTIKEUT
```

3139; <compound> 1105 1160 # HANGUL LETTER RIEUL
313A; <compound> A964 1160 # 1105 1100 1160, HANGUL LETTER RIEUL-KIYEOK [the mapping
uses a N3242 character]
313B; <compound> A968 1160 # 1105 1106 1160, HANGUL LETTER RIEUL-MIEUM [the mapping
uses a N3242 character]
313C; <compound> A969 1160 # 1105 1107 1160, HANGUL LETTER RIEUL-PIEUP [the mapping
uses a N3242 character]
313D; <compound> A96C 1160 # 1105 1109 1160, HANGUL LETTER RIEUL-SIOS [the mapping
uses a N3242 character]
**313E; <compound> A97E 1160 # 1105 1110 1160, HANGUL LETTER RIEUL-THIEUTH [the mapping
uses a NEW proposed character]**
**313F; <compound> A97F 1160 # 1105 1111 1160, HANGUL LETTER RIEUL-PHIEUPH [the mapping
uses a NEW proposed character]**
3140; <compound> 111A 1160 # 1105 1112 1160, HANGUL LETTER RIEUL-HIEUH
3141; <compound> 1106 1160 # HANGUL LETTER MIEUM
3142; <compound> 1107 1160 # HANGUL LETTER PIEUP
3143; <compound> 1108 1160 # 1107 1107 1160, HANGUL LETTER SSANGPIEUP
3144; <compound> 1121 1160 # 1107 1109 1160, HANGUL LETTER PIEUP-SIOS
3145; <compound> 1109 1160 # HANGUL LETTER SIOS
3146; <compound> 110A 1160 # 1109 1109 1160, HANGUL LETTER SSANGSIOS
3147; <compound> 110B 1160 # HANGUL LETTER IEUNG
3148; <compound> 110C 1160 # HANGUL LETTER CIEUC
3149; <compound> 110D 1160 # 110C 110C 1160, HANGUL LETTER SSANGCIEUC
314A; <compound> 110E 1160 # HANGUL LETTER CHIEUCH
314B; <compound> 110F 1160 # HANGUL LETTER KHIEUKH
314C; <compound> 1110 1160 # HANGUL LETTER THIEUTH
314D; <compound> 1111 1160 # HANGUL LETTER PHIEUPH
314E; <compound> 1112 1160 # HANGUL LETTER HIEUH

314F; <compound> 115F 1161 # HANGUL LETTER A
3150; <compound> 314F 1175 # 115F 1161 1175, HANGUL LETTER AE (A-I)
3151; <compound> 115F 1163 # HANGUL LETTER YA
3152; <compound> 3151 1175 # 115F 1163 1175, HANGUL LETTER YAE (YA-I)
3153; <compound> 115F 1165 # HANGUL LETTER EO
3154; <compound> 3151 1175 # 115F 1165 1175, HANGUL LETTER E (EO-I)
3155; <compound> 115F 1167 # HANGUL LETTER YEO
3156; <compound> 3155 1175 # 115F 1167 1175, HANGUL LETTER YE (YEO-I)
3157; <compound> 115F 1169 # HANGUL LETTER O
3158; <compound> 3157 1161 # 115F 1169 1161, HANGUL LETTER WA (O-A)
3159; <compound> 3158 1175 # 115F 1169 1161 1175, HANGUL LETTER WAE (O-A-I)
315A; <compound> 3157 1175 # 115F 1169 1175, HANGUL LETTER OE (O-I)
315B; <compound> 115F 116D # HANGUL LETTER YO
315C; <compound> 115F 116E # HANGUL LETTER U
315D; <compound> 315C 1165 # 115F 116E 1165, HANGUL LETTER WEO (U-EO)
315E; <compound> 315D 1175 # 115F 116E 1165 1175, HANGUL LETTER WE (U-EO-I)
315F; <compound> 315C 1175 # 115F 116E 1175, HANGUL LETTER WI (U-I)
3160; <compound> 115F 1172 # HANGUL LETTER YU
3161; <compound> 115F 1173 # HANGUL LETTER EU
3162; <compound> 3161 1175 # 115F 1173 1175, HANGUL LETTER YI (EU-I)
3163; <compound> 115F 1175 # HANGUL LETTER I

3164; <compound> 115F 1160 # HANGUL FILLER

3165; <compound> 1114 1160 # 1102 1102 1160, HANGUL LETTER SSANGNIEUN
3166; <compound> 1115 1160 # 1102 1103 1160, HANGUL LETTER NIEUN-TIKEUT
3167; <compound> 115B 1160 # 1102 1109 1160, HANGUL LETTER NIEUN-SIOS [the mapping
uses a N3242 char, 115B]
**3168; <compound> D7C7 1160 # 1102 1140 1160, HANGUL LETTER NIEUN-PANSIOS [the mapping
uses a NEW proposed character]**
**3169; <compound> D7C8 1160 # 1105 1100 1109 1160, HANGUL LETTER RIEUL-KIYEOK-SIOS [the
mapping uses a NEW proposed character]**
316A; <compound> A966 1160 # 1105 1103 1160, HANGUL LETTER RIEUL-TIKEUT [the mapping
uses a N3242 character]
**316B; <compound> D7C9 1160 # 1105 1107 1109 1160, HANGUL LETTER RIEUL-PIEUP-SIOS [the
mapping uses a NEW proposed character]**
**316C; <compound> D7CA 1160 # 1105 1140 1160, HANGUL LETTER RIEUL-PANSIOS [the mapping
uses a NEW proposed character]**
**316D; <compound> D7FC 1160 # 1105 1159 1160, HANGUL LETTER RIEUL-YEORINHIEUH [the
mapping uses a NEW proposed character]**

316E; <compound> 111C 1160 # 1106 1107 1160, HANGUL LETTER MIEUM-PIEUP
316F; <compound> A971 1160 # 1106 1109 1160, HANGUL LETTER MIEUM-SIOS [the mapping
uses a N3242 character]
**3170; <compound> D7FD 1160 # 1106 1140 1160, HANGUL LETTER MIEUM-PANSIOS [the mapping
uses a NEW proposed character]**
3171; <compound> 111D 1160 # 1106 110B 1160, HANGUL LETTER KAPYEOUNMIEUM
3172; <compound> 111E 1160 # 1107 1100 1160, HANGUL LETTER PIEUP-KIYEOK
3173; <compound> 1120 1160 # 1107 1103 1160, HANGUL LETTER PIEUP-TIKEUT
3174; <compound> 1122 1160 # 1107 1109 1100 1160, HANGUL LETTER PIEUP-SIOS-KIYEOK
3175; <compound> 1123 1160 # 1107 1109 1103 1160, HANGUL LETTER PIEUP-SIOS-TIKEUT
3176; <compound> 1127 1160 # 1107 110C 1160, HANGUL LETTER PIEUP-CIEUC
3177; <compound> 1129 1160 # 1107 1110 1160, HANGUL LETTER PIEUP-THIEUTH
3178; <compound> 112B 1160 # 1107 110B 1160, HANGUL LETTER KAPYEOUNPIEUP
3179; <compound> 112C 1160 # 1107 1107 110B 1160, HANGUL LETTER KAPYEOUNSSANGPIEUP
317A; <compound> 112D 1160 # 1109 1100 1160, HANGUL LETTER SIOS-KIYEOK
317B; <compound> 112E 1160 # 1109 1102 1160, HANGUL LETTER SIOS-NIEUN
317C; <compound> 112F 1160 # 1109 1103 1160, HANGUL LETTER SIOS-TIKEUT
317D; <compound> 1132 1160 # 1109 1107 1160, HANGUL LETTER SIOS-PIEUP
317E; <compound> 1136 1160 # 1109 110C 1160, HANGUL LETTER SIOS-CIEUC
317F; <compound> 1140 1160 # HANGUL LETTER PANSIOS
3180; <compound> 1147 1160 # 110B 110B 1160, HANGUL LETTER SSANGIEUNG
3181; <compound> 114C 1160 # HANGUL LETTER YESIEUNG
3182; <compound> 1145 1160 # 114C 1109 1160, HANGUL LETTER YESIEUNG-SIOS
3183; <compound> 1146 1160 # 114C 1140 1160, HANGUL LETTER YESIEUNG-PANSIOS
3184; <compound> 1157 1160 # 1111 110B 1160, HANGUL LETTER KAPYEOUNPHIEUPH
3185; <compound> 1158 1160 # 1112 1112 1160, HANGUL LETTER SSANGHIEUH
3186; <compound> 1159 1160 # HANGUL LETTER YEORINHIEUH

3187; <compound> 115F 1184 # 115F 116D 1163, HANGUL LETTER YO-YA
3188; <compound> 115F 1185 # 115F 116D 1163 1175, HANGUL LETTER YO-YAE (YO-YA-I)
3189; <compound> 115F 1188 # 115F 116D 1175, HANGUL LETTER YO-I
318A; <compound> 115F 1191 # 115F 1172 1167, HANGUL LETTER YU-YEO
318B; <compound> 115F 1192 # 115F 1172 1167 1175, HANGUL LETTER YU-YE (YU-YEO-I)
318C; <compound> 115F 1194 # 115F 1172 1175, HANGUL LETTER YU-I
318D; <compound> 115F 119E # HANGUL LETTER ARAEA
318E; <compound> 115F 11A1 # 115F 119E 1175, HANGUL LETTER ARAEAE (ARAEAE-I)

Not that some of the auxiliary decomposition mappings use Hangul Jamo characters already proposed elsewhere (see N3242).

Proposed additional Hangul Jamo

For reasons given above, to facilitate the mapping between logically equivalent representations of the same Hangul basic letter sequences, nine new characters, Hangul lead consonant Jamos, are hereby proposed:

A97D; <compound> 1100 1109 # HANGUL CHOSEONG KIYEOK-SIOS [NEW PROPOSED
CHARACTER; used for the mapping of 3133]
A97E; <compound> 1105 1110 # HANGUL CHOSEONG RIEUL-THIEUTH [NEW PROPOSED
CHARACTER; used for the mapping of 313E]
A97F; <compound> 1105 1111 # HANGUL CHOSEONG RIEUL-PHIEUPH [NEW PROPOSED
CHARACTER; used for the mapping of 313F]
D7C7; <compound> 1102 1140 # HANGUL CHOSEONG NIEUN-PANSIOS [NEW PROPOSED
CHARACTER; used for the mapping of 3168]
D7C8; <compound> A964 1109 # 1105 1100 1109, HANGUL CHOSEONG RIEUL-KIYEOK-
SIOS [NEW PROPOSED CHARACTER; used for the mapping of 3169]
D7C9; <compound> A969 1109 # 1105 1107 1109, HANGUL CHOSEONG RIEUL-PIEUP-
SIOS [NEW PROPOSED CHARACTER; used for the mapping of 316B]
D7CA; <compound> 1105 1140 # HANGUL CHOSEONG RIEUL-PANSIOS [NEW PROPOSED
CHARACTER; used for the mapping of 316C]

D7FC; <compound> 1105 1159 # HANGUL CHOSEONG RIEUL-YEORINHIEUH [NEW PROPOSED CHARACTER; used for the mapping of 316D]
D7FD; <compound> 1106 1140 # HANGUL CHOSEONG MIEUM-PANSIOS [NEW PROPOSED CHARACTER; used for the mapping of 3170]

This proposal does not propose to rearrange the Hangul Jamo characters already proposed (N3242). One may still interfile the code points for the here proposed characters with the ones already proposed, assuming that the code positions for previously proposed Hangul Jamos are not already fixed.

References

- [1] A history of Korean alphabet and movable types, Ministry of Culture and Information, Republic of Korea, 1970. (Part 1 reproduces the 1444 official design document for the Hangul alphabet.)
- [2] The Korean language, Ho-Min Sohn, Cambridge University Press, ISBN 0-521-36123-0 or 0-521-36943-6, 1999. (Section 6.3 gives a translation to English of the 1444 design document for the Hangul alphabet.)
- [3] The Korean alphabet of 1446 – Expositions, OPA, The visible speech sounds, Annotated translation, Future applicability; Hwun Min Ceng Um, Sek Yen Kim-Cho, Humanity Books and AC Press, New York, ISBN 89-428-1587-1, 2002. (Reproduces, translates and analyses (in English) the 1446 official design document for the Hangul alphabet.)
- [4] The Korean alphabet, its history and structure, ed. Young-Key Kim-Renaud, University of Hawai'i Press, ISBN 0-824-81989-6, 1997.
- [5] The Korean language, structure, use and context, Jae Jung Song, ISBN 0-415-32802-0, 2005.
- [6] Hangul, Korea background series, Korean overseas information service, Seoul, Korea, 1973.
- [7] ISO/IEC JTC1/SC2/WG2 N3168, A Proposal to add new Hangul Jamo extended characters to BMP of UCS, Kim Kyongsok and Dae Hyuk Ahn, 2006-09-27.
- [8] ISO/IEC JTC1/SC2/WG2 N3242, Proposal for allocation of proposed Hangul Jamos in the BMP, Irish National Body, 2007-04-16.