

Proposal to add nine multi-letter Jamo characters

Kent Karlsson
2008-02-01

Hangul Compatibility Jamo and new Hangul Jamo

TUS 5.0, page 436 says:

=====

Hangul Compatibility Jamo: U+3130–U+318F

This block consists of spacing, nonconjoining Hangul consonant and vowel (jamo) elements. These characters are provided solely for compatibility with the KS X 1001:1998 standard. Unlike the characters found in the Hangul Jamo block (U+1100..U+11FF), the jamo characters in this block have no conjoining semantics.

The characters of this block are considered to be fullwidth forms in contrast with the halfwidth Hangul compatibility jamo found at U+FFA0..U+FFDF.

Standards. The Unicode Standard follows KS X 1001:1998 for Hangul Jamo elements.

Normalization. When Hangul compatibility jamo are transformed with a compatibility normalization form, NFKD or NFKC, the characters are converted to the corresponding conjoining jamo characters. Where the characters are intended to remain in separate syllables after such transformation, they may require separation from adjacent characters. This separation can be achieved by inserting any non-Korean character.

- U+200B zero width space is recommended where the characters are to allow a line break.
- U+2060 word joiner can be used where the characters are not to break across lines.

=====

I don't think this is the best way to view the Hangul Compatibility Jamo. I think it would be better to view them as Hangul Syllables with Filler.

E.g., U+3131, HANGUL LETTER KIYEOK, and <U+1100, U+1160> (<HANGUL CHOSEONG KIYEOK, HANGUL JUNGSEONG FILLER>) should look and behave (except, unfortunately, for normalisation) just about the same.

Similarly for U+314F, HANGUL LETTER A, and <U+115F, U+1161> (<HANGUL CHOSEONG FILLER, HANGUL JUNGSEONG A>) should look and behave (except, unfortunately, for normalisation) just about the same.

The full list of auxiliary decomposition mappings for compatibility Jamos is as follows in the list below. Note that some of the auxiliary decomposition mappings here refer to proposed characters (L2/07-103), and some even to characters otherwise not yet proposed (but are proposed below, nine of them). These decomposition mappings, together with similar decomposition mappings for multi-letter Jamos, can be used by processes where one wishes to map between different Jamo sequence representations for

the same text. And that with the same method as for canonical decomposition and composition as is used for normalisation.

With (a certain few) characters already proposed (see L2/07-103) all of the multi-letter Jamos can get auxiliary decomposition mappings similar to canonical decomposition mappings (these not listed in this document).

NOTE, HOWEVER, THAT THE AUXILIARY DECOMPOSITION 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 auxiliary
decomp. mapping uses a NEW proposed char]
3134; <compound> 1102 1160 # HANGUL LETTER NIEUN
3135; <compound> 115C 1160 # 1102 110C 1160, HANGUL LETTER NIEUN-CIEUC [the auxiliary
decomp. mapping uses a proposed char]
3136; <compound> 115D 1160 # 1102 1112 1160, HANGUL LETTER NIEUN-HIEUH [the auxiliary
decomp. mapping uses a proposed char]
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 auxiliary
decomp. mapping uses a proposed char]
313B; <compound> A968 1160 # 1105 1106 1160, HANGUL LETTER RIEUL-MIEUM [the auxiliary
decomp. mapping uses a proposed char]
313C; <compound> A969 1160 # 1105 1107 1160, HANGUL LETTER RIEUL-PIEUP [the auxiliary
decomp. mapping uses a proposed char]
313D; <compound> A96C 1160 # 1105 1109 1160, HANGUL LETTER RIEUL-SIOS [the auxiliary
decomp. mapping uses a proposed char]
313E; <compound> A97E 1160 # 1105 1110 1160, HANGUL LETTER RIEUL-THIEUTH [the auxiliary
decomp. mapping uses a NEW proposed char]
313F; <compound> A97F 1160 # 1105 1111 1160, HANGUL LETTER RIEUL-PHIEUPH [the auxiliary
decomp. mapping uses a NEW proposed char]
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, should NOT display as a space, but as some
blotted glyph (e.g. the .notdef glyph), as should half-width 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, [see 115B] HANGUL LETTER NIEUN-SIOS [the
auxiliary decomp. mapping uses a proposed char]
3168; <compound> D7C7 1160 # 1102 1140 1160, HANGUL LETTER NIEUN-PANSIOS [the auxiliary
decomp. mapping uses a NEW proposed char]
3169; <compound> D7C8 1160 # 1105 1100 1109 1160, HANGUL LETTER RIEUL-KIYEOK-SIOS [the
auxiliary decomp. mapping uses a NEW proposed char]
316A; <compound> A966 1160 # 1105 1103 1160, HANGUL LETTER RIEUL-TIKEUT [the auxiliary
decomp. mapping uses a proposed char]
316B; <compound> D7C9 1160 # 1105 1107 1109 1160, HANGUL LETTER RIEUL-PIEUP-SIOS [the
auxiliary decomp. mapping uses a NEW proposed char]
316C; <compound> D7CA 1160 # 1105 1140 1160, HANGUL LETTER RIEUL-PANSIOS [the auxiliary
decomp. mapping uses a NEW proposed char]
316D; <compound> D7FC 1160 # 1105 1159 1160, HANGUL LETTER RIEUL-YEORINHIEUH [the
auxiliary decomp. mapping uses a NEW proposed char]
316E; <compound> 111C 1160 # 1106 1107 1160, HANGUL LETTER MIEUM-PIEUP
316F; <compound> A971 1160 # 1106 1109 1160, HANGUL LETTER MIEUM-SIOS [the auxiliary
decomp. mapping uses a proposed char]
3170; <compound> D7FD 1160 # 1106 1140 1160, HANGUL LETTER MIEUM-PANSIOS [the auxiliary
decomp. mapping uses a NEW proposed char]
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 L2/07-103).

Proposed additional Hangul Jamo

While I don't think one really should use multi-letter Jamos:

- 1) many are already encoded,
- 2) more are on the way (L2/07-103),
- 3) they are likely to be used, even though Hangul is an alphabetic script and all Hangul texts can be expressed using only Hangul single-letter Jamos,
- 4) the (few) multi-letter Jamos that occur in decompositions of precomposed Hangul syllables (U+AC00—U+D7A3) will occur in NFD texts formed from texts that use precomposed Hangul syllables,
- 5) they can be used for ad hoc decomposition and composition to/from single-letter Jamos representation of the text (using the same method as for NFD/NFC).

It is point five here that can cannot be fulfilled with the same restrictions as used for canonical decomposition mappings without allocating nine new (not already proposed) multi-letter Hangul Jamos.

In order to get these auxiliary decomposition mappings (not involved in normalisation) to be left-to-right recomposable (after full auxiliary decomposition), and only have two characters in the auxiliary decomposition, nine new characters, Hangul lead consonant Jamos, are hereby proposed:

```
A97D; <compound> 1100 1109 # NEW proposed HANGUL CHOSEONG KIYEOK-SIOS [NEW
PROPOSED CHARACTER; used for the auxiliary decomp. mapping of 3133]
A97E; <compound> 1105 1110 # NEW proposed HANGUL CHOSEONG RIEUL-THIEUTH [NEW
PROPOSED CHARACTER; used for the auxiliary decomp. mapping of 313E]
A97F; <compound> 1105 1111 # NEW proposed HANGUL CHOSEONG RIEUL-PHIEUPH [NEW
PROPOSED CHARACTER; used for the auxiliary decomp. mapping of 313F]

D7C7; <compound> 1102 1140 # NEW proposed HANGUL CHOSEONG NIEUN-PANSIOS [NEW
PROPOSED CHARACTER; used for the auxiliary decomp. mapping of 3168]
D7C8; <compound> A964 1109 # 1105 1100 1109, NEW proposed HANGUL CHOSEONG
RIEUL-KIYEOK-SIOS [NEW PROPOSED CHARACTER; used for the auxiliary decomp.
mapping of 3169]
D7C9; <compound> A969 1109 # 1105 1107 1109, NEW proposed HANGUL CHOSEONG
RIEUL-PIEUP-SIOS [NEW PROPOSED CHARACTER; used for the auxiliary decomp.
mapping of 316B]
D7CA; <compound> 1105 1140 # NEW proposed HANGUL CHOSEONG RIEUL-PANSIOS [NEW
PROPOSED CHARACTER; used for the auxiliary decomp. mapping of 316C]

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

This proposal does not propose to rearrange the Hangul Jamo characters already proposed (L2/07-103). 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.