

Hangul Jamo and Hangul precomposed syllable description

A response to ISO/IEC JTC1/SC2/WG2 N3095 (L2/06-286) and
ISO/IEC JTC1/SC2/WG2 N3113 (L2/06-289)

Kent Karlsson
2006-08-09

The assumption in documents N3095 and N3113 seems to be that a proposed alignment of the text in clause 26.1 of 10646 with Unicode would be the reflection of a new misunderstanding. However, the Unicode specification, including the canonical decomposition of Hangul precomposed syllable characters have been fixed for many years now. Indeed, the decompositions and equivalences are now stabilised and cannot be changed. Thus many of the suggestions in N3095 and N3113 cannot be carried through. Indeed, I would suggest not carrying through any of the suggestions in N3095 and N3113. Furthermore, I would argue that the only Hangul Jamos to be encoded in a new version of 10646 (or an amendment) are *single-letter* Jamos that are not already encoded. No more multi-letter Jamos, and many of the existing ones should not be used.

The lack of canonical decompositions for multi-letter Hangul Jamos unfortunately leads to multiple possible spellings, which are not canonically equivalent, of exactly the same words (written in Hangul). Since new canonical decompositions cannot be added (to Unicode) for the existing multi-letter Hangul Jamos, a usage recommendation should be made instead. I would recommend the following one: *Only the multi-letter Jamos that are used in the decompositions of precomposed Hangul syllable characters should be used (and then preferred over using the individual letters of the multi-letter Jamos); in all other cases, use just the single letter Jamos; and then of course allow canonical equivalents (precomposed Hangul syllables) to that. That rule can be refined, if desired.*

For decompositions and equivalences, in particular the arithmetic decomposition of precomposed Hangul syllable characters, a reference to the Unicode standard should be made.

In order to clarify the conjoining behaviour of Hangul Jamo and Hangul precomposed syllables, I suggest that the following text replacement is done. This clarifies "conjoining" in terms of conditional combining behaviour. This avoids explaining "conjoining" in isolation, and instead relates it to the notion of "combining". The same formulation can be used for the Unicode standard (without any factual change, just a change in exposition).

=====

-----OLD TEXT TO BE REPLACED-----

An incomplete syllable which starts with a Jungseong or a Jongseong shall be preceded by a CHOSEONG FILLER (0000 115F).

An incomplete syllable composed of a Choseong alone shall be followed by a JUNGSEONG FILLER (0000 1160).

The implementation level 3 shall be used for the Hangeul syllable composition method.

NOTE 1 - Hangeul Jamo are not combining characters.

NOTE 2 - When a combining character such as HANGUL SINGLE DOT TONE MARK (0000 302E) is intended to apply to a sequence of Hangeul Jamo it should be placed at the end of the sequence, after the Hangeul Jamo character which completes the syllable block.

=====

-----NEW, REPLACEMENT, TEXT-----

The implementation level 3 shall be used for the Hangeul syllable composition method.

The Hangeul Jamo characters and the precomposed Hangeul syllable characters are conjoining, which more precisely means that:

- A *Choseong Jamo* character that immediately follows a Choseong Jamo character shall be treated as a combining character.
- A *Jungseong Jamo* character that immediately follows a Choseong Jamo character, a Jungseong Jamo character, or a precomposed Hangeul syllable character that has no Jongseong part shall be treated as a combining character.
- A *Jongseong Jamo* character that immediately follows a Jungseong Jamo character, a Jongseong Jamo character, or a precomposed Hangeul syllable character shall be treated as a combining character.
- A *precomposed Hangeul syllable* character that immediately follows a Choseong Jamo character shall be treated as a combining character.
- All other instances of a Choseong Jamo character, a Jungseong Jamo character, a Jongseong Jamo character, or a precomposed Hangeul syllable character shall be treated as a base character (i.e. not combining).

Thus, a Hangul syllable consists of a base (by the above rules) Hangul Jamo character or a base (by the above rule) precomposed Hangul syllable character followed by zero or more combining (by the above rules) Hangul Jamo and combining (by the above rule) precomposed Hangul syllable characters.

When a combining (not by the above rules) character, such as HANGUL SINGLE DOT TONE MARK (0000 302E), is applied to a sequence of conjoining Hangul characters, it is placed after the conjoining Hangul character which completes the syllable block.

NOTE 1 - Hangul "LETTER" characters are base characters in all contexts. Though not equivalent, a Hangul "letter" can be represented using Hangul Jamos and Hangul Jamo FILLER characters. Strictly speaking, the Jamo FILLER characters are needed only if several Hangul "letters" are adjacent. This is useful when talking about isolated Hangul letters (not part of a word) or incomplete orthographic syllables, as well as for the historic "linear Hangul" writing, also when using letters that are only represented as Jamo and do not have any "LETTER" character allocated.

NOTE 2 - The recommended spelling using Hangul Jamo is to use only single-letter Jamos, except for those multi-letter Jamos that are used in precomposed Hangul syllables (five double-letter initial consonant clusters, eleven vowel clusters, and thirteen final consonant clusters).

NOTE 3 - In practice, in a Hangul syllable block there will be at most three Choseong Jamos in a row, at most three Jungseong Jamos in a row, at most three Jongseong Jamos in a row, and horizontal Jungseong precede vertical Jungseong.

NOTE 4 - Decomposition and composition of precomposed Hangul syllable characters to/from Hangul Jamo characters is specified in the Unicode standard.

=====