

# A consequence of being the first to adopt a standard is that you may end up being the only one to adopt it: The sad story of Korean jamo

 [devblogs.microsoft.com/oldnewthing/20201009-00](https://devblogs.microsoft.com/oldnewthing/20201009-00)

October 9, 2020



Raymond Chen

If you ask Windows to break the Korean string U+1100 U+1161 into graphemes, it will get broken up into two characters. U+1100 is HANGUL CHOSEONG KIYEOK ( ͆ ) and U+1161 is HANGUL JUNGSEONG A ( ͇ ).

Korean is written in the Hangul alphabet, and characters are composed of units known as *jamo*. In the above example, the two jamo combine to form the single syllable 가.

If the two code points combine to form a single character, why are they treated as separate graphemes? ICU treats them as a single grapheme. iOS treats them as a single grapheme. Android treats them as a single grapheme. Everybody treats them as a single grapheme, except Windows. Why does Windows do things wrong?

This is another case where Windows adopted a standard before anybody else and ended up suffering from the first-mover curse. In this case, Windows is following the Korean standard KS X 1026 and treating the characters as separate. (Indeed, the case of U+1100 U+1161 is the example used in the specification.) So the question isn't why Windows is doing things wrong. The question is why *everybody else* is doing things wrong.

Everybody else does things wrong because everybody else ignores the standard. But if you're the only one doing things right, then you end up looking wrong.

In practice, therefore, there are two competing standards. You have the *de jure* standard, which says that the characters are separate, and the *de facto* standard, which says that the characters form a single grapheme.

If you are interoperating with other systems, you would be best served by following the conventions that those other systems follow when communicating with them. In practice, this will usually mean that you need to ignore what the Unicode and Korean standards committees recommend, and instead do “what everybody else is doing.” Since ICU is one of those “everybody else”s, you can switch to using ICU to decompose your strings.

Today is Hangeul Day, a Korean national holiday commemorating the invention of the Hangeul alphabet.

**Bonus reading:** [Frequently Asked Questions about Korean and Unicode.](#)

[Raymond Chen](#)

**Follow**

