

Subject: Inconsistency in default value for BMG
From: Mark Davis
Date: 2012-05-09
Live: <http://goo.gl/iiGeK>

We have a contradiction in the definition of the default value for Bidi_Mirroring_Glyph in 6.1 that needs to be resolved.

Proposal

My recommendation is to change the following line in the next version of the UCD:

```
PropertyAliases-6.1.0d20.txt
...
117: # @missing: 0000..10FFFF; Bidi_Mirroring_Glyph; <none>
=> 117: # @missing: 0000..10FFFF; Bidi_Mirroring_Glyph; <code point>
```

and delete the note:

```
# BidiMirroring-6.1.0.txt
...
# Note: (2011-12-19) There is an inconsistency between the
# following statement about the default value
# of the Bidi_Mirroring_Glyph property and the
# value of the @missing line for Bidi_Mirroring_Glyph in
# PropertyValueAliases.txt. This inconsistency was discovered too
# late in the release process to be resolved by
# the UTC. The inconsistency will be resolved in a future revision.
```

Problem

In BidiMirroring.txt (in U6.1 and U6.0) we say:

```
# BidiMirroring-6.1.0.txt
...
# Formally, the default value of the Bidi_Mirroring_Glyph property
# for each code point is the code point itself, unless a mapping to
# some other character is specified in this data file. When a code
# point has the default value for the Bidi_Mirroring_Glyph property,
# that means that no other character exists whose glyph is suitable
# for character-based mirroring.
```

That is also consistent with UAX 44 (in U5.2-U6.1):

4.2.9 Default Values

- ...
- For string properties, including the definition of foldings, the default value is the code point of the character itself.

However, the following data in PropertyValueAliases (in U5.1-U6.1) is inconsistent with that principle.

```
PropertyAliases-6.1.0d20.txt
...
117: # @missing: 0000..10FFFF; Bidi_Mirroring_Glyph; <none>
```

We found this out when testing against ICU, which is where this property comes from, and which has always treated it as codepoint. It came too late in the release of U6.1 to make a change, and instead the editorial committee just recorded the issue in the note cited above:

Actions: Ken to remove the line in BidiMirroring, add a line of documentation to note the problem (but not slant the solution either way: that will be up to the UTC to decide what to do).

We need to resolve the issue one way or another.

Reasoning

If Bidi_Mirroring_Glyph had the default value of <none>, it would be the only String property that does so; all others default to <code point>. String properties are all made to be applied to strings (directly or indirectly): that's the reason that in UAX 44 we have the rule:

- For string properties, including the definition of foldings, the default value is the code point of the character itself.

That is by design, so that when the properties are successively applied to characters in a string, the characters without a mapping map to themselves. For example, if we apply the Simple_Uppercase_Mapping mapping to the string “A(s)”, we expect to get “A(s)”—*not* nuking the characters that didn't have a mapping, getting “S”.

Now, some string properties explicitly map some characters to “” as for U+00AD below:

```
00AA          ; NFKC_CF; 0061          # Lo          FEMININE ORDINAL INDICATOR
00AD          ; NFKC_CF;              # Cf          SOFT HYPHEN
```

However, those are explicit; they are not the default value for the property, which is to map to <code point>. This is so that (for example) reserved codepoints are maintained—not destroyed—in applying the property to a string. So we have the data line indicating that the default value is the code point:

```
# @missing: 0000..10FFFF; NFKC_CF; <code point>
```

Now, for a lot of implementations, this will not matter; BMG is normally just applied to characters in the context of the BIDI algorithm that mirror. But for those implementations that *do* have separate APIs (like ICU), it would be much better to not break the general principle that string properties have a default = code point.

The other alternatives for resolving this inconsistency would be:

1. to break the general principle for string properties (which I'm strongly against), or
2. to change BidiMirroringGlyph to be a Misc property (like, say, Jamo_Short_Name).