

Property Issues for U6.2

Mark Davis

Live doc: <http://goo.gl/zfBQU>

Proposal

The following is a proposal to adjust certain property values for consistency, or annotate the reason for the inconsistency.

1. **Require that General_Category=Number implies a Numeric_Type** (that is, NT≠None). Change the 4 characters that are inconsistent with that policy to have either (a) GC=So, or (b) NT=Numeric & NV=<whatever is appropriate>: U+12432, U+12433, U+12456, U+12457. Add an invariant test, and request a stability policy.
2. **Setting Consistent Uppercase Properties.** Change [U+1F150..U+1F169](#), [U+1F170..U+1F18A](#) to have the property Other_Uppercase=True in PropList.txt
3. **Annotating or fixing inconsistent properties.** Annotate why [U+2129](#), [U+2118](#), [U+213A](#), [U+2141..U+2144](#) have unusual Upper/Lowercase properties, or fix them.

Details and Background

1. Require that General_Category=Number implies a Numeric_Type

We should have the invariant that if a codepoint has General_Category=Number, then it has a Numeric_Type (that is, one not equal to None). It is completely counter-intuitive that a character has GC=Number, but has no Numeric_Type or value. (There *is* good reason for the reverse; having characters with Numeric_Type but not GC=Number: GC is the “predominant” category, and can’t always make the distinctions that other properties like Numeric_Type can make.)

We have the following 4 characters for which is not true:

```
U+12432 (    ) CUNEIFORM NUMERIC SIGN SHAR2 TIMES GAL PLUS DISH
U+12433 (    ) CUNEIFORM NUMERIC SIGN SHAR2 TIMES GAL PLUS MIN
U+12456 (    ) CUNEIFORM NUMERIC SIGN NIGIDAMIN
U+12457 (    ) CUNEIFORM NUMERIC SIGN NIGIDAESH
```

We should change either:

1. the GC (by making these characters not be numeric) or
2. the NT and NV (by making them numeric)

It is not terribly important which way we go, since the number of general-purpose implementations that care about the numeric status of these characters is small; and anyone who is really handling these as numbers would know about them and have to handle them specially.

2. Setting Consistent Uppercase Properties

The following characters sort like [U+24B6](#) (Ⓐ) CIRCLED LATIN CAPITAL LETTER A does, but [U+24B6](#) (Ⓐ) CIRCLED LATIN CAPITAL LETTER A has Uppercase=True and they don't. They otherwise are very similar, and have the same primary/secondary collation weights. There is no good reason for them to not also have

Uppercase=True.

[U+1F150](#) () NEGATIVE CIRCLED LATIN CAPITAL LETTER A

...

[U+1F169](#) () NEGATIVE CIRCLED LATIN CAPITAL LETTER Z

[U+1F170](#) () NEGATIVE SQUARED LATIN CAPITAL LETTER A

...

[U+1F189](#) () NEGATIVE SQUARED LATIN CAPITAL LETTER Z

[U+1F18A](#) () CROSSED NEGATIVE SQUARED LATIN CAPITAL LETTER P

UCA sample data

24B6 ; [.15D5.0020.000C.24B6] # CIRCLED LATIN CAPITAL LETTER A

1F150 ; [.15D5.0020.000C.1F150] # NEGATIVE CIRCLED LATIN CAPITAL LETTER A

The change would be to add lines like the ones already in PropList.txt

24B6.24CF ; Other_Uppercase # So [26] CIRCLED LATIN CAPITAL LETTER A..CIRCLED LATIN CAPITAL LETTER Z

3. Annotating or fixing inconsistent properties

The following seem inconsistent. We should either annotate why they are exceptional, or change their properties (by adding to Other_Uppercase or Other_Lowercase).

Characters without Lowercase=true:

[U+2129](#) (ι) TURNED GREEK SMALL LETTER IOTA

Characters without Uppercase=true

[U+2118](#) (Ⓟ) SCRIPT CAPITAL P

[U+213A](#) (Ⓚ) ROTATED CAPITAL Q

[U+2141](#) (Ⓛ) TURNED SANS-SERIF CAPITAL G

[U+2142](#) (Ⓛ) TURNED SANS-SERIF CAPITAL L

[U+2143](#) (Ⓛ) REVERSED SANS-SERIF CAPITAL L

[U+2144](#) (Ⓛ) TURNED SANS-SERIF CAPITAL Y

When similar characters are Uppercase=true

[U+2102](#) (Ⓒ) DOUBLE-STRUCK CAPITAL C

[U+2107](#) (Ⓔ) EULER CONSTANT

[U+210B](#) (Ⓗ) SCRIPT CAPITAL H

[U+210C](#) (Ⓗ) BLACK-LETTER CAPITAL H

[U+210D](#) (Ⓗ) DOUBLE-STRUCK CAPITAL H

[U+2110](#) (Ⓘ) SCRIPT CAPITAL I

[U+2111](#) (Ⓘ) BLACK-LETTER CAPITAL I

[U+2112](#) (Ⓛ) SCRIPT CAPITAL L

[U+2115](#) (Ⓖ) DOUBLE-STRUCK CAPITAL N

[U+2119](#) (Ⓟ) DOUBLE-STRUCK CAPITAL P

[U+211A](#) (Ⓟ) DOUBLE-STRUCK CAPITAL Q

[U+211B](#) (Ⓡ) SCRIPT CAPITAL R

[U+211C](#) (Ⓡ) BLACK-LETTER CAPITAL R

[U+211D](#) (Ⓡ) DOUBLE-STRUCK CAPITAL R

[U+2124](#) (Ⓔ) DOUBLE-STRUCK CAPITAL Z

[U+2128](#) (Ⓕ) BLACK-LETTER CAPITAL Z

[U+212C](#) (*ℬ*) SCRIPT CAPITAL B
[U+212D](#) (*ℭ*) BLACK-LETTER CAPITAL C
[U+2130](#) (*ℰ*) SCRIPT CAPITAL E
[U+2131](#) (*ℱ*) SCRIPT CAPITAL F
[U+2133](#) (*ℳ*) SCRIPT CAPITAL M
[U+213E](#) () DOUBLE-STRUCK CAPITAL GAMMA
[U+213F](#) () DOUBLE-STRUCK CAPITAL PI