

Title: Standardized Variation Sequences stability

Source: Michel Suignard

Status: Expert contribution

Distribution: L2

Summary:

This document discusses stability consideration concerning existing Standardization Variation Sequences and is seeking clarification about policy concerning possible removal of existing sequences. It is suggested that removal of existing standardized variants should be possible, provided that the removal is documented as comments in the relevant data file. In parallel, the Unicode core specification should clarify which type of variants sequences can be subject to such policy.

General

Version 15.0 of the Unicode Standard added 94 Standardized Variation Sequences representing rotation variants of existing Egyptian Hieroglyphs. These sequences are documented in <https://www.unicode.org/L2/L2022/22012r-hieroglyph-rotations.pdf>. Note that only sequences using the Variation Selectors U+FE00 to U+FE02 were included (rotation angles multiple of 90 degree). The document also showed other sequences with other Variation Selectors to indicate other angles, but these sequences were not included in Unicode 15.0 or 15.1.

After further consideration, Egyptologist have determined that three of these sequences related to U+13092, U+130A9 and U+13403 should be deprecated or better removed. If they are just deprecated, there would still appear in the code charts because the process of showing standardized variants in the names list is driven by their existence in the StandardizedVariants.txt file part of the Unicode UCD data set; current version in <https://unicode.org/Public/UNIDATA/StandardizedVariants.txt>.

However, there is no clear policy concerning the removal of standardized variation sequences. They are not in scope of the stability policy as in https://unicode.org/policies/stability_policy.html. There is also a precedent of removing standardized variation sequences as attested in the file itself:

```
# The following two entries were originally defined for Unicode 3.2
# but were determined to be in error and were removed from the list
# of standardized variation sequences. The entries are left commented out
# in the file for the historical record of changes made to the data.
#2278 FE00; with vertical stroke; # NEITHER LESS-THAN NOR GREATER-THAN
#2279 FE00; with vertical stroke; # NEITHER GREATER-THAN NOR LESS-THAN
```

The removal goes back to Unicode 4.0 (2003); prior, they were added in Unicode 3.2 (2002). However, the situation is a bit similar, because the interval between the addition of the new sequences in Unicode 15.0 (2022) and their possible removal (as of now) is also short. (The author has not verified whether other sequences were removed from the data file in other versions of Unicode without any remaining comments.)

Another useful reference is section 23.4 ‘Variation Selectors’ of the Unicode Standard (see <https://www.unicode.org/versions/Unicode15.0.0/ch23.pdf> and especially the following paragraph (underline added):

The standardization or support of a particular variation sequence does *not* limit the set of glyphs that can be used to represent the base character alone. If a user *requires* a visual distinction between a character and a particular variant of that character, then fonts must be used to make that distinction. The existence of a variation sequence does not preclude the later encoding of a new character with distinct semantics and a similar or overlapping range of glyphs.

While the section clearly allows the addition of characters that could be represented by an existing variation sequence, it does not make global statements about stability or immutability of these sequences, with the possible exception of standardized variant sequences for CJK compatibility ideographs which are defined as ‘*normalization-stable representations of the CJK compatibility ideographs*’.

Rationale for removal of these three standardized variation sequences

U+13092 has currently the following related variation sequence:

13092 FE00; rotated 90 degrees; # EGYPTIAN HIEROGLYPH D027A

13092  EGYPTIAN HIEROGLYPH D027A
13092
 rotated 90 degrees
13092 FE00

Because U+13092 is itself a non-core character (U+13091 is the preferred representation), it does not need a variation sequence.

U+130A9 has currently the following related variation sequence:

130A9 FE01; rotated 180 degrees; # EGYPTIAN HIEROGLYPH D047

130A9  EGYPTIAN HIEROGLYPH D047
130A9
 rotated 180 degrees
130A9 FE01

The issue is related to U+130C6:  which is a mirrored image of the variant. Both U+130A9 and U+130C6 have the same function value: dr.t (HAND), but the first code point is a classifier while the second is a logogram. Therefore the need for the variant is questionable.

Concerning U+13404 which has currently the following related variation sequence:

13403 FE00; rotated 90 degrees; # EGYPTIAN HIEROGLYPH Z015I

13403  EGYPTIAN HIEROGLYPH Z015I
13403
 rotated 90 degrees
13403 FE00

Egyptologists have now determined that a separate encoding of the rotated form would be preferable. Without going too much in details, it is related to the fact that the two groups of numerals Z015x (from U+133FA to U+13403) and Z016x (from U+13404 to U+1340C) are used to count different categories, and rotating U+13403 (Z015I) would put the rotated version in the other group, which is not advisable.

----end