

WG2 N4792R
L2/17-076R
Date: 2017-03-24

Title: Revised proposal for the encoding of an Egyptological YOD

Source: Michel Suignard

Status: Individual Contribution

Distribution: UTC, WG2

References: WG2 N2043 (L2/99-246), N2241(L2/00-293), N3382R(L2/08-049R), N3431(L2/08-087), N3432(L2/08-146), N3487(L2/08-272)

Summary: This document contains a draft proposal for the encoding of an Egyptological YOD in the Latin Extended-D block. While this character has been proposed before and rejected, the proposed alternative solutions have not been implemented by the vendors, therefore harming the user community and it is time to come back to an easier solution which is the encoding of a non-decomposable character.

Introduction

The story of the Egyptian transliteration characters goes a long way back. These characters were first introduced in document WG2 N2043 and finally proposed for encoding in document N2241, almost 17 years ago. Four of the six proposed characters were encoded in the Latin Extended-D block, but the proposed Egyptological YOD was not accepted because it was thought that it could be represented by a combining sequence using the Latin Letter I followed by either U+0313 COMBINING COMMA ABOVE or U+0357 COMBINING RIGHT HALF RING ABOVE. None of them fully represented the visual aspect of the character. Then it was suggested to use U+0486 COMBINING CYRILLIC PSILI PNEUMATA as the combining diacritic for this purpose, making its script property 'Inherited' (instead of 'Cyrillic'), and adding an annotation to the code point in the name list (ref N3382R). These suggestions were consequently rejected by document N3431, endorsed by N3432, recommending '*a specialized font for Egyptian transliteration be created that would permit the appropriate placement of the diacritic*'.

Then a new document N3487 was created requesting the additions of 6 characters (LATIN LETTER A, I, and U with SPIRITUS LENIS in both capital and small forms), with the LATIN SMALL LETTER I WITH SPIRITUS LENIS annotated: "used for Egyptological yod". The whole set was to be used for transliteration of the Ugaritic writing system. This was not accepted by the UTC, but instead, the previously rejected script property change for U+0486 COMBINING CYRILLIC PSILI PNEUMATA was accepted (but no added annotation).

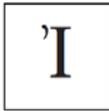
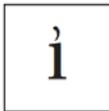
The forth and back has apparently confused implementers because there is no commonly accepted representation of the letter (nothing worse than having an undefined representation model for the diacritic mark).

Making the matter worse is the fact most Egyptian transliteration are done in an italicized style, and none of the commonly available commercial fonts produce a result which is acceptable when combining an Italic 'i' with any of the suggested diacritics. Clearly the Cyrillic PSILI PNEUMATA has the right shape in term of diacritic, but the confusion brought by documents N3431, 3432, and the reversal done later by the UTC has ensured that there is no commercial solution. Furthermore, a specialized font for Egyptian transliteration with no clear guidance for the encoding of the sequence has no chance to be used in an interoperable fashion.

The author faced the issue when creating an extension of the Egyptian Hieroglyphs and trying to transliterate the ideograms and logograms associated with the proposed characters and add them in a database that can be shared with the user community. There is simply no good solution. There are fonts using either incorrect code points or PUA for the YOD. Alternatively, many users just go around and use the letter 'j'.

1) By far the simplest solution is to revisit the issue and re-introduce the concept of a non-decomposable character representing the YOD.

We could simply use the original glyph as proposed in N2241:

	0236	LATIN CAPITAL LETTER EGYPTOLOGICAL YOD
	0237	LATIN SMALL LETTER EGYPTOLOGICAL YOD

It should be noted that the common representations look more like *I* and *i* because Egyptian transliteration is commonly italicized. This fact could be mentioned in an annotation in the name list because the code chart glyphs would de facto be extremely rare. In addition, the mark is always on the left of the I in capital form, unlike what would be typically done for a mark added to a capital I (on top).

The code positions could be U+A7BA and U+A7BB (next available code positions), properties as follow:

```
A7BA;LATIN CAPITAL LETTER EGYPTOLOGICAL YOD;Lu;0;L;;;;N;;;A7BB;  
A7BB;LATIN SMALL LETTER EGYPTOLOGICAL YOD;Ll;0;L;;;;N;;;A7BA;A7BA
```

There are not confusable with any existing code points, but the new characters could be confused with the sequence of the Latin letter I, Latin letter IOTA, Greek letter IOTA, followed by a diacritical mark resembling the top of the glyph. Given the existing confused situation, this does not exacerbate the risk in a meaningful way. There is a limited need to use a YOD in general purpose identifiers; its main purpose is to represent transliteration of historic Egyptian text originally written using Egyptian hieroglyphs.

In the Egyptian alphabet, the Egyptological yod is sorted after the Egyptological aleph (U+A723), and behaves like a normal letter. Obviously, these proposed characters would have no decompositions into a base letter I followed by a diacritic.

2) **Alternatively, we could encode all 6 characters originally proposed by N3487 (suggested code points are not available anymore)**

À	A790	LATIN CAPITAL LETTER A WITH SPIRITUS LENIS
à	A791	LATIN SMALL LETTER A WITH SPIRITUS LENIS
		• used in transliteration of Ugaritic
Ĭ	A792	LATIN CAPITAL LETTER I WITH SPIRITUS LENIS
ĭ	A793	LATIN SMALL LETTER I WITH SPIRITUS LENIS
		• used for Egyptological yod
Û	A794	LATIN CAPITAL LETTER U WITH SPIRITUS LENIS
û	A795	LATIN SMALL LETTER U WITH SPIRITUS LENIS
		• used in transliteration of Ugaritic

With the understanding that LATIN SMALL LETTER I WITH SPIRITUS LENIS would be the Egyptian YOD. Although 'SPIRITUS LENIS' is a rather rarely Latin-based term used for 'smooth breathing', which would be 'Psili pneumata' in Greek. Obviously PSILI PNEUMATA' could never be used in the name of a non-decomposable character.

3) **The other alternative would be to encode a single COMBINING SPIRITUS LENIS.**

There is obviously a risk of some confusion with the already encoded U+0486 COMBINING CYRILLIC PSILI PNEUMATA, but the main point of adding a new diacritical mark would be to provide strict guidance concerning the rendering of the 'smooth' breathing in conjunction with the 3 vowels A, I, and U, including the italicized form of the small letter 'i' used for Egyptological YOD.

The author just wants to do its research work on Egyptian Hieroglyphs extension for which a stable solution for the Egyptian YOD is required, so any of these 3 alternative solutions would be acceptable

The author would be ready to create a full summary form when one of the alternative is selected.
