ISO/IEC JTC1/SC2/WG2 N3487
L2/08-272
2008-08-04

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: **Working Group Document**
Title: **Proposal to encode Egyptological Yod and similar characters in the UCS**
Source: **Michael Everson**
Status: **Individual Contribution**
Date: **2008-08-04**
Replaces: **N3382 (2008-04-08), N2241 (2000-08-27)**

This proposal requests the encoding of six Latin characters for Egyptological and Ugariticist use. The characters are atomic characters with *no decomposition*. If this proposal is accepted, the following characters will exist:

Ꞔ	A790	LATIN CAPITAL LETTER A WITH SPIRITUS LENIS
ꞔ	A791	LATIN SMALL LETTER A WITH SPIRITUS LENIS
		• used in transliteration of Ugaritic
ꞕ	A792	LATIN CAPITAL LETTER I WITH SPIRITUS LENIS
ꞕ	A793	LATIN SMALL LETTER I WITH SPIRITUS LENIS
		• used for Egyptological yod
Ꞗ	A794	LATIN CAPITAL LETTER U WITH SPIRITUS LENIS
ꞗ	A795	LATIN SMALL LETTER U WITH SPIRITUS LENIS
		• used in transliteration of Ugaritic

The encoding of Egyptological yod has been a topic of discussion for nearly a decade. N2048, 1999-07-24, "On the apostrophe and quotation mark, with a note on Egyptian transliteration characters" discusses the issues in general, and was followed by N2241, 2000-08-27, "Proposal to add 6 Egyptological characters to the UCS", which requested the encoding of EGYPTOLOGICAL ALEF (now at U+A722..A723) and EGYPTOLOGICAL AIN (now at U+A724..A725). It also requested the Egyptological yod proposed here as LATIN LETTER I WITH SPIRITUS LENIS. In 2000, there was no consensus about how to encode Egyptological yod.

Document N3382R, 2008-04-08 "EGYPTOLOGICAL YOD and Cyrillic breathings" proposed to use the existing character U+0486 COMBINING CYRILLIC PSILI PNEUMATA by changing its character property from "Script=Cyrillic" to "Script=Inherited" and by specifying its positioning behaviour when used with Latin characters. No action was taken, in part because there did not appear to be consensus amongst experts in North America and experts in Europe.

At the July 2008 meeting of the IAE Computer Group (Informatique et Egyptologie, I&E) in Vienna, I gave a talk entitled "Yod, Unicode, and future options for encoding Egyptian". In that talk I outlined the options available to Egyptologists, first noting a number of preliminary points:

- A number of transliteration characters are already encoded: ḥ, ḫ, š, ḳ, ṯ, ḏ (or can be represented as a base character plus a diacritical mark).
- Recently alef and ayin were encoded: ꜣ U+A722, ꜣ U+A723, ꜥ U+A724, ꜥ U+A725.
- Egyptological Yod ꜧ ꜧ was proposed to be written with U+0486 ◌҆ COMBINING CYRILLIC PSILI PNEUMATA (also used for Ugaritic ꜣ ꜣ and ꜧ ꜧ).
- Experts in the U.S. and Europe could not come to consensus about how to encode Egyptological Yod, though a proposal was put forward after the 2006 I&E meeting.

Options which can be implemented immediately:

A.1: Use the existing U+0313 ◌̓ COMBINING COMMA ABOVE.
- Advantage: No need to wait as the character already exists.
- Disadvantage: Character is not really a spiritus lenis (though it is used for that in Greek). Sits *above* Latin capital letters—the *wrong* behaviour for us.
- Disadvantage: Obliges us to have special Egyptologist-only fonts rather than generic fonts like Gentium.

A.2: Use the existing U+0357 ◌͗ COMBINING RIGHT HALF RING ABOVE.
- Advantage: No need to wait as the character already exists.
- Disadvantage: Character is not really a spiritus lenis, but a half-ring. Sits *above* Latin capital letters—the *wrong* behaviour for us.
- Disadvantage: Obliges us to have special Egyptologist-only fonts rather than generic fonts like Gentium.

A.3: Use the existing U+0486 ◌҆ COMBINING CYRILLIC PSILI PNEUMATA by changing its script property to "Common" from "Cyrillic" specific. (The script property change would not harm people using it already for Cyrillic as for example they already use "Common" characters.)
- Advantage: No need to wait as the character already exists.
- Disadvantage: Requires UTC to agree to change the script property.
- Disadvantage: Obliges us to ensure that font developers support the combination.

Options which will take time (perhaps up to two years):

B.1: Encode new letters ꜧ ꜧ LATIN LETTER I WITH SPIRITUS LENIS alongside Ugariticist ꜣ ꜣ LATIN LETTER A WITH SPIRITUS LENIS and ꜧ ꜧ LATIN LETTER U WITH SPIRITUS LENIS.
- Advantage: Avoids problems of font suppliers missing these letters.
- Disadvantage: Standardization process takes about two years.

B.2: Encode a new diacritic COMBINING SPIRITUS LENIS for Egyptological and Ugariticist use.
- Advantage: Provides a non-Cyrillic solution with a dedicated and appropriate diacritical mark.
- Disadvantage: This character would be effectively identical with the Cyrillic character, only would not have a Cyrillic character property.
- Disadvantage: Standardization process takes about two years.

The report of the Closing Session of the I&E meeting took the following view:

> The meeting expressed a preference for B1, primarily as it was felt that a discrete character would be preferable to assuming the existence of another font on any particular Egyptologist's computer....

> The point was well-made that I&E does not have any formal powers to make these decisions on behalf of the world Egyptological community. However, it is the closest to a formal group on the subject of Informatique and so it was suggested that Nigel Strudwick write to James Allen, the new President of the IAE [International Association of Egyptologists], saying that it has made these suggestions to the Unicode group.

In a follow-up to this, Nigel Strudwick reported:

> This has been done. Jim Allen has responded positively and says he is happy for the group's collective response (as "those who know") to be passed on to the Unicode Consortium with the knowledge of the IAE.

Accordingly, I make this proposal on behalf of the International Association of Egyptologists.

Unicode Character Properties. Character properties are proposed here.

```
A790;LATIN CAPITAL LETTER A WITH SPIRITUS LENIS;Lu;0;L;;;;;N;;;;A791;
A791;LATIN SMALL LETTER A WITH SPIRITUS LENIS;Ll;0;L;;;;;N;;;A790;;A790
A792;LATIN CAPITAL LETTER I WITH SPIRITUS LENIS;Lu;0;L;;;;;N;;;;A793;
A793;LATIN SMALL LETTER I WITH SPIRITUS LENIS;Ll;0;L;;;;;N;;;A792;;A792
A794;LATIN CAPITAL LETTER U WITH SPIRITUS LENIS;Lu;0;L;;;;;N;;;;A795;
A795;LATIN SMALL LETTER U WITH SPIRITUS LENIS;Ll;0;L;;;;;N;;;A794;;A794
```

Bibliography

Gardiner, Alan. 1966. *Egyptian grammar: being an introduction to the study of hieroglyphs*. 3rd edition. London: Oxford University Press.

Hetzron, Robert, ed. 1997. *The Semitic Languages*. London: Routledge.

Ritter, R. M. 2002. *The Oxford guide to style*. Oxford: Oxford University Press. ISBN 0-19-869175-0

Examples.

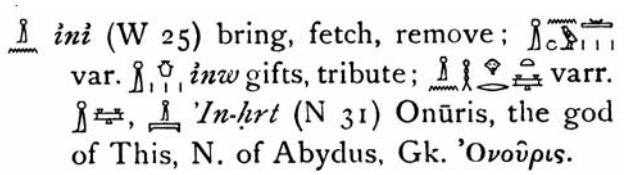

Figure 1. Example from Gardiner 1966, showing LATIN CAPITAL LETTER I WITH SPIRITUS LENIS in Ỉn-ḥrt and LATIN SMALL LETTER I WITH SPIRITUS LENIS in ỉnỉ and ỉnw.

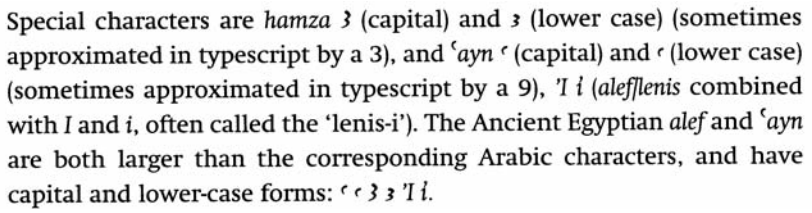

Figure 2. Example from Ritter 2002, showing three characters used in Egyptology with casing pairs.

The Abecedaries and the Consonantal Alphabet

The order of the alphabet is known from abecedaries and is similar to that of the later Northwest Semitic languages:

UG	ʾa	b	g	ḫ	d	h	w	z	ḥ	ṭ	y	k	š	l	m	δ	n	ẓ	s	ʿ	p	ṣ	q	r	θ	γ	t	ỉ	ủ	š
NWS	ʾ	b	g		d	h	w	z	ḥ	ṭ	y	k		l	m		n		s	ʿ	p	ṣ	q	r	š		t			

The basic consonantal inventory consisted of twenty-seven phonemes; the origin of the last three signs is in dispute. The three *ʾaleph* signs are used to indicate /ʾ/ plus following vowel (e.g., ⟨ʾa⟩ = /ʾa/), with ⟨ỉ⟩ used for syllable-final /ʾ/.

Figure 3. Chart from Hetzron 1997 showing Ugaritic vowels with *spiritus lenis*.

Table 37. Northwest Semitic Scripts

Value	Ugaritic	Hebrew	Estrangelo	Serto	Nestorian	Numeric Value
ʾ (ʾa)	𐎀	א	ܐ	ܐ	ܐ	1
b	𐎁	ב	ܒ	ܒ	ܒ	2
g	𐎂	ג	ܓ	ܓ	ܓ	3
(ḫ)	𐎃					
d	𐎄	ד	ܕ	ܕ	ܕ	4
h	𐎅	ה	ܗ	ܗ	ܗ	5
w	𐎆	ו	ܘ	ܘ	ܘ	6
z	𐎇	ז	ܙ	ܙ	ܙ	7
ḥ	𐎈	ח	ܚ	ܚ	ܚ	8
ṭ	𐎉	ט	ܛ	ܛ	ܛ	9
y	𐎊	י	ܝ	ܝ	ܝ	10
k	𐎋	כ ך	ܟ ܟ̣	ܟ ܟ̣	ܟ ܟ̣	20
(š)	𐎌					
l	𐎍	ל	ܠ	ܠ	ܠ	30
m	𐎎	מ ם	ܡ ܡ̣	ܡ ܡ̣	ܡ ܡ̣	40
(δ)	𐎏					
n	𐎐	נ ן	ܢ ܢ̣	ܢ ܢ̣	ܢ ܢ̣	50
ẓ	𐎑					
s	𐎒	ס	ܣ	ܣ	ܣ	60
ʿ	𐎓	ע	ܥ	ܥ	ܥ	70
p	𐎔	פ ף	ܦ	ܦ	ܦ	80
ṣ	𐎕	צ ץ	ܨ	ܨ	ܨ	90
q	𐎖	ק	ܩ	ܩ	ܩ	100
r	𐎗	ר	ܪ	ܪ	ܪ	200
ś		שׂ				
š (θ)	𐎘	שׁ	ܫ	ܫ	ܫ	300
(γ)	𐎙					
t	𐎚	ת	ܬ	ܬ	ܬ	400
(ỉ)	𐎛					
(ủ)	𐎜					
(ṡ)	𐎝					

Where two forms are shown, the one on the right occurs at the end of a word. Ugaritic values are given in parentheses.

Figure 4. Chart showing Ugaritic vowels with *spiritus lenis*.

A. Administrative

1. Title

Proposal to encode Egyptological Yod and similar characters in the UCS

2. Requester's name

Michael Everson

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2008-08-04

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

No.

1b. Proposed name of script

1c. The proposal is for addition of character(s) to an existing block

Yes

1d. Name of the existing block

Latin Extended-D

2. Number of characters in proposal

6.

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category A.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes.

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

Yes; cf. N2241 and 3382.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?
The Informatique & Égyptologie group and the International Association of Egyptologists.
2c. If YES, available relevant documents
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?
Egyptologists, Copticists, Semiticists, and other scholars.
4a. The context of use for the proposed characters (type of use; common or rare)
Used historically and in modern editions.
4b. Reference
5a. Are the proposed characters in current use by the user community?
Yes.
5b. If YES, where?
Scholarly publications.
6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?
Yes.
6b. If YES, is a rationale provided?
Yes.
6c. If YES, reference
Accordance with the Roadmap. Keep with other Egyptological transliteration characters.
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
No.
8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
No. No existing combining character yields the correct display.
8b. If YES, is a rationale for its inclusion provided?
8c. If YES, reference
9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?
No. These characters are proposed to be atomic with no decomposition.
9b. If YES, is a rationale for its inclusion provided?
9c. If YES, reference
10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?
No.
10b. If YES, is a rationale for its inclusion provided?
10c. If YES, reference
11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?
No.
11b. If YES, is a rationale for such use provided?
11c. If YES, reference
11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
No.
11e. If YES, reference
12a. Does the proposal contain characters with any special properties such as control function or similar semantics?
No.
12b. If YES, describe in detail (include attachment if necessary)
13a. Does the proposal contain any Ideographic compatibility character(s)?
No.
13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?