ISO/IEC JTC1/SC2/WG2 N1636

DATE: 1997-08-25

ISO/IEC JTC1/SC2/WG2
Universal Multiple-Octet Coded Character Set (UCS) – ISO/IEC 10646
Secretariat: ANSI

DOC TYPE:	Expert contribution
TITLE:	**Encoding Egyptian Hieroglyphs in ISO/IEC 10646-2**
SOURCE:	Michael Everson
PROJECT:	JTC1.02.18.02
STATUS:	Discussion paper
ACTION ID:	FYI
DUE DATE:	--
DISTRIBUTION:	Worldwide
MEDIUM:	Paper and web
NO. OF PAGES:	5 (printed at 85%)

1. Basic encoding principles

To encode Egyptian hieroglyphs, we should adopt the existing encoding principles which have been developed since computers were introduced into Egyptology in the late 1970s and the early 1980s, and which are widely if not universally employed by Egyptologists today. The *Manuel de codage* (Paris: 1988) and the Macintosh and Windows implementations (MacScribe and WinGlyph) which support it are sufficiently well-developed that they can be adopted with a minimum of effort into ISO/IEC 10646 -- optimizing the transfer from existing encoded texts to 10646-coded texts.

2. Character set repertoire

2.1 Repertoire

To begin with, two distinct repertoires should be encoded. The first, smaller set, called BASIC EGYPTIAN HIEROGLYPHS, consists of the essential Gardiner Middle Egyptian sign list, and comprises 761 characters. The second, much larger set, called EXTENDED EGYPTIAN HIEROGLYPHS, as presented in *Hieroglyphica* (Utrecht and Paris: 1993), comprises 4872 characters. Additionally, 37 Alternate Format characters will be needed in order to map currently-encoded texts to 10646. It is expected that in the next decade the final number of hieroglyphs known and analyzed will total some 8000, which means that we can expect, eventually, further extensions (such as EXTENDED EGYPTIAN HIEROGLYPHS-A) to be added to the standard.

## 2.2. Source of the repertoire

To achieve the maximum ease of mapping the characters, the 23 eight-bit fonts (4 Basic and 19 Extended) of the Centre for Computer-Aided Egyptological Research should be taken as the base for encoding, and the characters should be arranged serially in 10646. This means that the 166 characters in Glyph Basic A will occupy positions x000 to x0B5, the 192 characters in Glyph Basic B will occupy positions x0B6 to x175, the 185 characters in Glyph Basic C will occupy positions x176 to x22E, and the 202 characters in Glyph Basic D will occupy positions x22F to

x2F8. And so on for the 19 fonts of the Glyph ExtLib A - S set.

3. Character names

There is a number of ways to designate Egyptian hieroglyphic characters:

- by conventional sound (**w**, **m**, **wn**) which works for only a very few characters
- by complex description (baby chick, owl, hare, standing monkey holding severed head, recumbant crocodile with cobra headdress and flagellum, two crouching men holding hands, skirted man holding jackal aloft by the throat, seated Thoth, seated Thoth with flagellum, seated Thoth with ankh, Dhwty or Thoth)
- by standardized sign number (G43, G17, E34, E58A, I98, A470, A248A, C3, C3A, C3B, G26B)
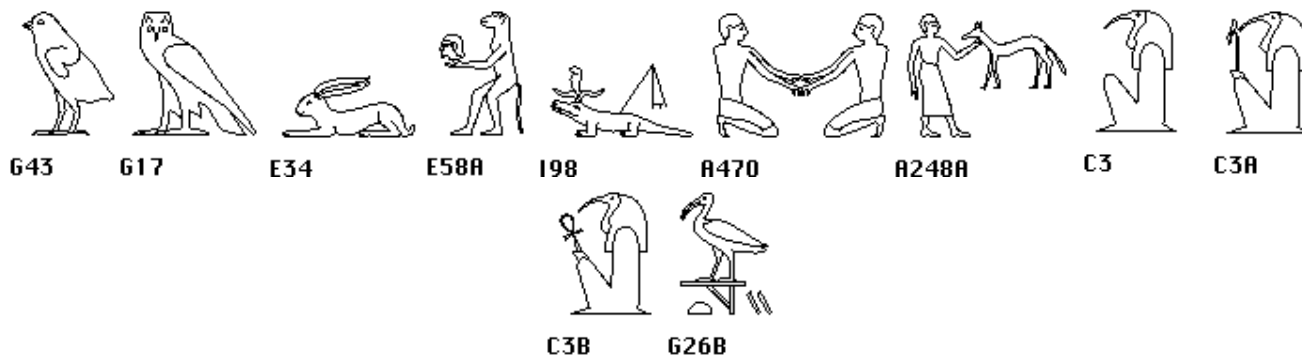
Obviously the only sensible way of naming the characters is to identify them with their standard Egyptological catalogue numbers, since otherwise the character names would be picturesque but overlong. The name EGYPTIAN HIEROGLYPHIC SIGN E34 refers uniquely and unambiguously to the hare character used for the sound **wn**.

4. Alternate format characters

The complexity of the repertoire is largely a problem for the Egyptologists in the first place and for the preparer of the encoding proposal in the second; but the set is finite and well-defined. More subtly difficult, however, is mapping the formatting characters of the *Manuel de codage* to 10646. The chief question is, how closely must the 10646 encoding reflect them? My view is that it should reflect as much of it as possible.

4.1 Plain text and Egyptian hieroglyphs

Positioning is an essential part of Egyptian hieroglyphic text processing, and ISO/IEC 10646 must represent positioning and directionality accurately and completely. It should be assumed that Egyptian has a basic left-to-right directionality; both left-to-right and right-to-left are used. The existing U+200E LEFT-TO-RIGHT MARK and U+200F RIGHT-TO-LEFT MARK can be used to effect whatever directional requirements are needed. However, these marks should be used only for reversing the directionality of the entire text, not for individual signs, which may be mirrored without affecting the directionality of the text. For those, a MIRROR SIGN is proposed here, as are a number of rotation signs -- though it should be noted that other scripts, such as Runic, could benefit from such signs. The END OF LINE MARKER and END OF PAGE MARKER are essential for correct processing of columnar Egyptian and their *Manual de codage* equivalents are heavily relied upon in current encoded texts.

4.2. Beginnings and endings

Signs like BEGIN CARTOUCHE and END CARTOUCHE should be understood to function exactly as do LEFT PARENTHESIS and RIGHT PARENTHESIS (that, is, they take bi-directional context into account). Rendering should draw a continuous line between the two characters.

4.3 Fancy text

Colour is not normally considered to be a feature of plain text, though red (grey in the hardcopy of this document) is used meaningfully in the Coffin Texts and the New Kingdom funerary papyri:

Papyrus of Ani plate IV:26-27

Papyrus of Ani plate VI:6-7

Red colour and black colour are encoded with special locking-shift sequences in the *Manuel de codage*. It should be noted that the Naxi or Tompa script (in China) also uses colour to achieve semantic distinctions, and it may not be practical to proscribe colour as a quality of plain text. In the present proposal, an encoding is given for red hieroglyphs, because the following editorial practice supports it: I give below examples from Budge 1895 in which the rubric is represented not by colour but in another way:

Papyrus of Ani plate IV:26-27

Papyrus of Ani plate VI:6-7

It is proposed that BEGIN RUBRIC SIGN and END RUBRIC SIGN be encoded and that rendering as red or as a black line above be left to implementation. The RUBRIC SIGNs function just as the CARTOUCHE SIGNs do.

Issue: Have all the relevant Manuel de codage characters been accounted for in the list below?

4.4 Proposed Alternate Formatting and other characters

```
0001 x300        EGYPTIAN HIEROGLYPHIC SIGN SEPARATOR
0001 x301        EGYPTIAN HIEROGLYPHIC SIGN JUXTAPOSITIONER
0001 x302        EGYPTIAN HIEROGLYPHIC SIGN SUBORDINATOR
0001 x303        EGYPTIAN HIEROGLYPHIC BEGIN CLUSTER MARK
0001 x304        EGYPTIAN HIEROGLYPHIC END CLUSTER MARK
0001 x305        EGYPTIAN HIEROGLYPHIC END OF LINE MARKER
0001 x306        EGYPTIAN HIEROGLYPHIC END OF PAGE MARKER
0001 x307        EGYPTIAN HIEROGLYPHIC BEGIN CARTOUCHE
0001 x308        EGYPTIAN HIEROGLYPHIC END CARTOUCHE
0001 x309        EGYPTIAN HIEROGLYPHIC BEGIN REVERSED CARTOUCHE
0001 x30A        EGYPTIAN HIEROGLYPHIC END REVERSED CARTOUCHE
0001 x30B        EGYPTIAN HIEROGLYPHIC BEGIN CAPLESS CARTOUCHE
0001 x30C        EGYPTIAN HIEROGLYPHIC END CAPLESS CARTOUCHE
0001 x30D        EGYPTIAN HIEROGLYPHIC BEGIN HWT SIGN
0001 x30E        EGYPTIAN HIEROGLYPHIC END HWT SIGN
0001 x30F        EGYPTIAN HIEROGLYPHIC BEGIN LOW HWT SIGN
0001 x310        EGYPTIAN HIEROGLYPHIC END LOW HWT SIGN
0001 x311        EGYPTIAN HIEROGLYPHIC BEGIN HIGH HWT SIGN
0001 x312        EGYPTIAN HIEROGLYPHIC END HIGH HWT SIGN
0001 x313        EGYPTIAN HIEROGLYPHIC BEGIN RUBRIC SIGN
0001 x314        EGYPTIAN HIEROGLYPHIC END RUBRIC SIGN
0001 x315        EGYPTIAN HIEROGLYPHIC ROTATE TWO HUNDRED SEVENTY DEGREES SIGN
0001 x316        EGYPTIAN HIEROGLYPHIC ROTATE ONE HUNDRED EIGHTY DEGREES SIGN
0001 x317        EGYPTIAN HIEROGLYPHIC ROTATE NINETY DEGREES SIGN
```

```
0001 x318        EGYPTIAN HIEROGLYPHIC MIRROR SIGN
0001 x319        EGYPTIAN HIEROGLYPHIC MIRROR AND ROTATE NINETY DEGREES SIGN
0001 x31A        EGYPTIAN HIEROGLYPHIC MIRROR AND ROTATE ONE HUNDRED EIGHTY
DEGREES SIGN
0001 x31B        EGYPTIAN HIEROGLYPHIC MIRROR AND ROTATE TWO HUNDRED SEVENTY
DEGREES SIGN
0001 x31C        EGYPTIAN HIEROGLYPHIC COMBINING QUADRANT SHADING
0001 x31D        EGYPTIAN HIEROGLYPHIC COMBINING TOP HALF QUADRANT SHADING
0001 x31E        EGYPTIAN HIEROGLYPHIC COMBINING BOTTOM HALF QUADRANT SHADING
0001 x31F        EGYPTIAN HIEROGLYPHIC COMBINING LEFT HALF QUADRANT SHADING
0001 x320        EGYPTIAN HIEROGLYPHIC COMBINING RIGHT HALF QUADRANT SHADING
0001 x321        EGYPTIAN HIEROGLYPHIC COMBINING LEFT TOP QUADRANT SHADING
0001 x322        EGYPTIAN HIEROGLYPHIC COMBINING RIGHT TOP QUADRANT SHADING
0001 x323        EGYPTIAN HIEROGLYPHIC COMBINING LEFT BOTTOM QUADRANT SHADING
0001 x324        EGYPTIAN HIEROGLYPHIC COMBINING RIGHT BOTTOM QUADRANT SHADING
0001 x325        (This position shall not be used)
0001 x326        (This position shall not be used)
0001 x327        (This position shall not be used)
0001 x328        (This position shall not be used)
0001 x329        (This position shall not be used)
0001 x32A        (This position shall not be used)
0001 x32B        (This position shall not be used)
0001 x32C        (This position shall not be used)
0001 x32D        (This position shall not be used)
0001 x32E        (This position shall not be used)
0001 x32F        (This position shall not be used)
```

0001	x30	x31	x32
0	⊡	⊡	◨
1	⊡	⊡	◨
2	⊡	⊡	◨
3	⊡	⊡	◨
4	⊡	⊡	◨
5	⊡	⊡	
6	⊡	⊡	
7	⊡	⊡	
8	⊡	⊡	
9	⊡	⊡	
A	⊡	⊡	
B	⊡	⊡	
C	⊡	◨	
D	⊡	◨	
E	⊡	◨	
F	⊡	◨	

5. Allocation in ISO/IEC 10646

Egyptian hieroglyphs should be encoded in Plane 1 of ISO/IEC 10646. Three rows are required for Basic Egyptian Hieroglyphs and nineteen rows are required for Extended Egyptian Hieroglyphs. The Alternate Format Characters are a bit difficult to place; since three rows are required for the Basic characters, there are only 7 empty spaces at the end of the row. Assuming, a as shown in the table below, that the Basic characters occupy positions 0001 x000, it would be convenient to begin the Extended characters at position 0001 x400. This leaves space in row 0001 x300 available for future standardization (such as additional Alternate Format characters, Meiroitic alphabetic characters (which were taken directly from the hieroglyphic alphabet), etc.).

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
x0	¿Basic Egyptian Hieroglyphs?															
x1	¿Basic Egyptian Hieroglyphs?															
x2	¿Basic Egyptian Hieroglyphs?															
x3	¿Alt. Form.?			???												¿Meroitic?
x4	¿Egyptian Hieroglyphs Extended?															
x5	¿Egyptian Hieroglyphs Extended?															
x6	¿Egyptian Hieroglyphs Extended?															
x7	¿Egyptian Hieroglyphs Extended?															
x8	¿Egyptian Hieroglyphs Extended?															
x9	¿Egyptian Hieroglyphs Extended?															
xA	¿Egyptian Hieroglyphs Extended?															
xB	¿Egyptian Hieroglyphs Extended?															
xC	¿Egyptian Hieroglyphs Extended?															
xD	¿Egyptian Hieroglyphs Extended?															
xE	¿Egyptian Hieroglyphs Extended?															
xF	¿Egyptian Hieroglyphs Extended?															

y0	¿Egyptian Hieroglyphs Extended?
y1	¿Egyptian Hieroglyphs Extended?
y2	¿Egyptian Hieroglyphs Extended?
y3	¿Egyptian Hieroglyphs Extended?
y4	¿Egyptian Hieroglyphs Extended?
y5	¿Egyptian Hieroglyphs Extended?
y6	¿Egyptian Hieroglyphs Extended?
y7	¿Egyptian Hieroglyphs Extended-A?
y8	¿Egyptian Hieroglyphs Extended-A?
y9	¿Egyptian Hieroglyphs Extended-A?
yA	¿Egyptian Hieroglyphs Extended-A?
yB	¿Egyptian Hieroglyphs Extended-A?
yC	¿Egyptian Hieroglyphs Extended-A?
yD	¿Egyptian Hieroglyphs Extended-A?
yE	¿Egyptian Hieroglyphs Extended-A?
yF	¿Egyptian Hieroglyphs Extended-A?
z0	¿Egyptian Hieroglyphs Extended-A?
z1	¿Egyptian Hieroglyphs Extended-A?
z2	¿Egyptian Hieroglyphs Extended-A?
53	¿Egyptian Hieroglyphs Extended-A?

Michael Everson, everson@indigo.ie, http://www.indigo.ie/egt, Dublin, 1997-08-25