

Proposal to encode three control characters for Egyptian Hieroglyphs

Bob Richmond
bobqq at live.co.uk


4th May, 2015

1. Introduction

Egyptian hieroglyphs were added to Unicode in version 5.2 (October 2009) on the basis of the Everson and Richmond *Proposal to encode Egyptian Hieroglyphs in the SMP of the UCS* (L2/07-097; N3237). This basic collection of hieroglyphs is mostly scoped to the *List of Hieroglyphic signs* from Gardiner’s *Egyptian Grammar* (1957). However, at the present time, Egyptian hieroglyphics cannot be displayed in plain text using the Egyptian quadrat (clustering) format that is a signature feature integral to the script. Therefore instead of the standard format,



, non-specialist software such as web browsers or word processors can only

express this text in linear form, . While this is readable it not the way the writing system was or is intended to be used. This situation has resulted in very limited use of Unicode for hieroglyphs since they became available in 2009.

Egyptian hieroglyphic has been used in typographic form in modern publications since the mid-19th century. For example, the Theinhardt font was designed for Karl Lepsius (1810–1884). A new typeface was designed for Gardiner for use in *Egyptian Grammar* and the *Topographical Bibliography of Ancient Egypt*. Hand drawn hieroglyphs continued to be used in publications such as the comprehensive dictionary *Wörterbuch der Ägyptischen Sprache* (Erman et al, 1926-1963) and the *Concise Dictionary of Middle Egyptian* (Faulkner, 1962). A LaserComp version of the Oxford Gardiner font was created in the early 1980s. Since then computer based technology has become the norm for publishing hieroglyphs as text, for example *Middle Egyptian: an introduction to the language and culture of Hieroglyphs* (Allen, 1999), *Großes Handwörterbuch Ägyptisch-Deutsch* (Hannig, 1995), *Topographical Bibliography vol. 8* (Malek, 2000). Modern digital encodings largely use the Manuel de Codage (MdC) scheme. The *Hieroglyphica* (1993, 2000) sign list and font is used as a reference catalogue of signs in MdC applications. Current practice for online publication is largely to generate images from MdC and publish images rather than text. More complex documents often incorporate MdC renderings in PDF using embedded fonts.

The fact that specialist software is required to render Egyptian hieroglyphic text correctly means that content being produced by specialists is siloed in proprietary software encodings, and thus misses out on the benefits of being encoded in Unicode. That lack of a standard way of encoding Egyptian hieroglyphs in quadrat format effectively blocks the broader adoption of Unicode Egyptian by specialists. This proposal requests the addition of three control characters corresponding to the Manuel de Codage (MdC) control codes ‘*’, ‘:’, and ‘&’ to generate a wide range of quadrats.

Having dedicated control characters for Egyptian hieroglyphics will allow rendering engines to treat quadrat formation in the same way as syllable clusters in other complex scripts. This would allow standardized Egyptian hieroglyphic fonts to be produced using OpenType rules that render quadrats in the same way that Egyptian typefaces have been used for the past 150 years.

2 Proposed characters

Three control characters are proposed:

Characters

Default glyph	Code point	Character name
	134xx	EGYPTIAN HIEROGLYPHIC SIGN JUXTAPOSITIONER
	134xx	EGYPTIAN HIEROGLYPHIC SIGN SUBORDINATOR
	134xx	EGYPTIAN HIEROGLYPHIC SIGN LIGATOR

Properties

134xx;EGYPTIAN HIEROGLYPHIC SIGN JUXTAPOSITIONER;Mn;9;NSM;;;;;N;;;;;

134xx;EGYPTIAN HIEROGLYPHIC SIGN SUBORDINATOR;Mn;9;NSM;;;;;N;;;;;

134xx;EGYPTIAN HIEROGLYPHIC SIGN LIGATOR;Mn;9;L;;;;;NSM;;;;;

3. Mode of use

EGYPTIAN HIEROGLYPHIC SIGN JUXTAPOSITIONER

JUXTAPOSITIONER is the equivalent of MdC ‘*’ and is placed between hieroglyphs to state a preference for rendering side by side in a single quadrat. E.g., < , JUXTAPOSITIONER, , JUXTAPOSITIONER, >. In such cases, caret positioning should follow the quadrat boundaries in the same way as a complex syllable cluster.

EGYPTIAN HIEROGLYPHIC SIGN SUBORDINATOR

SUBORDINATOR is the equivalent of MdC ‘.’ and is placed after a hieroglyph indicate that the following hieroglyph renders below the preceding hieroglyph in a quadrat. E.g., < , SUBORDINATOR, > means render as .

SUBORDINATOR may be used in combination with EGYPTIAN HIEROGLYPHIC SIGN JUXTAPOSITIONER. E.g., < , JUXTAPOSITIONER, , SUBORDINATOR, > means .

EGYPTIAN HIEROGLYPHIC SIGN LIGATOR

LIGATOR is the equivalent of MdC ‘&’ and is placed between hieroglyphs to signal that the sequence forms a ligature.

E.g., < , LIGATOR, > means the very common phonetic combination . This method is necessary to render clusters that cannot be encoded using JUXTAPOSITIONER and/or SUBORDINATOR. It may also be used in combination with JUXTAPOSITIONER and/or SUBORDINATOR. E.g. < , LIGATOR, , SUBORDINATOR, > means .

4. Other MdC features not proposed for plaintext

Quadrat boundaries

Like the Indic Virama model, quadrat boundaries occur wherever basic characters have not been explicitly joined using one of these control characters. Therefore no equivalent to MdC ‘-’ or ‘space’ is required to signal a quadrat boundary.

Complex quadrats

MdC supports sub-quadrats using parentheses ‘(, ’). E.g., MdC F9*(X1:Z4):D40 means quadrat . This form can be expressed more simply via the three control characters, F9*X1&Z4:D40. An extensive survey of the Middle Egyptian

texts indicates there is no need to support parenthetical expressions for quadrats at this stage of the language, i.e., the stage of language currently supported by the repertoire of Egyptian hieroglyphs encoded in Unicode.

More complex quadrats mostly occur in vertical writing and in some styles of Late Egyptian writing. The rationale for excluding this feature from the proposed plain text encoding is that normal scholarly practice is to transliterate vertical text into horizontal text in modern publications. In the process, the transcription can adopt a quadrat style more suitable for horizontal writing. A good example is seen in the following illustration of directionality in Hieroglyphic from *Egyptian Grammar* (p25):



Note that the complex cluster in this example is essentially produced by kerning two of the horizontal quadrats so that they overlap. Such cases could be accommodated via other contextual OT rules and need not be defined in the quadrat.

Hieroglyph variants

MdC allows for rotation and scaling of hieroglyphs. In practice scaling is most often seen as a workaround for flaws in quadrat rendering in specific implementations, in turn giving problems in other implementations. Rotation is mostly used for a small number of signs with horizontal and vertical variants, a situation easily dealt with by variation selectors in Unicode if necessary. MdC-like control codes for variants are therefore not proposed.

5. Selected references

Allen, James P. 1999. *Middle Egyptian: an introduction to the language and culture of Hieroglyphs*. Cambridge: Cambridge University Press. ISBN 0-521-77483-7

Collins, Lee. 2009. *Unicode TN#32 Unicode Technical Note #32 MAPPING BETWEEN MANUEL DE CODAGE AND UNICODE EGYPTIAN HIEROGLYPHS* <http://www.unicode.org/notes/tn32/>.

Erman, Adolf; Grapow, Herman (editors). 1971 (1926-1963). *Wörterbuch der Ägyptischen Sprache*.

Everson, Michael. 1997-08-25. *Encoding Egyptian Hieroglyphs in ISO/IEC 10646-2. N1636* <http://www.dkuug.dk/JTC1/SC2/WG2/docs/n1636/n1636.htm>.

Everson, Michael; Richmond, Bob. 2007-04-10. *Proposal to encode Egyptian Hieroglyphs in the SMP of the UCS*. <http://www.unicode.org/L2/L2007/07097-n3237-egyptian.pdf>.

Faulkner, Raymond O. 1986 (1962). *A concise dictionary of Middle Egyptian*. Oxford: Griffith Institute. ISBN 0-900416-32-7.

Hannig, Rainer. *Großes Handwörterbuch Ägyptisch-Deutsch*. 1995.

Gardiner, Alan H. 1957. *Egyptian Grammar: being an introduction to the study of hieroglyphs*. 3rd edition.

Grimal, Nicolas; Hallof, Jochen; van der Plas, Dirk. 2000. *Hieroglyphica*. 2nd edition.

Porter, Bertha; Moss, Rosalind; Malek, Jaromir (editors) *Topographical Bibliography of Ancient Egyptian Hieroglyphic Texts, Statues, Reliefs and Paintings* (1927-2000, ongoing). [<http://topbib.griffith.ox.ac.uk>].

van den Berg, Hans. 1997. *“Manuel de Codage”*: A standard system for the computer-encoding of Egyptian transliteration and hieroglyphic texts. [Leiden]: Centre for Computer-Aided Egyptological Research. (<http://www.catchpenny.org/codage/>)

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS

FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	Proposal to encode three control characters for Egyptian Hieroglyphs	
2. Requester's name:	Bob Richmond	
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution	
4. Submission date:		
5. Requester's reference (if applicable):		
6. Choose one of the following:		
This is a complete proposal:	Complete	
(or) More information will be provided later:		

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):		
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	13000–1342F	
Name of the existing block:	Egyptian Hieroglyphs	
2. Number of characters in proposal:	3	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
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	F	G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided?	Yes	
a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?	Yes	
b. Are the character shapes attached in a legible form suitable for review?	Yes	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	Bob Richmond	

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):

bobqq at live.co.uk

6. References:

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

Yes

b. Are published examples of use (such as samples from newspapers, magazines, or other sources)

of proposed characters attached?

Yes

7. Special encoding issues:

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

Shaping

8. Additional Information:

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/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	No
If YES explain	
2. 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
If YES, with whom?	Jaromir Malek, Vincent Razanajao, Mark-Jan Nederhof, Serge Rosmorduc
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?	Yes
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	Rare
Reference:	
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference:	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	No
If YES, is a rationale provided?	
If YES, reference:	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	No
If YES, is a rationale for such use provided?	
If YES, reference:	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	No
If YES, reference:	
12. Does the proposal contain characters with any special properties such as	

control function or similar semantics?	Yes
If YES, describe in detail (include attachment if necessary)	
See attached	
13. Does the proposal contain any Ideographic compatibility characters?	No
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	