L2/15-123

Proposal to encode three control characters for Egyptian Hieroglyphs

Bob Richmond
bobqq at live.co.uk

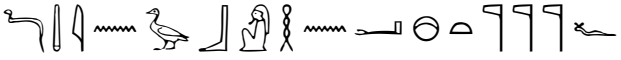
24th April, 2015

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 *Egyptian Grammar* (Gardiner, 1957).

Egyptian hieroglyphic is first and foremost an historic script but nevertheless has been used in modern publications since the rediscovery of the writing system in the early 19th century. Hieratic, the cursive form of the Egyptian writing system, is also usually transliterated into hieroglyphic for scholarly purposes. Several hieroglyphic typefaces such as the Theinhardt font (designed for Lepsius) emerged in the mid-19th Century although use of handwritten hieroglyphs in publishing continued through the 20th Century, the most recent well known example being the *Concise Dictionary of Middle Egyptian* (Faulkner, 1962). A LaserComp version of the Oxford Gardiner font had replaced older technology by the early 1980s and use of computer based technology has become the norm for publishing Hieroglyphs as text over the last 30 years.

Discussion of ancient sources often requires more accuracy than a text representation affords. Photography and redrawing or reconstruction of ancient texts are essential techniques for many scholarly purposes and it is important to recognize the limitations of fonts and text in this bigger picture. Nevertheless there are many publications in print, and now online, that are best served by hieroglyphic as text and that is the subject of this proposal.

A signature feature of the hieroglyphic script is the arrangement of hieroglyphs into Egyptian Quadrats: rectangular clusters of hieroglyphs positioned in an aesthetically pleasing arrangement. The hieroglyphs in a quadrat are read downwards and from left to right or right to left (according to text directionality) so the sequence of hieroglyphs is clear to the reader. The quadrat arrangements for vertical text in columns are often different to those used in horizontal writing simply because the width constraint encourages a different aesthetic.

Currently there is no way of using quadrats in Unicode plaintext so instead of  rendering of hieroglyphs in non-specialist software such as web browsers or word processors there is no alternative in plaint text to  Readable but not the way the writing system was or is intended to be used. This situation has resulted in very limited use of Unicode for since hieroglyphs became available in 2009.

The Unicode 7.0 description of *Egyptian Hieroglyphs* (section 11.4) summarizes Egyptian writing systems and the use of Manuel de Codage (MdC) encoding as the de-facto standard for rendering hieroglyphs outside Unicode. Additional information on MdC and data relating to Unicode is given in *Unicode Technical Note #32: MAPPING BETWEEN MANUEL DE CODAGE AND UNICODE EGYPTIAN HIEROGLYPHS*. MdC uses three control codes '*', ':', and '&' to generate a wide range of quadrats. The basic notation is outlined in the Unicode standard.

This proposal is about adding three MdC equivalent characters to Unicode to enable basic plaintext rendering of Egyptian hieroglyphic. This MdC compatible approach retains the order of hieroglyphs so does not raise any significant issues with search, sort or analysis.

The possible use of MdC codes in Unicode was described in the discussion paper *Encoding Egyptian Hieroglyphs in ISO/IEC 10646-2* (N1636, Everson, 1997). Everson described all MdC (1988) formatting codes for hieroglyphs including sign transformations and shading of uncertain or lost hieroglyphs. The proposal here is only concerned with quadrat rendering for plain text in horizontal writing, although it could be adapted to vertical writing if a need

for vertical plain text were established. Other MdC features can be addressed using specialist protocols or generic Unicode features.

In one sense, adopting the MdC approach to quadrats for Egyptian plaintext is straightforward enough as the practice is established and proven useful for over 20 years over its domain of applicability for plaintext rendering. However it is understood that more evidence may be required before this proposal can be accepted.

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;;;;;

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, > means preferably render all three signs in a quadrat rather than as 3 separate signs. Rendering software or font may then, for instance, place them closer together for aesthetic effect. Normally this character is used in conjunction with EGYPTIAN HIEROGLYPHIC SIGN SUBORDINATOR rather than by itself.

EGYPTIAN HIEROGLYPHIC SIGN SUBORDINATOR. SUBORDINATOR is the equivalent of MdC ':' and is placed after a hieroglyph to state a preference to render the following hieroglyphs below any preceding hieroglyphs in a quadrat. E.g. < , SUBORDINATOR, > means render as . SUBORDINATOR can be used alongside EGYPTIAN HIEROGLYPHIC SIGN JUXTAPOSITIONER. E.g. <  , JUXTAPOSITIONER, , SUBORDINATOR, > means .

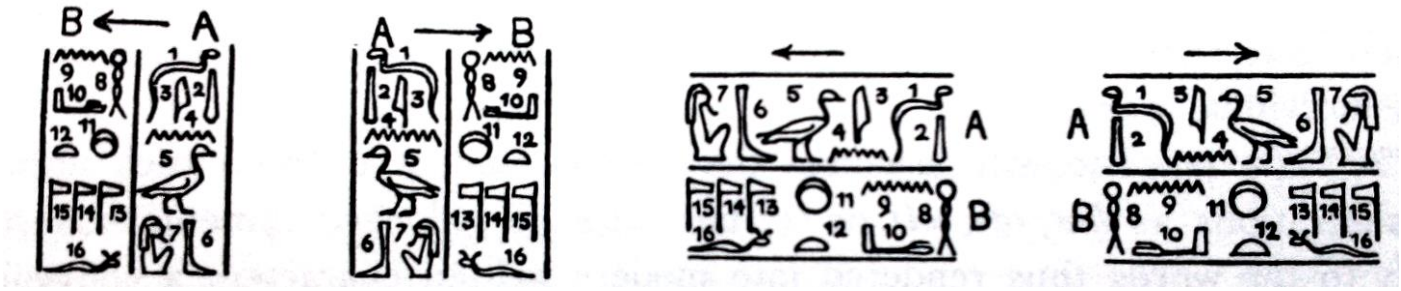
EGYPTIAN HIEROGLYPHIC SIGN LIGATOR. EHSS states a preference to join two or more hieroglyphs if a combination belongs to a known list. E.g. < , LIGATOR, > means the very common phonetic combination . This method is necessary to render clusters that cannot be encoded using JUXTAPOSITIONER and/or SUBORDINATOR. Most frequently, LIGATOR is used to create a single quadrat. It may also be used in combination with JUXTAPOSITIONER and/or SUBORDINATOR. E.g. < , LIGATOR, , SUBORDINATOR, > means .

Quadrat boundaries. The start of a multi-hieroglyph quadrat is signalled by the presence of JUXTAPOSITIONER, SUBORDINATOR, or LIGATOR following a hieroglyph in which case that hieroglyph begins the quadrat. The end of a quadrat is signalled by any character other than JUXTAPOSITIONER, SUBORDINATOR, or LIGATOR following a hieroglyph. If a sequence involving these control characters does not follow the MdC pattern then the control characters should be ignored. This is different to MdC which uses a character '-' or space to indicate quadrat

boundary because hieroglyphs are represented by ASCII ‘Gardiner’ codes. This is unnecessary in Unicode as each hieroglyph is a character.

3. Comparison with MdC quadrat encoding. The proposed plain text encoding follows the MdC pattern but there are some differences.

Complex quadrats. MdC supports sub-quadrats using ‘(‘ and ‘)’ parentheses. Sub-quadrats are used to form complex quadrats. E.g. MdC F9*(X1:Z4):D40 means quadrat  which cannot be expressed using the three control characters (unless X1&Z4 were a known ligature in which case the quadrat could be encoded F9*X1&Z4:D40). 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 the normal scholarly practice of transliterating vertical text into horizontal text in modern publications, and in the process 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):



Nothing is lost from the reading of the text and a higher level protocol can be used in instances where the quadrat structure is somehow important.

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.

Implementation of typographies. Implementations of MdC often have limits such as the maximum of three SUBORDINATOR codes in a quadrat in WinGlyph software. There is no standard layout protocol for MdC, the typography is left to the implementation. However implementations generally allow for any quadrat structure however fanciful the hieroglyph arrangement aside from any such limits on complexity. For Unicode plain text, it is proposed that well defined quadrat possibilities are used and these be specified by the user community for specific purposes. For instance a ‘Basic Typography’ for the Middle Egyptian style of quadrat used in *Egyptian Grammar* and many other publications can be defined from analysis of the quadrats used in these sources. Fonts supporting ‘Basic Typography’ need then only interpret sequences using the control characters compatible with a well-defined quadrat specifications. The three proposed control characters enable the user community to decide what makes a satisfactory approach to quadrats for specific specialist purposes outside the scope of ‘Basic Typography’. This reference ‘Basic Typography’ for Unicode can be defined in this wider context to be made available to address the majority of non-specialist plain text requirements should this proposal be accepted.

Extensions to MdC. MdC as described represents the last official version of the encoding. However software released since then, such as MacScribe and JSesh, support application-specific extensions to quadrat construction including font-dependent absolute positioning of hieroglyphs in quadrats and mechanisms that attempt to be less font dependent in quadrat construction. These extensions can be useful, especially for vertical text, and necessary if it is important to highlight nuances when transcribing certain texts but are unnecessary for the vast majority of publishing encoding requirements where horizontal plain text is sufficient.

4. Summary. Three control characters are proposed for use in encoding Egyptian Hieroglyphs as plaintext. These characters enable rendering of hieroglyphs in quadrats in a way consistent with the established MdC protocol. The

degree to which specific implementations such as fonts support use of the control characters for specific purposes is proposed as a matter for the user community.

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 [ME]

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

Erman, Adolf (editor), and Herman Grapow (editor), 1971 (1926-1963). *Wörterbuch der Ägyptischen Sprache. Berlin: Akademie Verlag*. [WB]

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 and Bob Richmond. 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. [DME]

Gardiner, Alan H. 1957. *Egyptian Grammar: being an introduction to the study of hieroglyphs*. 3rd edition. London: Oxford University Press. [EG]

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/)

ISO/IEC JTC 1/SC 2/WG 2

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS

FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]

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: