An Extension to the three control characters for Egyptian Hieroglyphs and some additional remarks

Bob Richmond; 2016-08-01

1 Introduction

The *Proposal to encode three control characters for Egyptian Hieroglyphs* L2/16-018 was adopted as a UTC recommendation in January 2016 and is now under ballot.

Three documents from Egyptologists and others were submitted together in April 2016 as *Comments on three control characters for Egyptian Hieroglyphs* L2/16-190.

The proposal and comments were discussed at length at the *Informatique et Egyptologie* Cambridge workshop July 10-11, 2016 A different approach was introduced by Nederhof as *A comprehensive system of control characters for Ancient Egyptian hieroglyphic text (preliminary version)* L2/16-177.

Concerns raised in *Comments* primarily centred on EGYPTIAN HIEROGLYPH LIGATURE JOINER ⌗ not being defined in a geometric sense. Discussions raised questions about vertical text and 'tall group' orthographies as well as a need to represent a wider range of arrangements of hieroglyphs in some circumstances.

Since the I&E meeting, data has been supplied by the *Ramses* and *Thesaurus Linguae Aegyptiae* (TLA) database projects (contributors to *Comments*). This has enabled some analysis of cluster issues in those databases to help inform this proposal.

Having considered all comments and feedback it seemed necessary to provide a clarification of how the EGYPTIAN HIEROGLYPH LIGATURE JOINER is to be used and limit its functionality slightly. Two additional controls are proposed: EGYPTIAN HIEROGLYPH GROUP HORIZONTAL JOINER and EGYPTIAN HIEROGLYPH GROUP VERTICAL JOINER. These two characters simplify vertical and 'tall group' writing.

Some further additions have been suggested during the consultation process. These possible extensions are summarized in the section 4 of this proposal. Some are closely linked to decisions yet to be made about future extensions to the hieroglyph repertoire. Others, on evidence presented so far, are not safe to add until better understood and have little impact on the vast majority of applications.

The expert user base, understandably, has little experience with using complex scripts in Unicode and it as important experience is gained using a basic system before considering the best way to make specialist additions. However, before the basic system can be agreed there still remains work to be done around EGYPTIAN HIEROGLYPH LIGATURE JOINER.

A crucial goal of L2/16-018 is to enable Egyptologists and others to work with the hieroglyphic writing system in a straightforward way, taking into account familiar ways of working such as MdC as well as modern techniques for working with complex scripts. The proposed changes described here maintain the essential simplicity and usability while adding useful extra functionality requested by members of the user base we have consulted.

This note is intended for consideration by UTC and the expert user base prior to any submission of a formal proposal.

2 The two new joiners

Some feedback from the expert user base amounts to <u>reliance on OpenType features for Complex Quadrats (see L2/16-018 p9-10) is unsatisfactory both for vertical writing and 'tall groups'.</u>

This proposal addresses these concerns and provides a more structured cluster model by adding two more control characters to those in L2/16-018:

⊠   EGYPTIAN HIEROGLYPH GROUP HORIZONTAL JOINER

⊟   EGYPTIAN HIEROGLYPH GROUP VERTICAL JOINER

These new control characters enable clusters to be built explicitly from arrangements of the basic 3-control groups in vertical text and 'tall groups' in horizontal text. This method increases usability, for example in software that supports complex scripts. It is easier to understand, control and read than reliance on behind the scenes OpenType or other rendering methods. Some features that used the LIGATURE control no longer do. The controls also simplify font development so this is a win-win for implementers and users alike. Typical Middle Egyptian style transcriptions used in textbooks and many printed references that follow the traditional font approach retain a simple to use system as before.

2.1 Vertical Text

Vertical text in columns follows a similar orthography to horizontal text hieroglyphic in terms of groups. However, the fact that each cluster width is fixed to the same value (while cluster height is variable) means the scribe may choose different arrangements of the same sequence of hieroglyphs to better fit the available space.

It is important to recognize that active intervention is normally required in order to transcribe vertical text into horizontal or vice-versa. They do not use identical coding sequences.

A common way of utilizing space is to place what would be two or more groups in horizontal writing side by side in a cluster. The control ⊠, EGYPTIAN HIEROGLYPH GROUP HORIZONTAL JOINER, is used to accomplish this.

For example, the cluster 𓏤𓆟 is written as the sequence 𓏤 ⊠ 𓆟 ⊡ 𓈖 (or for discussion simply 𓏤 ⊠ 𓆟).

For those familiar with MdC notation, ⊠ replaces the use of * and ( ) brackets: 𓏤 * (𓆟 : 𓈖) for the 𓏤𓆟 example.

Aside. Much work by Egyptologists involving vertical text in print has involved transcription to horizontal text. Where vertical text is required, first generation digital tools such as MdC based editors are often used to create a graphic to be modified in a graphics editor to better arrange the hieroglyphs. JSesh provides MdC extensions to allow absolute positioning and sizing of hieroglyphs o allow something similar to be encoded in MdC. Not an ideal situation and Unicode plain text is no silver bullet. Nevertheless, the control-based coding scheme enables font-based techniques to be developed to improve over traditional layout of vertical text.
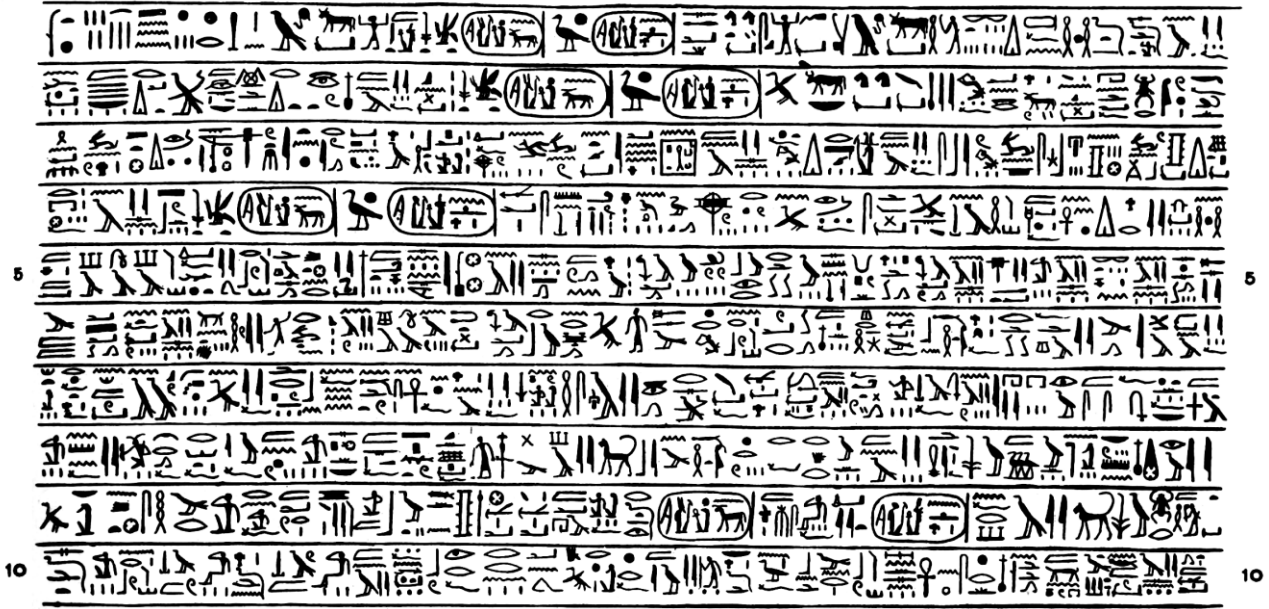
2.2 Horizontal Text involving 'tall groups'

Tall groups are part of a variant orthography found in all periods but best known from Late Egyptian inscriptions such as the 'Israel' stela of Merenptah:
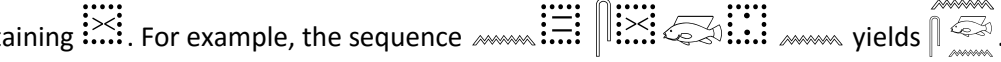

1:8   THEBES.   INSCRIPTION ON STELE OF MERENPTAH.   XIV.

It can be seen that some hieroglyphs are arranged in the more traditional way but some arrangements form narrower, tall groups with up to six hieroglyphs arranged one above the other. Inspection of these 'tall groups' show they follow an orthography similar to vertical text in columns.

As with vertical text, these groups may contain rows like ⌐⌐ so the ⊠ GROUP HORIZONTAL JOINER is used to create sequences in exactly the same way described above.

The ⊞ GROUP VERTICAL JOINER is used in tall groups instead of a ⋮ vertical joiner as a row break around a group containing ⊠. For example, the sequence ∿ ⊞ ║⊠⊂ ⊞ ∿ yields ║⊂.

The use of joiners rather than an MdC-like bracketing system provides simpler and more robust cluster sequences (e.g. errors when start and end brackets are missing simply can't occur). Editing in word processors and other software with general support for complex scripts works naturally.

2.3 Implementing the Cobra pattern

The cobra sign ⌐ carries a distinctive role in hieroglyphic orthography where the empty space may be used to contain one or more hieroglyphs in a sequence. By far the most common examples are a fairly small number of simple combinations such as ⌐, ║, and ⌐ which in basic horizontal writing behave like monograms. However, especially in vertical writing and 'tall groups', a more diverse selection is found (about 200 attested so far). The ⊞ GROUP VERTICAL JOINER can be used to compose these sequences at a quadrat level (rather than the LIGATURE method previous used in groups).

Using this new method, 🐍 now has the sequence 🐍⬚ 🐍⬚ 〰 rather than the ligature sequence 🐍⬚ 🐍⬚ 〰 used previously. The VERTICAL JOINER only forms the ligature when prefixed by specific signs that follow this use-pattern namely ⟋, ⟋, ⟋ and ⟋ (and probably ⟋, ⬚, ⬚, ⟋, ⟋, and ⟋ subject to further analysis and consultation).

The raw code sequence remains humanly readable and straightforward to use.

3 The basic three joiners of L2/16-018

L2/16-018 provides three joiner characters used to arrange hieroglyphs into a cluster/group. These are:

⊞   EGYPTIAN HIEROGLYPH LIGATURE JOINER

✳   EGYPTIAN HIEROGLYPH HORIZONTAL JOINER

⦂   EGYPTIAN HIEROGLYPH VERTICAL JOINER

3.1 EGYPTIAN HIEROGLYPH HORIZONTAL JOINER and EGYPTIAN HIEROGLYPH VERTICAL JOINER

These ✳ and ⦂ joiners arrange hieroglyphs in a grid. For instance, code sequences 🐟⦂〰 and ☐✳⬭⦂⬭ render as 🐟〰 and ☐⬭. In conjunction with ⊞, when necessary, these form basic groups of hieroglyphs

Comments have been generally positive about EGYPTIAN HIEROGLYPH HORIZONTAL JOINER and EGYPTIAN HIEROGLYPH VERTICAL JOINER. They remain unchanged in function in this extended system.

3.2 EGYPTIAN HIEROGLYPH LIGATURE JOINER

*Comments* raised a concern that the LIGATURE JOINER is not geometrical. The sequence ⟋⊞🐍⊞〰 was simply defined to be 🐍 in the attested list; LIGATURE does not explicitly state 🐍 is above 〰 under ⟋. 🦩 is defined to be ⬭⊞🦩⊞⬭. Also, some users would like to be able to write $x⊞y$ for a group containing **x** and **y** but not yet in the attested set. And still know roughly how x and y are meant to be arranged.
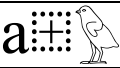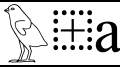
The two new joiners eliminate use of LIGATURE in forming quadrats for column writing and 'tall groups'. The cobra patterns like 🐍 are also moved out of scope of LIGATURE.

This simplifies discussion of groups formed using LIGATURE. It does not address all concerns about edge cases.

One way to address the geometrical concern is to use patterns that limit correct use of LIGATURE. One example is bird patterns.

There are several commonplace ligatures involving birds such as , and . My working dataset contains over 400 examples from 700 total making these patterns the most common use of LIGATURE. Many of these 400 are fairly rare, especially in printed works such as textbooks, but are an important characteristic of hieroglyphic orthography.

		Pattern	Example
		a⊞𓅓	
		𓅓⊞a	
		a⊞𓅓⊞a	

By defining a set of 10-15 patterns including these three examples it is possible to make LIGATURE deterministic (effectively geometrical). The price to means that means some clusters such as (previously sequence) may no longer have a valid sequence when pattern constrained unless an alternate LIGATURE is added. These patterns can be used to guide font and software development as well as specialist working with as yet unattested sequences.

This pattern-constrained LIGATURE approach continues to work well for the vast majority of use cases including traditional typesetting of hieroglyphs.

It does not address the TLA/Ramses desire for greater precision in the future so is unlikely to achieve consensus without extensions/modifications.

The simple solution of moving ahead with LIGATURE and adding extensions for rare cases is problematic because it potentially opens the door to having more than one sequence encoding the same cluster. This means extensions (see 4.5) need to be sufficiently understood so we know this situation is avoided. A solution may imply replacing the single ligature with 2 or more joiners.

One difficulty with coming to a firm conclusion on this point is the rarity of unusual clusters. I've been actively investigating and data so far indicates that we are looking at about 0.01% of corpus' such as Ramses and TLA databases (although that may not be representative of the entire corpus).

There needs to be a clear set of examples agreed by Ramses/TLA Egyptologists as elements that need to be in plain text but cannot be represented yet before we can progress.

4 Potential extensions

These are all elements I regard as viable extensions in some form but not for inclusion at this first stage.

4.1 Monogram joiner (STACK)

The idea of using a control to place one hieroglyph above another as a monogram (as Gardiner calls them) or for another purpose has been a discussion point for quite some time. Something like:

 -> .

Unicode (2009) encodes these kinds of combinations as distinct characters and this policy remains. See *Preliminary draft of the extended Egyptian Hieroglyphs repertoire* L2/16-028 by Michel Suignard.

Technically, this functionality could be added to the current proposal but I recommend the topic is fully researched as part of work on expanding the repertoire. If seen as desirable, there will need to be a clear specification of when such a control applies or when a separate character is required.

4.2 Centre-insert

This control has been suggested in L2/16-177 for use in sequences for combinations such as with one sign centred inside another. is already featured in the L2/16-028 extended repertoire. If such a control were to be considered its scope and significance will need to be fully understood and documented. As with monogram, this is something that should be considered as part of work on the repertoire at this time.

4.3 Geometric ligatures

Certain hieroglyph clusters are attested that cannot be represented by the first stage sequences given here. Data kindly provided by the Ramses project shows a relatively small number of examples occur in their large corpus of texts but it has been pointed out that the scale of the problem may prove more significant when a greater variety of texts are digitally coded.

Various suggestions have been made and I have looked at possibilities from 2 to 10 new control characters with geometric functions such as "place a hieroglyph/group of hieroglyphs in the top right corner of a base sign". A "four-corner" control scheme was proposed at I&E Cambridge using top/bottom left/right corners of the base sign but the theory is as yet unproven. My conclusion so far is that these low frequency issues and potential solutions need to be better characterized before it is safe to define an extension to the system.

Meanwhile projects such as Ramses that would like to evaluate options can choose appropriate simple higher level protocols for these rare instances. For instance, the cluster could be represented using brackets by .

4.4 Cluster kerning

A font could decide to allow quadrats or groups to overlap in certain situations and use ZWNJ or other to prevent that happening. An alternative would be to make cluster joining explicit by use of another control or controls. The danger here is increased complexity of implementation and/or coding sequences with unclear intent. Options could be considered once there is sufficient detailed understanding how Unicode hieroglyphic writing is being used.

4.5 EMPTY sign

A new EMPTY sign may prove useful for hieroglyph spacing, or a generic sign could be re-purposed. EMPTY is known from MdC where it is often used as a fudge rather than with logical intent. EMPTY needs a well-defined understanding of purpose before consideration as an addition.