To: UTC
From: Debbie Anderson, SEI, UC Berkeley
Subject: Summary of Meeting Report on Egyptian Hieroglyphs by Michel Suignard (L2/19-315) and Additional Comments
Date: 4 October 2019

The following is a summary of the results of the Meeting on Egyptian Hieroglyphs (in section A) and additional thoughts by Michel Suignard on how to proceed on the encoding of additional Egyptian Hieroglyphs (in section B).

A. Highlights from Meeting Report on Egyptian Hieroglyphs (L2/19-315)
1. Comments on the current code chart (pages 1-6 of Meeting Report)
Changes:

- Based on feedback from the meeting, some glyphs will be modified:
    o Revert to 2-stroke version for 1321A (N037A) (page 1)
    o Slightly decrease size of 13092 (D027A) (page 2)
    o Slightly reduce the size of 13092 (D027A) (page 2)
    o Slightly modify 13097 (D031A), so the size of the middle protuberance is between the current glyph and the glyph shown in N3237 (page 3)
    o Modify glyphs at 13377/8/9/A/B so they are full height (page 4).
- One character may be added, since the source image of a click beetle provided by Jorke G. is quite different from the current glyph for 131AC (which may not be an insect) (page 4). (An appropriate glyph is needed.)
- Annotations may be added to 1310B/C, 13319/A, 13214/5, 1337D/E (page 5).

Comments by Michel Suignard:

- Some modifications to the sizes of glyphs will be made (see above), but the representative glyphs are not prescriptive (page 2).
- To change the orientation of a Unicode character, use a higher-level protocol, unless the various orientations appear in the same text or can be shown to have distinct pronunciation, semantics, etc. in which case new characters could be proposed (page 6).
- Use a different font to represent the figure with a beard or without a beard in A43. (page 7)
- The new taxonomy will apply to new additions. Had the new taxonomy been used in the original proposal, some of the characters in the encoded repertoire would have been classified differently.

2. Comments on Revised draft for the encoding of an extended Egyptian Hieroglyphs repertoire (L2/19-220) (pages 8-14 of Meeting Report)

Changes:

- Move 14026/7 and 1340F to AA category (page 11).
- Move 1400F and 1341D to AA category (page 12)
- Consider an annotation for 13F9D (pages 11-12)
- Consider removing the cursive forms 1379A/AA (although it was noted that they are present in Hieroglyphica and JSesh)

Comments on control characters:

- The majority view of the group was that existing set of format controls should be used to avoid unnecessary encoding (page 13).
- Using control characters to create signs dynamically, instead of encoding all the variants (page 10), is very challenging because of current font technology. In many cases an exact glyph representing the sequences needs to be contained in the font, and hence a list of known sequences would need to be registered and maintained (cf. RGI for emoji). It may be simpler to encode them atomically, at least for complex cases. (p. 10 with more details on the top of page 11).
- An alternative is to develop a character description language (like IDC for CJK ideographs) to describe the contents of a hieroglyph. They do not enforce a particular layout system or describe a formal decomposition, so they are less constrained (page 11).
- See comments in section B below, on page 3 ("Additional thoughts on Egyptian Hieroglyphs")

3. Responses to Bob Richmond's documents (pages 14-19 of Meeting Report)

Comments based on discussion of "On extending the Egyptian Hieroglyphic repertoire in Unicode" ([Link](#)) and "Egyptian Hieroglyph Overlays from Hieroglyphica" ([Link](#)):

- The proposed name of the new database being maintained by Michel is "Unikemet2"
- In the discussion of staging of proposals, meeting participants preferred to see whole repertoire (A to Z, Aa) proposed at once, instead of in tranches, even if later additions are made to the sub-groups (pp. 14-15).
- The "[Egyptian Hieroglyph Overlays](#)" doc lists 200+ 2-character combinations of hieroglyphs in Hieroglyphica that are candidates for encoding using the OVERLAY MIDDLE format control. (Both characters must already be in Unicode. Over 100 additional combinations are possible, if one includes the characters in the new repertoire.) Bob R recommends maintaining a list of attested overlay combinations outside the Unicode Standard, since it would be easier to add to the list without going through standards process.
  - Meeting participants liked the idea outlined above, although Michel recommended going through Unicode, cf. Unicode Named Character Sequences ([UAX 34](#)). (p. 15)
  - Adding components to the proposed repertoire in order to cover all the sequences with overlays is ok with Michel (page 19).
  - Bob R mentions the sequence order is crucial for OVERLAY MIDDLE, since A+B and B+A are different. However, Michel reports there is no clear connection between z-order and reading. Most meeting members disagreed sequence order was crucial or reading order can be derived from original carving (which can be ambiguous). If, however, it was crucial, both sequences should be created, which would create disunification issues (page 19).
- Changing the IFAO taxonomy was suggested by Bob R, but Michel prefers to not change the taxonomy (page 16) or to modify the IFAO naming conventions that have been adopted by Michel (pages 17-18).
- A transitional PUA system, organized by taxonomy, could be defined. Once the repertoire is published, a mapping table of the PUA and published code points could be made available, to switch between PUA and encoded code points. In this way, font developers could build a release

font from a PUA version. Michel has been considering this approach and is willing to do the transition (pages 16-17).

4. Other topics in Egyptian Meeting agenda (L2/19-314**) (pages 19-20 of Meeting Report)**

- Svenja Gülden and her team will provide input on corrections to Möller in the L2/19-220 spreadsheet (page 19).
- There was no consensus to include all IFAO characters in the repertoire. No one seemed interested in creating a digitized version of IFAO (page 19).
- Timeline for submission of the proposal L2/19-220 was not discussed, other than the preference for a proposal covering the full repertoire, rather than in tranches (page 20).
- The idea of shading control character was not discussed at length. Michel reports defining a syntax using the overlay control character and an existing geometric shade character is possible, but it is not clear it would be widely implemented (and it is not trivial with current font technology) (page 20).

B. Additional Thoughts by Michel Suignard

- OVERLAY as sequence descriptors

  Michel would prefer use of OVERLAY MIDDLE and other similar operators (such as proposed INSERT CONTROL) as sequence descriptors rather than layout operators (cf. CJK IDS). Overlay operators require outline contours merge, which can't be done with current font technology. INSERT CENTER has additional problems, including grouping and recursion.

  Ken Whistler thinks that use of the overlay operator as a sequence descriptor would require agreement on a change of documentation and properties.

- Registered sequences

  Attendees at the meeting liked Bob Richmond's suggestion about having explicit registration of all known sequences, though Michel currently thinks atomic encoding may be better (but will oblige if consensus emerges in support of sequences). He notes that not all hieroglyphs with overlay can be handled this way (i.e., those with more than 2 members).

  Though use of overlay operator as a productive way to create combinations may not be feasible, Ken thinks the UTC will want to consider the option of listing useful combinations explicitly.

  Michel agrees it is possible to create a list of useful combinations and restrict the combinations to such a list. This would make implementation easier. However, users will have the expectation the overlay is productive and could be used for more complex combinations (3 or more signs). Of course, some combinations have already been encoded atomically.

  For Michel, stability is a concern. If there is consensus that the overlay operator (and possibly INSERT CENTER) is used like the RGI emoji sequences, he is ok with the idea, as long as all sequences are predefined. Also, the UTC must be firmly in control of the list of sequences. Still, atomic encoding is simpler to manage, as the actual sets are reasonably stable.