L2/21-248

Additional control characters for Ancient Egyptian hieroglyphic texts

Andrew Glass, Jorke Grotenhuis, Mark-Jan Nederhof,
Stéphane Polis, Serge Rosmorduc, Daniel A. Werning

Dec 22, 2021

This is a request to add a total of 30 additional Egyptian Hieroglyph characters (listed in Table 2). In order to accommodate the new characters and future additions, we request the current Egyptian Hieroglyph Format Controls block be extended to U+1345F.

Background documents

- L2/17-112R summarizes and finalizes ideas from earlier proposals, and is the most self-contained description of the existing nine control characters for Ancient Egyptian hieroglyphic text [17].
- L2/18-236 discusses details of the syntax [27].
- L2/19-331 discusses details of the semantics [28].
- L2/20-176 discusses recent progress with implementation in OpenType [16].
- L2/21-208 provisional proposal for discussion with the Script Ad Hoc Committee.

Summary

Ancient Egyptian hieroglyphic texts consist of signs (hieroglyphs) arranged next to one another, above one another, or in a variety of other spatial arrangements. Text can be horizontal (in rows), or vertical (in columns). Since Unicode 12 (March 2019), there are nine control characters to express the relative positioning of signs, listed in Table 1 together with their syntax.

The two most self-evident control characters are the VERTICAL JOINER and the HORIZONTAL JOINER, which arrange signs above one another and beside one another, respectively. The BEGIN SEGMENT and END SEGMENT are needed to nest horizontal and vertical arrangements. Such nested arrangements are relatively common in Ancient Egyptian inscriptions, but there are few parallels in other writing systems, and consequently, introduction of such controls was a significant novelty in Unicode. Signs can also be overlayed. Some combinations of overlayed signs are very common, and exist as atomic code points in Unicode. Because overlaying is largely productive, it is not feasible to enumerate all possible overlayed arrangements, which motivated OVERLAY MIDDLE. The four remaining control characters represent insertion of a sign, or of a nested group of signs, in the empty corner of a larger sign.

The present document proposes three additional insertion controls, revisits the encoding of cartouches and similar enclosing shapes, and proposes a control for mirroring. Also the need to encode rotation is discussed. We further propose controls to encode damaged and incomplete text. The new primitives are summarized in Table 2 and Table 3. The properties are given at the end of this document, in Tables 29–31.

U+13430 EGYPTIAN HIEROGLYPH VERTICAL JOINER
U+13431 EGYPTIAN HIEROGLYPH HORIZONTAL JOINER
U+13432 EGYPTIAN HIEROGLYPH INSERT AT TOP START
U+13433 EGYPTIAN HIEROGLYPH INSERT AT BOTTOM START
U+13434 EGYPTIAN HIEROGLYPH INSERT AT TOP END
U+13435 EGYPTIAN HIEROGLYPH INSERT AT BOTTOM END
U+13436 EGYPTIAN HIEROGLYPH OVERLAY MIDDLE
U+13437 EGYPTIAN HIEROGLYPH BEGIN SEGMENT
U+13438 EGYPTIAN HIEROGLYPH END SEGMENT

fragment ::= (group)$^+$
group ::= vertical_group | horizontal_group | basic_group

vertical_group ::= ver_subgroup (: ver_subgroup)$^+$
ver_subgroup ::= horizontal_group | basic_group
horizontal_group ::= hor_subgroup (* hor_subgroup)$^+$
hor_subgroup ::= (vertical_group) | basic_group

basic_group ::= core_group | insertion_group
insertion_group ::= core_group insertions
insertions ::= in_group [in_group] [in_group] [in_group] |
in_group [in_group] [in_group] |
in_group [in_group] |
in_group
in_group ::= (vertical_group) | (horizontal_group) | (insertion_group) |
core_group
core_group ::= flat_hor_group flat_ver_group | **sign**
flat_hor_group ::= (**sign** (* **sign**)$^+$) | **sign**
flat_ver_group ::= (**sign** (: **sign**)$^+$) | **sign**

Table 1: The existing control characters for Ancient Egyptian and their syntax.

1 Background

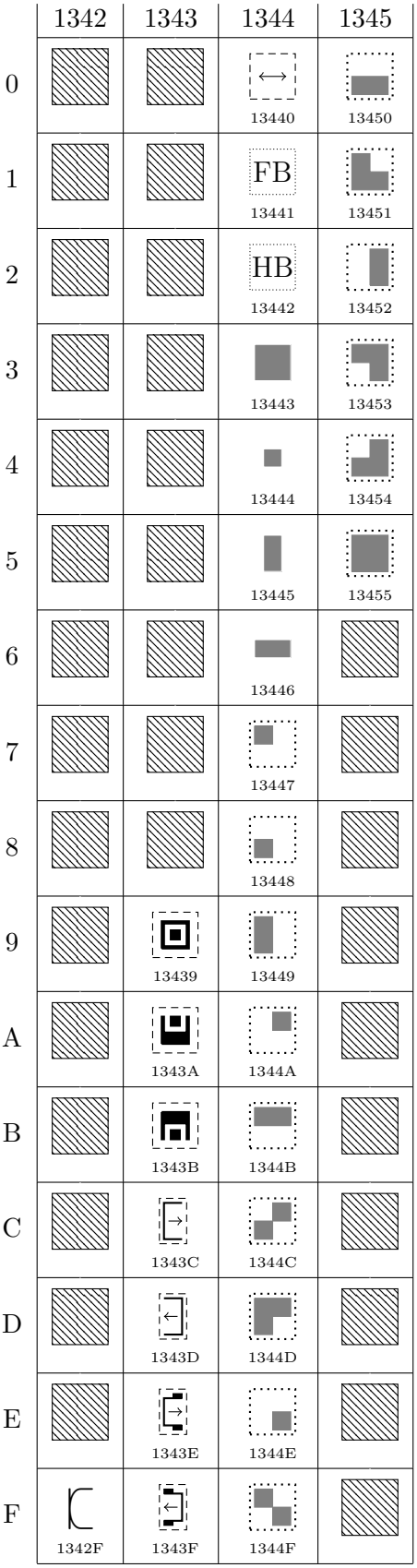
The design of the existing nine control characters in 2016/2017 followed three main principles:

(1) Inspiration can be gleaned from existing forms of encoding of hieroglyphic text, but one cannot blindly adopt their primitives in Unicode.

(2) The Unicode controls must have a straightforward meaning independent from the platform, font, and rendering engine.

(3) There should be as few primitives as possible.

The first principle is motivated by the awareness that existing forms of encoding of hieroglyphic text, in particular the Manuel de Codage (MdC) [9], were designed for the purpose of preparing *printed* publications. A typical use case is as follows: For a hieroglyphic or hieratic text of interest, a specific tool, such as Glyph [8] or JSesh [32], is used to produce a transcription, using the primitives of the MdC, with custom settings



Unclassified

1342F EGYPTIAN HIEROGLYPH V011D
- knotted beginning of cartouche
- not to be confused with 13378

The following glyphs and format controls are used to render groups of Egyptian hieroglyphs.

Sign insertion controls

13439 EGYPTIAN HIEROGLYPH INSERT AT MIDDLE
1343A EGYPTIAN HIEROGLYPH INSERT AT TOP
1343B EGYPTIAN HIEROGLYPH INSERT AT BOTTOM

Enclosure controls

1343C EGYPTIAN HIEROGLYPH BEGIN ENCLOSURE
1343D EGYPTIAN HIEROGLYPH END ENCLOSURE
1343E EGYPTIAN HIEROGLYPH BEGIN WALLED ENCLOSURE
1343F EGYPTIAN HIEROGLYPH END WALLED ENCLOSURE

Mirror control

13440 EGYPTIAN HIEROGLYPH MIRROR HORIZONTALLY

Blank and lost signs

13441 EGYPTIAN HIEROGLYPH FULL BLANK
13442 EGYPTIAN HIEROGLYPH HALF BLANK
13443 EGYPTIAN HIEROGLYPH LOST SIGN
13444 EGYPTIAN HIEROGLYPH HALF LOST SIGN
13445 EGYPTIAN HIEROGLYPH TALL LOST SIGN
13446 EGYPTIAN HIEROGLYPH WIDE LOST SIGN

Damage modifiers

13447 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP START
13448 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM START
13449 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT START
1344A EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP END
1344B EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP
1344C EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM START AND TOP END
1344D EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT START AND TOP
1344E EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM END
1344F EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP START AND BOTTOM END
13450 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM
13451 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT START AND BOTTOM
13452 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT END
13453 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP AND END
13454 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM AND END
13455 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED

Table 2: Portion of the code chart pertaining to the new characters.

```
fragment ::= ( group | singleton_group )+
horizontal_group ::= ⟨[⟩ hor_subgroup [ ⟨]⟩ ] ( 𓐱 [ ⟨[⟩ ] hor_subgroup [ ⟨]⟩ ] )*  |
                     hor_subgroup ⟨]⟩ ( 𓐱 [ ⟨[⟩ ] hor_subgroup [ ⟨]⟩ ] )*  |
                     hor_subgroup ( 𓐱 [ ⟨[⟩ ] hor_subgroup [ ⟨]⟩ ] )+
basic_group ::= core_group | insertion_group | placeholder | enclosure
insertions ::= [ 𓐲 in_group ] [ 𓐳 in_group ] [ 𓐴 in_group ] [ 𓐵 in_group ]
               [ 𓐹 in_group ] [ 𓐺 in_group ] [ 𓐻 in_group ]
          /* as in Table 1, at least one of these 7 optional elements must be present */
in_group ::= 𓐷 vertical_group 𓐸 | 𓐷 horizontal_group 𓐸 | 𓐷 insertion_group 𓐸 |
             core_group | placeholder
core_group ::= flat_hor_group 𓐶 flat_ver_group | literal
flat_hor_group ::= 𓐷 literal ( 𓐱 literal )+ 𓐸 | literal
flat_ver_group ::= 𓐷 literal ( 𓐰 literal )+ 𓐸 | literal
literal ::= sign [ VS1 | VS2 | VS3 ] [ 𓑀 ] [ damaged ]
placeholder ::= 𓑁 | 𓑂 | ( 𓑃 | 𓑄 | 𓑅 | 𓑆 ) [ VS1 ]
enclosure ::= [ opening ] inner_enclosure [ closing ]
inner_enclosure ::= 𓐼 ( group )* 𓐽 | 𓐾 ( group )* 𓐿
opening ::= opening_delimiter [ damaged ]
closing ::= closing_delimiter [ damaged ]
singleton_group ::= delimiter [ damaged ]
damaged ::= 𓑇 | 𓑈 | 𓑉 | 𓑊 | 𓑋 | 𓑌 | 𓑍 | 𓑎 | 𓑏 | 𓑐 | 𓑑 | 𓑒 | 𓑓 | 𓑔 | 𓑕
```

Table 3: The syntax of newly proposed control characters, as far as it differs from Table 1. The ⟨[⟩ and ⟨]⟩ stand for any opening and closing brackets that may interact with hieroglyphic groups.

of the tool, possibly extended with custom glyphs. One then lets the tool export an image, which can be embedded in a scientific article. After this, the encoding can be discarded.

The considerations for Unicode are very different however, with its emphasis on *interchange* of textual data, which leads us to the second of the above principles. A Unicode encoding of hieroglyphic texts should not include any element whose meaning changes significantly depending on the chosen tool, its custom settings, or the choice of font, as this makes interchange of the encoding impossible; see [26] for further discussion.

A case in point is the '&' operator, which is widely used in existing MdC tools such as JSesh. This operator can be applied on two or more signs to obtain a particular relative positioning of those signs. The problem is that the exact relative positioning for a particular choice of signs is determined by the used tool, possibly by customizing its settings. For example, an expression of the form A & B could mean that sign A should be placed in the upper-right corner of sign B in one tool, but in another tool, or in the same tool with different custom settings, it could mean something entirely different, perhaps that A should be placed in the lower-left corner of B. Whereas this is harmless if the encoding is solely used to produce a printed publication and is then discarded, the '&' operator precludes interchange of the encoding without losing its meaning, all the more as the spatial arrangement may affect the linguistic interpretation of the encoding.

The third principle is motivated by the significant challenges that Ancient Egyptian poses to common font technology such as OpenType. In particular, the productive use of controls to form deeply nested groups stretches the capabilities of such technology to its limits. One should therefore find a compromise

	PLOTTEXT	Unicode
	`zA;/ra;`	
	`F4;t//;`	
	`w;t/;`	
	`D;n,d;`	
	`A17;//Z2b;`	
	`w;t/;;/t;`	

Table 4: Corner insertions.

Table 5: Groups with corner insertion rendered by the OpenType implementation discussed by [16].

between expressiveness and technical feasibility. For example, there should be primitives to distinguish such spatial arrangements as overlay and corner insertions, as satisfactory transcriptions cannot be obtained without them, but we would not aim to fine-tune scaling of signs or the distances between them. This also means that we cannot expect Unicode to replace tools such as JSesh, as the power of such tools is beyond the capabilities of common font technology. For faithful transcriptions of hieroglyphic texts that are to be included in printed publications, JSesh and other such tools will still be needed in the future.

A fourth principle of lesser importance is:

(4) Let each rendered text be expressible by only one encoding.

This principle is generally desirable in Unicode, as it simplifies text search, among other things. For many reasons however, this principle cannot be consistently applied on Ancient Egyptian. One reason is that some groups of signs that can be obtained using controls since Unicode 12 have also existed as atomic code points since Unicode 5.2.

Given the gains made by applying the above principles, which resulted in a simple but powerful set of controls with well-defined meanings, we propose a continuation in the same direction.

2 Three more insertion controls

Our very first proposal in 2016, L2/16-177, had 9 insertion controls, namely four corner insertions, one middle insertion and four *side* insertions. As this was thought to be too ambitious at the time, later proposals were simplified to only include the four corner insertions, as these exist in PLOTTEXT and are easiest to justify. One of the four side insertions, the 'insert at start'  , which is mainly needed to insert signs above the feet of a bird sign, and which also has an analogue in PLOTTEXT, was merged into 'insert at bottom start'  . Table 4 provides examples, with their analogues in PLOTTEXT.

Since then, it was shown that the four corner insertions can be implemented in OpenType (Table 5). Additional insertion controls could be implemented in the same way. Moreover, we have systematically

(a)

	PLOTTEXT	Unicode
	%010;/f/;	

(b)

(c)

(d) (e) (f) (g)

Table 6: Middle insertion, (a) the syntax in PLOTTEXT, and how it could be represented in Unicode, (b) examples of groups with middle insertion that exist as atomic signs in Unicode, (c) examples from [24], (d) examples from [20], (e) [21, Plate XXII], (f) Louvre = DZA 29.228.980, (g) [6, Plate 32].

Table 7: Combination of middle insertion and corner insertion [7, Plate 49]; cf. Table 6 (d) and (g).

matched the existing controls against inscriptions and found that at least three more insertion controls would significantly increase the coverage of the repertoire of controls. These are the middle insertion and two of the four side insertions, namely 'insert at top' and 'insert at bottom'. We have found few cases of 'insert at start' and 'insert at end', and are not proposing these as controls at this time.

2.1 Middle insertion

Next to the corner insertion primitives, PLOTTEXT also includes a middle insertion, illustrated by Table 6 (a). Lack of a corresponding insertion control in Unicode has so far left a significant gap in coverage. Several groups involving middle insertion currently exist as atomic signs, such as those in Table 6 (b). The phenomenon is highly productive however, and enumerating all such groups is not feasible. A small selection from a few sources is listed in Table 6 (c-g). This motivates adding the 'insert at middle' control to Unicode. Its use is illustrated in Table 6 (a). As with the existing insertion controls, some signs allow for more than one insertion; Table 7 illustrates a combination of middle insertion and corner insertion.

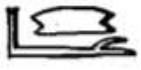
One slight complication is that some common signs such as and may need to be included in a font in a second, larger version, such as and , for use with middle insertion. This is easy to realize within the existing implementations of Ancient Egyptian control characters in OpenType.

2.2 Top insertion and bottom insertion

Also insertion of a sign or of a group of signs at the top side, or at the bottom side, of a base sign is common, which motivates and , respectively. Examples are listed in Tables 8 and 9.

2.3 Special cases

Inevitably some cases will remain that cannot be handled perfectly with the discussed insertion controls, such as the insertions into O013 illustrated in Table 10 (a). Some of these have a group inserted into the left

(a)  (b)  (c)  (d)  (e)  (f) 

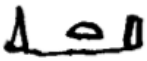
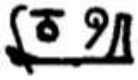
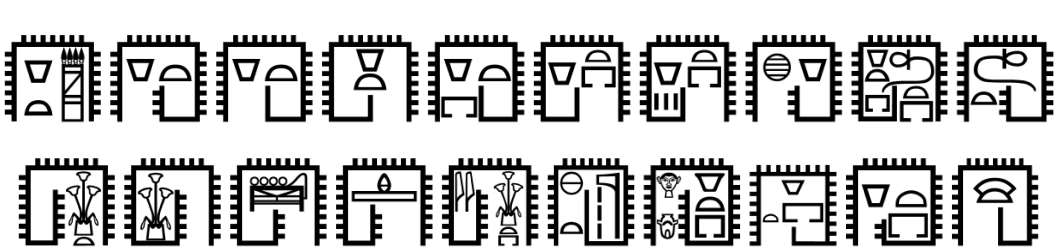
(g)  (h)  (i)  (j)  (k)  (l) 

Table 8: Examples of top insertion, (a) [10, Taf. 2], (b) BM EA 1242 [21], (c) [19, Plate 27], (d) [4, Plate 86], (e) [6, Plate 37], (f) [6, Plate 20], (g) [6, Plate 21], (h) BM EA 143 [13, p. 110], (i) [19, Plate 7], (j) [19, Plate 9], (k) [19, Plate 23], (l) [19, Plate 24].

(a)  (b)  (c)  (d)  (e)  (f)  (g)  (h) 

Table 9: Examples of bottom insertion, (a) [6, Plate 37], (b) [7, Plate 31], (c) [6, Plate 35], (d) [6, Plate 24], (e) BM EA 571 [13, p. 77], (f) [15, p. 348], (g) [19, Plate 3], (h) [21, Plate XIII].



(a) (b)

Table 10: (a) Special cases of middle insertion in variants of O013. (b) Single generic area for middle insertion.

half, some have a group inserted into the right half, and some have a combination of these. As no repertoire of controls can realistically be expected to faithfully capture all such cases, it seems best to opt for a shape of O013 in which there is a single large rectangle, indicated as the shaded area in Table 10 (b), which allows middle insertion of a single group.

3 Cartouches and other enclosures

A *cartouche* encloses a name (in fact two out of the five names) of a king. The enclosed name can be lengthy, and consequently the cartouche has, in principle, unrestricted width (or height in the case of vertical text). There are a number of other enclosing shapes whose dimensions vary in the same manner, as illustrated in Table 11. These also include the *ḥwt* sign discussed in Section 2.1, and hence one may consider an encoding with middle insertion. However, this is not feasible if several groups are enclosed. The first issue is that the encoding becomes quite unwieldy, due to the need for several pairs of parentheses and and horizontal joiners to connect subsequent subgroups. The second issue has to do with conversion between horizontal and vertical text to suit the purposes of publication. In particular, in grammar books, text direction is generally normalized to be horizontal. A *ḥwt* sign, a cartouche or other enclosing shape

(a)  (b)  (c)  (d) 

Table 11: Enclosures other than cartouches, (a) example from [24], (b) examples from [20], (c) example from [15], (d) *ḥwt* enclosure on the statue of royal scribe Youpa [38, Plate 22]; 19th dynasty.

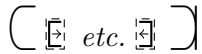
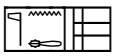
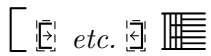
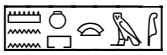
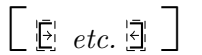
	PLOTTEXT	MdC/JSesh	RES	Unicode
	`%Z1` *etc.* `%Z2`	`<-` *etc.* `->`	`cartouche(` *etc.* `)`	
	`%Z3v` *etc.* `%033` (only in vertical text)	`<S-` *etc.* `->`	`serekh(` *etc.* `)`	
	(not available)	`<F-` *etc.* `->`	`inb(` *etc.* `)`	
	`%Z3` *etc.* `%Z3`	`<h1-` *etc.* `-h1>`	`rectangle(` *etc.* `)`	
	`%Z3` *etc.* `%O6`	`<h1-` *etc.* `-h3>`	`Hwtcloseover(` *etc.* `)`	

Table 12: Encoding of cartouches and similar shapes, in PLOTTEXT, Manuel de Codage (MdC) [9, 32], RES [29], and as proposed in Unicode.

in vertical text enclosing several groups of signs underneath each other would be printed with those groups beside each other instead. The third and final issue is the difficulty of stretching out the enclosing shape depending on the dimensions of the enclosed groups. These considerations suggest an alternative encoding.

Traditionally, cartouches and similar shapes have been encoded by pairs of symbols, delimiting the start and the end of the enclosed text; see Table 12 for examples of encodings in PLOTTEXT and MdC, which contrast with the syntax of RES.

In the past few years, pairs of delimiters have already been used in some prototype implementations to render cartouches and similar enclosing shapes, using the code points listed in Table 13. An additional opening delimiter V011D 𓍿 is needed to represent a mirrored cartouche; cf. Table 14. This is distinct from the independent hieroglyph V011 𓍾.

Before fixing the syntax of the use of the delimiters, a few issues deserve consideration. First, the delimiters 𓍿 and 𓎁 of cartouches also exist as separate, stand-alone signs in hieratic. Users who transcribe hieratic in Unicode will generally prefer that pairs of 𓍿 and 𓎁 are *not* interpreted as delimiters of 'full-form' cartouches that enclose text. Also signs such as 𓏛 have stand-alone occurrences, as in Table 15. Further note that the stand-alone delimiters are different from normal independent signs in that their orientation should depend on the text direction; they are rotated by a quarter turn in vertical text. Because of this, stand-alone occurrences of these signs must form a group (in the sense of Table 1) on their own, without further joiners, which becomes a singleton_group in Table 3.

One should be aware that O033 𓉐 is also an independent hieroglyph, used without a matching opening delimiter, as determiner or logogram [20, p. 731]. In this use, we would not wish the sign to rotate spontaneously upon a change of text direction, which sets it apart from O033A 𓉑.

A different issue is an enclosing shape of which the closing or opening is damaged (see also Section 5), which could be printed as for example  or  . Here we would need a mechanism to signal the start or the end of the enclosing shape without a delimiter. The most economical way to deal with all of the above is to introduce two new characters 𓍼 and 𓍽 that are solely responsible for drawing the connecting lines of

opening	closing
U+13258 O006A	U+1325B O006D
U+13259 O006B	U+1325C O006E
U+1325A O006C	U+1325D O006F
	U+13282 O033A
U+13379 V011A	U+1337A V011B
(**) V011D	U+1337B V011C
U+13286 O036A	U+13287 O036B
U+13288 O036C	U+13289 O036D

Table 13: List of code points since Unicode 5.2 (2009) that could act as opening and closing delimiters, and one desired additional such delimiter marked with '(**)'.

Table 14: Mirrored cartouche; stela of Intef (Leiden V 6) [25, 190-192 (Nr. 56), line 1].

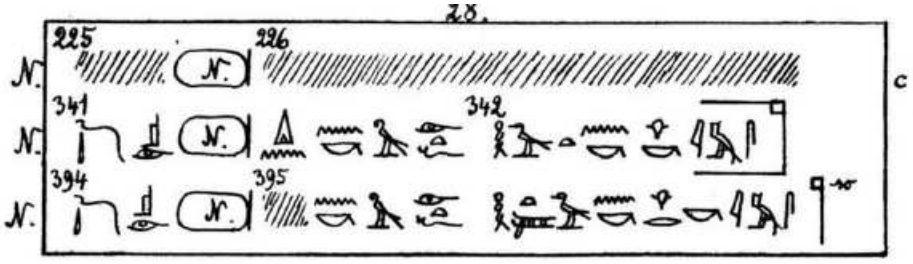
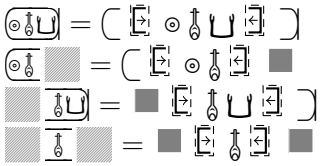


Table 15: Partial enclosure and stand-alone delimiter 𓍘 [35, p. 12].

an enclosure. If used in combination with, for example, 𓍑 𓍒 then this creates a full-form enclosing cartouche, but 𓍑 or 𓍒 or both may be missing. If 𓍑 or 𓍒 are used without 𓍗 and 𓍘, then they act as stand-alone signs. The characters 𓍗 and 𓍘 must always occur in pairs to ensure an unambiguous syntax. To ensure unambiguity, the tokenizer is further required to analyze an **opening_delimiter** immediately preceding 𓍗 to open an enclosure, and a **closing_delimiter** immediately following 𓍘 to close it; see Table 3.

With the proposed encoding, we get for example:



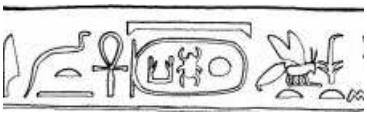
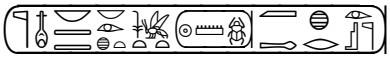


Table 16: Complex enclosure; stela of Ity And Iuri (British Museum EA586) [13, p. 90].

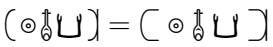
Table 17: A 'vertical' cartouche in horizontal text; [21, Plate VI]

(a) 

(b) 

(c) 

Table 18: Nested enclosures: (a) the original is vertical right-to-left [3, Plate V], (b) the original is vertical left-to-right (Tomb of Thutmosis III, Safari Afrika, https://images.app.goo.gl/9xS7mFHtakuVEA6v8), (c) *ḥwt* enclosure within cartouche [2, p. 194].



(The character ■ for representing lost signs is discussed in Section 5.)

The pair  and  can be used for cartouches, serekhs and *ḥwt* enclosures. We need a separate pair of characters  and  for walled enclosures. It is conceivable that more such characters are needed in the future, but what we described fails to cover only extremely rare cases of enclosures, such as the 'heaven' sign above the cartouche in Table 16, where both shapes are stretched out together to the full length of the enclosed text, so that they could potentially be analyzed as a single complex enclosing shape.

middle insertion versus pairs of delimiters

It is tempting to *disallow* use of middle insertion (Section 2.1) where the same appearance can also be achieved using an enclosure as above. We would hesitate to impose that restriction at this time however, as middle insertion has a number of advantages over enclosures.

First, the delimiters of an enclosure are interpreted according to text direction, that is, the delimiters are rotated by a quarter turn if the direction is vertical. Connected to this, we assume that the implementation in say OpenType does not require remembering the opening delimiter until the moment when the closing shape is being drawn. However, it is generally undesirable that a simple composition like  would turn into  upon a change of text direction, which would unavoidably be the case if pairs of delimiters are used. A related issue is the existence of 'vertical' cartouches in horizontal text, as in Table 17, where use of the middle insertion, rather than a pair of delimiters, would prevent an undesirable normalization to a horizontal cartouche.

Second, we would not expect *nested* enclosures to be implemented, of the kind exemplified in Table 18 (a-b), as this would require considerable effort to handle a case that is very rare. However, corner and middle

insertion would normally be available within enclosures, so that for example Table 18 (c) could be encoded in any case, even if  were not an atomic code point.

Allowing say  to be expressed using middle insertion, next to encoding as enclosure, does not require additional effort, beyond the implementation of middle insertion in general.

In practice, most users may prefer middle insertion over enclosure if only one small group is enclosed. If several groups are enclosed, then middle insertion is not applicable.

Next to O006  , we would need graphical variants with the square in the other three corners. We will leave this to the work on the sign list that is going on in parallel.

4 Mirroring and rotation

A small number of signs change their meaning when mirrored, becoming thereby essentially distinct signs. For example,  is among other things a logogram in *jw*, 'come', whereas  is determinative in words related to 'going backwards'. The majority of signs however can be mirrored for reasons of symmetry and of interactions between the animated hieroglyphs, frequently in cartouches as in Table 19 (a-b). A number of signs that represent inanimate objects and that have no clear 'front' or 'back' can be arbitrarily mirrored even outside cartouches; cf. Table 19 (c). In the Thesaurus Linguae Aegyptiae[1] (TLA) we counted over 230 distinct signs that were mirrored at least once. The most frequent ones are listed in Table 19 (d), excluding signs that are not in the current Unicode repertoire.

Introducing an additional code point for each sign that has been attested to have been mirrored in some text would be problematic, as it appears that a fair portion of signs, perhaps as much as one tenth, has a known occurrence in mirrored form.

Similarly, a small number of signs change their meaning when rotated by an eighth, a quarter or half a turn, becoming thereby essentially distinct signs. Cf. , 'mace', versus  , 'smite', and  , 'wall', versus  , 'tilt'. A great number of signs however can be rotated for aesthetical reasons, especially if their shape is narrow and tall, or low and wide, so they may fit better into the composition of a group. This means that the rotated forms are mere graphical variants, and introducing additional code points would be problematic. At the same time, using the wrong orientation of such a sign yields an unsatisfactory composition, which moreover typical users would consider to be an incorrect representation of the original inscription; cf. Table 20 (a). More examples of rotation are listed in Table 20 (b-e). In the TLA, about 80 signs occur in rotated form, some with more than one rotation, and some in combination with mirroring. One strong piece of evidence that signs could be rotated for purely aesthetical reasons is found in Table 21.

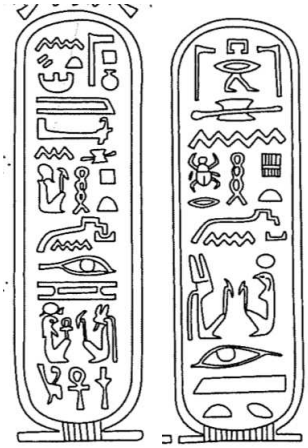
Most frequent seem to be rotations by a quarter turn clockwise of low and wide signs, and rotations by a quarter turn counter-clockwise of narrow and tall signs (assuming a left-to-right text direction). Other rotations exist however, such as rotations by half a turn, which for a portion of signs does not change the meaning. It is further worth noting that signs such as  ('harpoon') and  ('platform') are rotated *despite*, rather than because of, their iconic value.

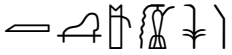
We propose one control character  for horizontal mirroring of a sign. For rotation, variation selectors can be used, as indicated in Table 22. We assume that if a sign is both rotated and mirrored, then rotation is done before mirroring.

Implementing the mirroring control has a relatively small cost in terms of the size of fonts. The reason is that mirrored copies of all signs need to be included in any case in order to render both left-to-right and right-to-left text. For rotation however, one can limit implementation to those signs that are known to occur frequently in rotated form, or that could reasonably be expected to occur in rotated form, such as inanimate objects that are low and wide or narrow and tall. Lists of signs for which certain rotations are assumed to be implemented can be maintained in a central place, analogous to the tables mentioned by [16].

[1] http://aaew.bbaw.de/tla/

(a) 

(b) 

(c) 

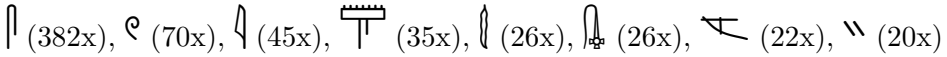
(d) 

Table 19: (a) Mirroring for reasons of symmetry [30, p. 333], and (b) [5, p. 34 and p. 66], (c) a selection of signs that appear mirrored in [36] relative to their representation in Unicode, (d) the signs most frequently mirrored in the TLA.

5 Text-critical marks

In the corpus of Ancient Egyptian hieroglyphic inscriptions, undamaged and complete texts from beginning to end are the exception. The current set of Unicode characters is inadequate therefore to express the great majority of texts and parts of texts, on stone fragments, potsherds and snippets of papyrus.

For encoding an incomplete text, we first need to be able to indicate surfaces within an inscription where signs are no longer legible. Such a surface is normally drawn as a rectangle, filled by a light gray pattern, or by drawing diagonal lines across the surface, which is known as *shading* or *hatching*. For typical users, it is highly informative to know how large such a surface is, to make an educated guess about the signs that are lost. Furthermore, signs occur in groups, and the relative size of the shaded area can significantly affect the scaling and positioning of the remaining signs. This suggests the need to form shaded areas of different dimensions. Examples are found in Table 24 (b), (c), (e), (f) and (j).

We propose introducing a LOST SIGN character ■ to represent a lost sign of approximate size 1 em × 1 em, where 1 em represents the unscaled height of the common sign 𓀀. Next to this there is a HALF LOST SIGN ▪ of size 0.5 em × 0.5 em, a TALL LOST SIGN ▮ of size 0.5 em × 1 em, and a WIDE LOST SIGN ▬ of size 1 em × 0.5 em. Use is demonstrated in Table 23. (At least some part of the user community has indicated a requirement for additional 'lost sign' characters, at the granularity of 0.25 em. This awaits further discussion.)

By default, the above characters are realized by shaded squares or rectangles that should not touch neighboring such squares or rectangles, nor the shading of damaged signs (to be discussed below). In order to achieve 'continuous shading', exemplified in the first groups of Table 24 (b) and (f), where a block of shading seamlessly connects to other areas of shading, the variation selector **VS1** can be applied on any of

(a)

best achievable	original
	and

(b) (21x), (20x), (7x), (6x), (5x), (4x)

(c) (18x), (18x), (5x), (2x)

(d) (4x), (2x), (2x), (2x), (1x)

(e) (4x), (3x), (2x), (1x), (1x)

Table 20: (a) Rotated signs within a group from [36], preceded by the best achievable encoding at present. (b) Some of the most frequent clockwise quarter-turn rotations in the TLA, (c) the same counter-clockwise, (d) the same half a turn. Omitted are signs not in Unicode and rotations in combination with mirroring. (e) A selection of signs in the TLA that are rotated by an eighth turn or less, clockwise or counter-clockwise.

the 'lost sign' characters, to make the shading expand to fill the entirety of the available surface, as opposed to just the rectangle of the exact chosen aspect ratio.

A second requirement is to be able to indicate that extant signs are damaged or partly illegible. This is once more done by shading, but now the shading overlays the affected signs, as exemplified once more by Table 24. Where this is combined with a pair of square brackets, this generally indicates that the enclosed shaded signs are a reconstruction, as opposed to being individually discernible.

It is important to be able to indicate that a part of a sign is damaged. For example, suggests that the actual sign may be , or , and suggests that the actual sign may be or , which are two different signs with non-overlapping meanings, most easily distinguishable by the shape of the tail.

A convenient granularity for indicating damage to a sign divides its surface into four quarters, and a single post-modifying character indicates which quarters are damaged: , , , , , , , , , , , , , , . In the naming, we use 'START' and 'END' rather than 'LEFT' and 'RIGHT', because hieroglyphic text can be written left-to-right or right-to-left, with mirrored signs and mirrored composition of groups, so that 'START' corresponds to either 'left' or 'right' depending on the text direction; a similar naming was adopted in the past for the four corner insertion controls. The ordering is chosen such that, after subtraction of 0x13446, each of the four least significant bits represents damage to one of the four quarters.

An alternative encoding would have been to combine a hieroglyph with a *sequence* of up to four post-modifying characters taken from the list , , , . This idea was rejected in favour of a single post-modifying character per hieroglyph as above, for the following reasons:

Table 21: Tomb of Horemheb. Compare the red and black text of the left column: the scribe/draughtsman rotated 𓏺 for the final version of the layout, presumably as that would fit better in the composition.

VS1	(= U+FE0)	90°
VS2	(= U+FE1)	180°
VS3	(= U+FE2)	270°

Table 22: Variation selectors for signs rotated by a quarter turn or half a turn. The direction of rotation is clockwise for left-to-right text and counter-clockwise for right-to-left text. One could reserve VS4 – VS7 for other rotations by a multiple of an eighth turn.

- This is needed to safely stay within the OpenType spec's limit of 31 characters per cluster[2], considering hieroglyphic groups can contain 8 or more hieroglyphs.
- The cluster size could be reduced in some cases by introducing a fifth character that could be used in place of the full sequence . This however creates encoding ambiguity without **Do Not Use** tables that ban the full sequence, which causes further complications.
- A single post-modifying character avoids the need to normalize non-standard ordering of the up to four post-modifying characters.

We further want to allow square brackets as in Table 24 (a), (d), (g) and (h), and in Table 25 without shading. These brackets would ideally take up no additional space as far as the scaling and positioning of signs within groups is concerned. Next to square brackets, a small number of other bracket pairs may be

[2] https://docs.microsoft.com/en-us/typography/script-development/use#cluster-length

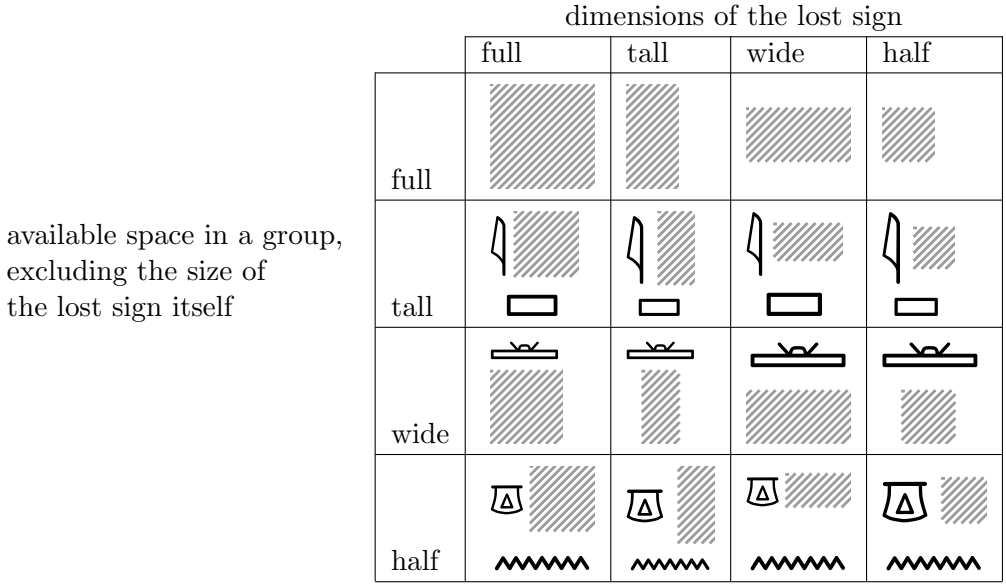
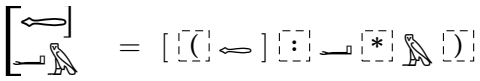


Table 23: Use of the four LOST SIGN characters within groups.

allowed as part of hieroglyphic groups. Brackets can precede or follow a basic_group (in the simplest case a single sign), or precede or follow a vertical_group, as specified in Table 3. An example is:



A third requirement is to be able to indicate *absence* of signs on some surface area of an inscription, which is different from damage to the surface area. There are several cases where this is needed, in transcriptions of both hieroglyphic and hieratic texts. A non-exhaustive list is:

1. A blank in the middle of a text was left in order to eventually fill in a name or a date, which never happened, as in phrases of the form "This book belongs to [ ]" or "In year 3, [ ] month of the inundation, day [ ]".

2. A cartouche 𓍯 can contain blank space as placeholder for a king's name, as in Table 26 (b). This is very different from the independent cartouche sign U+13377 𓍯.

3. A scribe copying a text left a corresponding amount of blank space where the original was damaged, as in Table 26 (c), where the blank space corresponds to the missing grammatical subject in "[...] Ra! [ ] says to Ra: [...]". The name of a goddess is expected in the light of the feminine suffix pronoun.

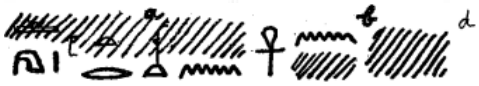
4. Hieroglyphs on walls of tombs were first drawn in ink and later carved. Sometimes the artists forgot to carve a sign and the ink faded away or was painted over with a background color. This resulted in a blank in place of a sign, as in Table 26 (d).

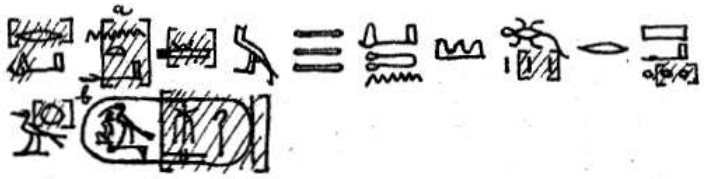
5. An empty space refers to the object or material on which the inscription is written, as in the enigmatic spelling near the centre of Table 26 (a) of 'loving myrrh', with *mr* ('loving') written with a blank, referring to the palette made of softwood, which is also *mr* in Ancient Egyptian.

In all these cases, the exact size of a blank surface is suggestive of its interpretation in context. Relying on existing space or blank symbols in Unicode is little effective, as we need to rely on definite sizes relative

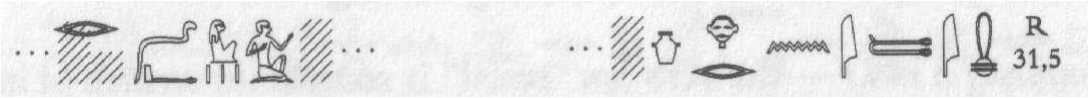
(a) 

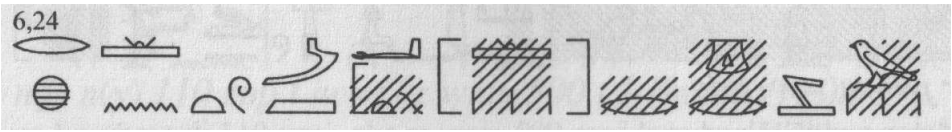
(b) 

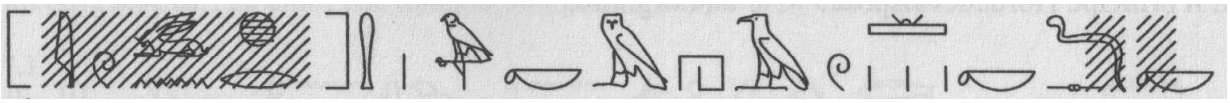
(c) 

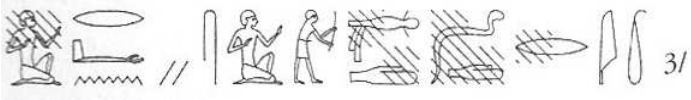
(d) 

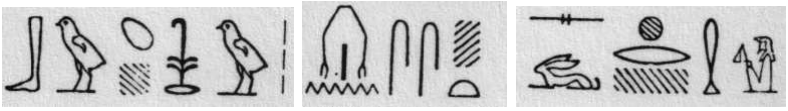
(e) 

(f) 

(g) 

(h) 

(i) 

(j) 

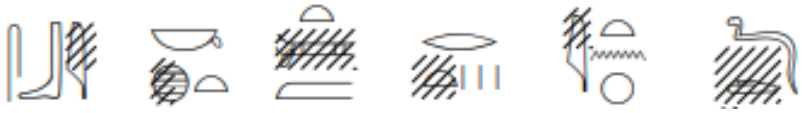
(k) 

Table 24: Examples of shading: (a-c) [37], (d) [34], (e-f) [1], (g-h) [12], (i) [14], (j) [31], (k) [39].

to 1 em (the height of the unscaled 𓀀 in the font). Furthermore, the blank signs are specifically expected to interact with the horizontal and vertical joiners and other control characters of Ancient Egyptian, and would be subject to the same rules of scaling as hieroglyphs would. A pragmatic choice is to introduce two blank characters, namely a FULL BLANK FB of size 1 em × 1 em and a HALF BLANK HB of size 0.5 em × 0.5 em. (These characters are to be rendered as whitespace, without the dotted squares and without the letters.) Note that the LOST SIGN characters indicate areas where text had existed before it was destroyed, whereas the BLANK characters indicate areas where text never existed.)

See Table 27 for examples of encodings in PLOTTEXT, MdC, RES and prospectively Unicode. The extant documentation of PLOTTEXT does not provide a fully unambiguous description of the syntax, but it appears that shading is available for individual signs, by enclosing them within a pair of `%S`, as well as for parts of a group, by using an overlay with a shading symbol. There are four such shading symbols `%S1`, `%S2`, `%S3`, `%S4`, and there are four blank signs `%B1`, `%B2`, `%B3`, `%B4`, of different sizes and aspect ratios. Similarly,

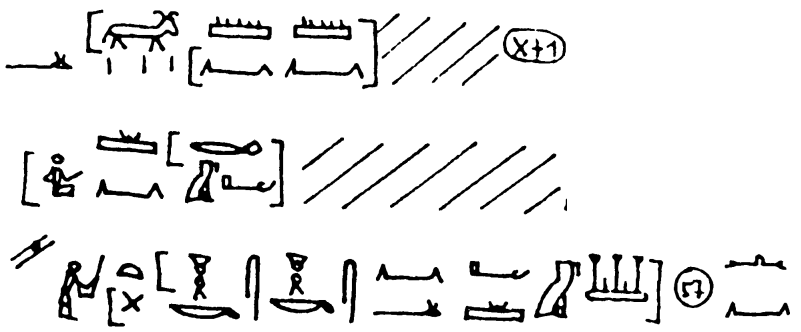


Table 25: Square brackets in transcriptions [23].

the MdC allows shading of individual quarters of signs, as well as shading of quarters of groups. It has four individual shading symbols `//`, `/`, `h/`, `v/`, of the same dimensions as those of PLOTTEXT. It has two blank signs `..` and `.` comprising a full square and a quarter square, respectively. RES allows shading of glyphs and of the spaces between glyphs, to unbounded granularity.

References

[1] J.P. Allen. *Middle Egyptian Literature*. Cambridge University Press, 2015.

[2] J. von Beckerath. *Handbuch der ägyptischen Königsnamen*. Deutscher Kunstverlag, 1984.

[3] H. Beinlich and M. Saleh. *Corpus der hieroglyphischen Inschriften aus dem Grab des Tutanchamun*. Griffith Institute, Ashmolean Museum, Oxford, 1989.

[4] M.L. Bierbrier. *Hieroglyphic texts from Egyptian stelae, etc., Part 10*. British Museum, 1982.

[5] S. Biston-Moulin and C. Thiers. *Le temple de Ptah à Karnak*, volume 49 of *Bibliothèque générale*. Institut français d'archéologie orientale, Le Caire, 2016.

[6] E.A. Wallis Budge. *Hieroglyphic texts from Egyptian stelae, etc., Part II*. British Museum, 1912.

[7] E.A. Wallis Budge. *Hieroglyphic texts from Egyptian stelae, etc., Part III*. British Museum, 1912.

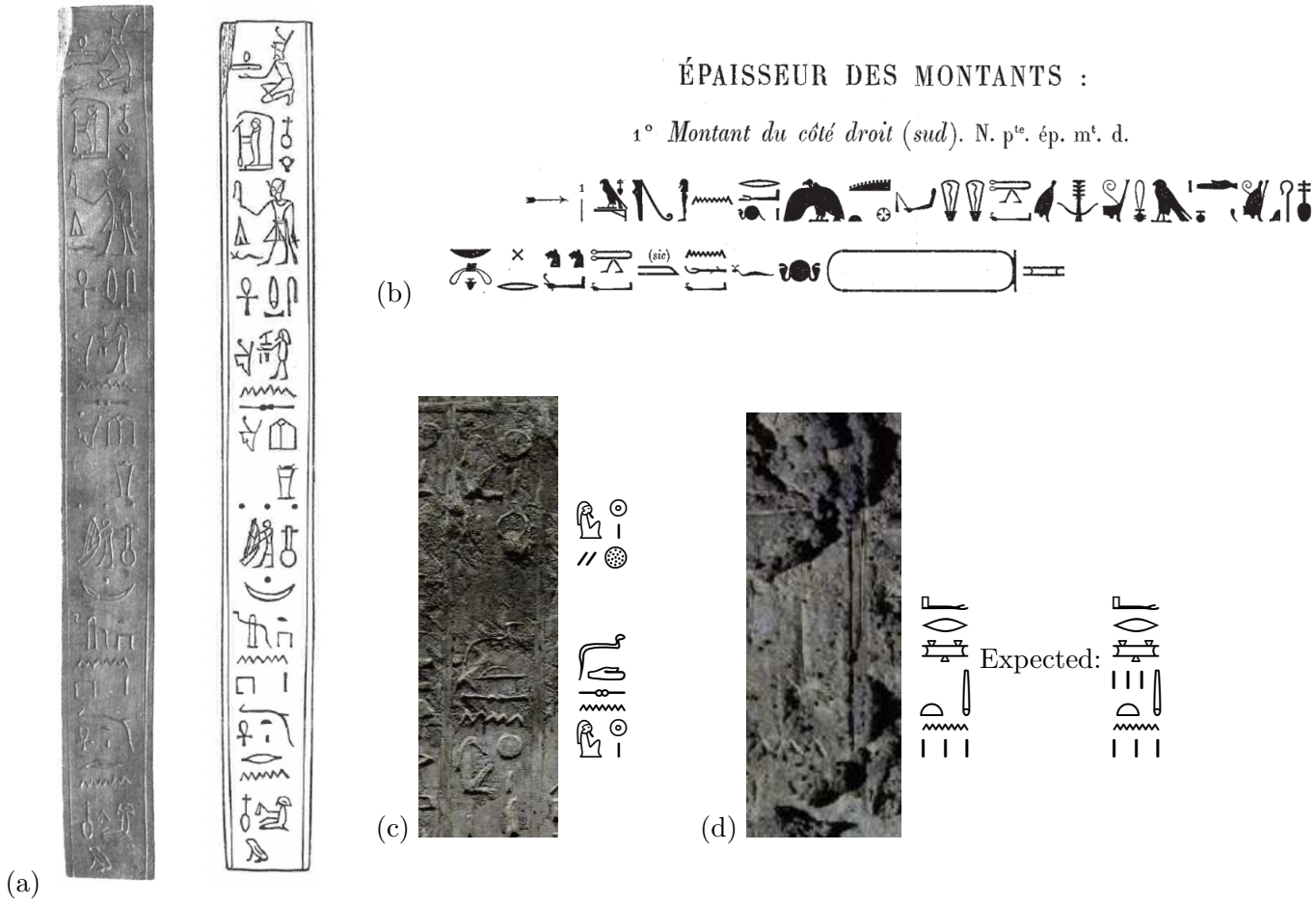
[8] J. Buurman and E. de Moel. GLYPH – computer programs for Egyptologists. In *Informatique et Égyptologie 3*, pages 4–7, 1987.

[9] J. Buurman, N. Grimal, M. Hainsworth, J. Hallof, and D. van der Plas. *Inventaire des signes hiéroglyphiques en vue de leur saisie informatique – Informatique et Égyptologie 2*. Institut de France, Paris, 3rd edition, 1988.

[10] J. Capart. Une liste damulettes. *Zeitschrift für Ägyptische Sprache und Altertumskunde*, 45:pp. 14–21 & pl. 1–2, 1908/1909.

[11] E. Chassinat. *Le temple de Dendara*. Institut français d'archéologie orientale du Caire, 1934.

[12] M.E. Chioffi and G. Rigamonti. *I Racconti di Re Kheope – The Tales of king Kheops*. Duat Edizioni, 2005.



(a) (b) (c) (d)

Table 26: Blank spaces in inscriptions: (a) from the scribal palette Frankfurt Liebighaus IN 1899 [33, p. 321], (b) temple of Dendara [11], (c) Theban Tomb 33, Book of Caverns, col. 1287, Hb. 103,21, photo DOI 10.17171/2-8-358-2, [39, p. 251], (d) Theban Tomb 33, Book of Caverns, col. 809, Hb. 76.37,Z, photo DOI 10.17171/2-8-334-2, [39, p. 183].

[13] M. Collier and B. Manley. *How to read Egyptian hieroglyphs: A step-by-step guide to teach yourself.* British Museum Press, 1998.

[14] M. Dessoudeix. *Lettres égyptiennes*, volume 2. Actes Sud, 2012.

[15] A. Erman and H. Grapow. *Wörterbuch der Ägyptischen Sprache.* Akademie-Verlag, Berlin, 1926–1961.

[16] A. Glass. Cluster model for Egyptian hieroglyphic quadrats. http://www.unicode.org/L2/L2020/20176-hierogyph-cluster.pdf, 2020.

[17] A. Glass, I. Hafemann, M.-J. Nederhof, S. Polis, B. Richmond, S. Rosmorduc, and S. Schweitzer. A method for encoding Egyptian quadrats in Unicode. http://www.unicode.org/L2/L2017/17112r-quadrat-encoding.pdf, 2017.

[18] G. Goyon. *Nouvelles inscriptions rupestres du Wadi Hammamat.* Imprimerie Nationale, Paris, 1957.

[19] H.R. Hall. *Hieroglyphic texts from Egyptian stelae, etc., Part VII.* British Museum, 1925.

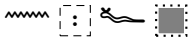
	PLOTTEXT	MdC/JSesh	RES	Unicode
(a) 	`N35,%S I9 %S` (?)	`n:f\shading1234`	`n:f[shade]`	
(b) 	`G17+%S3` (?)	`m\shading12`	`m[t]`	
(c) 	`N35,%S4` (?)	`n://`	`n:empty[shade]`	
(d) 	`"t,r"+%S3` (?)	`t:r#12`	`t[shade]:[t]r`	

Table 27: Encoding of damaged and lost signs.

[20] R. Hannig. *Grosses Handwörterbuch Ägyptisch-Deutsch: die Sprache der Pharaonen (2800-950 v.Chr.)*. Verlag Philipp von Zabern, Mainz, 1995.

[21] T.G.H. James. *Hieroglyphic texts from Egyptian stelae, etc., Part I*. British Museum, 2nd edition, 1961.

[22] K.A. Kitchen. *Ramesside Inscriptions*, volume 6. B.H. Blackwell, 1983.

[23] R. Koch. *Die Erzählung des Sinuhe*. Fondation Égyptologique Reine Élisabeth, 1990.

[24] D. Kurth. *Einführung ins Ptolemäische – Eine Grammatik mit Zeichenliste und Übungsstücken*. Backe-Verlag, 2008.

[25] R. Landgrafova. *It is My Good Name that You Should Remember*. Charles University in Prague, 2011.

[26] M.-J. Nederhof. The Manuel de Codage encoding of hieroglyphs impedes development of corpora. In S. Polis and J. Winand, editors, *Texts, Languages & Information Technology in Egyptology*, pages 103–110. Presses Universitaires de Liège, 2013.

[27] M.-J. Nederhof. A note on the syntax of Ancient Egyptian hieroglyphic control characters. https://www.unicode.org/L2/L2018/18236-nederhof.pdf, 2018.

[28] M.-J. Nederhof. A note on OpenType implementation of Ancient Egyptian hieroglyphic text. https://www.unicode.org/L2/L2019/19331-egyptian-opentype.pdf, 2019.

[29] M.-J. Nederhof. RES (Revised Encoding Scheme). http://mjn.host.cs.st-andrews.ac.uk/egyptian/res/, 2019.

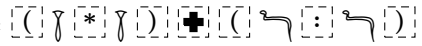
[30] S. Polis. The functions and toposyntax of Ancient Egyptian hieroglyphs. *SIGNATA*, 9:291–363, 2018.

[31] G.A. Reisner and M.B. Reisner. Inscribed monuments from Gebel Barkal. *Zeitschrift für Ägyptische Sprache und Altertumskunde*, 69:24–39, 1933.

[32] S. Rosmorduc. JSesh. http://jseshdoc.qenherkhopeshef.org, 2020.

[33] S.J. Seidlmayer. Eine Schreiberpalette mit änigmatischer Aufschrift. *Mitteilungen des deutschen archäologischen Instituts, Abteilung Kairo*, 47:319–330, 1991.

[34] K. Sethe. *Urkunden der 18. Dynastie, Fascicle 8*. Hinrichs, Leipzig, 1906.

(a)  (b)  = 

(c) 

Table 28: Problems for overlays. For (a) see [22, p. 1], [18, pp. 103-104, Plate 29]. Cf. (c) [21, Plate XXI].

[35] K. Sethe. *Die altaegyptischen Pyramidentexte*, volume 1. Hinrichs, Leipzig, 1908.

[36] K. Sethe. *Urkunden der 18. Dynastie, Fascicle 1.* Hinrichs, Leipzig, 1927.

[37] K. Sethe. *Urkunden der 18. Dynastie, Fascicle 2.* Hinrichs, Leipzig, 1927.

[38] R.A. Wahid, M.-C. Bruwier, N. Gauthier, and M. Haggag. *Antiquités égyptiennes de la Préhistoire à la Basse Époque.* Centre d'Etudes Alexandrines, 2019.

[39] D.A. Werning. *Das Höhlenbuch im Grab des Petamenophis (TT33).* Propylaeum, 2020.

A Overlays revisited

Rare special cases generally deserve a pragmatic solution allowing them to be encoded with the existing controls and syntax. An example is the problem posed to overlays by Table 28 (a) and (b). The issue is that in the past it was assumed that overlays are restricted to being applied on one 'flat' horizontal group and one 'flat' vertical group (cf. Table 1). This restriction was intended to keep implementations simple. Here however, one may argue that the two cobras are arranged by corner insertion. Rather than generalizing the syntax of overlays, a more straightforward solution is to analyze the two cobras in Table 28 (a) and (b) in terms of a flat vertical group, with 'kerning' to leave as little space in between as possible. This is in keeping with the observation that the flat horizontal and vertical groups in overlays are as a rule squeezed together. Moreover, the original inscription in (a) looks less like corner insertion and more like two equally sized cobras that are shifted together; see also the three cobras underneath each other with 'kerning' in Table 28 (c).

```
# ================================================

# GraphemeBreakProperty.txt

13430..13440  ; Control # Cf  [17] EGYPTIAN HIEROGLYPH VERTICAL JOINER..EGYPTIAN HIEROGLYPH MIRROR HORIZONTALLY
13441..13446  ; Control # Cf   [6] EGYPTIAN HIEROGLYPH FULL BLANK..EGYPTIAN HIEROGLYPH WIDE LOST SIGN

# ================================================

# SentenceBreakProperty.txt

13430..13440  ; Format # Cf  [17] EGYPTIAN HIEROGLYPH VERTICAL JOINER..EGYPTIAN HIEROGLYPH MIRROR HORIZONTALLY
13441..13446  ; Format # Cf   [6] EGYPTIAN HIEROGLYPH FULL BLANK..EGYPTIAN HIEROGLYPH WIDE LOST SIGN
...
13000..1342F  ; OLetter # Lo [1072] EGYPTIAN HIEROGLYPH A001..EGYPTIAN HIEROGLYPH V011D

# ================================================

# WordBreakProperty.txt

13430..13440  ; Format # Cf  [17] EGYPTIAN HIEROGLYPH VERTICAL JOINER..EGYPTIAN HIEROGLYPH MIRROR HORIZONTALLY
13441..13446  ; Format # Cf   [6] EGYPTIAN HIEROGLYPH FULL BLANK..EGYPTIAN HIEROGLYPH WIDE LOST SIGN
...
13000..1342F  ; ALetter # Lo [1072] EGYPTIAN HIEROGLYPH A001..EGYPTIAN HIEROGLYPH V011D

# ================================================

# Blocks.txt

13430..1345F; Egyptian Hieroglyph Format Controls

# ================================================

# EastAsianWidth-14.0.0.txt

13000..1342F;N   # Lo  [1072] EGYPTIAN HIEROGLYPH A001..EGYPTIAN HIEROGLYPH V011D
13430..13440;N   # Cf    [17] EGYPTIAN HIEROGLYPH VERTICAL JOINER..EGYPTIAN HIEROGLYPH MIRROR HORIZONTALLY
13441..13446;N   # Lo     [6] EGYPTIAN HIEROGLYPH FULL BLANK..EGYPTIAN HIEROGLYPH WIDE LOST SIGN
13447..13455;N   # Mn    [15] EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP START..
                 #               EGYPTIAN HIEROGLYPH MODIFIER DAMAGED

# ================================================

# Linebreak data: LineBreak.txt

1337C..1342F;AL   # Lo   [180] EGYPTIAN HIEROGLYPH V012..EGYPTIAN HIEROGLYPH V011D
...
13439..1343B;GL   # Cf     [3] EGYPTIAN HIEROGLYPH INSERT AT MIDDLE..EGYPTIAN HIEROGLYPH INSERT AT BOTTOM
1343C;OP          # Cf         EGYPTIAN HIEROGLYPH BEGIN ENCLOSURE
1343D;CL          # Cf         EGYPTIAN HIEROGLYPH END ENCLOSURE
1343E;OP          # Cf         EGYPTIAN HIEROGLYPH BEGIN WALLED ENCLOSURE
1343F;CL          # Cf         EGYPTIAN HIEROGLYPH END WALLED ENCLOSURE
13440;GL          # Cf         EGYPTIAN HIEROGLYPH MIRROR HORIZONTALLY
13441..13446;GL   # Lo     [6] EGYPTIAN HIEROGLYPH FULL BLANK..EGYPTIAN HIEROGLYPH WIDE LOST SIGN
13447..13455;CM   # Mn    [15] EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP START..
                  #               EGYPTIAN HIEROGLYPH MODIFIER DAMAGED
```

Table 29: Properties of the new code points (1/3).

```
# ================================================

# NamesList.txt

1342F EGYPTIAN HIEROGLYPH V011D
* knotted beginning of cartouche
* not to be confused with 13378

@ Sign insertion controls
13439 EGYPTIAN HIEROGLYPH INSERT AT MIDDLE
1343A EGYPTIAN HIEROGLYPH INSERT AT TOP
1343B EGYPTIAN HIEROGLYPH INSERT AT BOTTOM
@ Enclosure controls
1343C EGYPTIAN HIEROGLYPH BEGIN ENCLOSURE
1343D EGYPTIAN HIEROGLYPH END ENCLOSURE
1343E EGYPTIAN HIEROGLYPH BEGIN WALLED ENCLOSURE
1343F EGYPTIAN HIEROGLYPH END WALLED ENCLOSURE
@ Mirror control
13440 EGYPTIAN HIEROGLYPH MIRROR HORIZONTALLY
@ Blank and lost signs
13441 EGYPTIAN HIEROGLYPH FULL BLANK
13442 EGYPTIAN HIEROGLYPH HALF BLANK
13443 EGYPTIAN HIEROGLYPH LOST SIGN
13444 EGYPTIAN HIEROGLYPH HALF LOST SIGN
13445 EGYPTIAN HIEROGLYPH TALL LOST SIGN
13446 EGYPTIAN HIEROGLYPH WIDE LOST SIGN
@ Damage modifiers
13447 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP START
13448 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM START
13449 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT START
1344A EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP END
1344B EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP
1344C EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM START AND TOP END
1344D EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT START AND TOP
1344E EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM END
1344F EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP START AND BOTTOM END
13450 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM
13451 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT START AND BOTTOM
13452 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT END
13453 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP AND END
13454 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM AND END
13455 EGYPTIAN HIEROGLYPH MODIFIER DAMAGED

# ================================================

# Script data: Scripts.txt

13000..1342F  ; Egyptian_Hieroglyphs # Lo [1072] EGYPTIAN HIEROGLYPH A001..EGYPTIAN HIEROGLYPH V011D
13430..13440  ; Egyptian_Hieroglyphs # Cf   [17] EGYPTIAN HIEROGLYPH VERTICAL JOINER..
                                     #               EGYPTIAN HIEROGLYPH MIRROR HORIZONTALLY
13441..13446  ; Egyptian_Hieroglyphs # Lo    [6] EGYPTIAN HIEROGLYPH FULL BLANK..
                                     #               EGYPTIAN HIEROGLYPH WIDE LOST SIGN
13447..13455  ; Egyptian_Hieroglyphs # Mn   [15] EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP START..
                                     #               EGYPTIAN HIEROGLYPH MODIFIER DAMAGED
```

Table 30: Properties of the new code points (2/3).

```
# ================================================

# General properties UnicodeData.txt

1342F;EGYPTIAN HIEROGLYPH V011D;Lo;0;L;;;;;N;;;;;
...
13439;EGYPTIAN HIEROGLYPH INSERT AT MIDDLE;Cf;0;L;;;;;N;;;;;
1343A;EGYPTIAN HIEROGLYPH INSERT AT TOP;Cf;0;L;;;;;N;;;;;
1343B;EGYPTIAN HIEROGLYPH INSERT AT BOTTOM;Cf;0;L;;;;;N;;;;;
1343C;EGYPTIAN HIEROGLYPH BEGIN ENCLOSURE;Cf;0;L;;;;;N;;;;;
1343D;EGYPTIAN HIEROGLYPH END ENCLOSURE;Cf;0;L;;;;;N;;;;;
1343E;EGYPTIAN HIEROGLYPH BEGIN WALLED ENCLOSURE;Cf;0;L;;;;;N;;;;;
1343F;EGYPTIAN HIEROGLYPH END WALLED ENCLOSURE;Cf;0;L;;;;;N;;;;;
13440;EGYPTIAN HIEROGLYPH MIRROR HORIZONTALLY;Cf;0;L;;;;;N;;;;;
13441;EGYPTIAN HIEROGLYPH FULL BLANK;Lo;0;L;;;;;N;;;;;
13442;EGYPTIAN HIEROGLYPH HALF BLANK;Lo;0;L;;;;;N;;;;;
13443;EGYPTIAN HIEROGLYPH LOST SIGN;Lo;0;L;;;;;N;;;;;
13444;EGYPTIAN HIEROGLYPH HALF LOST SIGN;Lo;0;L;;;;;N;;;;;
13445;EGYPTIAN HIEROGLYPH TALL LOST SIGN;Lo;0;L;;;;;N;;;;;
13446;EGYPTIAN HIEROGLYPH WIDE LOST SIGN;Lo;0;L;;;;;N;;;;;
13447;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP START;Mn;0;NSM;;;;;N;;;;;
13448;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM START;Mn;0;NSM;;;;;N;;;;;
13449;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT START;Mn;0;NSM;;;;;N;;;;;
1344A;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP END;Mn;0;NSM;;;;;N;;;;;
1344B;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP;Mn;0;NSM;;;;;N;;;;;
1344C;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM START AND TOP END;Mn;0;NSM;;;;;N;;;;
1344D;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT START AND TOP;Mn;0;NSM;;;;;N;;;;;;
1344E;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM END;Mn;0;NSM;;;;;N;;;;;
1344F;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP START AND BOTTOM END;Mn;0;NSM;;;;;N;;;;
13450;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM;Mn;0;NSM;;;;;N;;;;;;
13451;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT START AND BOTTOM;Mn;0;NSM;;;;;N;;;;;
13452;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT END;Mn;0;NSM;;;;;N;;;;;
13453;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP AND END;Mn;0;NSM;;;;;N;;;;;
13454;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT BOTTOM AND END;Mn;0;NSM;;;;;N;;;;;
13455;EGYPTIAN HIEROGLYPH MODIFIER DAMAGED;Mn;0;NSM;;;;;N;;;;;

# ================================================

# VerticalOrientation-14.0.0.txt

13000..1342F   ; U  # Lo  [1072] EGYPTIAN HIEROGLYPH A001..EGYPTIAN HIEROGLYPH V011D
13430..13440   ; U  # Cf    [17] EGYPTIAN HIEROGLYPH VERTICAL JOINER..EGYPTIAN HIEROGLYPH MIRROR HORIZONTALLY
13441..13446   ; U  # Lo     [6] EGYPTIAN HIEROGLYPH FULL BLANK..EGYPTIAN HIEROGLYPH WIDE LOST SIGN
13447..13455   ; U  # Mn    [15] EGYPTIAN HIEROGLYPH MODIFIER DAMAGED AT TOP START..
                    #               EGYPTIAN HIEROGLYPH MODIFIER DAMAGED
13456..1345F   ; U  # Cn    [10] <reserved-13456>..<reserved-1345F>
```

Table 31: Properties of the new code points (3/3).

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:** **Additional control characters for Ancient Egyptian hieroglyphic texts**
2. Requester's name: Andrew Glass, Jorke Grotenhuis, Mark-Jan Nederhof, Stéphane Polis, Serge Rosmorduc, Daniel A. Werning
3. Requester type (Member body/Liaison/Individual contribution): Liaison contribution
4. Submission date: *December 21, 2021*
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: 13430–1345F
      Name of the existing block: Egyptian Hieroglyph Format Controls
2. Number of characters in proposal: 29
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?
   Andrew Glass
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   Andrew Glass, asg@uw.edu
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.

[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)

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?	Co-authors
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:	*Versions of these control characters are used in current typesetting systems for Egyptian.*
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?	Yes
If YES, is a rationale for such use provided?	*Yes*
If YES, reference:	See attached proposal
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	Yes
If YES, reference:	See attached proposal
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 proposal	
13. Does the proposal contain any Ideographic compatibility characters?	No
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	