L2/20-176

# Cluster model for Egyptian Hieroglyphic Quadrats

Andrew Glass

## Introduction

The Unicode Egyptian Hieroglyph Format Controls (U+13430–U+1344F) were added in Unicode 12.1. These control glyphs enable an author to specify the arrangement of Egyptian Hieroglyphic signs in a quadrat. For example:

| | Linear | Quadrat |
|---|---|---|
| Vertical join | 𓏏 𓄁 | 𓏏𓐰𓄁 |
| | U+133CF U+13101 | U+133CF U+13430 U+13101 |
| Vertical and Horizontal joining | 𓂧 𓇳 𓏤 | 𓂧𓐰𓇳𓐱𓏤 |
| | U+130A7 U+131F3 U+133E4 | U+130A7 U+13430 U+131F3 U+13431 U+133E4 |
| Corner insertion | 𓄿 𓏤 | 𓄿𓐴𓏤 |
| | U+1313F U+133E4 | U+1313F U+13434 U+133E4 |

Such quadrats can be rendered correctly when a suitable font is used in conjunction with text shaping capabilities provided by a platform or application. Platform or application rendering support needs to be updated in order to provide full support for shaping an Egyptian Hieroglyphic quadrat. This document describes the changes necessary to provide that support.

## Cluster model changes

An arrangement of Egyptian Hieroglyphic signs in a single quadrat should be treated by a shaping engine as a valid cluster. A shaping engine based on the Universal Shaping Engine specification[1] or equivalent can be updated to support treating Egyptian quadrats in the following way:

| Sigla | USE class | Derivation |
|---|---|---|
| H | HIEROGLYPH | Egyptian Hieroglyphic signs, viz.: U+13000–U+1342E |
| J | JOINER | Format controls that join Egyptian Hieroglyphic signs: U+13430–U+13436 |
| SB | SEG_BEGIN | The begin segment formatting control: U+13437 |
| SE | SEG_END | The end segment formatting control: U+13438 |
| VS | VARIATION_SELECTOR | U+FE00–U+FE0F |

A single, unjoined, Hieroglyph forms a cluster by itself:

H

A pair of Hieroglyphs may be joined to form a single cluster:

H J H

[1] See https://docs.microsoft.com/en-us/typography/script-development/use, accessed 9/29/2019.

A cluster is still considered complete when a Joiner has been applied, even if there is not yet a following hieroglyph. This may occur during input.

H J

Additional hieroglyphs may be joined up to a maximum of 16 to a cluster:

H (J H)*

Segment begin and end controls may surround a joining group to nest it within an outer group:

SB H J H SE J H

Taken all together this gives the following cluster model:

```
    [SB] H [SE] ( J [SB] H [SE] )*
```

### Notes

- Current shaping engines support cluster sizes up to 32 characters. This is sufficient for clusters with up to 16 signs in a single cluster. Corpus analysis of Middle Egyptian texts suggests that the attested maximum number of signs in a single quadrat is seven.
- Corpus analysis of Egyptian texts done in preparation for L2/17-112R[2] suggests that supporting three levels of embeddings is sufficient for most known texts of all periods.
- Superfluous use of embeddings may prevent text matching in collation scenarios. However, this cluster model does not check for or enforce the minimal use of embeddings. Marking superfluous use of embeddings is left to a higher-level protocol such as orthographic checking.
- To support input of quadrats character-by-character, implementations should support a quadrat terminated by a joiner so the cluster is valid prior to input of the following hieroglyph character.

## Unicode property updates

The above cluster model is generalizable to other hieroglyphic writing systems (e.g., Mayan Hieroglyphs), therefore the model should be supported by properties analogous to ISC properties:

```
# Extended_Syllabic_Category=Hieroglyph
  13000..1342E ; Hieroglyph  # Lo EGYPTIAN HIEROGLYPH A001..EGYPTIAN HIEROGLYPH AA032

# Extended_Syllabic_Category=Hieroglyph_Joiner
  13430.. 13436 ; Hieroglyph_Joiner # Cf EGYPTIAN HIEROGLYPH VERTICAL JOINER..EGYPTIAN
  HIEROGLYPH OVERLAY MIDDLE

# Extended_Syllabic_Category= Hieroglyph_Segment_Begin
  13437 ; Hieroglyph_Segment_Begin # Cf EGYPTIAN HIEROGLYPH BEGIN SEGMENT

# Extended_Syllabic_Category= Hieroglyph_Segment_End
  13438 ; Hieroglyph_Segment_End   # Cf EGYPTIAN HIEROGLYPH END SEGMENT
```

These properties could be maintained by Unicode, or they could be maintained in another location, such as https://github.com/microsoft/font-tools/tree/master/USE. My preference would be to see shaping properties standardized through Unicode.

[2] See https://www.unicode.org/L2/L2017/17112r-quadrat-encoding.pdf accessed 9/29/2019.

## Mirroring

A font should allow for mirroring all non-symetical signs when a run of text is forced RTL. This may be done via the OpenType RTLM feature so that the font can select mirrored forms glyphs and also mirror the overall structure of a quadrat.

A data file enumerates the signs that mirror among currently encoded Egyptian Hieroglyphs. There are two lists: *internal mirrors* maps pairs of encoded Egyptian Hieroglyphs signs that mirror each other; *mirroring* lists signs that need to be supported with mirrored glyphs within the font. This data file is available here: [https://github.com/microsoft/font-tools/blob/master/EgyptianOpenType/mirroring.txt](https://github.com/microsoft/font-tools/blob/master/EgyptianOpenType/mirroring.txt)

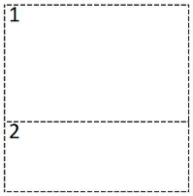
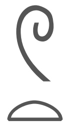
Excerpt from data file:

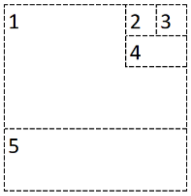
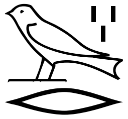
```
# Internal mirrors
130BB; 130BD # D54  𓂻 𓂽 D55
130BD; 130BB # D55  𓂽 𓂻 D54
13132; 13134 # F46  𓄲 𓄴 F47
13133; 13135 # F46a 𓄳 𓄵 F47a
13134; 13132 # F47  𓄴 𓄲 F46
…
# Mirroring
13000 # A1 𓀀
13001 # A2 𓀁
13002 # A3 𓀂
13003 # A4 𓀃
13004 # A5 𓀄
…
```

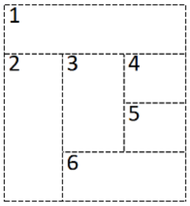
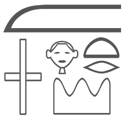
At this time, no mechanism for flipping individual signs is proposed.

## Examples

### Single clusters

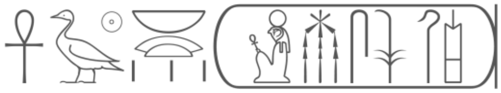
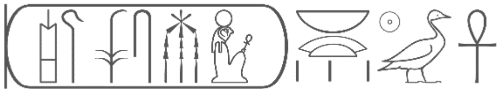
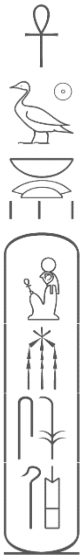
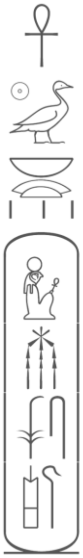
| | | | |
|---|---|---|---|
|  | Source | Rosetta Stone l. 3 |  |
| | Symbolic | Z7 : X1 | |
| | Characters | 𓏲 𓏏 | |
| | Cluster definition | H J H | |
| | Characters | U+133F2 U+13430 U+133CF | |

| | | | |
|---|---|---|---|
|  | Source | Rosetta Stone l. 4 |  |
| | Symbolic | G36 𓐴 ( Z1 * Z1 : Z1) : D21 | |
| | Characters | 𓅨 𓏤𓏤𓏤 𓂋 | |
| | Cluster definition | H J SB H J H J H SE J H | |
| | Characters | U+13168 U+13434 U+13437 U+133E4 U+13431 U+133E4 U+13430 U+133E4 U+13438 U+13430 U+1308B | |

| | | | |
|---|---|---|---|
|  | Source | Abydos temple of Ramesses II, p. 531–2 |  |
| | Symbolic | J15 : Z11 * (D2 * (D21 : X1) : N25) | |
| | Characters | 𓐝𓏶𓁷𓂋𓏏𓈉 | |
| | Cluster definition | H J H J SB H J SB H J H SE J H SE | |
| | Characters | U+1341D U+13430 U+133F6 U+13431 U+13437 U+13077 U+13431 U+13437 U+133CF U+13430 U+1308B U+13438 U+13430 U+13209 U+13438 | |

## Layout

Proper clustering in the rendering engine helps applications respect cluster boundaries when rendering RTL layout and when doing line breaking.

| | |
|---|---|
| **Text rendered left-to-right**<br>The standard direction for Egyptian Hieroglyphs in modern works and the direction specified by Unicode. | |
|  | |
| **Text rendered right-to-left**<br>Text may be forced RTL using directional overrides. In this case, all signs and structures in the run should be mirrored by applying OpenType's RTLM feature in a suitably encoded font. | |
|  | |
| **Text rendered vertically left-to-right**<br>Quadrats stack vertically and orient LTR. Cartouche terminals need to be rotated using OpenType's VERT feature. | **Text rendered vertically right-to-left**<br>Vertical text can also be forced RTL using directional overrides. In this case, both VERT and RTLM should be applied to ensure mirroring of signs and structures and rotation of terminals of enclosing forms. |
|  |  |