

New Unicode control characters for Ancient Egyptian hieroglyphic text

Mark-Jan Nederhof

Andrew Glass

Jorke Grotenhuis

Stéphane Polis

Serge Rosmorduc

Daniel Werning

ICE 2023



Overview

- motivations
- 2019: first 9 controls
- process
- 2022: further 29 controls
- implications
- limitations
- conclusion

Unicode

Unicode can represent characters in (almost) any script for any language, in a unified way:

- visible characters
- control characters

Strengths of Unicode:

- universal
- interoperable
- search

Requirements of encoding of hieroglyphic text

Requirements from Unicode

- standard **not tool**
- characters **not glyphs, not palaeography**
- an encoding should have unique, well-defined meaning
- no two encodings should have same meaning
- implementable in common font technology (OpenType)
- uniformity with other scripts

Requirements from Egyptology

- faithful appearance
- wide coverage achieved by few primitives
- familiarity

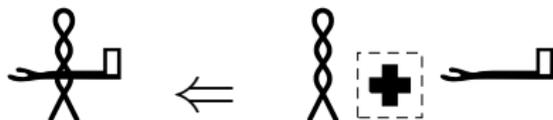
Unicode 12 (March 2019)

Unicode committees were reluctant to grant more than 9 control characters:

- horizontal and vertical joiners
- pair of brackets



- overlay



- 4 corner insertions

Inspired by PLOTTEXT, RES

Corner insertions versus MdC/JSesh

	MdC[1]	MdC[2]	Unicode
	Ax & x	Ax &&& x	  
	x & Ax	x ^^^ Ax	  
	x & Ax & t	x ^^^ Ax &&& x	    
	(?)	(?)	    
	Ax & x & t (?)	(?)	    

Corner insertions versus MdC/JSesh (cont.)

	MdC[1]
(?)	w & D40 & 3
(?)	ir & t & f
(?)	P5 & W & 3
(?)	mH & i & i
(?)	ns & z & n & Z2



MdC[3]

$A_{x\{0,2,81\}}**x\{555,0,58\}**t\{549,564,58\}$

- fine for printed publications
- detrimental to standardization, search, interoperability

Process: Unicode working group on Ancient Egyptian

Main activity from April 2020 to December 2021

Final proposal on control characters authored by:

Andrew Glass	<i>Unicode</i>
Jorke Grotenhuis	<i>Egyptology</i>
Mark-Jan Nederhof	<i>Computer Science & Egyptology</i>
Stéphane Polis	<i>Egyptology</i>
Serge Rosmorduc	<i>Egyptology & Computer Science</i>
Daniel Werning	<i>Egyptology</i>

Further contributions from: Deborah Anderson, Peter Dils, Svenja Gülden, Bob Richmond, Michel Suignard

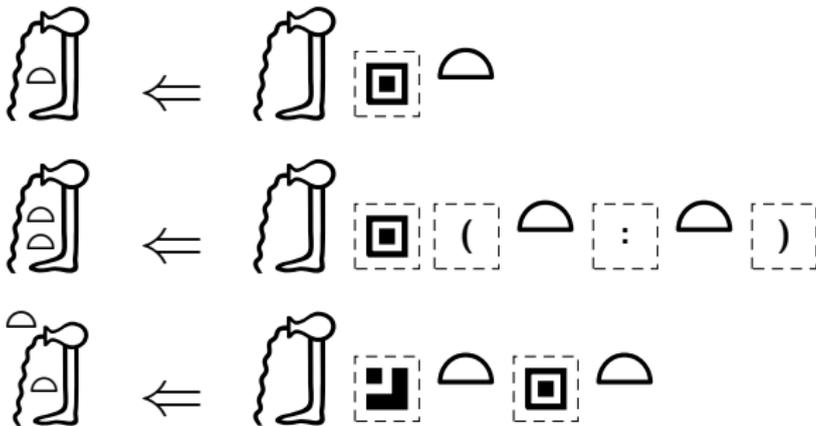
“Design by committee”

Unicode 15 (September 2022)

29 more control characters:

- 3 more insertions
- cartouches and other enclosures
- missing and lost signs
- damaged signs
- mirroring and rotation
- syntax now includes philological brackets

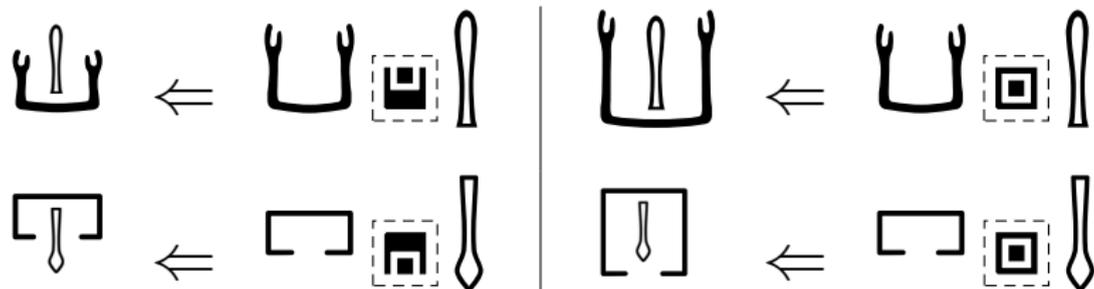
Middle insertion



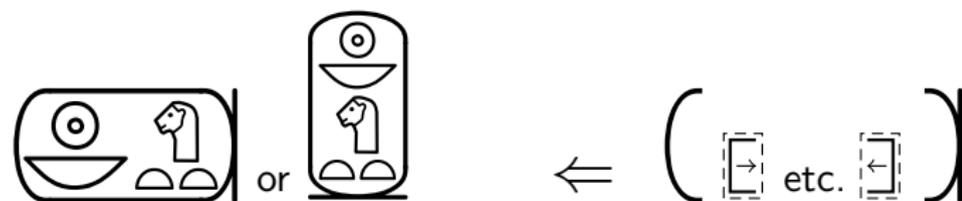
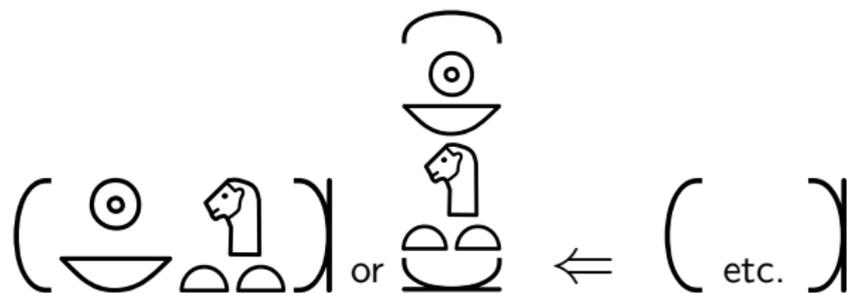
Insertion at top and insertion at bottom



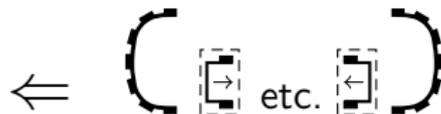
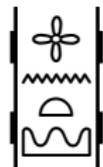
Compare:



Cartouches



Other enclosures



Middle insertion instead of enclosure

... for few inserted signs (*hwt* enclosure)

or if there is clash with text direction



Missing and lost signs

Full & half blank for missing signs:



Four shapes/sizes for lost signs:



Option to 'expand' to cover all available space

Damaged signs

15 modifiers:



Shading of damaged signs connect to expanding 'lost' signs

Note:

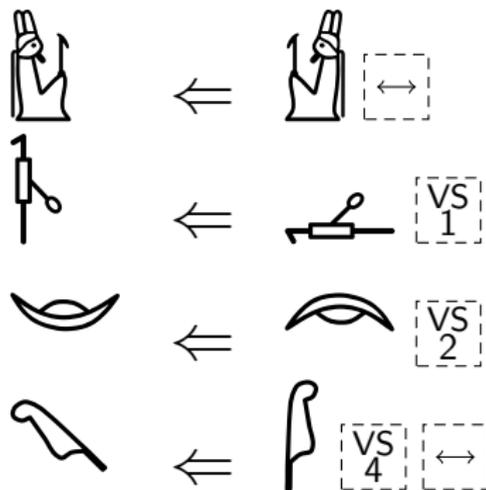
JSesh can also shade quarters of a top-level group ('quadrat')

Unicode can only shade quarters of a sign

Mirroring and rotation

 for horizontal mirroring

7 variation selectors for rotation (roughly multiples of 45 degrees)
(Only defined for certain combinations of signs and rotations)

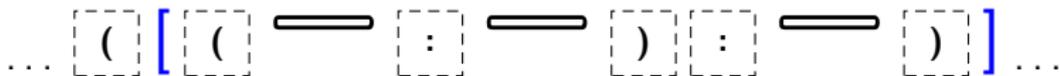


Philological brackets

Some brackets are to interact with hieroglyphs, for now:

(), [], { }, < >, 「 」

Ideally implemented to not take up space for themselves



Implementation

HieroJax

Full implementation for web pages, with graphical editor and more:

<https://nederhof.github.io/hierojax/>

OpenType implementation and input method under development
by Andrew Glass

Are the controls adequate?

Answer depends on requirements

Automatic conversion MdC/JSesh → Unicode

170+ JSesh files considered

70+ of these could be converted almost exactly to Unicode

Obstacles:

- missing signs
- Latin text between hieroglyphs (e.g. “sic”)
- absolute positioning

Also automatic conversion RES → Unicode

120+ hieroglyphic texts currently exist in Unicode

Limitation: Sign list

Currently only 1072 Ancient Egyptian hieroglyphs in Unicode

Extension of Unicode sign list now firmly in hands of Egyptologists:

Peter Dils

Svenja Gülden

Jorke Grotenhuis

Stéphane Polis

Daniel Werning

Substantial progress towards much larger sign list is being made

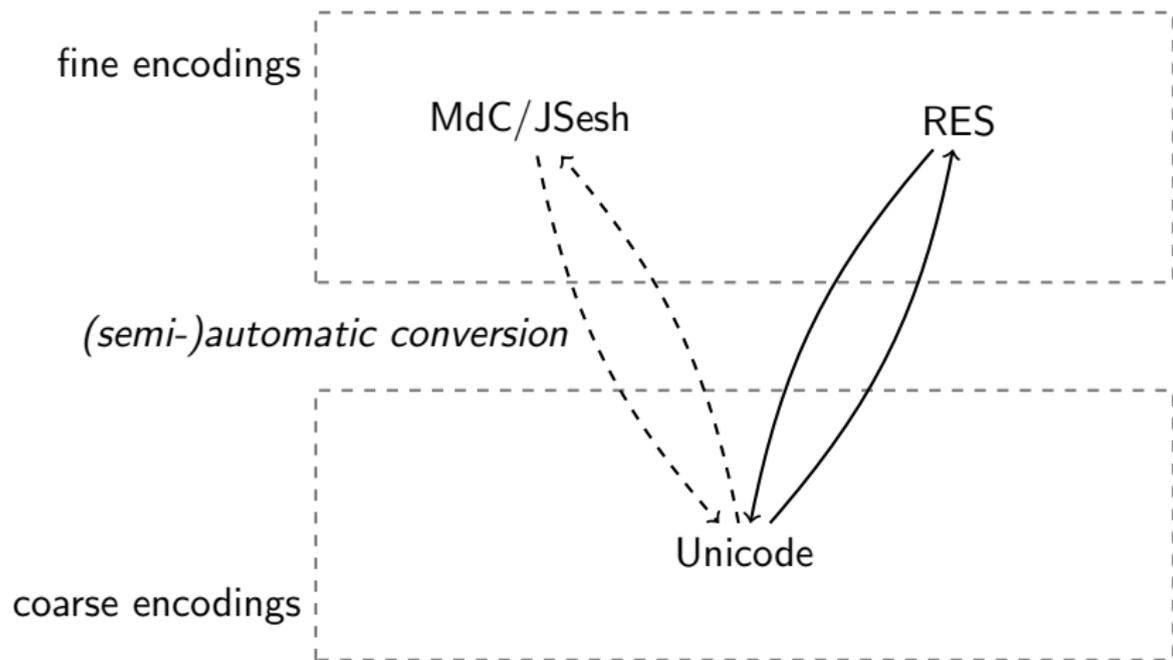
Limitation: Colors

Color (black/red) is not encoded in Unicode

Higher-level protocols (HTML/CSS) can encode colors

But what if one hieroglyphic group includes two colors?

Limitation: Coarse versus fine



Conclusion

Demo:

- graphical editor (as web application)
- Ancient Egyptian in web pages (HieroJax)
- fine-tuning of appearance
- search
- conversion MdC → Unicode
- input method (Windows)
- interoperability