

Computerizing Hieroglyphic Scripts

History, Challenges, and Opportunities
from a Computer Scientist's Perspective

Kevin Graaf

23 June 2026

13.0.13.12.12, 9 Eb', 5 Sek

100% Human-made

Everything in this presentation is 100% human-made.

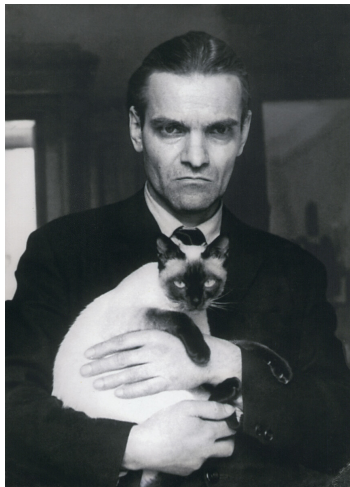
No “AI.”

No LLMs.

- Computer Scientist/Independent Researcher
- Interested in digital typography and text encoding, especially as related to ancient and minoritized scripts

Some History

- Linguists, humanists, etc have made use of digital computers since their invention
- Computers used in Soviet Maya studies since Knorozov's initial decipherment!



Some History

- Linguists, humanists, etc have made use of digital computers since their invention
- Computers used in Soviet Maya studies since Knorozov's initial decipherment!
- Ancient scripts initially excluded from Unicode, this restriction was later relaxed
- Computers have proved valuable in (e.g.) assembling corpora, however...
- We have only begun to properly assess the computational character of scripts *themselves*

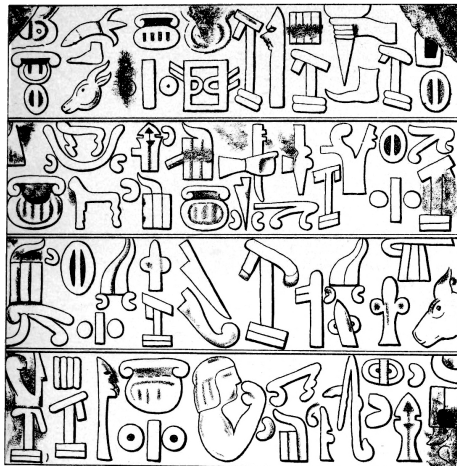


Dramatis Personae

- Luwian
- Egyptian
- and especially, Maya

Anatolian Hieroglyphs

- Used to write Luwian, an ancient Anatolian language



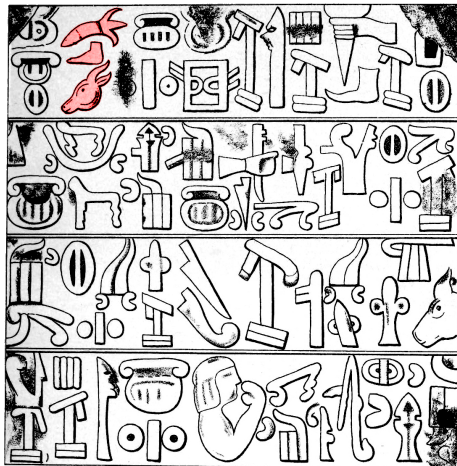
Anatolian Hieroglyphs

- Used to write Luwian, an ancient Anatolian language
- Typically written in *boustrophedon* order



Anatolian Hieroglyphs

- Used to write Luwian, an ancient Anatolian language
- Typically written in *boustrophedon* order
- However, *within* a line, glyphs may vertically stack



- Needs no introduction!



Egyptian

- Needs no introduction!
- Organized into quadrats



- Organized into quadrats (like Egyptian, kind of)
- Typically (though not always) read in zig-zag columns of two. If there are an odd number of columns, the last is vertical.



- Organized into quadrats (like Egyptian, kind of)
- Typically (though not always) read in zig-zag columns of two. If there are an odd number of columns, the last is vertical.
- Now consider that each Maya quadrat *also* has an *internal* reading order which varies depending on the quadrat!

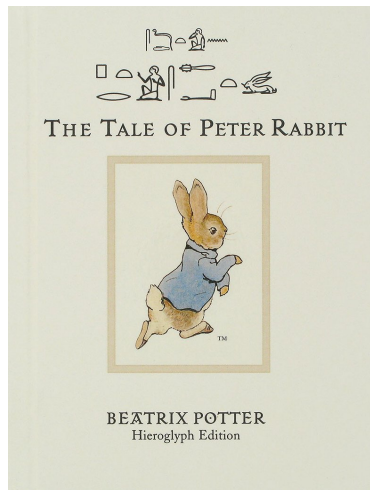


Why Computerize Hieroglyphs?

- Serious need for typesetting in Egyptology

Why Computerize Hieroglyphs?

- Serious need for typesetting in Egyptology
- Occasionally, there's even new texts!



Why Computerize Hieroglyphs?

- Serious need for typesetting in Egyptology
- Occasionally, there's even new texts!
- A typographic presentation could, in my view, help tremendously in teaching the Maya script
- Periodic interest in revival of Maya for modern languages



Why Computerize Hieroglyphs?

- Serious need for typesetting in Egyptology
- Occasionally, there's even new texts!
- A typographic presentation could, in my view, help tremendously in teaching the Maya script
- Periodic interest in revival of Maya for modern languages
- Challenging cases give us a better theoretical understanding of what text is, viz. computers (my initial interest)
- Maybe I *do* want to send an email in Luwian?
- Most importantly, it's just interesting!



What are Hieroglyphs?

- Is "hieroglyphic" a real category?
- I answer, YES!
- In non-hieroglyphic scripts, a "neutral" layout (in a given typeface/hand, for a given page size, etc etc) exists and can be broadly determined from a sequence of graphemes
- In hieroglyphic scripts, we cannot derive any singular, hard-and-fast, or normative rule about how to place/arrange/layout individual glyphs from some sequence of individual graphemes themselves—layout is fundamentally underdetermined
- This "formatting" is inextricably part of an individuated text to no less a degree than punctuation is of text written in contemporary alphabetic scripts
- Put another way—hieroglyphic scripts are irreducibly "two dimensional," even in "neutral" contexts

Yaxchilan Lintel 27 (Excerpt)



K'A'-yi u-[?]SAK-IK'-li

k'a'ay / k'a'aay u...? [u]sak ik'[i]l / ik'[aa]l

k'a'-ay-Ø / k'a'-aay-Ø u-? [u-]sak-ik'-il / -ik-aal

withér-MPAS-3SA 3SE-? [3SE-]white-wind?-POS

"It got withered, his/her ?, his/her white wind/breath"?

(Yaxchilan, Lintel 27: A2-B2)



K'A'-yi-u-[?]SAK-IK'

k'a'ay / k'a'aay u...? [u]sak ik'[il] / ik'[aal]

k'a'-ay-Ø / k'a'-aay-Ø u-? [u-]sak-ik'[-il] / -ik[-aal]

withér-MPAS-3SA 3SE-? [3SE-]white-wind?

"It got withered, his/her ?, his/her white wind/breath"?

(Yaxchilan, Lintel 27: F2)

Writing As a Linear Sequence of Symbols

Plain text is the underlying content stream to which formatting can be applied...[It] must contain enough information to permit the text to be rendered legibly, and nothing more

—The Unicode Specification

[Unicode] inherently assumes that the geometry of 1-dimensional graphetic sequences—as long as there is no 2- dimensional higher structure such as a list or a table—carries no syntactic or semantic information...[but]...[t]here are cases where human creativity has transcended this model...

—Yannis Haralambous and Martin Dürst,
“Unicode from a linguistic point of view”

The Status Quo

No *technical* barrier exists to individually handling hieroglyphic scripts (JSesh, WinGlyph, etc.), but a new, *universal* solution is desired because:

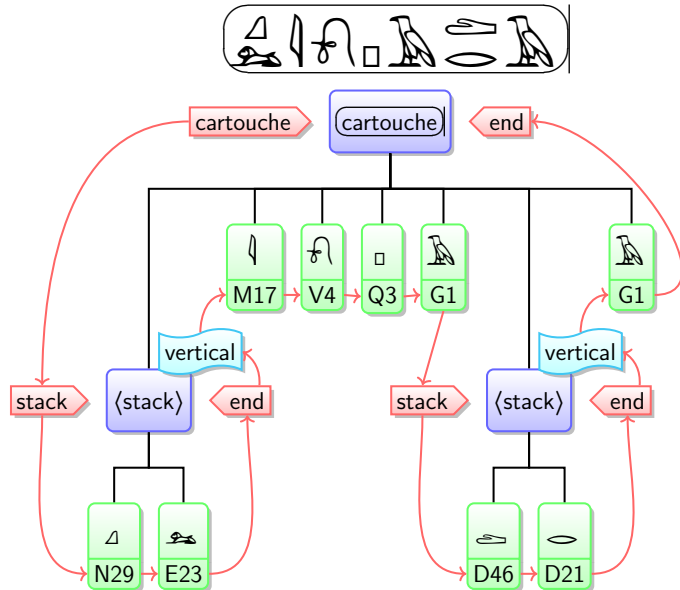
- Offline rendering of hieroglyphic texts renders it a second class citizen, normal methods of text manipulation no longer work properly
- Researchers are at the mercy of software vendors/maintainers, e.g. WinGlyph was retired
- Philosophically, if “normal” text handling software can’t handle text as a category, in what sense can it be said to be complete?
- The economic incentive will never be there: if hieroglyphic scripts are ever to be widely supported, we must identify abstractions which don’t require special treatment for commercially nonviable scripts

Towards Maya Computerization: Challenges

- Not fully deciphered, how do we encode it?
- How should backspace work?
- Should the cursor travel *inside* a quadrat?
- Seems that we need sub-quadrat editing!
- Imagine if alphabets could only be edited at the granularity of words!
- Unfortunately, we can't just ask an ancient Mayan
- Maya also had no print tradition, we can't consult that either
- How do we render it? Glyphs can be shrunk, squeezed, clipped, overlaid...a static set of shapes won't cut it
- Computer text is a new modality of writing, antecedent forms can't tell us the answers!

- A new, *structured* representation/encoding of text
- Telecode text is a sequence of ordered, labeled n -ary trees
- It has three kinds of entities
 - ① Base Characters: letters of the Latin alphabet, numerals, punctuation, Egyptian Hieroglyphs, &c (like regular, non-combining characters in Unicode)
 - ② Tags: introduce structures, e.g. Egyptian cartouches, Maya quadrats, multi-character diacritics (âe)
 - ③ Attributes: modify base characters and structures
- Base characters introduce leaf nodes, tags introduce branch nodes/delimit children, attributes modify nodes—closely related to XML data model
- Base characters are stored in *reading order* (where possible)
- Each category of entity is provisioned from a separate encoding space
- Crucially, this means it is easy to *ignore* structure we don't care about! What I call a *telescoping* representation

Telecode (continued)



The Dubsar Programmable Typography Engine



- After the Sumerian word for scribe (lit. “tablet-writer”)
- My claim: fonts are *functions* from abstract text to concrete forms
- My bigger claim: only *programmable* typography has the generality required to handle *all* categories of human writing
- A Dubsar font is a program in the language of the Dubsar VM bytecode (i.e. a basic programming language)
- A font’s code is constructed such that it fully embodies the “logic” of a writing system in a self-contained manner
- More general mechanisms can be paradoxically simpler
- *All* “advanced” typographic features fall out as modes of use of a single, general mechanism
- *Drastically* simpler: 2500 lines of Standard ML vs. hundreds of thousands (millions?) of lines for avg text handling/rendering stack

Computer Text Ecosystems

- How are texts made? Transformed? Transmitted? Edited? Input?
- Dynamic production of text is just as salient of a question as static representation!
- w.r.t. Maya, I am exploring two input methods:
 - ① A phonetic system
 - ② A “radical” based scheme based on visual components and textural signifiers (inspired by Cangjie, Wubi)
- I have a preliminary phonetic Maya keyboard here with me
- Phonetic input methods are easier, and perhaps more intuitive, but...
- Problematic for any script intended for multiple languages!
- This area is even more under-theorized; much more research needed



Enough talk, let's take a look at Dubsar in action!

There's a lifetime of work to be done here...

- Prepare code for a public release, lots of boring engineering work
- Add enough glyph coverage to typeset original Maya texts
- Formally specify the Maya Telecode encoding
- Develop fonts for other hieroglyphic scripts (esp. Egyptian)
- Integrate Dubsar into everyday application software

Questions?

email: `kevingraaf@kbg.sh`

website: `https://kbg.sh`

code: `https://fossil.kbg.sh`