HISTORICAL SCRIPTS AND INNOVATIONS TO UNICODE ENCODING MODELS

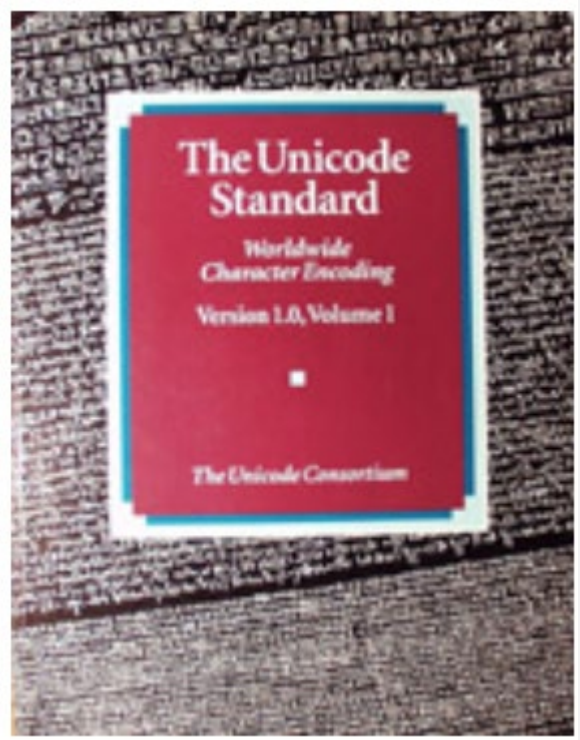
Deborah (Debbie) Anderson

Anshuman Pandey

Script Encoding Initiative, UC Berkeley

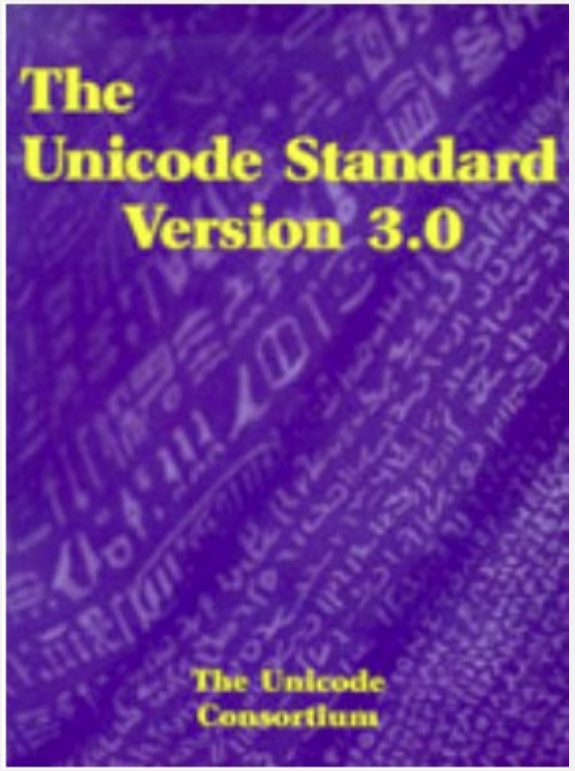
Internationalization & Unicode Conference 45 | 14 October 2021

HISTORICAL SCRIPTS IN UNICODE

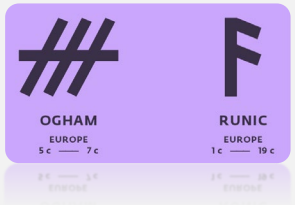


Unicode 1.0 (1991): “Other scripts…being considered for possible addition to the Unicode standard are: **Egyptian Hieroglyphics**, … Interest has also been expressed in including **Cuneiform, …** and **Glagolitic**”.

HISTORICAL SCRIPTS IN UNICODE

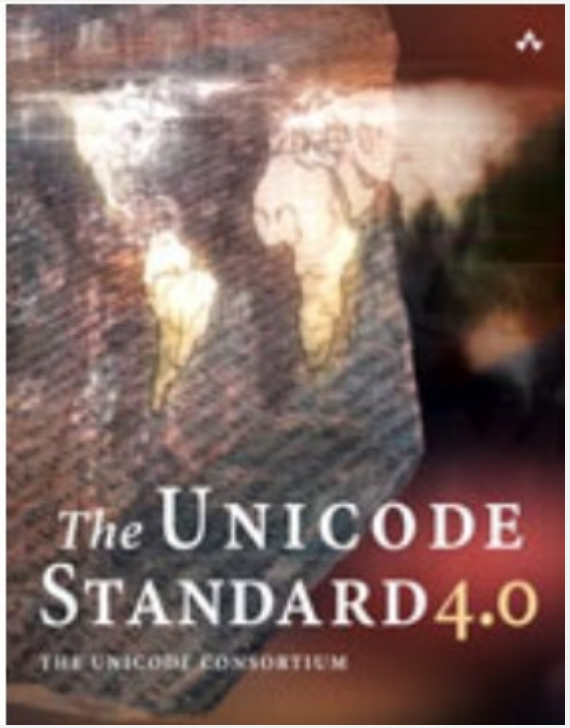


Unicode 1.1 (1993): [from other standards and industry priorities]
Unicode 2.0 (1996): [as above]
Unicode 3.0 (1999): [primarily national scripts]
+ Ogham, Runic



Glyphs from World's Writing System poster by Bergerhausen et al.

HISTORICAL SCRIPTS IN UNICODE



Unicode 1.1 (1993): --
Unicode 2.0 (1996): --
Unicode 3.0 (1999): Ogham, Runic

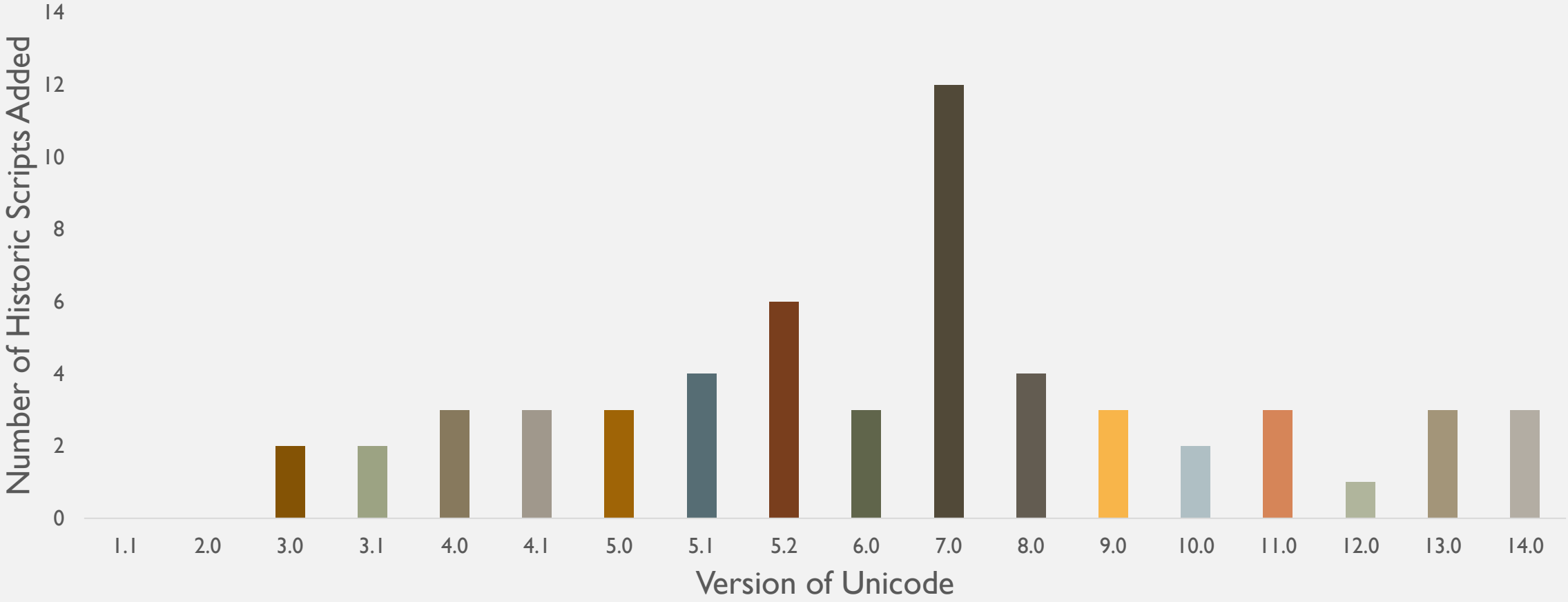
Unicode 3.1 (2001): Old Italic, Gothic

Unicode 4.0 (2003): Linear B, Cypriot, Ugaritic

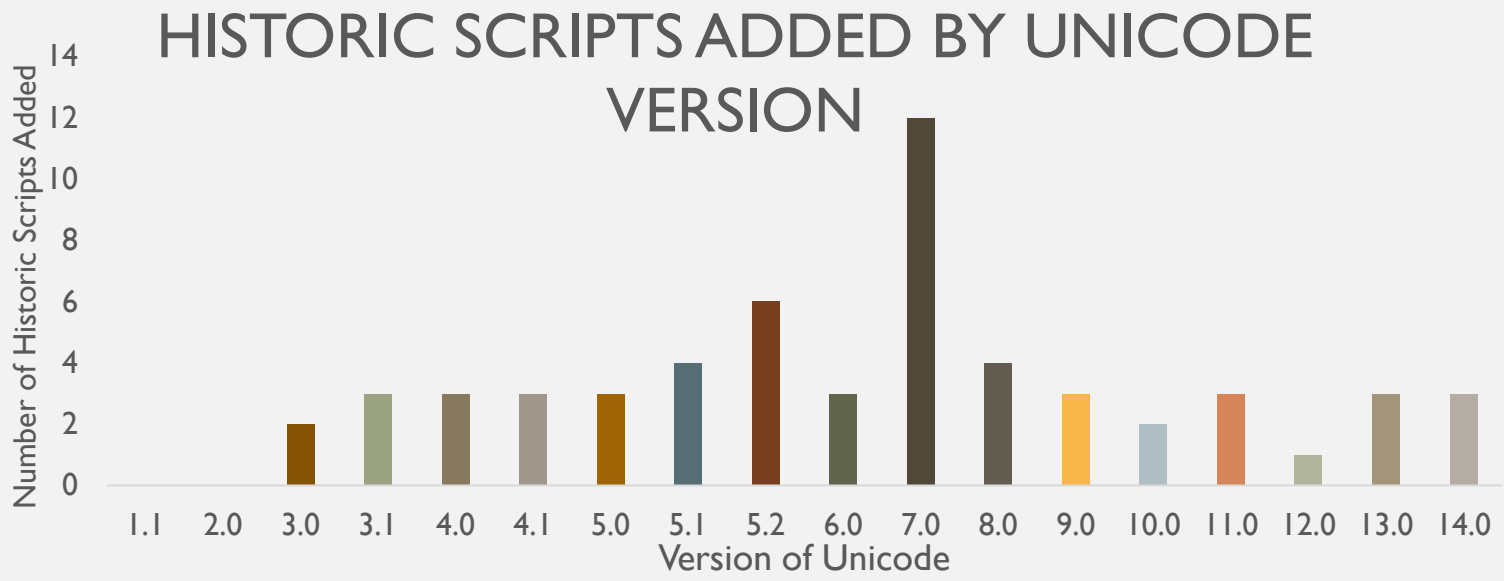
Unicode 4.1 (2005): Glagolitic, Old Persian, Kharoshthi

Glyphs from World's Writing System poster by Bergerhausen et al.

HISTORIC SCRIPTS ADDED BY UNICODE VERSION



36% of all encoded scripts are historic (58 out of 159)



SCRIPT ENCODING INITIATIVE

SEI

Script Encoding Initiative
Department of Linguistics
University of California, Berkeley

- **Home**
- Scripts to Encode
- Progress Overview
- UTC Reports
- News & Presentations
- Press
- About Us
- How to Donate
- Donors

WELCOME TO THE SCRIPT ENCODING INITIATIVE

The Script Encoding Initiative (SEI), established in the UC Berkeley Department of Linguistics in April 2002, is a project devoted to the preparation of formal proposals for the encoding of scripts and script elements not yet currently supported in Unicode (ISO/IEC 10646).

Unicode is the universal computing standard specifying the representation of text in all modern software. To date, Unicode has largely focused on the major modern scripts, particularly those scripts most widely used in business. Some minority and historic scripts have already been encoded, as well as historic characters of the major modern scripts.

The goal of the SEI project is to fund the preparation of script proposals that will be successfully approved by the Unicode Technical Committee and WG2 (ISO/IEC 10646) without requiring extensive revision or involvement of the committee itself.

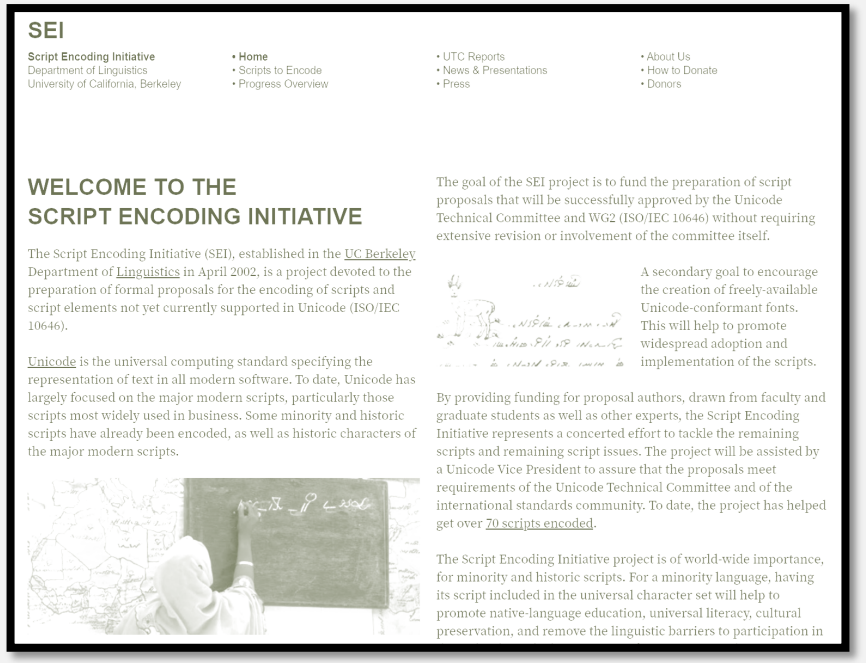
A secondary goal to encourage the creation of freely-available Unicode-conformant fonts. This will help to promote widespread adoption and implementation of the scripts.

By providing funding for proposal authors, drawn from faculty and graduate students as well as other experts, the Script Encoding Initiative represents a concerted effort to tackle the remaining scripts and remaining script issues. The project will be assisted by a Unicode Vice President to assure that the proposals meet requirements of the Unicode Technical Committee and of the international standards community. To date, the project has helped get over 70 scripts encoded.

The Script Encoding Initiative project is of world-wide importance, for minority and historic scripts. For a minority language, having its script included in the universal character set will help to promote native-language education, universal literacy, cultural preservation, and remove the linguistic barriers to participation in

SCRIPT ENCODING INITIATIVE

Over 80% of historic scripts encoded in Unicode have come through SEI



SEI

Script Encoding Initiative
Department of Linguistics
University of California, Berkeley

• Home
• Scripts to Encode
• Progress Overview

• UTC Reports
• News & Presentations
• Press

• About Us
• How to Donate
• Donors

WELCOME TO THE SCRIPT ENCODING INITIATIVE

The Script Encoding Initiative (SEI), established in the UC Berkeley Department of Linguistics in April 2002, is a project devoted to the preparation of formal proposals for the encoding of scripts and script elements not yet currently supported in Unicode (ISO/IEC 10646).

Unicode is the universal computing standard specifying the representation of text in all modern software. To date, Unicode has largely focused on the major modern scripts, particularly those scripts most widely used in business. Some minority and historic scripts have already been encoded, as well as historic characters of the major modern scripts.

The goal of the SEI project is to fund the preparation of script proposals that will be successfully approved by the Unicode Technical Committee and WG2 (ISO/IEC 10646) without requiring extensive revision or involvement of the committee itself.

A secondary goal to encourage the creation of freely-available Unicode-conformant fonts. This will help to promote widespread adoption and implementation of the scripts.

By providing funding for proposal authors, drawn from faculty and graduate students as well as other experts, the Script Encoding Initiative represents a concerted effort to tackle the remaining scripts and remaining script issues. The project will be assisted by a Unicode Vice President to assure that the proposals meet requirements of the Unicode Technical Committee and of the international standards community. To date, the project has helped get over 70 scripts encoded.

The Script Encoding Initiative project is of world-wide importance, for minority and historic scripts. For a minority language, having its script included in the universal character set will help to promote native-language education, universal literacy, cultural preservation, and remove the linguistic barriers to participation in

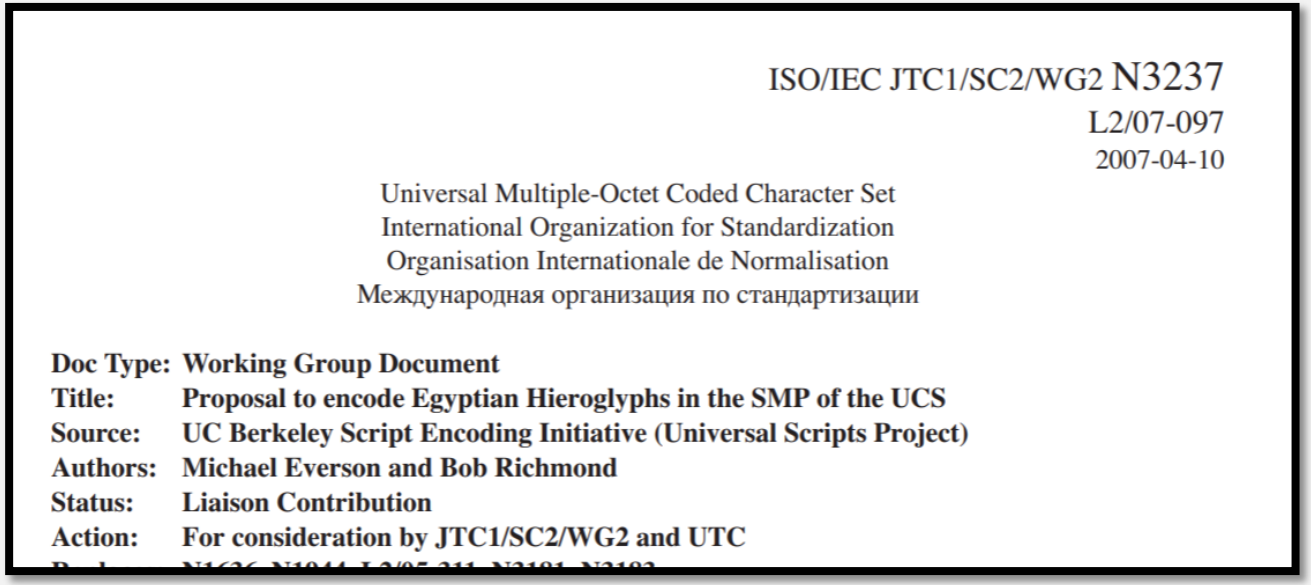
HIEROGLYPHS

Glyphs from World's Writing System poster by Bergerhausen et al.

EGYPTIAN HIEROGLYPHS

EGYPTIAN HIEROGLYPHS

- First proposed in 1997 by Michael Everson
- Comments from users 1999, meetings 2002 and 2006
- Final proposal 2007 by Everson/Richmond; script published Unicode 5.2, 2009



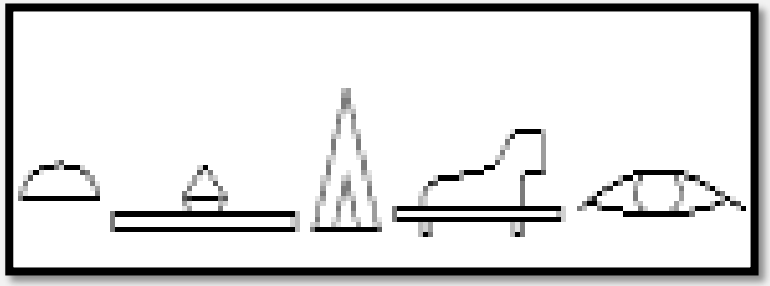
ISO/IEC JTC1/SC2/WG2 N3237
L2/07-097
2007-04-10

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: **Working Group Document**
Title: **Proposal to encode Egyptian Hieroglyphs in the SMP of the UCS**
Source: **UC Berkeley Script Encoding Initiative (Universal Scripts Project)**
Authors: **Michael Everson and Bob Richmond**
Status: **Liaison Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**

EGYPTIAN HIEROGLYPHS

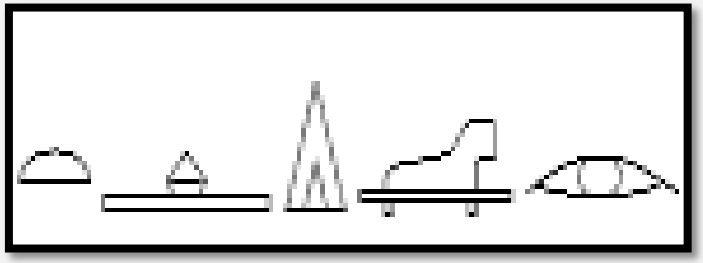
- First proposed in 1997 by Michael Everson
- Comments from users 1999, meetings 2002 and 2006
- Final proposal 2007 by Everson/ Richmond; script published Unicode 5.2, 2009

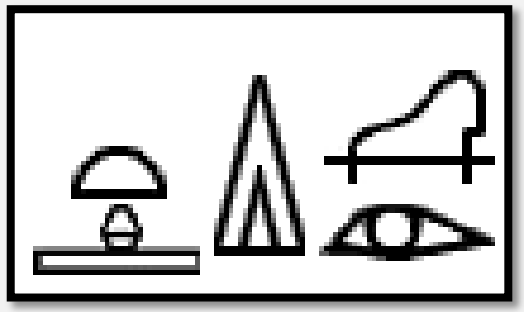


EGYPTIAN HIEROGLYPHS
ISSUES WITH ORIGINAL REPERTOIRE

After publication in Unicode, two issues arose for Egyptologists:

1. Positioning of glyphs





EGYPTIAN HIEROGLYPHS
ISSUES WITH ORIGINAL REPERTOIRE

After publication in Unicode, two issues arose for Egyptologists:

1. Positioning of glyphs
2. Limited repertoire

```
5.2 # [1071] EGYPTIAN HIEROGLYPH A001..EGYPTIAN HIEROGLYPH AA032
```

EGYPTIAN HIEROGLYPHS
ISSUES WITH ORIGINAL REPERTOIRE

1. Positioning of glyphs

Format characters proposed 2016 and 2017 by Andrew Glass et al.; published in Unicode 12.0 2019

13430	𓍼	EGYPTIAN HIEROGLYPH VERTICAL JOINER
13431	𓍽	EGYPTIAN HIEROGLYPH HORIZONTAL JOINER
13432	𓍾	EGYPTIAN HIEROGLYPH START AT TOP
13433	𓍿	EGYPTIAN HIEROGLYPH START AT BOTTOM
13434	𓎀	EGYPTIAN HIEROGLYPH END AT TOP
13435	𓎁	EGYPTIAN HIEROGLYPH END AT BOTTOM
13436	𓎂	EGYPTIAN HIEROGLYPH OVERLAY MIDDLE
13437	𓎃	EGYPTIAN HIEROGLYPH BEGIN SEGMENT
13438	𓎄	EGYPTIAN HIEROGLYPH END SEGMENT

EGYPTIAN HIEROGLYPHS
ISSUES WITH ORIGINAL REPERTOIRE

2. Limited repertoire

Large set of extensions (10,350 glyphs under review)

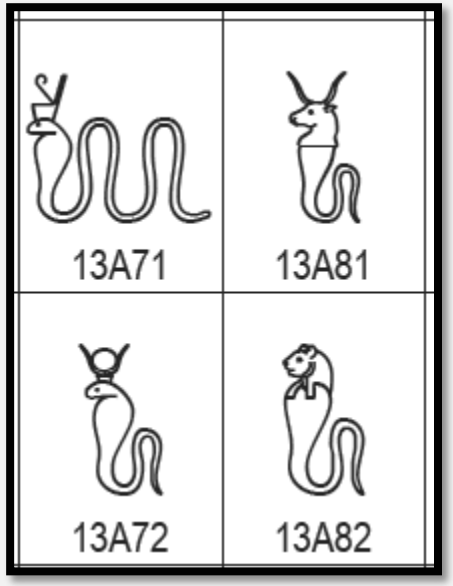
WG2 N5163
L2/21-108
Date: 2021-06-14

Title: Consideration for the encoding of an extended Egyptian Hieroglyphs repertoire

Source: Michel Suignard
Status: Individual Contribution
Distribution: UTC, WG2
Replace: N5128R (L2/20-068R)

Executive Summary:

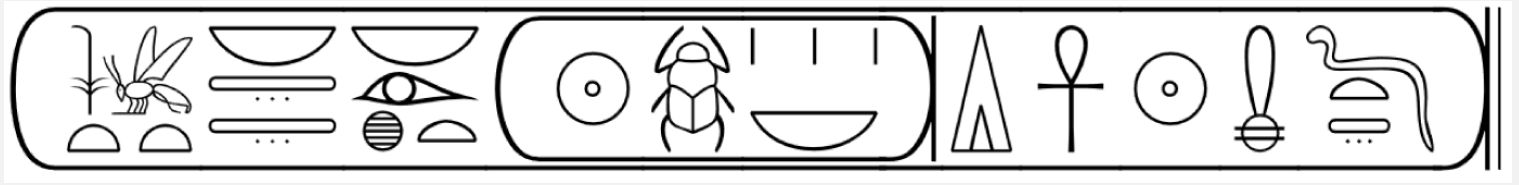
This document presents the current stage of research on the Egyptian Hieroglyph extensions as done by the author. Unlike previous version, it does not formally propose a set of characters for immediate encoding. Instead, it describes a large set of



EGYPTIAN HIEROGLYPHS ADVANCES

Until now, many Egyptologists have been using JSesh (etc.)

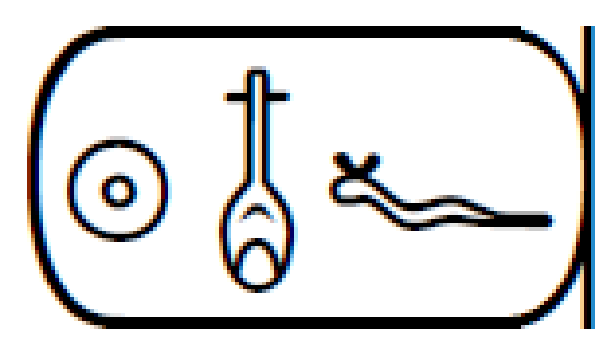
Now, work can take advantage of technical advances:
OpenType and Universal Shaping Engine (USE)



EGYPTIAN HIEROGLYPHS ADVANCES

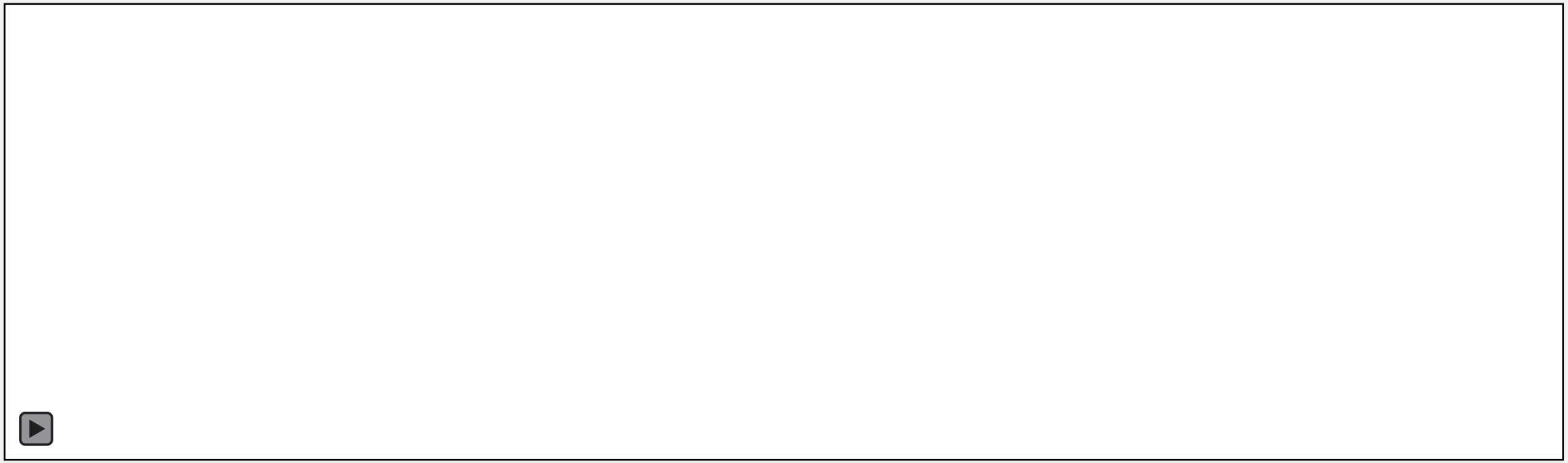
Forthcoming proposal for additional format controls:

- Cartouches and other enclosures (with format controls)



EGYPTIAN HIEROGLYPHS ADVANCES

Demo by Andrew Glass



EGYPTIAN HIEROGLYPHS ADVANCES

- Rotation (with format control or variation selector)

A1

A1 90

A1 180

A1 270

EGYPTIAN HIEROGLYPHS ADVANCES

- Rotation (with format control or variation selector)
- Mirroring (with format control)

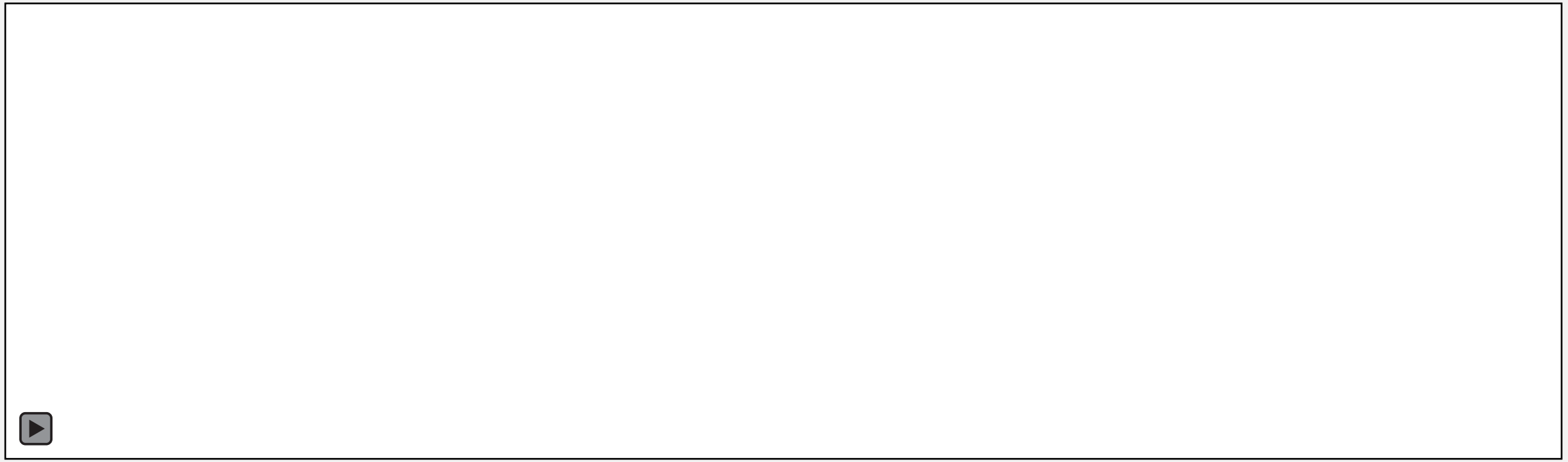
EGYPTIAN HIEROGLYPHS ADVANCES

- Rotation (with format control or variation selector)
- Mirroring (with format control)
- Shading to indicate damaged areas (with format control)

- Editorial marks (brackets, etc.)

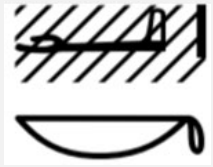
EGYPTIAN HIEROGLYPHS ADVANCES

Demo by Andrew Glass

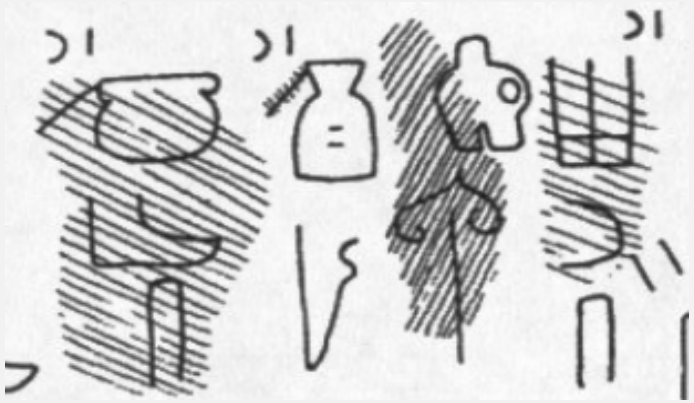


EGYPTIAN HIEROGLYPHS ADVANCES

Egyptian Hieroglyph



Anatolian Hieroglyphs

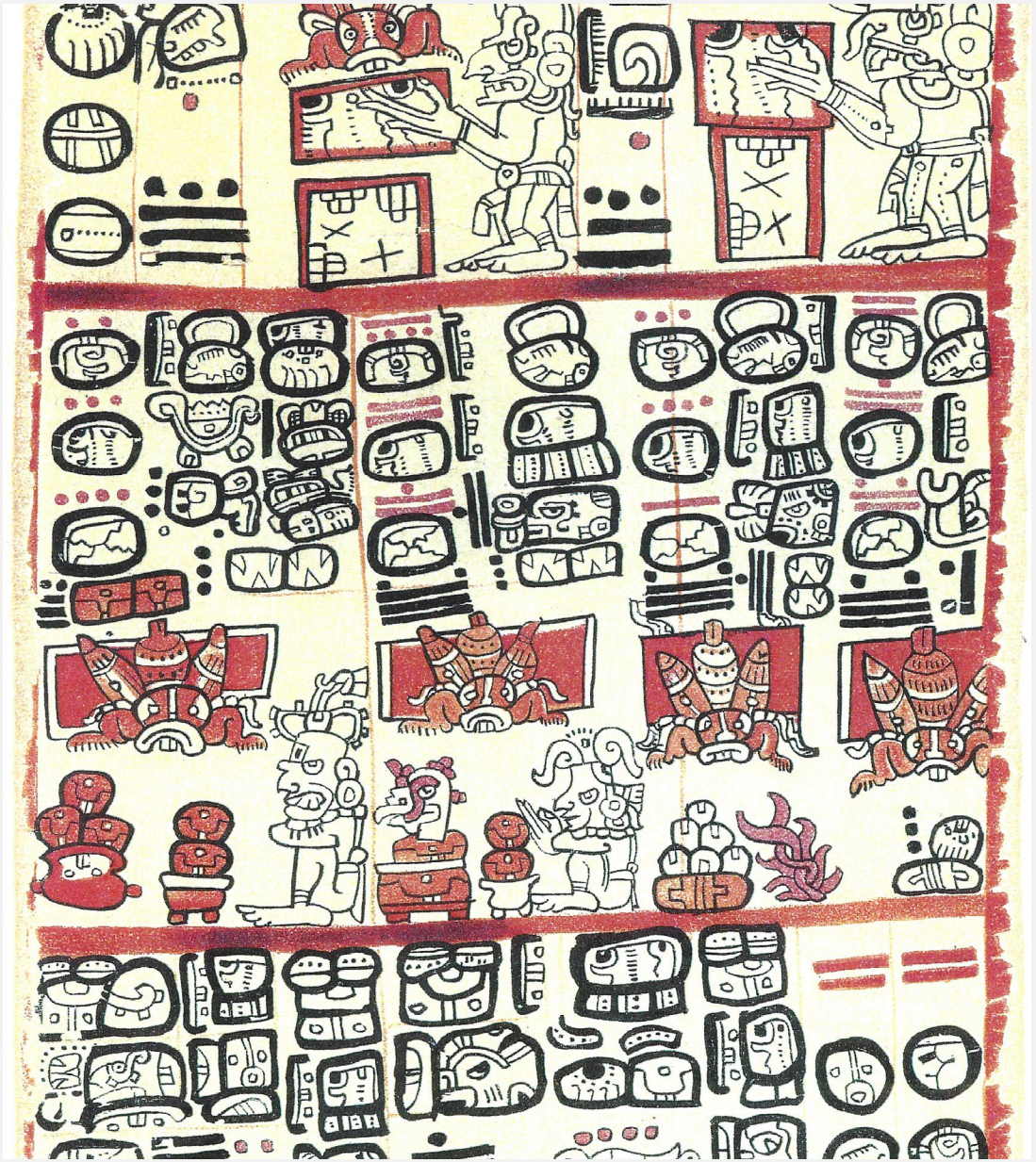


Cuneiform (Hittite)

EGYPTIAN HIEROGLYPHS
OTHER ISSUES: COLLABORATION

Need to involve script and Unicode experts

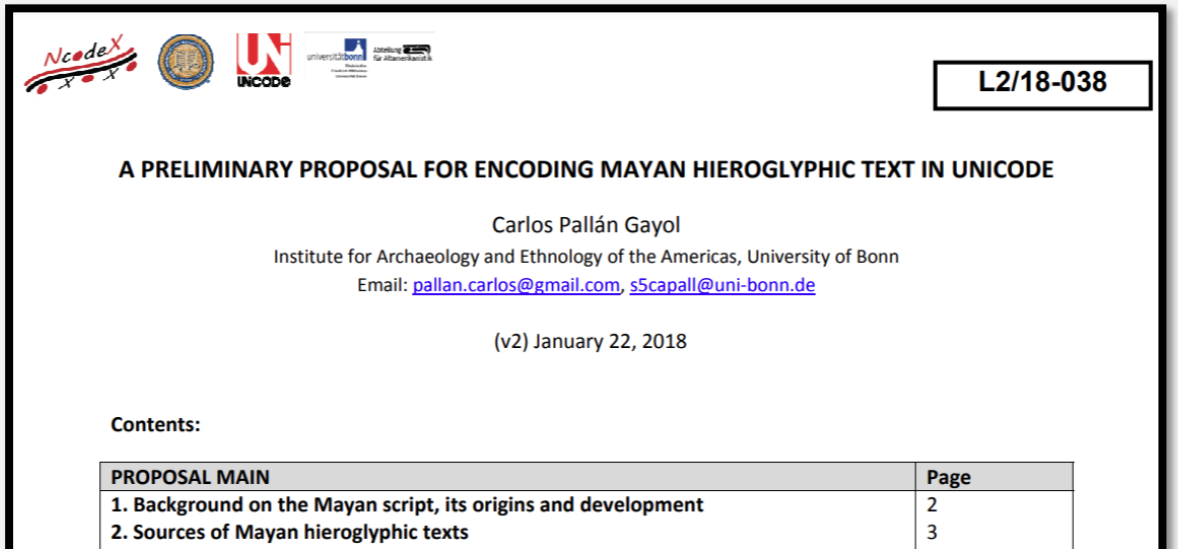
MAYA HIEROGLYPHS
(FROM POST-CLASSIC AND CLASSIC PERIODS)



Madrid Codex (from Post-Classic period [ca. 1100-1519 AD])

MAYA HIEROGLYPHS

- Discussions started 2015 with Carlos Pallán; preliminary proposal 2018 on signs from Codices by Pallán (later with 2020 update)
- Subsequent work on Classic period signs by Dr. Gabrielle Vail and colleagues
- Not yet in Unicode



L2/18-038

A PRELIMINARY PROPOSAL FOR ENCODING MAYAN HIEROGLYPHIC TEXT IN UNICODE

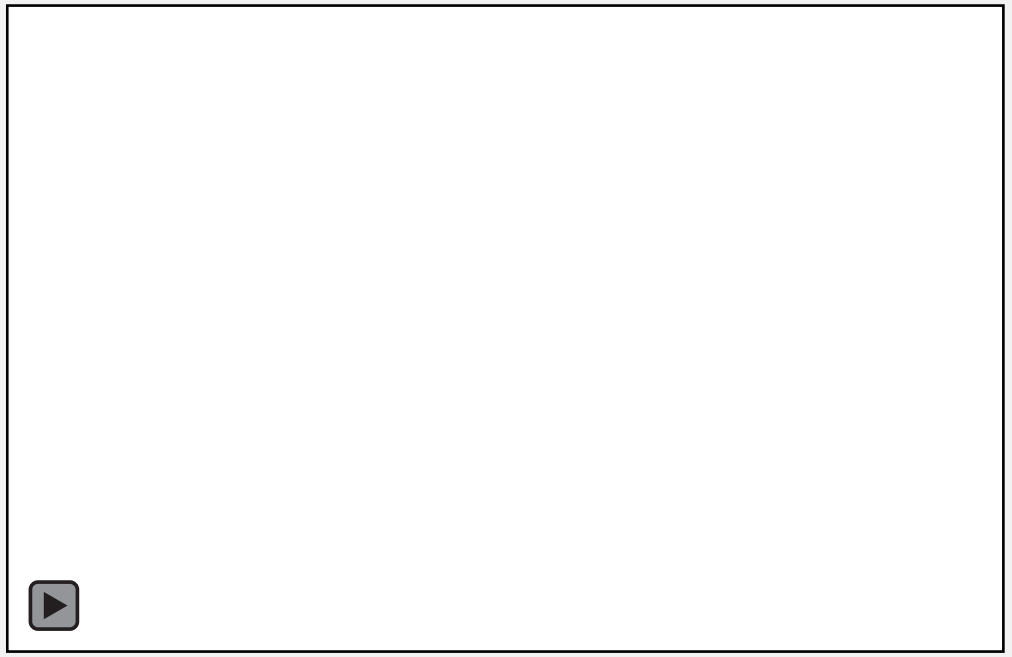
Carlos Pallán Gayol
Institute for Archaeology and Ethnology of the Americas, University of Bonn
Email: pallan.carlos@gmail.com, s5capall@uni-bonn.de

(v2) January 22, 2018

Contents:

MAYA HIEROGLYPHS ADVANCES

- Work on fonts and rendering can take advantage of advances done for Egyptian Hieroglyphs and be applicable to other hieroglyph scripts (i.e., Aztec)

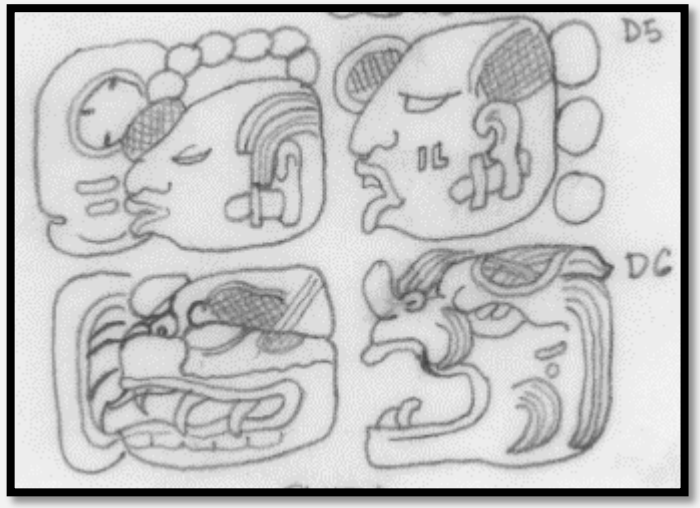


MAYA HIEROGLYPHS ISSUES

- Collaboration key to progress on proposal
- Results need to be useful for users and cover their needs

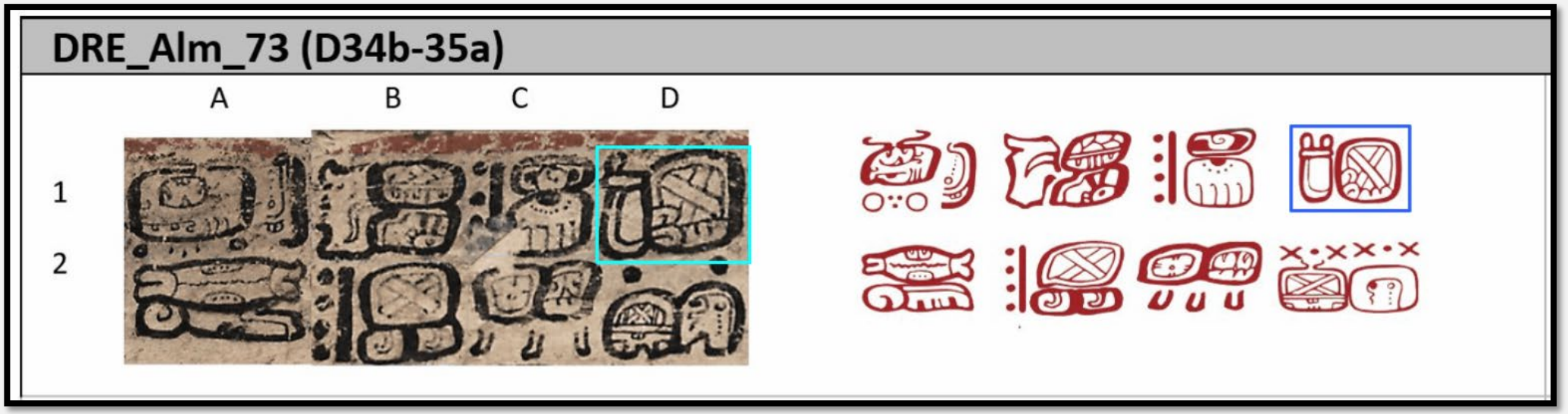
MAYA HIEROGLYPHS ISSUES

- Access to materials



Linda Schele drawing SD-7675
Source: LACMA with the permission of David Schele

MAYA HIEROGLYPHS
ISSUES



From demo by Carlos Pallán at Digital Humanities conference, Utrecht, July 2019 (https://ncodex.org/Demo_01_NcodeX_DRE_analysis_block_level.php)

THOUGHTS ON ENCODING HISTORIC SCRIPTS IN UNICODE

- Historic scripts were part of the original vision for Unicode and should be supported in fonts/implementations
- Historic scripts often have limited materials available; Egyptian and Maya H are exceptions – they have plentiful materials but this causes problems for review
- Timeline from proposal to approval/publication can be very long (but hieroglyphs now can take advantage of technical advances)
- Involve experts early on in Unicode discussions to set realistic expectations

ADDITIONAL THOUGHTS

- How to support undeciphered scripts that are actively used by scholars?
- How to balance character encoding processes and scholarly needs
  - Character encoding requires semantic understanding of signs being encoded
  - Scholars need to exchange data as part of decipherment processes
- Requires rethinking character encoding norms
  - Expansion of character-glyph model:
    - Encoding of variants, alternates
    - Encoding character / graphical primitives

PROTO-SINAITIC

SIDETIC

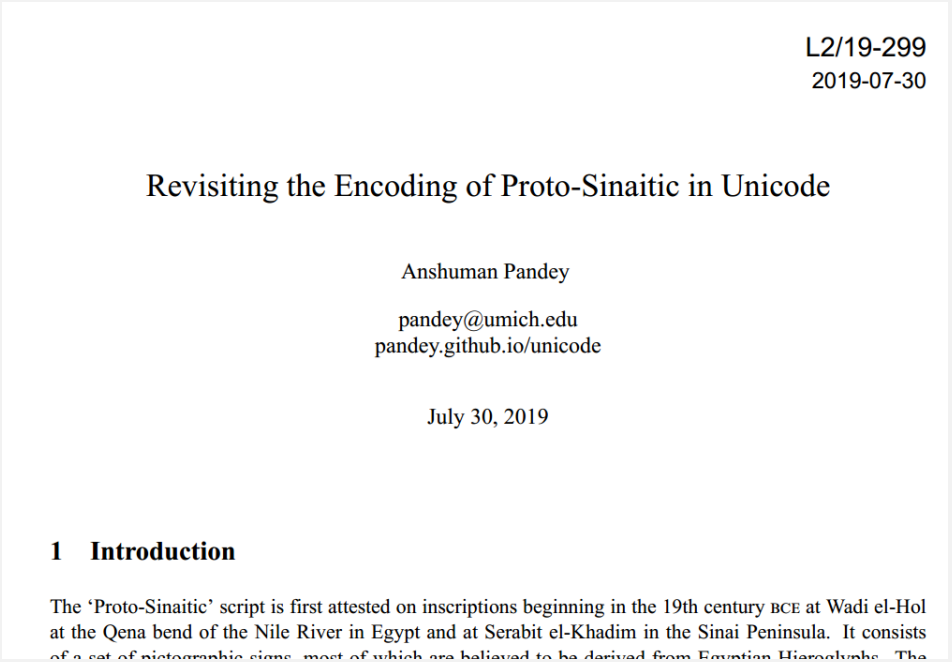
BOOK PAHLAVI

- Commonalities
  - Small repertoire
  - Large variation
  - Significant ambiguity
- Requirements
  - Active usage
  - Corpus development
  - Data interchange
- Challenges
  - No scholarly standards
  - No existing technical support

PROTO-SINAITIC

PROTO-SINAITIC

- Placed on "Not The Roadmap" in 2002
- Allocated to SMP in 2019



L2/19-299
2019-07-30

Revisiting the Encoding of Proto-Sinaitic in Unicode

Anshuman Pandey

pandey@umich.edu
pandey.github.io/unicode

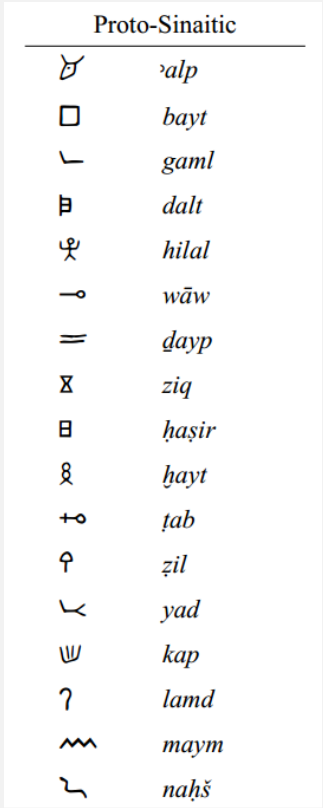
July 30, 2019

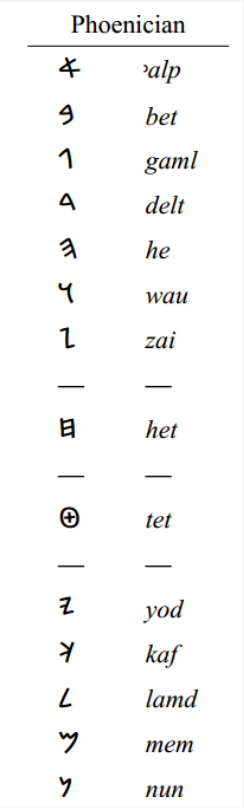
1 Introduction

The 'Proto-Sinaitic' script is first attested on inscriptions beginning in the 19th century BCE at Wadi el-Hol at the Qena bend of the Nile River in Egypt and at Serabit el-Khadim in the Sinai Peninsula. It consists

PROTO-SINAITIC BACKGROUND

- First attested alphabet
- Derived from hieroglyphs in 2nd millennium BCE
  - West Semitic miners? Literate officials?
- Used for inscriptions
  - Sinai and Wadi el-Hol in Egypt
- Spread throughout Eastern Mediterranean
  - Developed and standardized by Phoenicians
    - Ancestor of Latin, Greek, etc



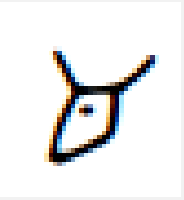


PROTO-SINAITIC BACKGROUND

k3

Semitic
"ox"
ʿalp̣u / ʿalaṗ

Semitic
⟨ʔ⟩



aleph
/ʾ/

Egyptian
3,000 BC

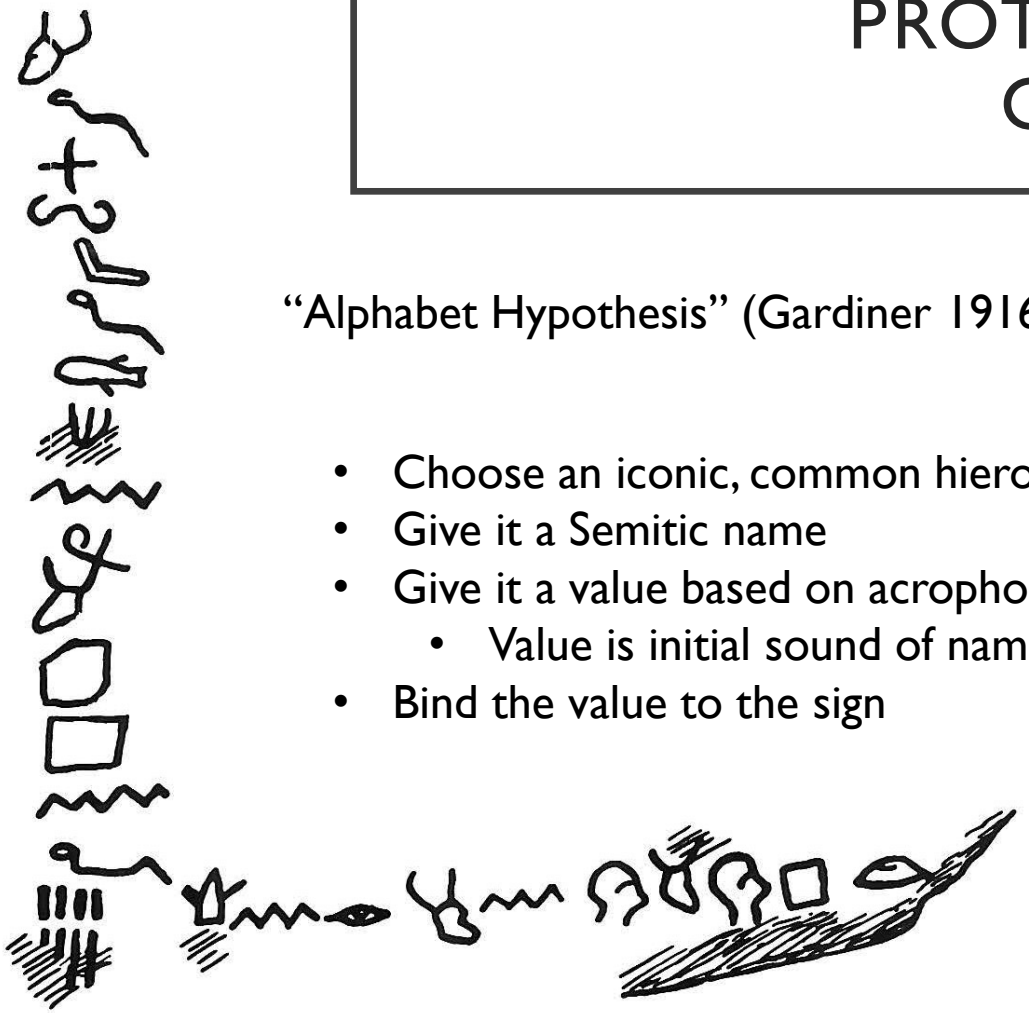
Sinai
1,850 BC

Phoenician aleph
1,200 BC

Greek alpha
600 BC

Roman A
114 AD

PROTO-SINAITIC ORIGINS



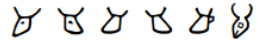
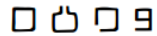
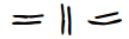
"Alphabet Hypothesis" (Gardiner 1916)

- Choose an iconic, common hieroglyph
- Give it a Semitic name
- Give it a value based on acrophony
  - Value is initial sound of name
- Bind the value to the sign

value	PS	EH	Gardiner	Egyptian	meaning
ʾ	𐤀	𓄿	F1	kꜣ	head of an ox
b	𐤁	𓉐	O1	pr	house, plan of house
h	𐤄	𓀀	A28	qꜣ, qꜣi	man-arms-uplifted
m	𐤌	𓈖	N35	n	ripple of water
n	𐤍	𓇋	I10	ḏ	cobra in repose
r	𐤓	𓂝	D19	fndj	nose

PROTO-SINAITIC ISSUES: VARIATION

- Repertoire
  - At least 27 attested signs
  - Variant forms of a single sign
    - Some used contemporaneously
- Challenges
  - Are directional variants semantically distinctive?
  - Are orientational variants semantically distinctive?
  - Character semantics not established

Letterforms	Name	Meaning	Value
	*ʾalp*	ox head	ʾ
	*bayt*	house	b
	*gaml*	throw stick	g
	*dalt*	door	d
	ḥe	fence	ḥ
	*ho*	man calling	h
	hll	jubilate	h
	*wāw*	hook	w
	*ḏayp*	eyebrow	ḏ

PROTO-SINAITIC ISSUES: VARIATION

Forms of *mayim*

Forms of *he*

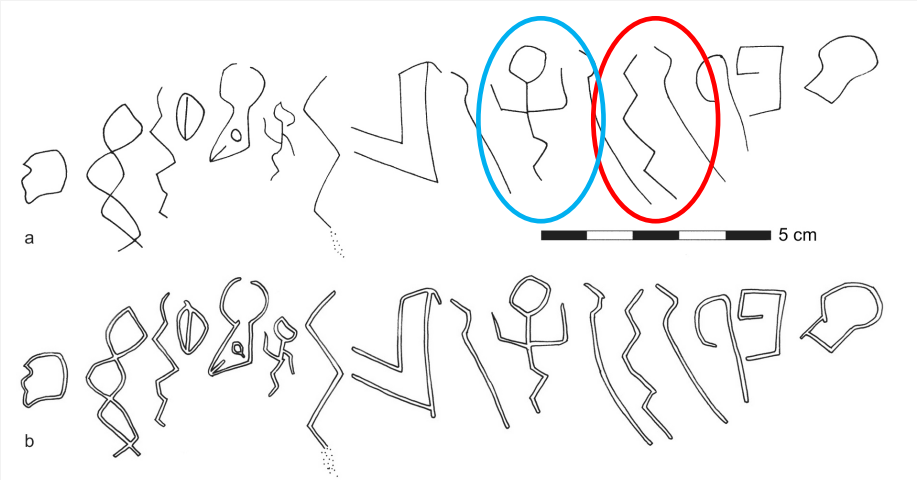


Fig. 2. *Wadi el-Hôl Alphabetic Inscription 1; the Horizontal Inscription. a) facsimile of the base of the incised lines; b) facsimile of the outer edges of the incised lines.*

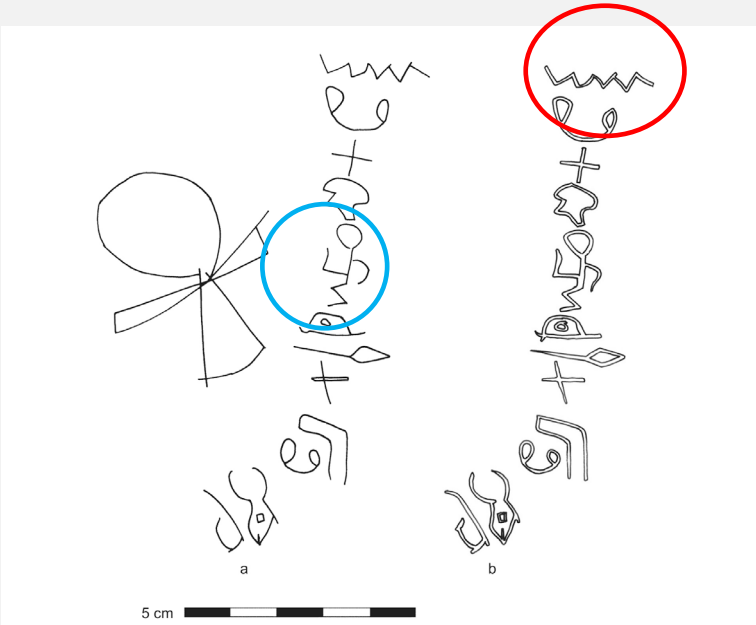


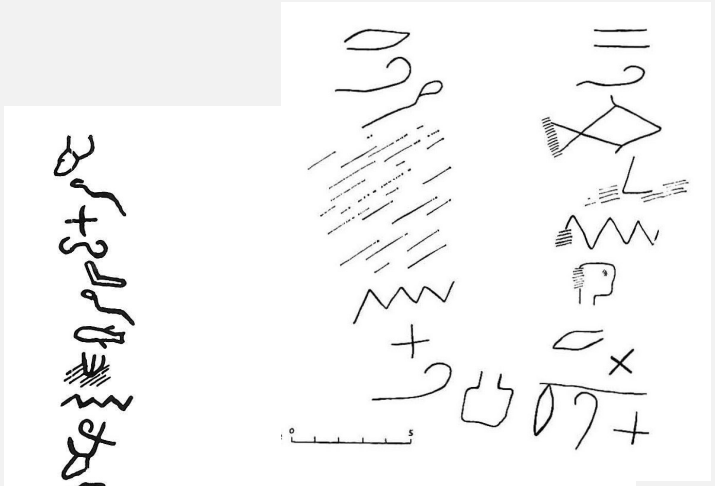
Fig. 16. *Wadi el-Hôl Alphabetic Inscription 2; the Vertical Inscription. a) facsimile of the base of the incised lines; b) facsimile of the outer edges of the incised lines.*

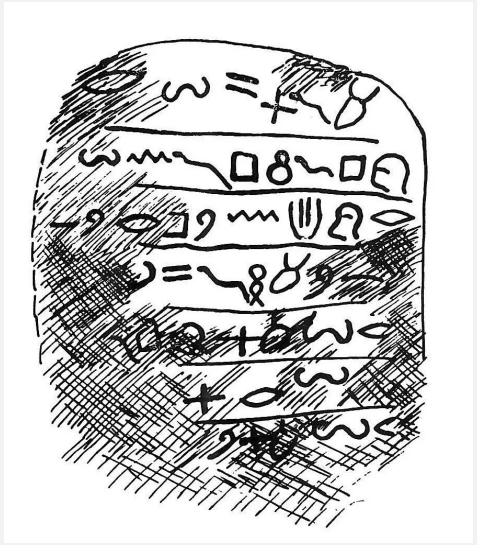
PROTO-SINAITIC
ISSUES: DIRECTIONALITY

- Orientation
  - Right to left
  - Left to right
  - Top to bottom
  - Boustrophedon
  - Random

- Challenges
  - Which default direction?

Top to bottom & right to left

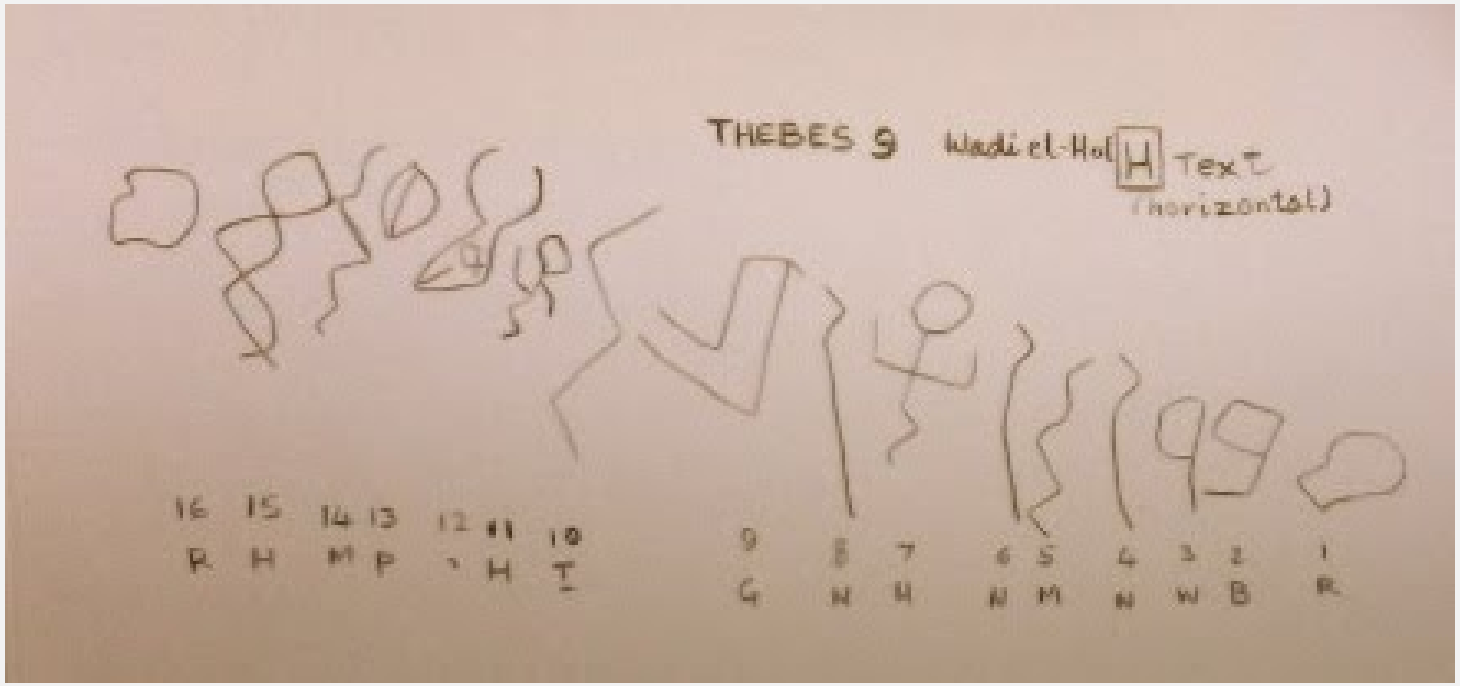




Right to left

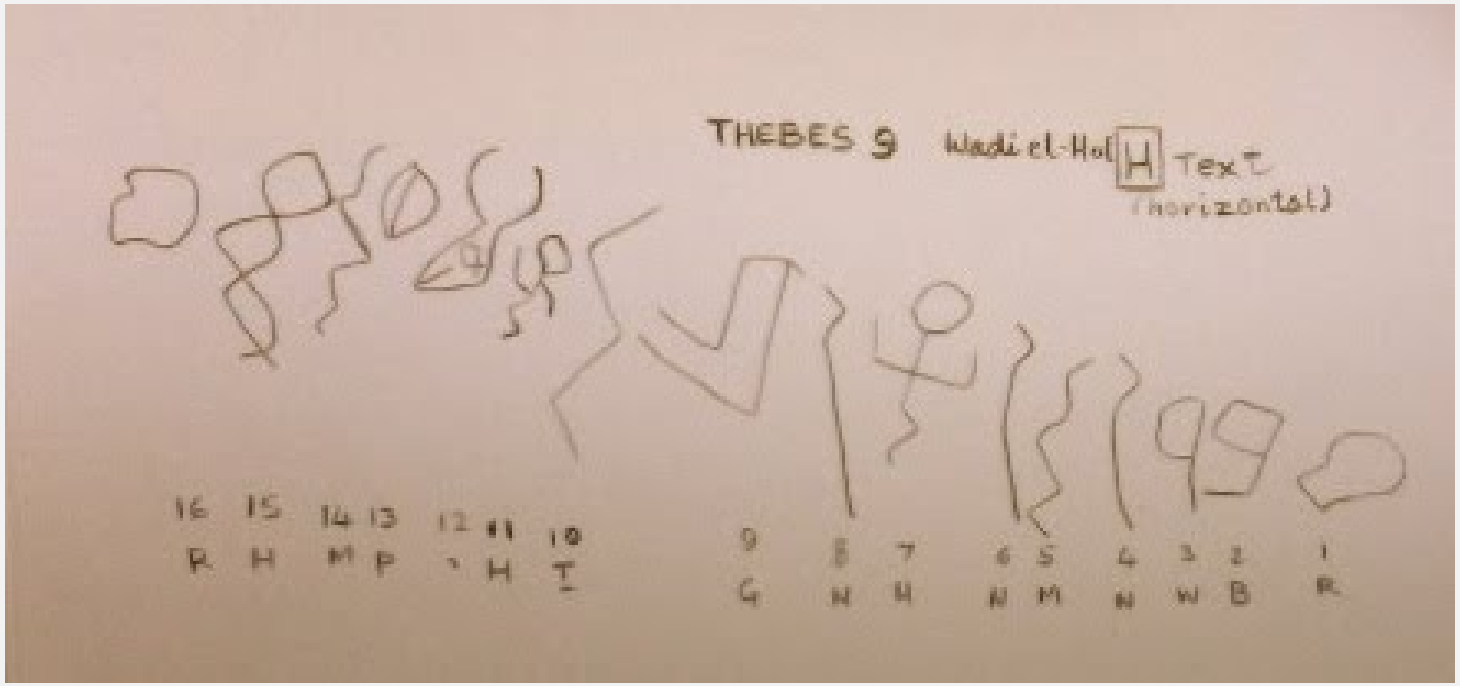
Top to bottom & left to right

PROTO-SINAITIC ISSUES: SEMANTICS



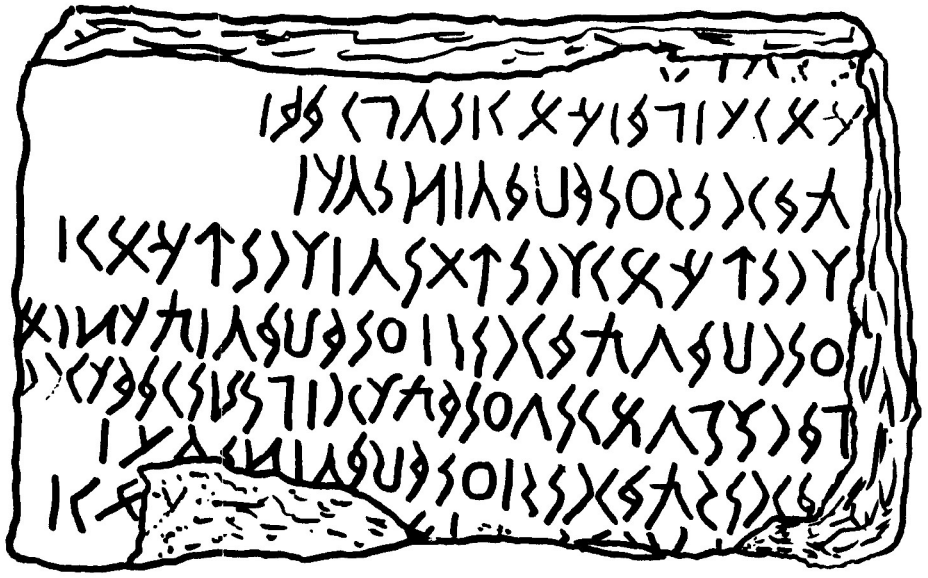
"First-class (R) feast (MShT) of the celebration (H) of `Anat (`NT). 'El ('L) will provide (YGSh) **[H]** plenty (RB) of wine (WN) and victuals (MN) for the celebration (H). We will sacrifice (NGTh) for her (H) an ox (') and (P) a prime (R) fatling (MKh)" (Colless 2019)

PROTO-SINAITIC ISSUES: SEMANTICS



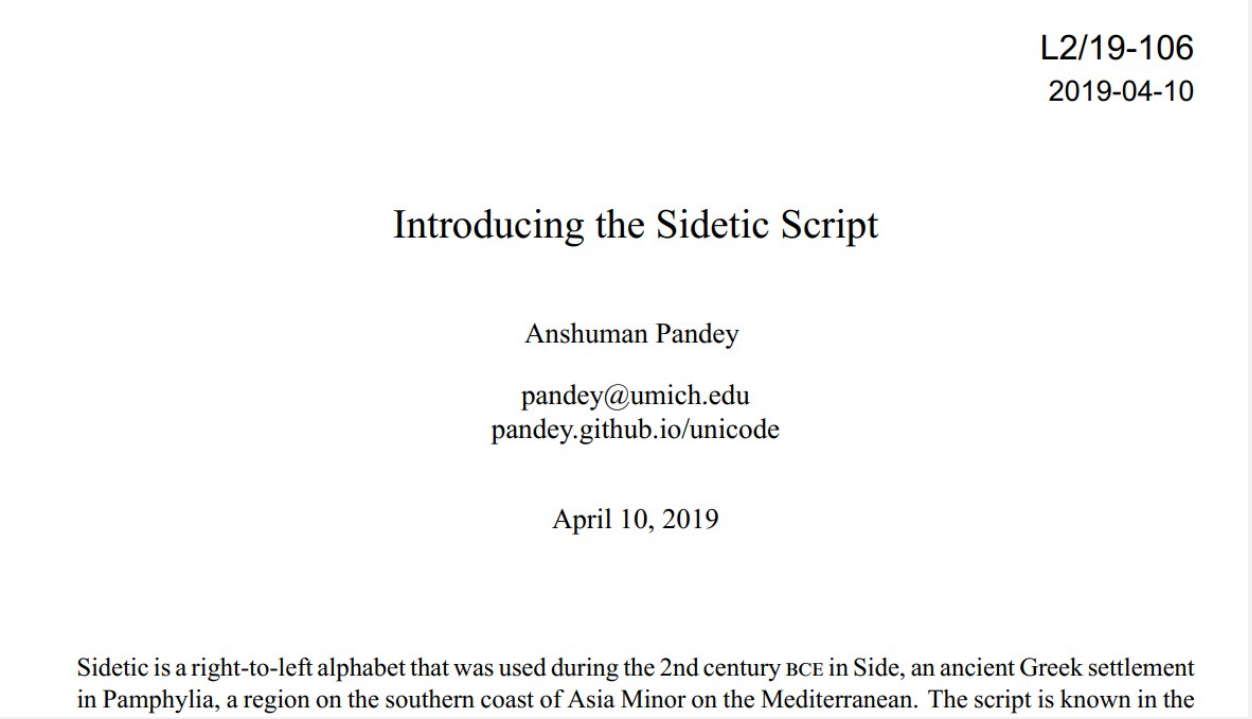
"First-class (R) feast (MShT) of the celebration (H) of `Anat (`NT). 'El ('L) will provide (YGSh) **[H]** plenty (RB) of wine (WN) and victuals (MN) for the celebration (H). We will sacrifice (NGTh) for her (H) an ox (') and (P) a prime (R) fatling (MKh)" (Colless 2019)

SIDETIC



SIDETIC

- Placed on “Not The Roadmap” in 2002
- Allocated to SMP in 2019



L2/19-106
2019-04-10

Introducing the Sidetic Script

Anshuman Pandey

pandey@umich.edu
pandey.github.io/unicode

April 10, 2019

Sidetic is a right-to-left alphabet that was used during the 2nd century BCE in Side, an ancient Greek settlement in Pamphylia, a region on the southern coast of Asia Minor on the Mediterranean. The script is known in the

SIDETIC BACKGROUND

- Anatolian alphabet
- Likely developed from Phoenician
  - Related to, but distinct from Greek
- Related to Carian, Lycian, Lydian

(Greek)	Phrygian[a]	Lydian[a]	Lycian[a]	Carian[b]	Sidetic[c]
Α	Α a	Α a	Α a	Α a	𐤀 a
			𐤀 e	𐤃 d	𐤄 e
Β	Β b	𐤁 b	Β b	Δ l	𐤉 i
			𐤅 β	Ε y	𐤊 o
Γ	Γ g	𐤂 g	𐤂 g	Γ r	𐤋 u
Δ	Δ d	𐤃 d	Δ d	Ι λ	𐤌 w
Ε	Ε e	𐤄 e	Ε i	Θ q	𐤍 j
F	F υ	𐤇 v	F w	Γ b	𐤎 p
Ζ	ſ z		Ι z	Ν m	𐤏 ç
Η			+ h	Ο o	𐤐 m
Θ		Ι i	Χ θ	Ρ t	𐤑 t
Ι	Ι i	𐤃 y	Ι y	𐤒 š	𐤓 d
Κ	Κ k	𐤊 k	κ k	Μ s	Ο θ
			⁎ q	Τ ?	Ι ś
Λ	Λ l	𐤋 l	Λ l	Υ u	Ν s
Μ	Μ m	𐤌 m	ᴍ m	Φ ñ	𐤔 n
Ν	Ν n	𐤍 n	ᴎ n	Χ ḱ	𐤕 l
			Χ m̃	Υ n	𐤖 t̲
Ξ			Ŧ ñ	Μ p	И g
Ο	Ο o	ο o	Ο o	Φ ś	Χ χ
Π	Ρ p		Ρ p	𐤃 i	Λ r
Ϙ		+ q	◇ κ	◻ e	𐤗 a/u
Ρ	Ρ r	𐤒 r	Ρ r	Ψ ý	Π k
Σ	ξ s	Ŧ s	ς s	▽ k	𐤁 b
		𐤔 ś		⩓ δ	𐤙 n
Τ	Τ t	Τ t	Τ t	Ш w	𐤊 z

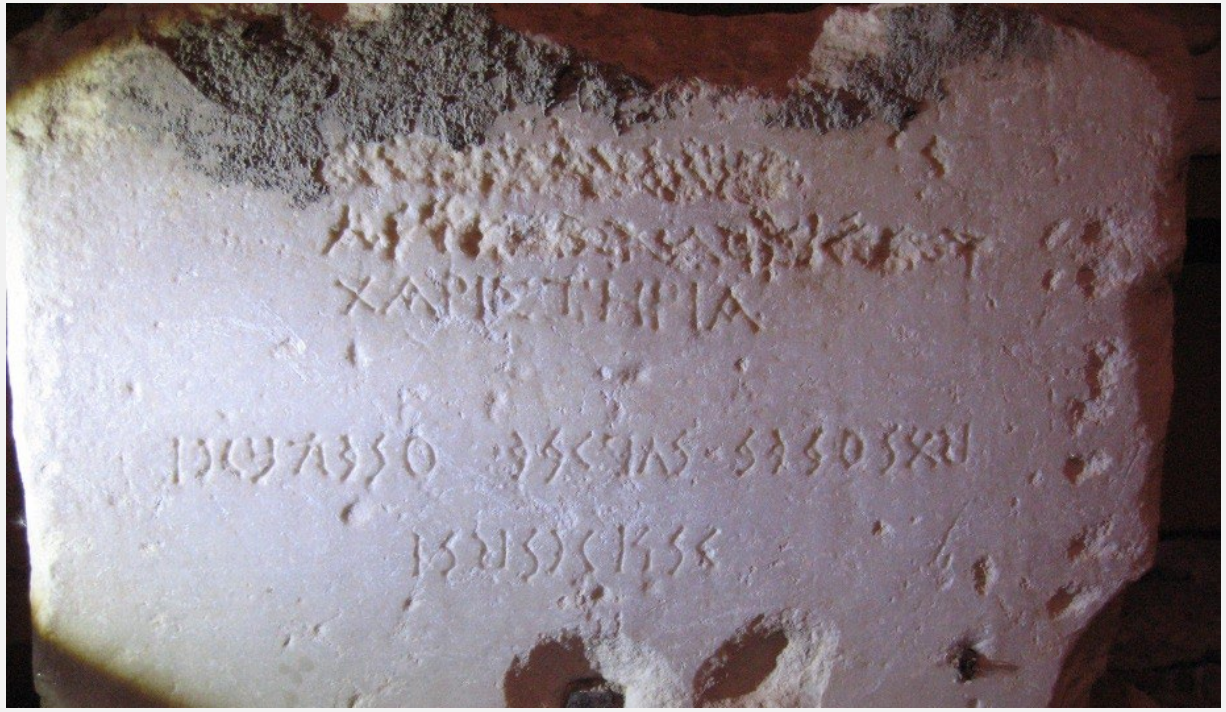
SIDETIC BACKGROUND

- Used concurrently with Greek

XXXXXXXXX

A[PTEM]ΩNAΘ[HNOBI]OY

XAPIΣTHPIA

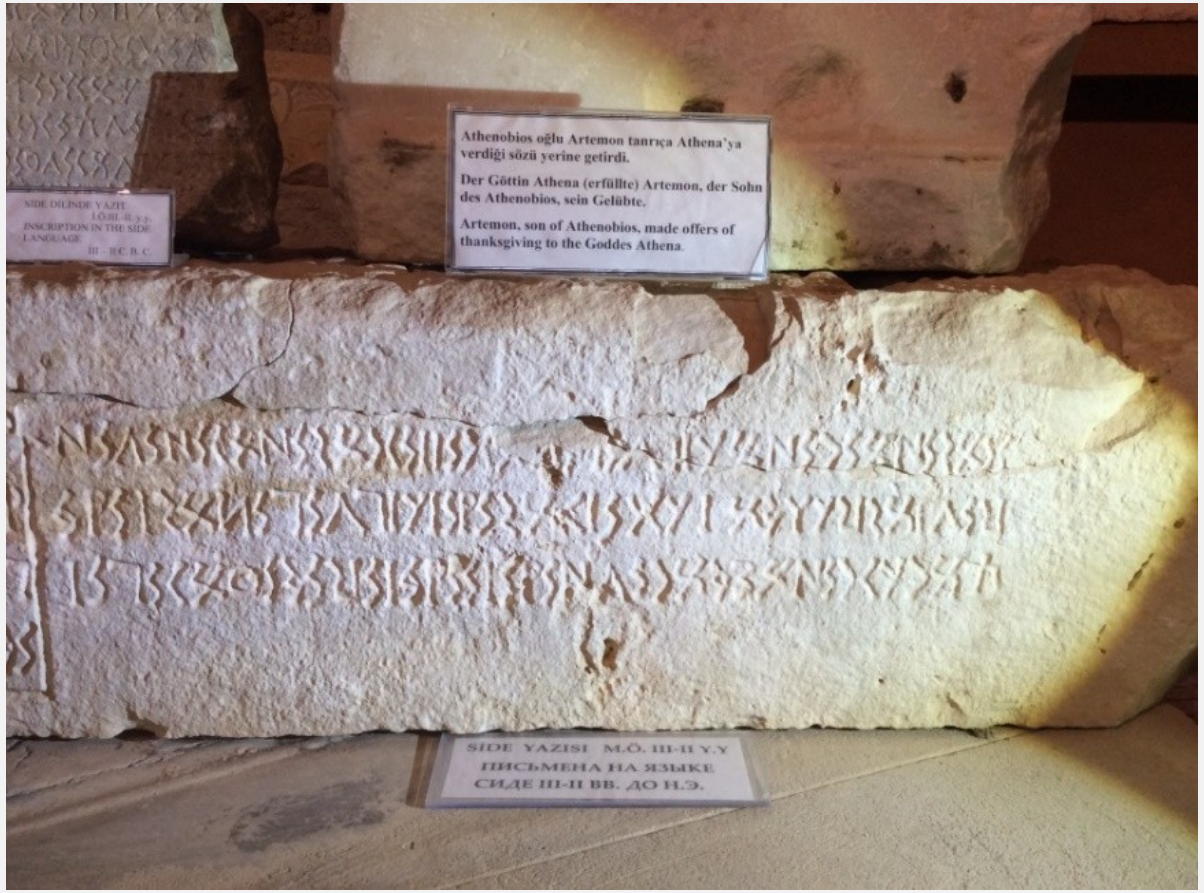


SIDETIC BACKGROUND

- Short texts on coinage

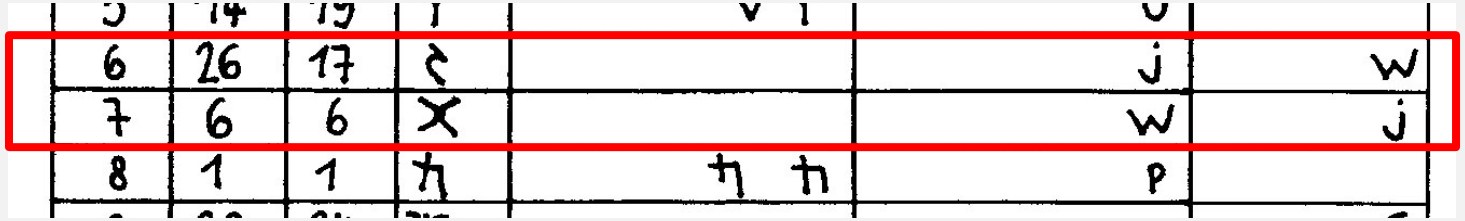
SIDETIC BACKGROUND

- Long inscriptions

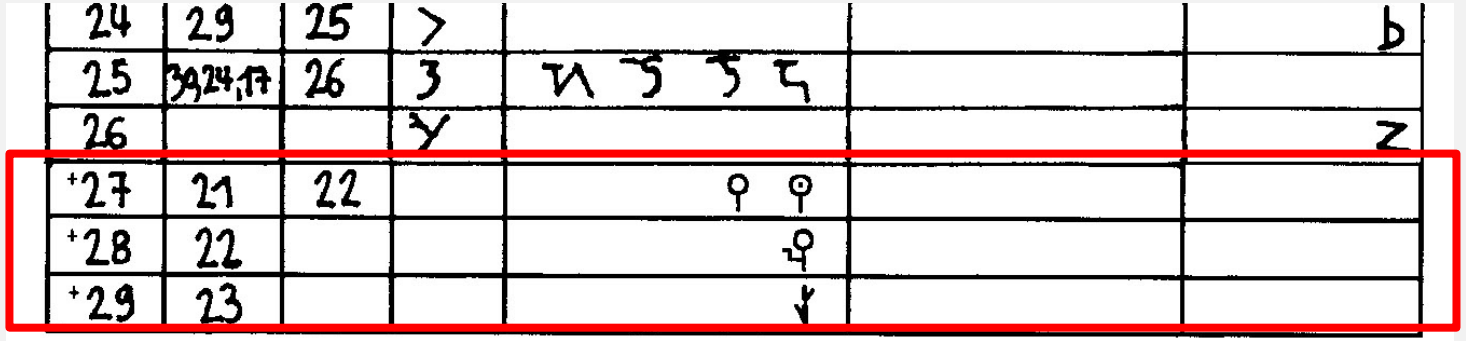


SIDETIC ISSUES

- Scholarly disagreement on letter values

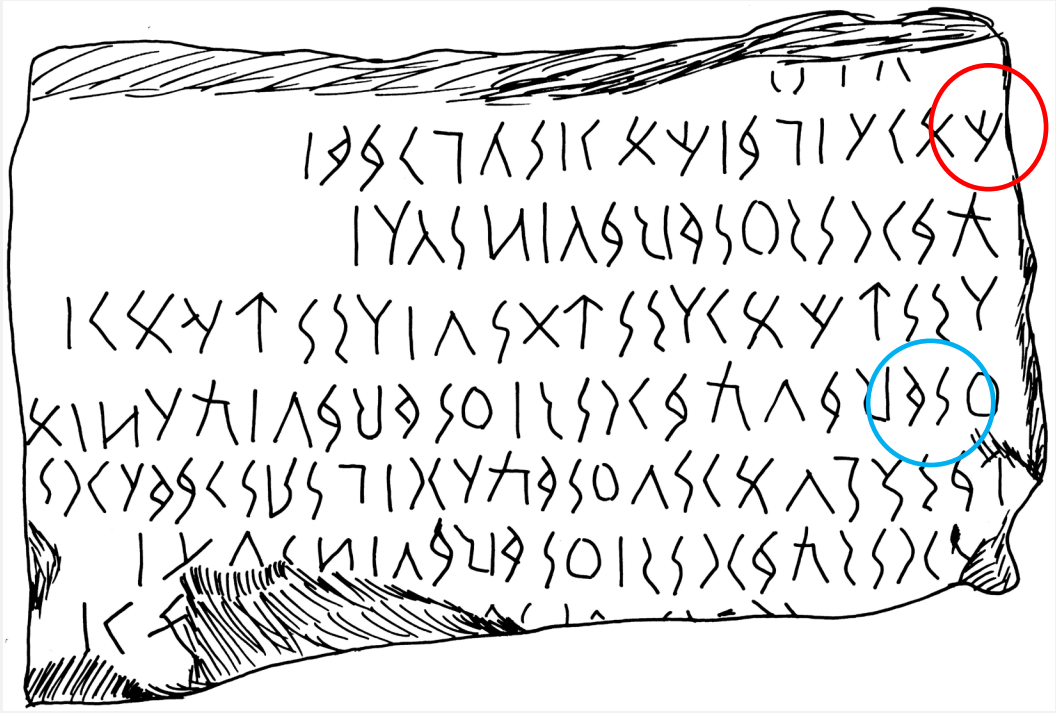


- Unknown values of common letters:



SIDETIC ISSUES

- Scholarly disagreement on character identity



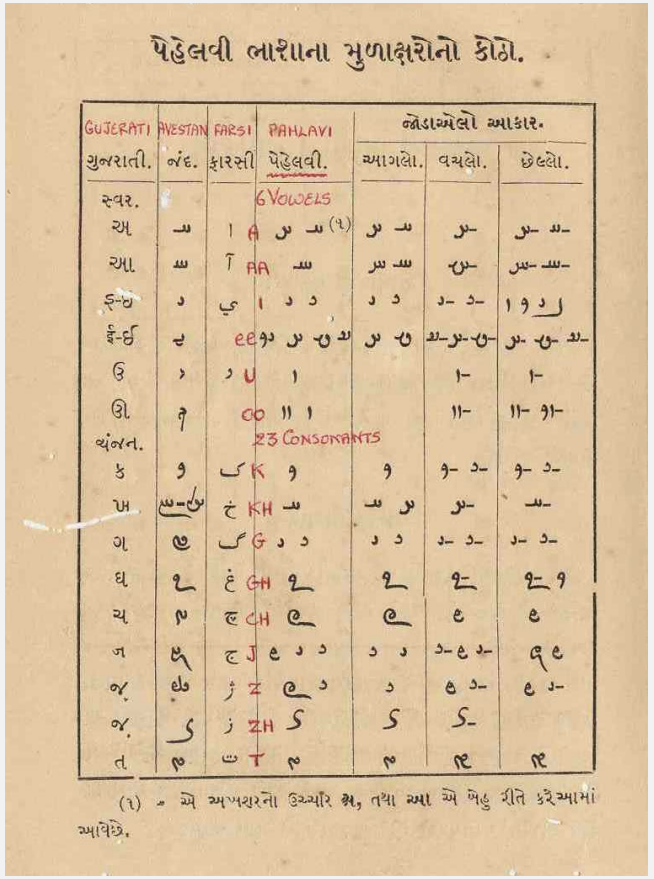
Zinko & Zinko (2019)

Nolle (2001)

BOOK PAHLAVI

BOOK PAHLAVI BACKGROUND

- Aramaic alphabet used in Iran
- Highly cursive, highly ambiguous
- Developed for writing Zoroastrian texts
- Currently used by Zoroastrian and Parsis
  - Concurrently with Gujarati
  - Actively studied and printed in India



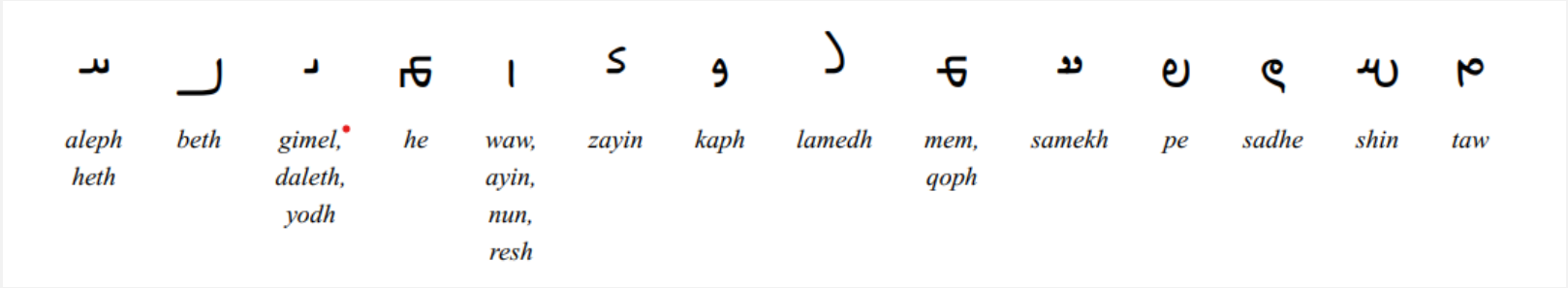
પેહેલવી ભાષાના મુળાક્ષરોનો કોઠો.

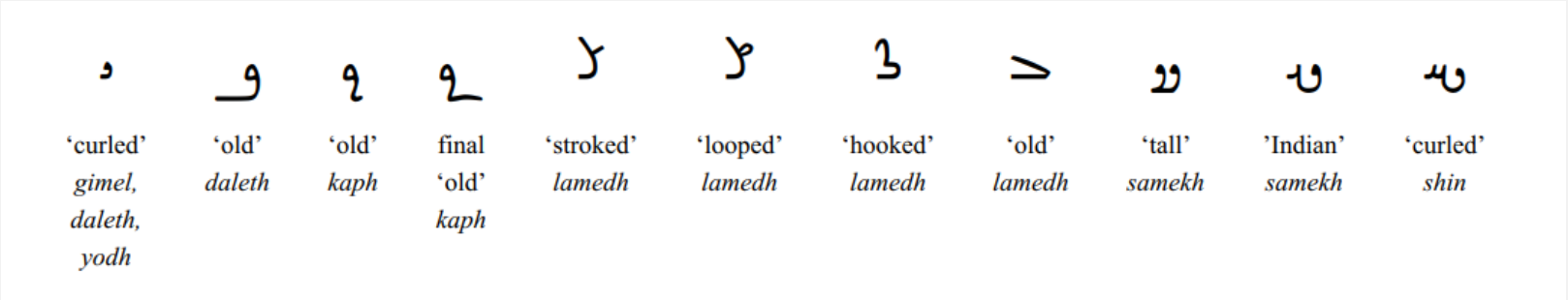
GUJERATI ગુજરાતી.	AVESTAN જંદ.	FARSI ફારસી	PAHLAVI પેહેલવી.	જોડાએલો આકાર. આગલો.	વચલો.	છેલ્લો.
સ્વર.			6 VOWELS			
અ			A (૧)			
આ			AA			
ઇ-ઇ						
ઈ-ઈ			EE			
ઉ			U			
ઊ			OO			
વ્યંજન.			23 CONSONANTS			
ક			K			
ખ			KH			
ગ			G			
ઘ			GH			
ચ			CH			
જ			J			
જ઼			Z			
જ઼			ZH			
ત			T			

(૧) * એ અક્ષરનો ઉચ્ચાર અ, તથા આ એ બેહુ રીતે કરવામાં આવેછે.

BOOK PAHLAVI BACKGROUND

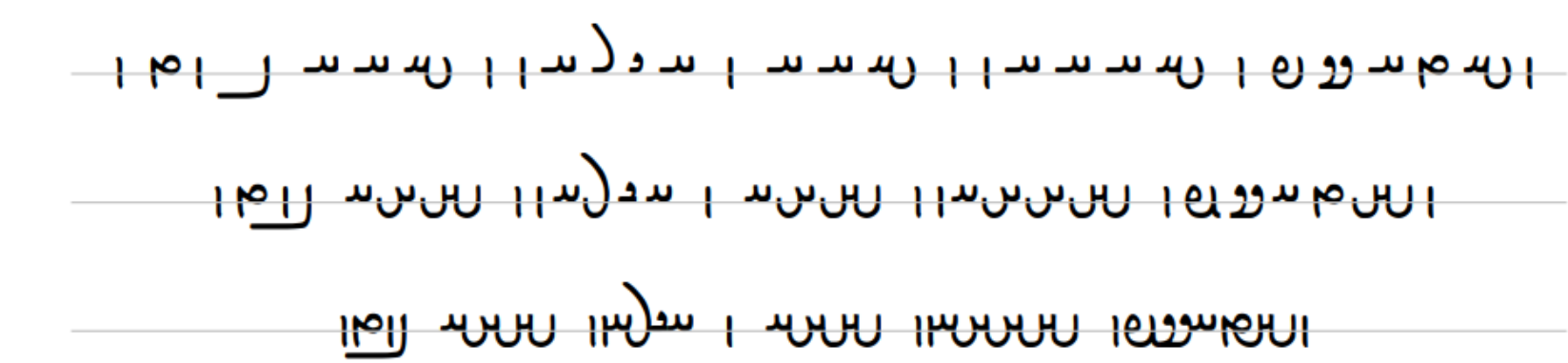
- 25 letters
- Multiple values per letter





BOOK PAHLAVI BACKGROUND

- Joining behavior



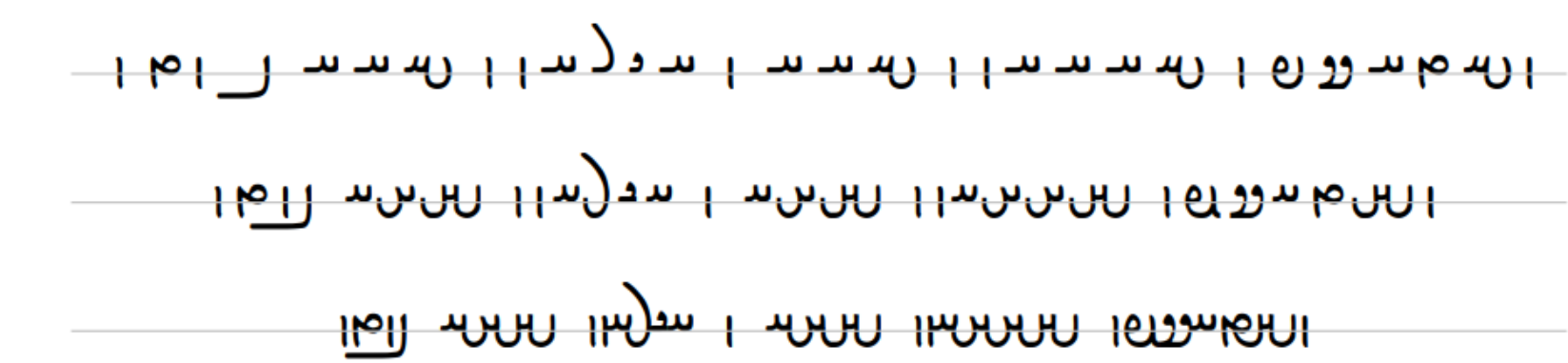
<wšt›sp′ š›h›n′ š›h w ›yl›n′ š›h bwt′>

wištāsp šāhān šāh ud ērān šāh būd

Wištāsp was the king of kings and the king of the Iranians.

BOOK PAHLAVI BACKGROUND

- Cursive joining behavior



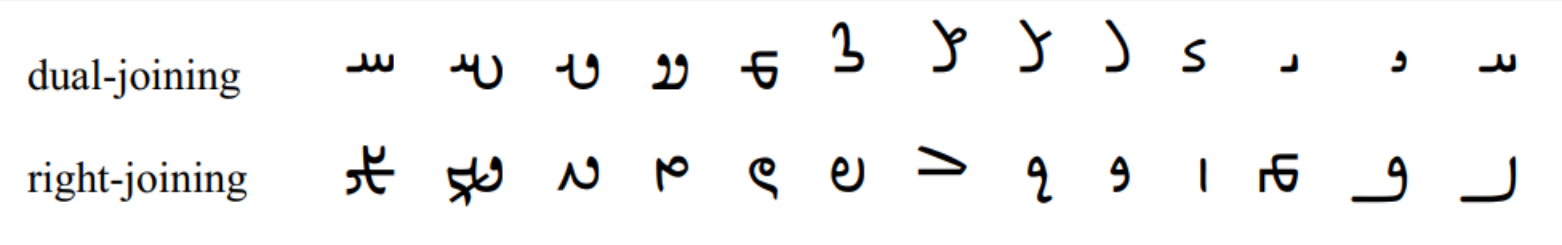
<wšt›sp′ š›h›n′ š›h w ›yl›n′ š›h bwt′>

wištāsp šāhān šāh ud ērān šāh būd

Wištāsp was the king of kings and the king of the Iranians.

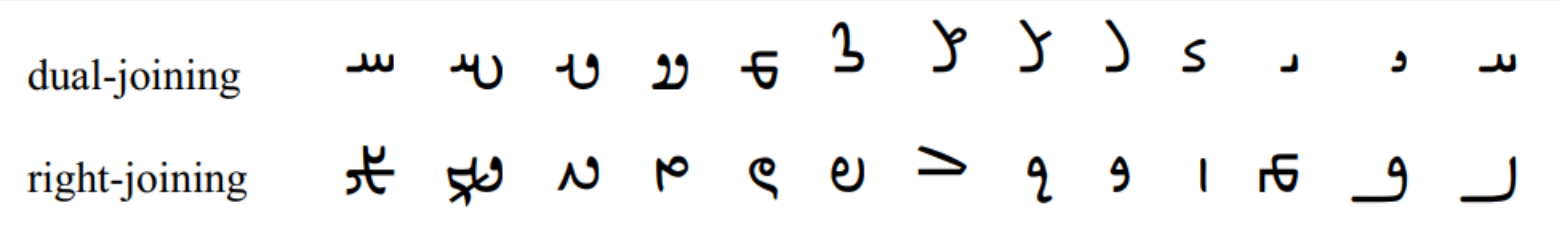
BOOK PAHLAVI BACKGROUND

- Joining categories



BOOK PAHLAVI BACKGROUND

- Joining categories



- Complex letters

BOOK PAHLAVI BACKGROUND

- Shaping behavior

		X_n	X_f	X_m	X_i
Truncation of strokes, or no change	*pe*	𐭯	𐭯 , 𐭯 , 𐭯	—	—
	sadhe	𐭰	𐭰 , 𐭰	—	—
	shin	𐭩	𐭩	𐭩 , 𐭩 , 𐭩	𐭩 , 𐭩 , 𐭩
Descent of terminal or no change	*aleph-heth*	𐭠	𐭠	𐭠 , 𐭠 , 𐭠	𐭠 , 𐭠 , 𐭠
	gimel-daleth-yodh	𐭣	𐭣 , 𐭣	𐭣 , 𐭣 , 𐭣 , 𐭣 , 𐭣	𐭣 , 𐭣 , 𐭣 , 𐭣 , 𐭣

BOOK PAHLAVI ISSUES

- Multiple representations

Some combinations have alternate, multiple valid representations, which cannot be predicted

aleph + gimel-daleth-yodh	𐭠𐭣 , 𐭠𐭤 , 𐭠𐭥 , 𐭠𐭦 , 𐭣 (medial)
aleph + shin	𐭠𐭧 , 𐭠𐭨
aleph + pe	𐭠𐭩 , 𐭠𐭪 , 𐭠𐭫
gimel-daleth-yodh + aleph-heth	𐭤 , 𐭬
gimel-daleth-yodh + kaph	𐭤 , 𐭤 , 𐭤
gimel-daleth-yodh + pe	𐭤 , 𐭤 , 𐭤
zayin + aleph	𐭥 , 𐭥
zayin + kaph	𐭥 , 𐭥

BOOK PAHLAVI ISSUES

- Ambiguity resulting from joining

𐭭	*aleph-heth*	𐭣 + 𐭣	*daleth-gimel-yodh* + *daleth-gimel-yodh*
𐭦	*he*	𐭩 + 𐭮	*mem-qoph* + height-adjusted *waw-nun-ayin-resh*
𐭬	*samekh*	𐭤 + 𐭤	'curled' *daleth-gimel-yodh* + 'curled' *daleth-gimel-yodh*
𐭬	'tall' *samekh*	𐭮 + 𐭮	descending 'curled' *daleth-gimel-yodh* + descending 'curled' *daleth-gimel-yodh*
𐭪	'Indian' *samekh*	𐭣 + 𐭪	descending 'curled' *daleth-gimel-yodh* + *daleth-gimel-yodh*
𐭩	'curled' *shin*	𐭭 + 𐭪	descending 'curled' *daleth-gimel-yodh* + *aleph-heth*

Joined sequences of certain letters resemble nominal forms of other letters

BOOK PAHLAVI ISSUES

- Ambiguity resulting from joining
- Poses issues for interpretation

𐭡𐭩	𐭡 + 𐭩	*samekh* + *pe*
	𐭡 + 𐭥 + 𐭥	descending ‘curled’ *daleth-gimel-yodh* + descending ‘curled’ *daleth-gimel-yodh* + *pe*
	𐭮 + 𐭩	*samekh* + *sadhe*
	𐭮 + 𐭣 + 𐭣	*daleth-gimel-yodh* + *daleth-gimel-yodh* + *sadhe*

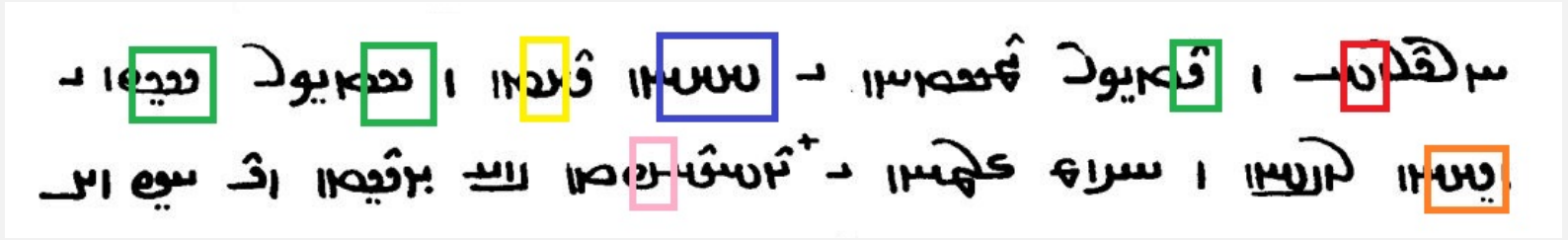
𐭠𐭭	𐭣 + 𐭭	*aleph-heth* + *daleth-gimel-yodh*
	𐭭 + 𐭣	*daleth-gimel-yodh* + *aleph*
𐭠𐭯	𐭡 + 𐭯	*aleph* + *pe*
	𐭮 + 𐭯	*aleph* + *sadhe*
𐭩𐭩	𐭩 + 𐭣	‘curled’ *daleth-gimel-yodh* + *samekth*
	𐭣 + 𐭩	*samekh* + medial *daleth-gimel-yodh*
𐭡𐭣	𐭡 + 𐭣	*daleth-gimel-yodh* + *pe*
	𐭮 + 𐭣	*daleth-gimel-yodh* + *sadhe*
𐭱𐭤	𐭠 + 𐭤	*shin* + *waw-nun-ayin-rest*
	𐭠 + 𐭭 + 𐭣	*daleth-gimel-yodh* + *aleph* + *waw-nun-ayin-rest*

BOOK PAHLAVI ISSUES

- Variations in print

Jamasp Asana (1913)

Nyberg (1964)



Lead to difficulties in determining conventional shapes

BOOK PAHLAVI PROGRESS

L2/20-246

Teeth and bellies: a proposed model for encoding Book Pahlavi

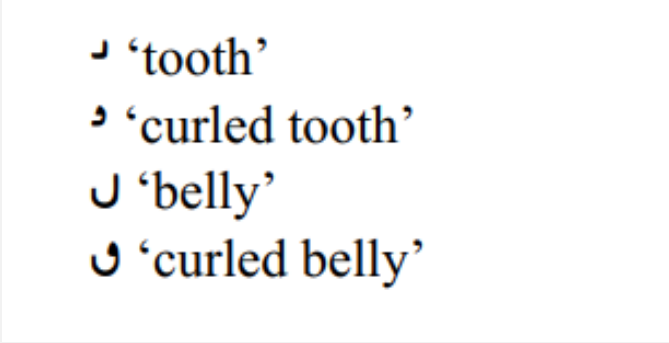
Roozbeh Pournader (WhatsApp)
September 7, 2020

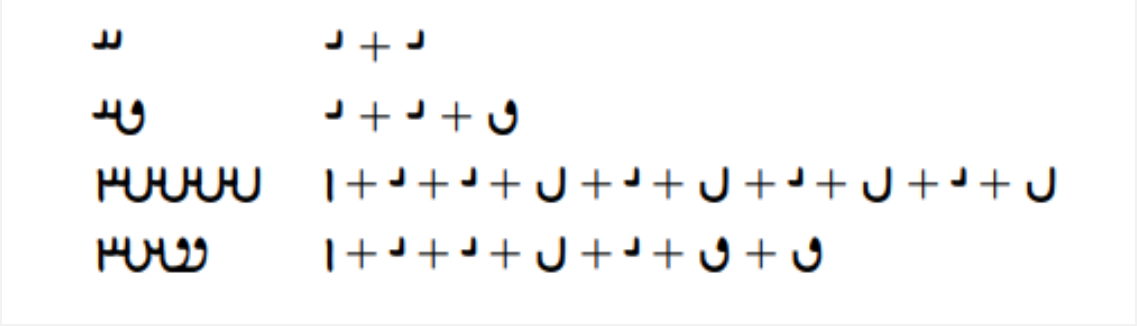
Background

In Everson 2002, a proposal was made to encode a unified Avestan and Pahlavi script in the Unicode Standard. The proposal went through several iterations, eventually leading to a separate encoding of Avestan as proposed by Everson and Pournader 2007a, in which Pahlavi was considered non-unifiable with Avestan due to its cursive joining property. The non-cursive Inscriptional Pahlavi (Everson and Pournader 2007b) and the cursive Psalter Pahlavi (Everson and Pournader 2011) were later encoded too. But Book Pahlavi, despite several attempts (see the Book Pahlavi Topical Document list at https://unicode.org/L2/topical/bookpahlavi/), remains unencoded.

BOOK PAHLAVI PROGRESS

- Example of possible model
- ‘Graphetic’ – Represent text using graphical elements

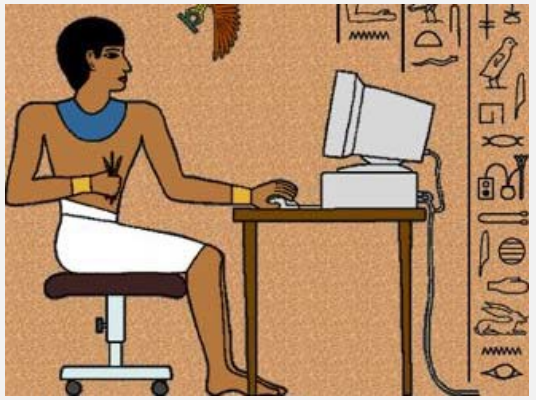




THANK YOU

Email: dwanders@berkeley.edu
Web page: linguistics.berkeley.edu/sei

Email: pandey@umich.edu
Web page: https://pandey.github.io/unicode/



SEI is supported by NEH grant PR-268710-20 and donations;
Classic Maya hieroglyph signs supported through NEH grant HAA-268887-20 and Unicode Adopt-a-Character grant

NATIONAL ENDOWMENT FOR THE Humanities