

Undeciphered Scripts in the Unicode Age

Challenges for encoding early writing systems of the Near East

Internationalization and Unicode Conference #42 • September 2018

Dr. Anshuman Pandey

Dr. Deborah (Debbie) Anderson

Script Encoding Initiative, UC Berkeley

WELL-KNOWN UNDECIPHERED SCRIPTS (OUTSIDE THE NEAR EAST)

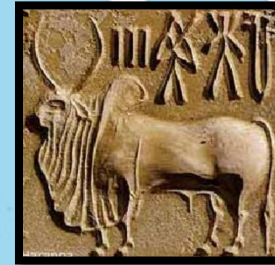


**MESO-AMERICAN
SCRIPTS**

RONGO RONGO



INDUS VALLEY



UNDECIPHERED (OR PARTLY DECIPHERED)
N.E. SCRIPTS NOT IN UNICODE

1. Proto-Cuneiform
2. Proto-Elamite
3. Linear Elamite
4. Cretan Hieroglyphs

5. Byblos

6. Proto-Sinaitic

7. Cretan Minoan

**UNDECIPHERED (OR PARTLY DECIPHERED)
N.E. SCRIPTS NOT IN UNICODE**

Cretan Hieroglyphs

Cypro-Minoan

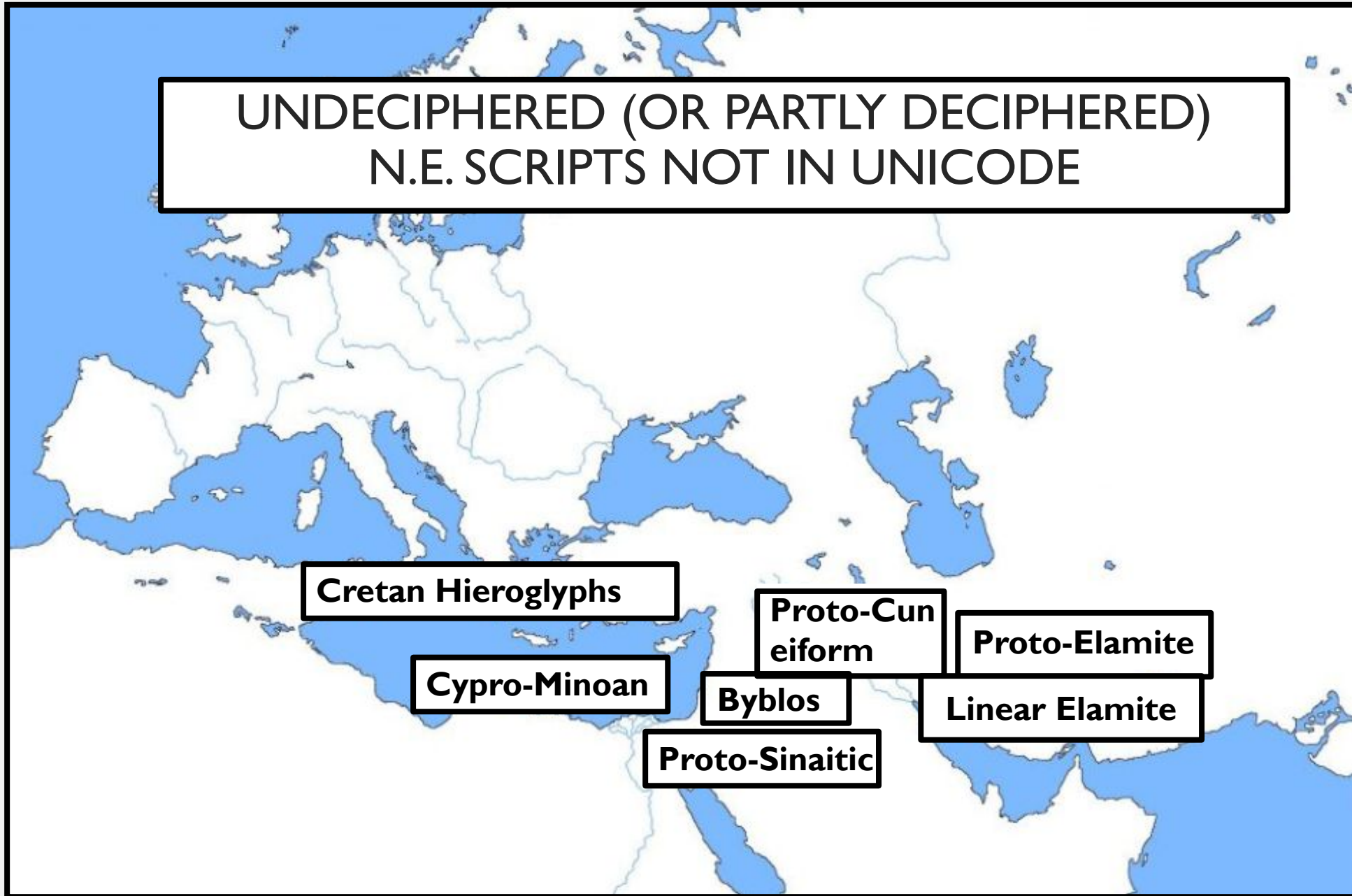
Proto-Cuneiform

Byblos

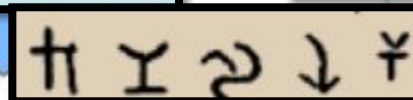
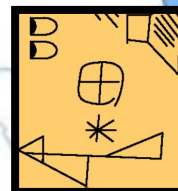
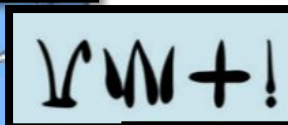
Proto-Sinaitic

Proto-Elamite

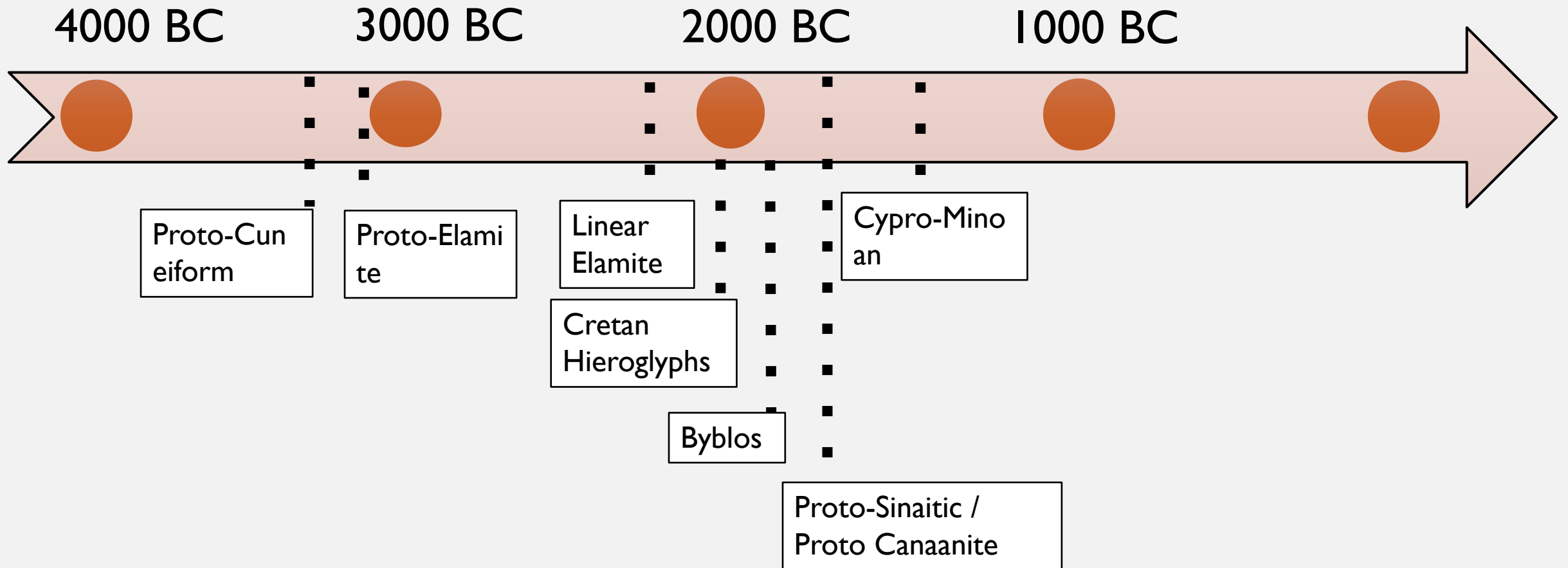
Linear Elamite



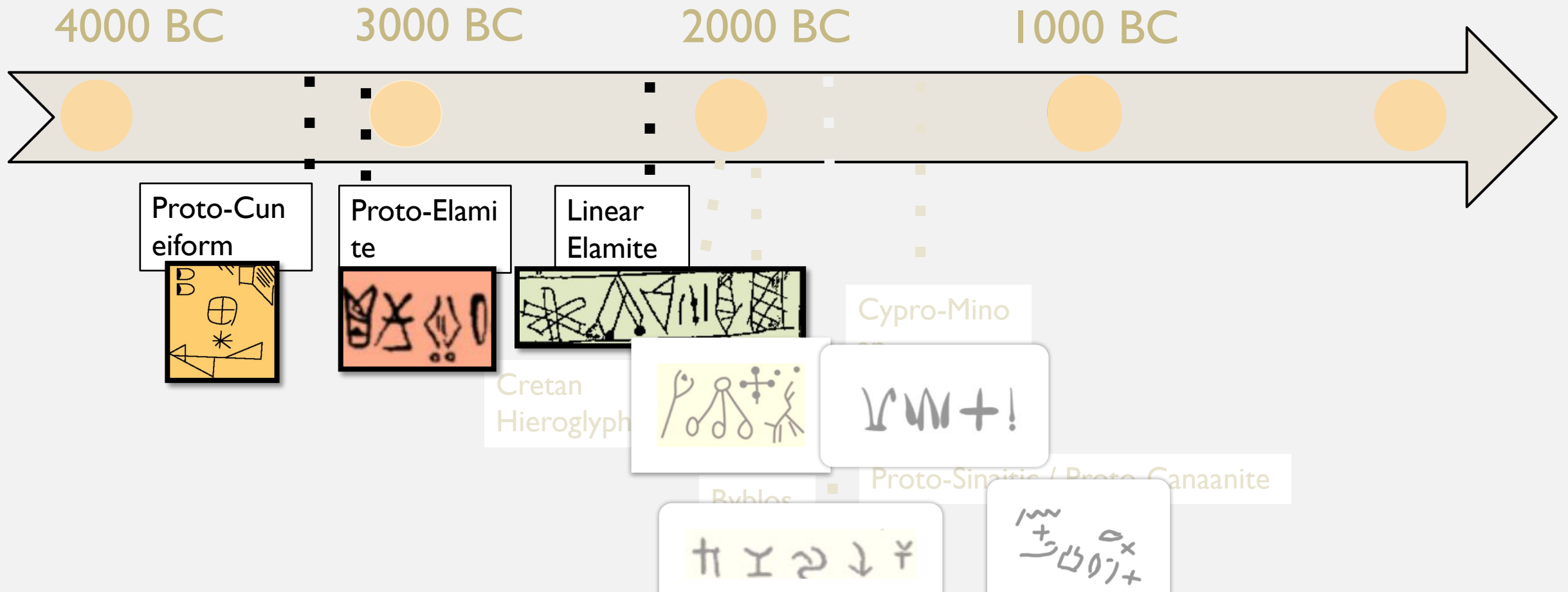
UNDECIPHERED (OR PARTLY DECIPHERED)
N.E. SCRIPTS NOT IN UNICODE



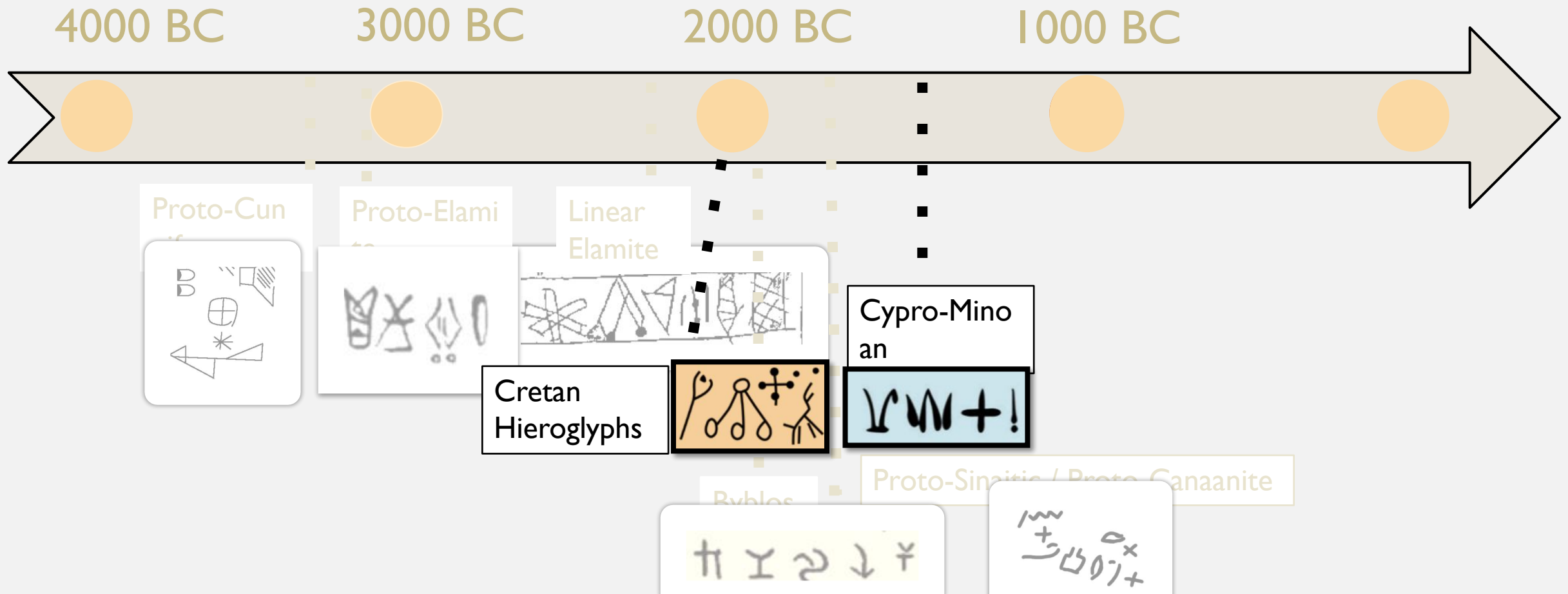
TIMELINE OF UNDECIPHERED/PARTLY DECIPHERED SCRIPTS OF THE NEAR EAST (3200-1500 BC)



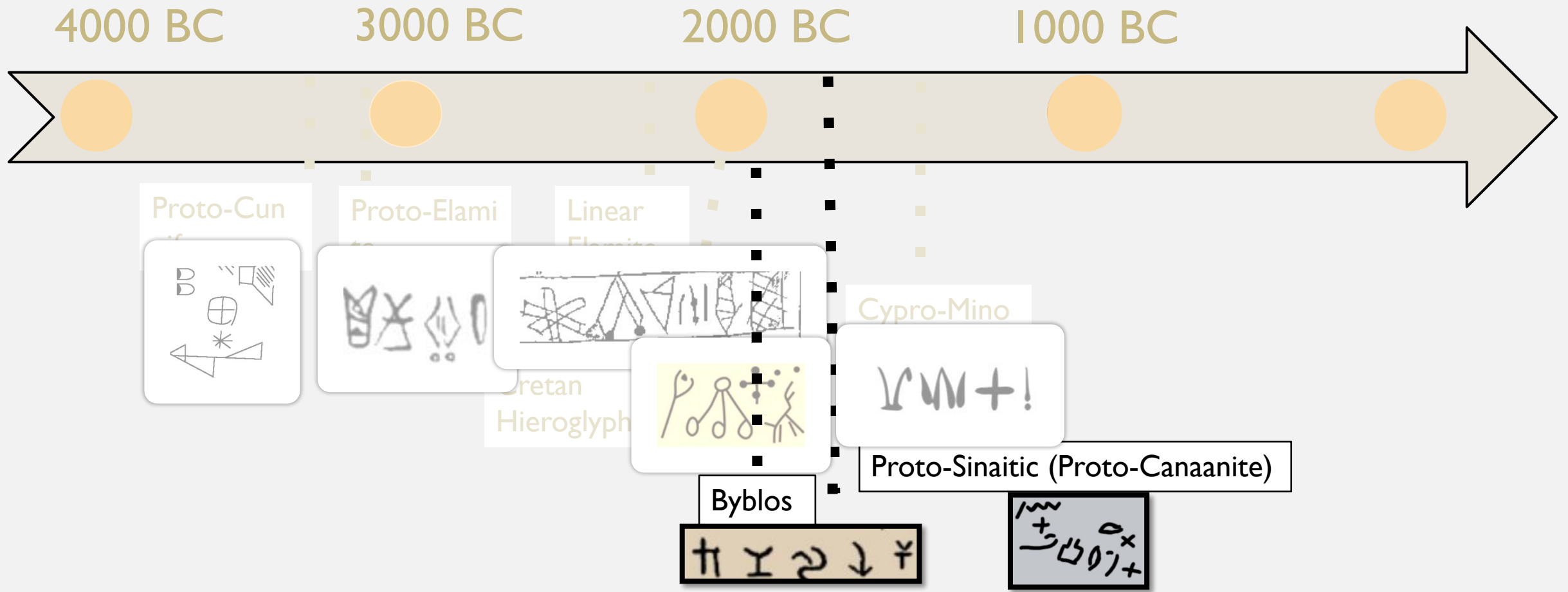
TIMELINE OF UNDECIPHERED/PARTLY DECIPHERED SCRIPTS OF THE NEAR EAST



TIMELINE OF UNDECIPHERED/PARTLY DECIPHERED SCRIPTS OF THE NEAR EAST



TIMELINE OF UNDECIPHERED/PARTLY DECIPHERED SCRIPTS OF THE NEAR EAST

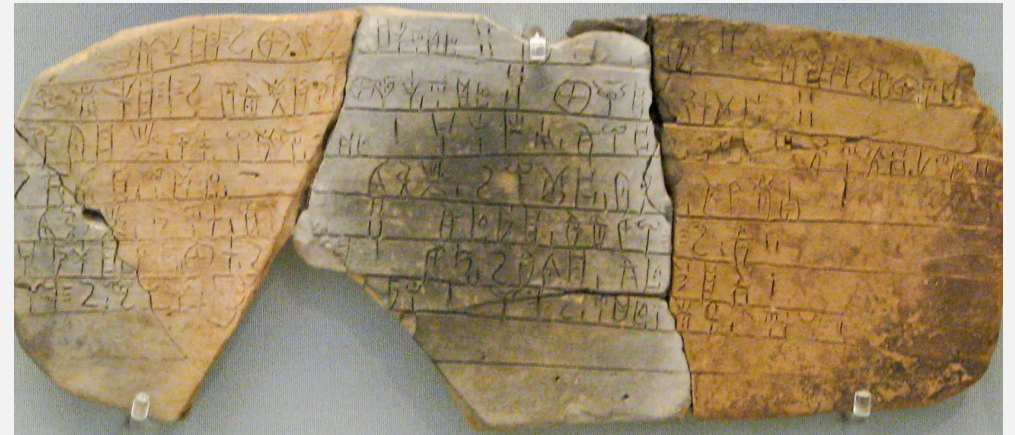


KEYS TO DECIPHERMENT



- Corpus large enough? Can the number of signs be determined?

EXAMPLE: LINEAR B
Over 5000 tablets with
inscriptions; 85-90 distinct signs



KEYS TO DECIPHERMENT



- Known relationship(s) to other scripts?

EXAMPLE: LINEAR B

Shown to have some relationship to the Cypriot Syllabary, which was deciphered earlier



Cypriot
Syllabary

KEYS TO DECIPHERMENT



- Underlying language (or language family) known?

EXAMPLE: LINEAR B

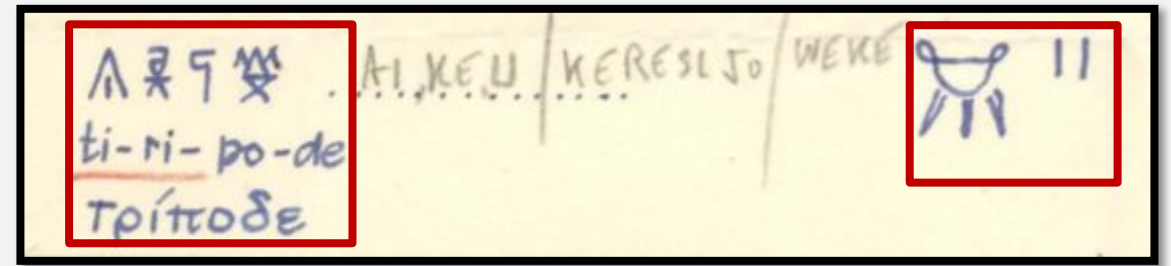
Nouns varied only by I sign, suggesting the language was inflected (and led to identifying it as an Indo-European lang.)

	Noun 1	Noun 2
Case I:	𐀀𐀃𐀆𐀇𐀏	𐀀𐀃𐀆𐀇𐀏
Case II:	𐀀𐀃𐀆𐀇𐀐	𐀀𐀃𐀆𐀇𐀐

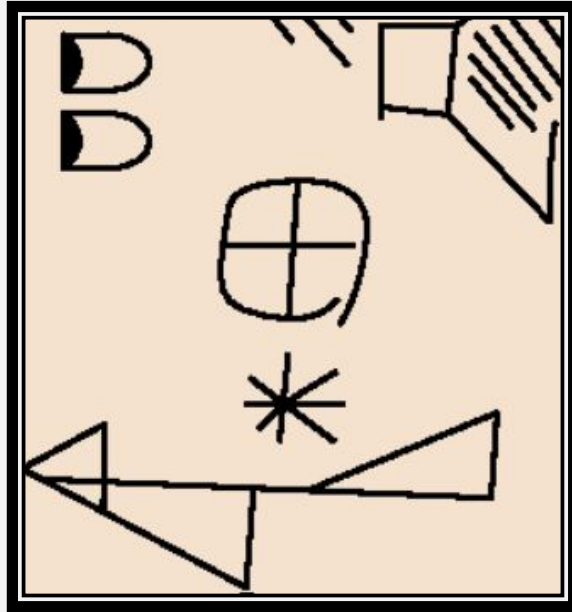
KEYS TO DECIPHERMENT



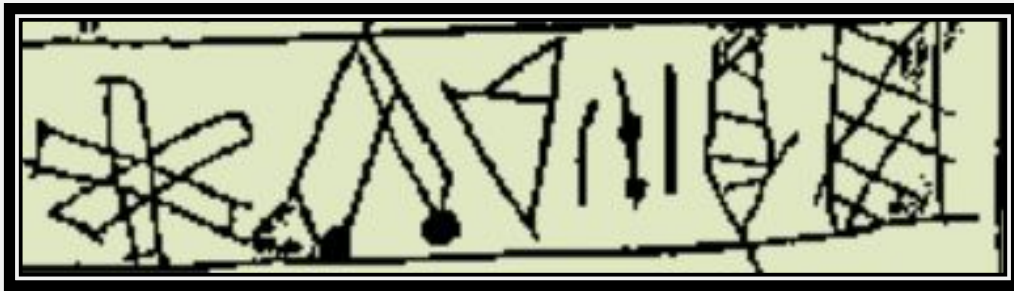
- Bilingual available? (Or are there other ways to be able to confirm a reading?)



Letter from Michael Ventris to
E. Bennett, May 1953



Proto-Cuneiform



Linear Elamite

COMMENTS ON N.E. UNDECIPHERED SCRIPTS

1. Scripts in this talk reflect a spectrum, ranging from partially deciphered to completely undeciphered.
2. May attract wide-ranging, unusual theories

ENCODING PROCESS

PROPOSAL IS WRITTEN

REVIEWED BY UNICODE
SCRIPT AD HOC

REVIEWED BY UNICODE
TECHNICAL COMMITTEE
AND APPROVED

REVIEWED BY ISO SC2
AND WORKING GROUP 2
AND PUT ON ISO BALLOT

ENCODING PROCESS

PROPOSAL IS WRITTEN

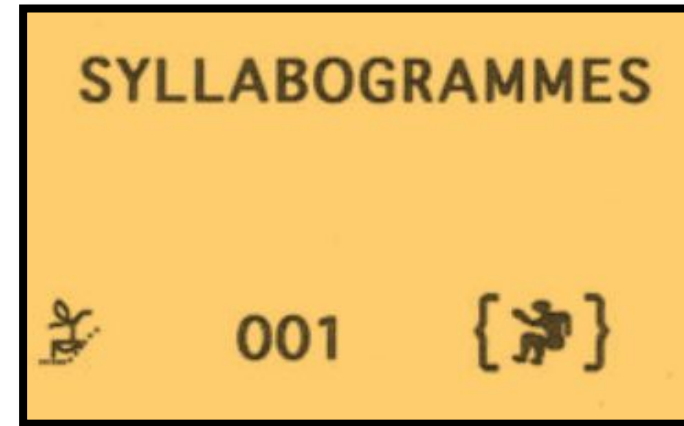
REVIEWED BY UNICODE
SCRIPT AD HOC

REVIEWED BY UNICODE
TECHNICAL COMMITTEE
AND APPROVED

REVIEWED BY ISO SC2
AND WORKING GROUP 2
AND PUT ON ISO BALLOT

FACTORS TO
CONSIDER
RE: UNDECIPHERED
SCRIPTS
(FROM THE SCRIPT
AD HOC)

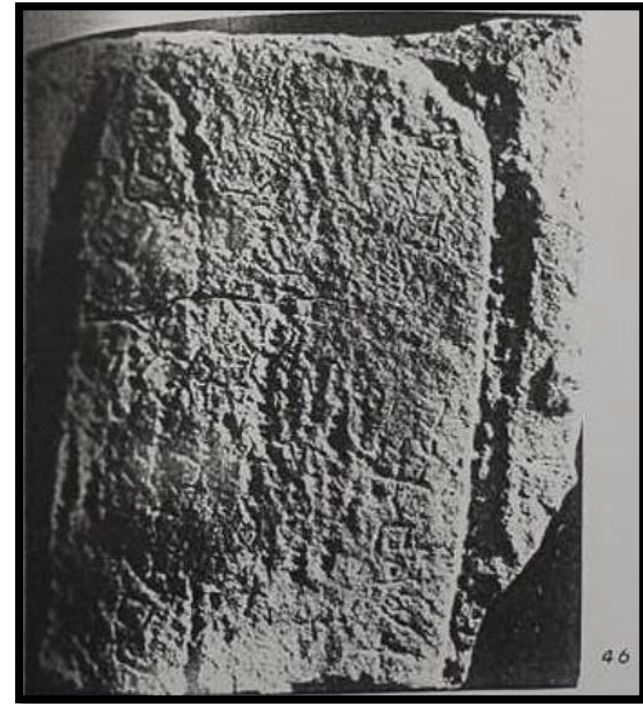
- Does the script have a stable list of the characters that scholars refer to?



From CHIC = Corpus
Hieroglyphicarum Inscriptionum
Cretae (Godart and Olivier)

FACTORS TO
CONSIDER
RE: UNDECIPHERED
SCRIPTS
(FROM THE SCRIPT
AD HOC)

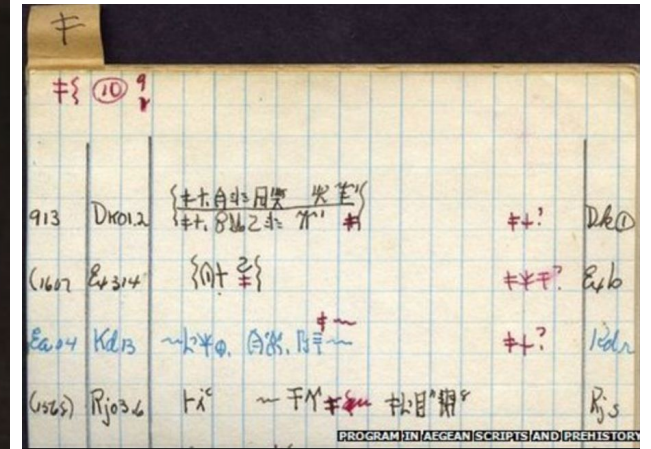
- How much material in the script exists today?



Proto-Sinaitic stele

FACTORS TO
CONSIDER
RE: UNDECIPHERED
SCRIPTS
(FROM THE SCRIPT
AD HOC)

- What is the state of decipherment?
- Is the underlying language known?



Alice Kober's files for Linear B

FACTORS TO
CONSIDER
RE: UNDECIPHERED
SCRIPTS
(FROM THE SCRIPT
AD HOC)

- Can a strong case be made to encode the script? Is text in script being interchanged?

The Ras Shamra Tablet Inventory
Tablets and Texts
an OCHRE Database Project



The Oriental Institute | The University of Chicago

CYPRO-MINOAN



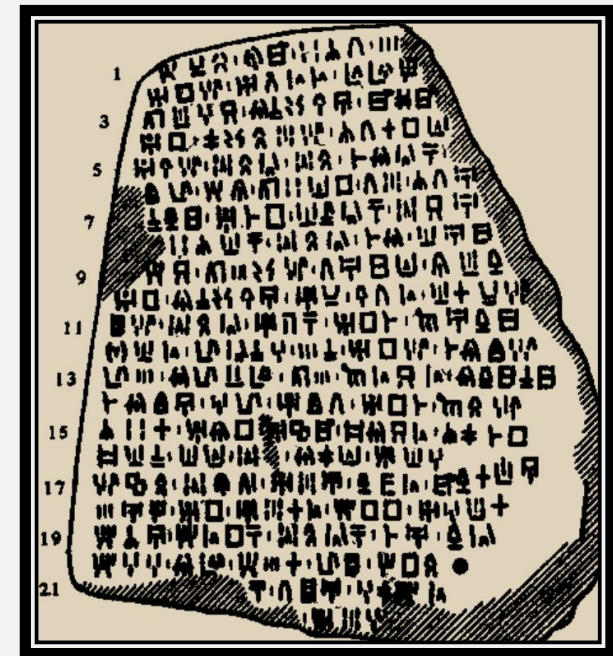
recto



recto

CYPRO-MINOAN

- Found on ca. 250 objects
- Current proposal is stalled:
 - Some characters in proposal are not regarded today by scholars as valid
 - Apparent duplicates in repertoire (from Enkomi tablet)

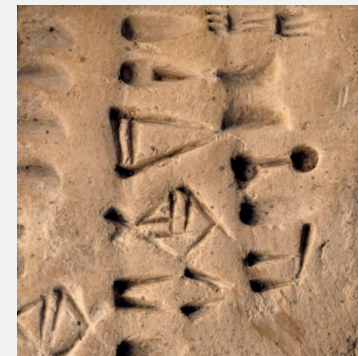


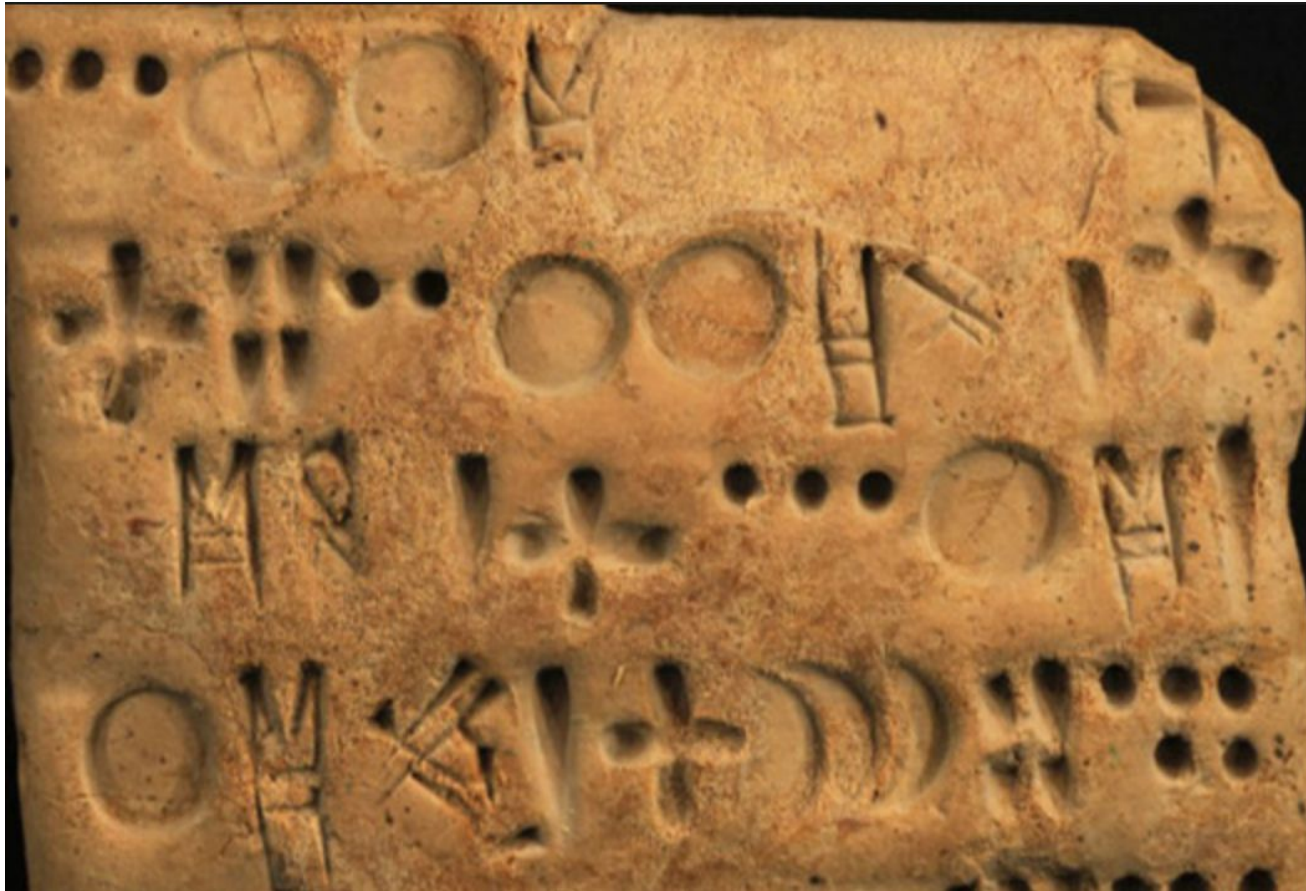
Enkomi tablet ENKO Atab 001

PROTO-ELAMITE



- Found on over 1600 tablets, most from Susa, in SW Iran
- A short-lived writing system (ca. 3100-2900 BC)





PROTO-ELAMITE

- Closed set of characters (300-400)
- Similar numerical system to Proto-Cuneiform
- New texts will be available from Tehran soon

PROTO-SINAITIC



- Also known as 'Early Alphabetic'
 - 'Proto' is misnomer
- ~18th-17th c. BCE
- Inspired by Egyptian Hieroglyphs
- First alphabetic script
 - Ancestor of Proto-Canaanite
 - in turn, Phoenician
 - and all organically evolved alphabets, abjads, abugidas

PROTO-SINAITIC



- ~50 inscriptions
- Serabit el-Khadim (Sinai), 17th c. BCE
- Wadi el-Hol (Egypt, Qena bend of the Nile)

PROTO-SINAITIC



- Repertoire
 - One sign for phoneme in target language (there are a variant sign or two)
 - Closed set of characters (20 signs)
 - No number signs or punctuation

PROTO-SINAITIC



- Users / Inventors
 - Two hypotheses:
 - Illiterate miner
 - Literate foreman

PROTO-SINAITIC



- Origin
 - Selection of hieroglyphs on the basis of shape of the sign with a familiar object
 - No semantic or phonetic connection between sign and hieroglyph

PROTO-SINAITIC

- Orthography

'alp	bêl	gaml	digg	hô	wô (waw)	zê(n)	hê(t)	yad	kapp
lamd	mêm	nahš	(šamk?)	'ên	pi't	ša(d)	qu(p)		
ra's	tann	tô (taw)							

PROTO-SINAITIC



- Specimen

PROTO-SINAITIC

Not The Roadmap

Revision	11.0.0
Authors	Michael Everson, Rick McGowan, Ken Whistler, V.S. Umamaheswaran
Date	2018-06-06
This Version	http://www.unicode.org/roadmaps/not-the-roadmap/not-the-roadmap-11-0-0.html
Previous Version	http://www.unicode.org/roadmaps/not-the-roadmap/not-the-roadmap-10-0-0.html
Latest Version	http://www.unicode.org/roadmaps/not-the-roadmap/

Summary

This page lists scripts which, for one or another reason, are not given tentative pre-allocations in the Roadmap tables for addition to ISO/IEC 10646 and the Unicode Standard. Several categories are provided, to indicate the reasons why a script might not be suitable for pre-allocation in the Roadmap tables. For known scripts for which there is enough information and sufficient reason to do some rough pre-allocation, see instead the Roadmap tables for the BMP, SMP, SIP, TIP, and SSP.

Known scripts which have been investigated, but which are unified with existing encoded scripts.

- **Christian Sogdian** is unified with Syriac

Scripts (or pseudoscripts) which have been investigated and rejected as unsuitable for encoding.

- **Proto-Sinaitic**
- **Klingon pIQaD**
- **Aiha**
- Jindai scripts

- Unicode status
- proposed, but rejected

PROTO-SINAITIC

Not The Roadmap

Revision	11.0.0
Authors	Michael Everson, Rick McGowan, Ken Whistler, V.S. Umamaheswaran
Date	2018-06-06
This Version	http://www.unicode.org/roadmaps/not-the-roadmap/not-the-roadmap-11-0-0.html
Previous Version	http://www.unicode.org/roadmaps/not-the-roadmap/not-the-roadmap-10-0-0.html
Latest Version	http://www.unicode.org/roadmaps/not-the-roadmap/

Summary

This page lists scripts which, for one or another reason, are not given tentative pre-allocations in the Roadmap tables for addition to ISO/IEC 10646 and the Unicode Standard. Several categories are provided, to indicate the reasons why a script might not be suitable for pre-allocation in the Roadmap tables. For known scripts for which there is enough information and sufficient reason to do some rough pre-allocation, see instead the Roadmap tables for the BMP, SMP, SIP, TIP, and SSP.

Known scripts which have been investigated, but which are unified with existing encoded scripts.

- **Christian Sogdian** is unified with Syriac

Scripts (or pseudoscripts) which have been investigated and rejected as unsuitable for encoding.

- **Proto-Sinaitic**
- **Klingon pIQaD**
- **Aiha**
- Jindai scripts

- Current usage
 - active scholarship
 - exchange of documents containing signs

PROTO-SINAITIC

Not The Roadmap

Revision	11.0.0
Authors	Michael Everson, Rick McGowan, Ken Whistler, V.S. Umamaheswaran
Date	2018-06-06
This Version	http://www.unicode.org/roadmaps/not-the-roadmap/not-the-roadmap-11-0-0.html
Previous Version	http://www.unicode.org/roadmaps/not-the-roadmap/not-the-roadmap-10-0-0.html
Latest Version	http://www.unicode.org/roadmaps/not-the-roadmap/

Summary

This page lists scripts which, for one or another reason, are not given tentative pre-allocations in the Roadmap tables for addition to ISO/IEC 10646 and the Unicode Standard. Several categories are provided, to indicate the reasons why a script might not be suitable for pre-allocation in the Roadmap tables. For known scripts for which there is enough information and sufficient reason to do some rough pre-allocation, see instead the Roadmap tables for the BMP, SMP, SIP, TIP, and SSP.

Known scripts which have been investigated, but which are unified with existing encoded scripts.

- **Christian Sogdian** is unified with Syriac

Scripts (or pseudoscripts) which have been investigated and rejected as unsuitable for encoding.

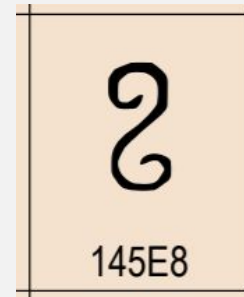
- **Proto-Sinaitic**
- **Klingon pIQaD**
- **Aiha**
- Jindai scripts

- Status of decipherment
 - not firmly deciphered
 - however, there is consensus on values of several signs

N.E. SCRIPTS IN THE UNICODE STANDARD NOT FULLY UNDERSTOOD

- Scripts in Unicode containing some characters whose values are still unknown:
 - Linear B
 - Carian
 - Anatolian Hieroglyphs

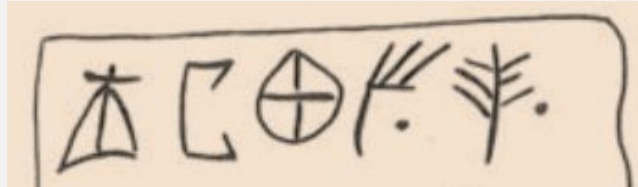
U+145E8 ANATOLIA
HIEROGLYPH A435
= syllabic a-x?



N.E. SCRIPTS IN THE UNICODE STANDARD NOT FULLY UNDERSTOOD

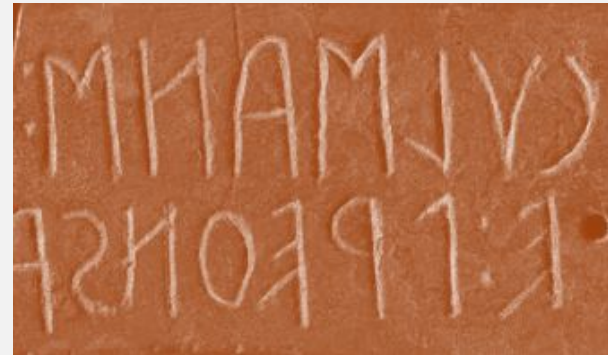
- Scripts that are partly deciphered/language not fully understood:

Linear A



- Underlying language not fully understood:

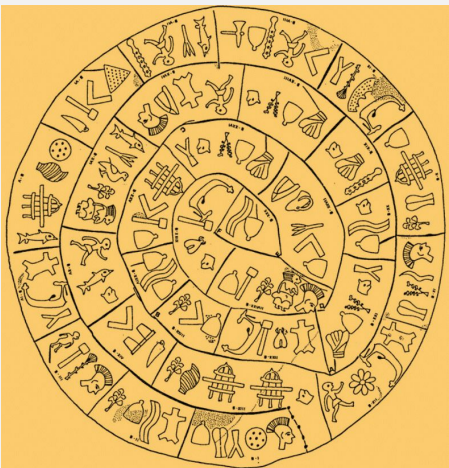
Etruscan, (Old Italic script)



Etruscan inscription in Old Italic script

APPROACHES TO UNDECIPHERED SCRIPTS

- Use image-based solutions/PUA until script is better understood, or a stronger case for encoding can be made
- Option: Encode characters as *symbols*, such as was done for Phaistos Disc symbols



APPROACHES TO UNDECIPHERED SCRIPTS – PROTO-SINAITIC

- Model A
 - Use encoding for Phoenician or Hebrew, as currently done for existing fonts
- Model B
 - Encoding Proto-Sinaitic (or, Sinaitic) as a separate script
 - Character repertoire may be chosen in consultation with scholars
 - Values assigned by scholars
 - Goal: interchange, not perfection