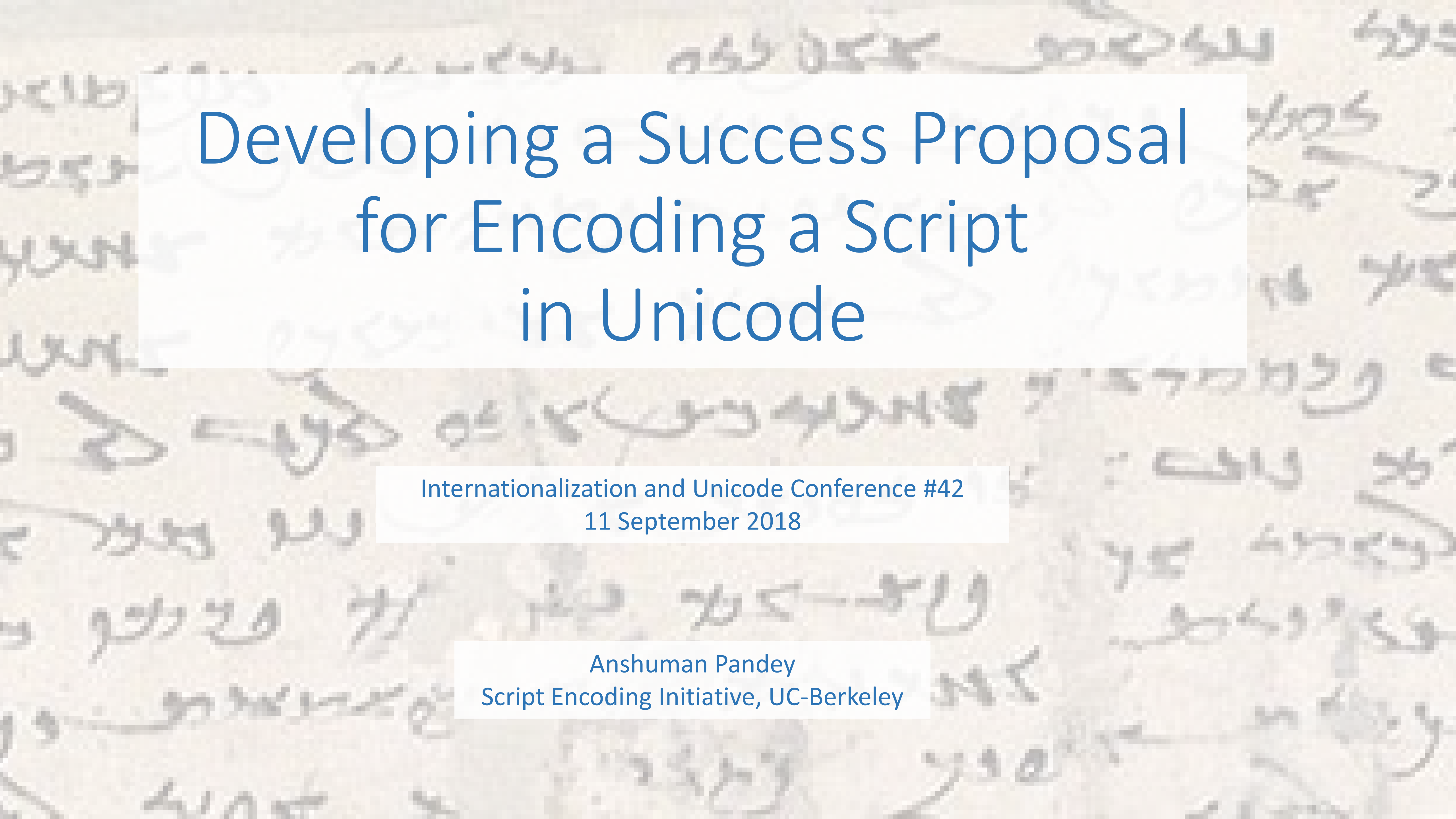


Developing a Success Proposal for Encoding a Script in Unicode

Internationalization and Unicode Conference #42
11 September 2018

Anshuman Pandey
Script Encoding Initiative, UC-Berkeley

About me

- Product Strategist, Avant, Chicago
  - Fin-tech analytics
- Researcher, Script Encoding Initiative, UC-Berkeley
  - Develop Unicode standards for writing systems of the world
- Education
  - Post-Doctoral Researcher (Linguistics), UC-Berkeley, USA
  - PhD (History), University of Michigan, Ann Arbor, USA

About me: Unicode

- Developed encodings for
  - 24 scripts
  - 4 number systems
  - various individual characters

- Work in progress
  - 25 script proposals
  - research on 100+ scripts

About me: Unicode

- Developed encodings for
  - 24 scripts
  - 4 number systems
  - various individual characters
  - 2 emoji
- Work in progress
  - 25 script proposals
  - research on 100+ scripts

auto rickshaw
(tuk tuk)

diya lamp

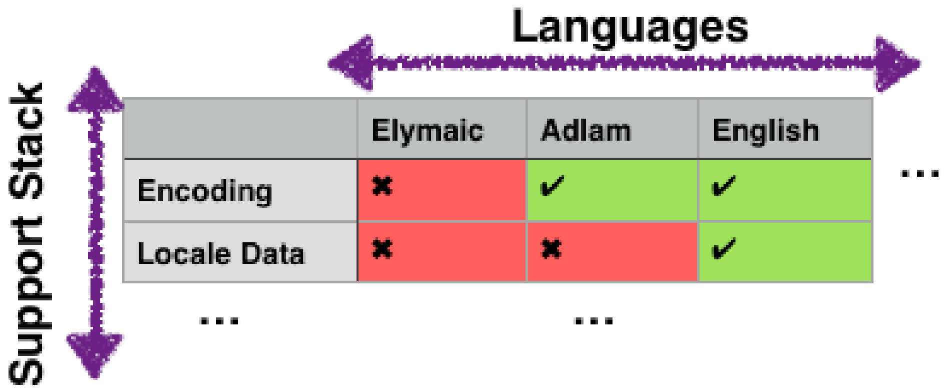
nazar amulet

Why?: Enable Digital Humanities Foundations

- Representation of ‘visible language’, eg. script / writing system
  - Often ignored in traditional and applied linguistics and language studies
  - Key to access and presentation of textual sources in the original
- Display and input of language (keyboard, stylus, etc.)
  - Enable users to input, search, and view text uniformly on devices
- Natural language processing (NLP) and machine learning (ML)
  - Information extraction and analysis using programmatic methods
- Goal: all languages as digitally native as English

‘Full Stack’ Digital Support for Languages

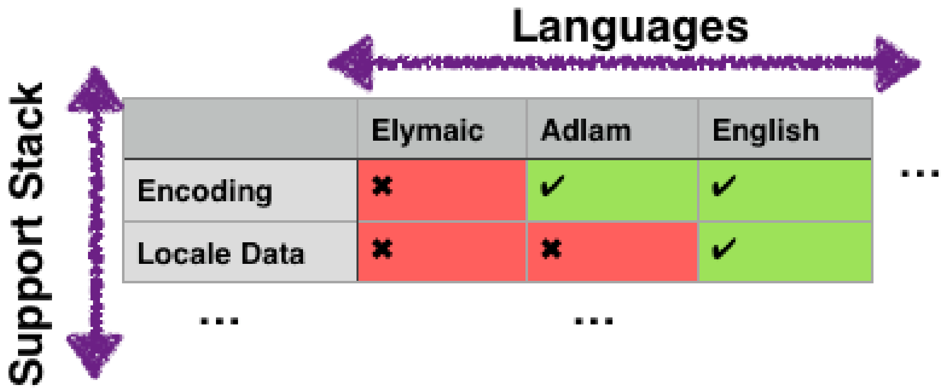
- Character-encoding standard
- Locale data
- Font / typeface
- Keyboard / input methods
- Text layout and formatting
- Database support
- System and browser support
- NLP tools and algorithms



	Elymaic	Adlam	English
Encoding	✖	✔	✔
Locale Data	✖	✖	✔

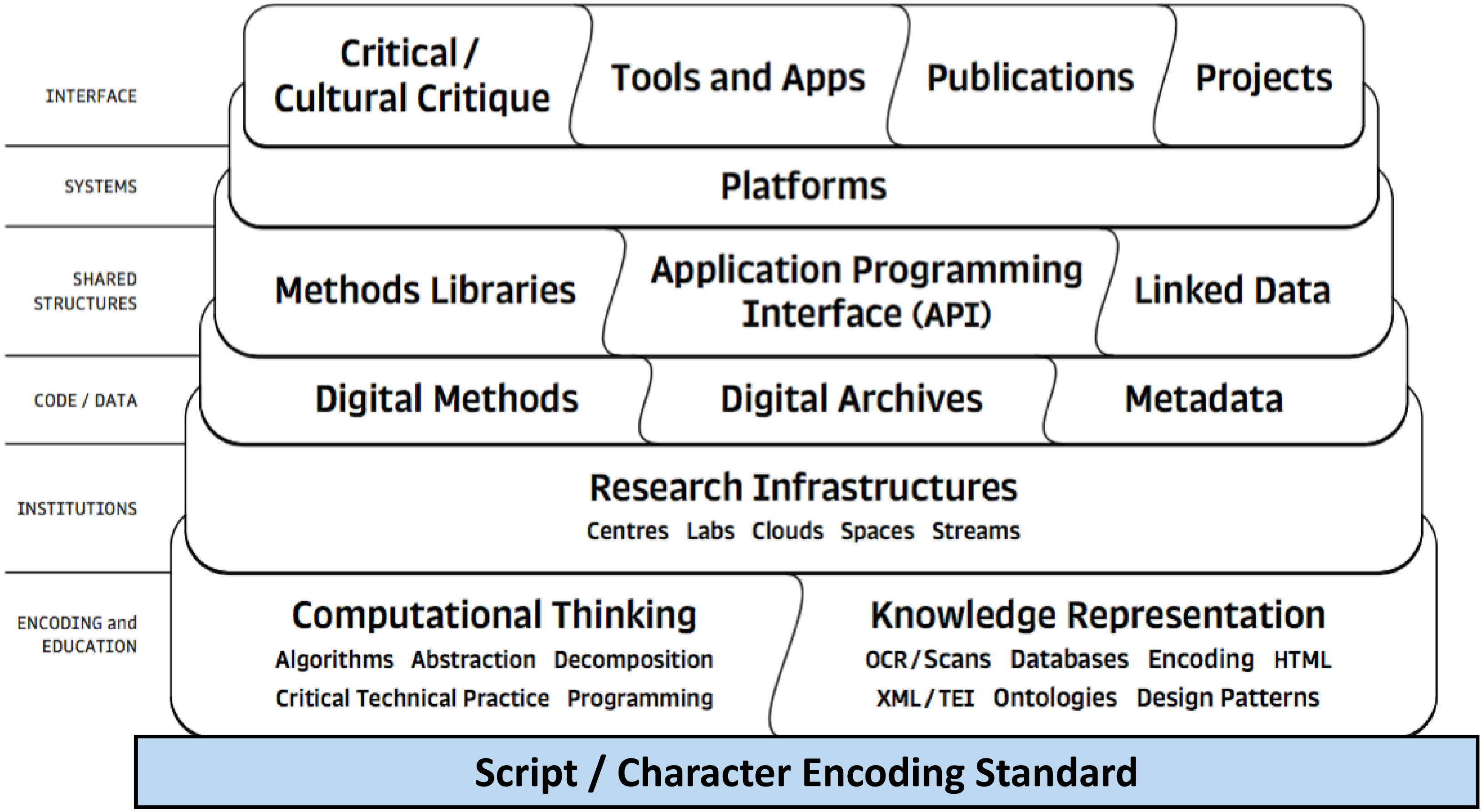
‘Full Stack’ Digital Support for Languages

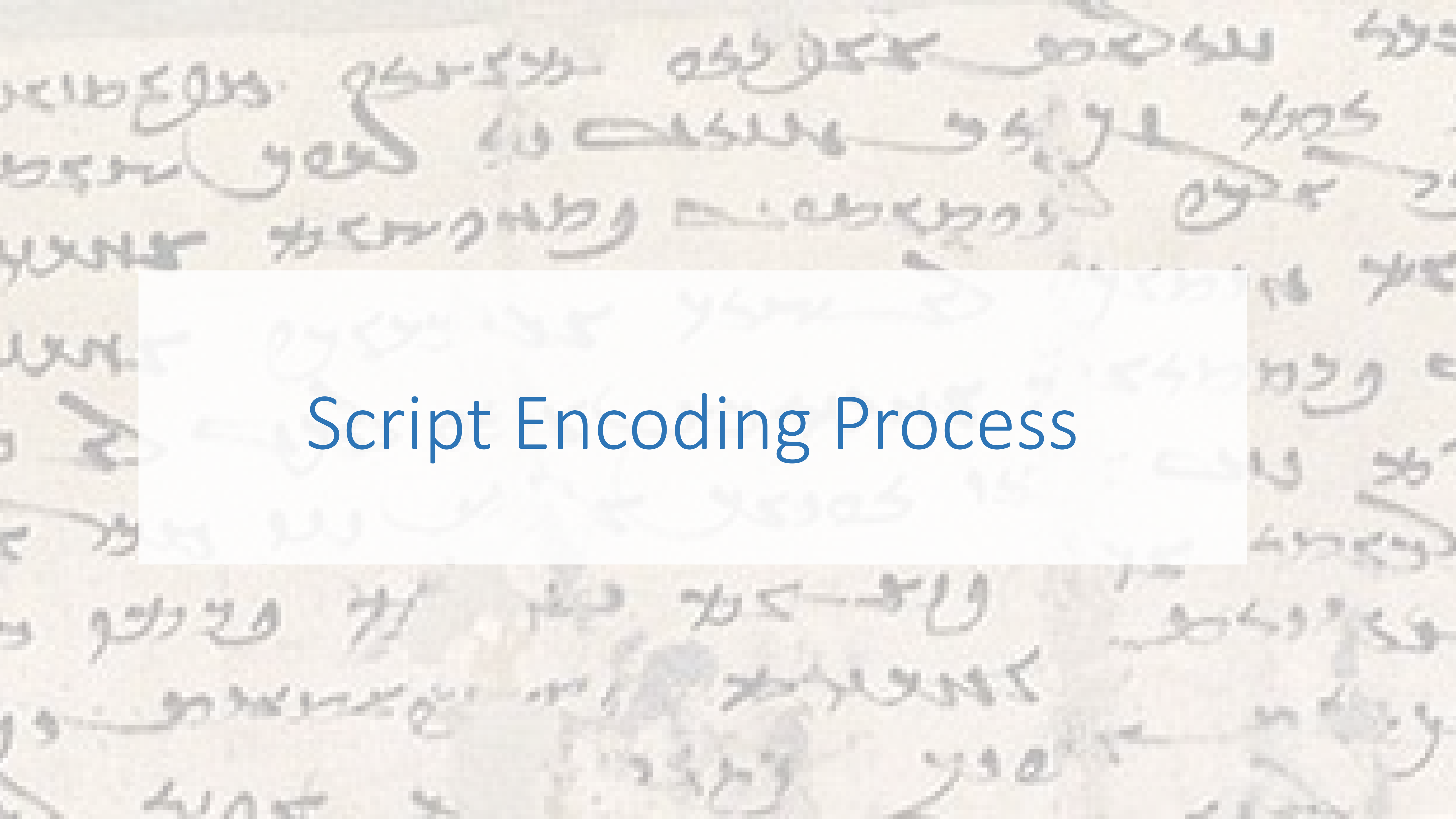
- Character-encoding standard
- Locale data
- Font / typeface
- Keyboard / input methods
- Text layout and formatting
- Database support
- System and browser support
- NLP tools and algorithms



	Elymaic	Adlam	English
Encoding	✖	✔	✔
Locale Data	✖	✖	✔

Script Encoding in the DH Stack


INTERFACE
Critical / Cultural Critique
Tools and Apps
Publications
Projects
SYSTEMS
Platforms
SHARED STRUCTURES
Methods Libraries
Application Programming Interface (API)
Linked Data
CODE / DATA
Digital Methods
Digital Archives
Metadata
INSTITUTIONS
Research Infrastructures
Centres Labs Clouds Spaces Streams
ENCODING and EDUCATION
Computational Thinking
Algorithms Abstraction Decomposition
Critical Technical Practice Programming
Knowledge Representation
OCR / Scans Databases Encoding HTML
XML / TEI Ontologies Design Patterns
Script / Character Encoding Standard



Script Encoding Process

Script Encoding Process: Overview

1. **Identification:** Native users, scholars, others identify a script not yet encoded in Unicode

2. **Development:** Research script and develop script proposal

3. **Review:** Unicode Technical Committee reviews proposals; may request changes; vote to approve or disapprove

4. **Publication:** Publication of script in Unicode standard

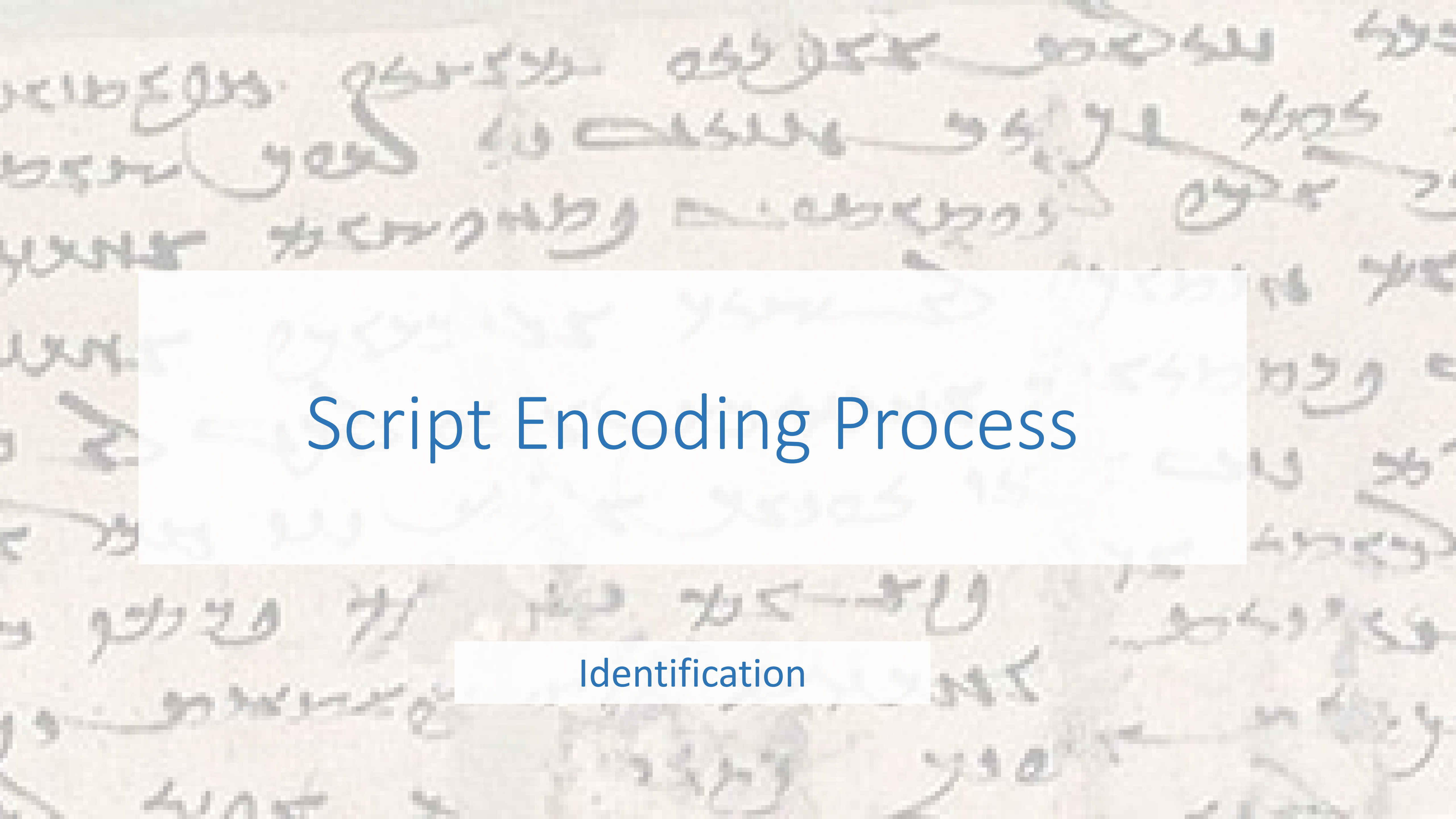
5. **Implementation:** Create fonts, keyboards, update software (Noto project; individuals)

6. **Iterative development:** Repeat for new character additions

Script Encoding Process: Focus

1. **Identification:** Identify a script not yet encoded in Unicode

2. **Development:** Techniques for researching scripts and models



Script Encoding Process

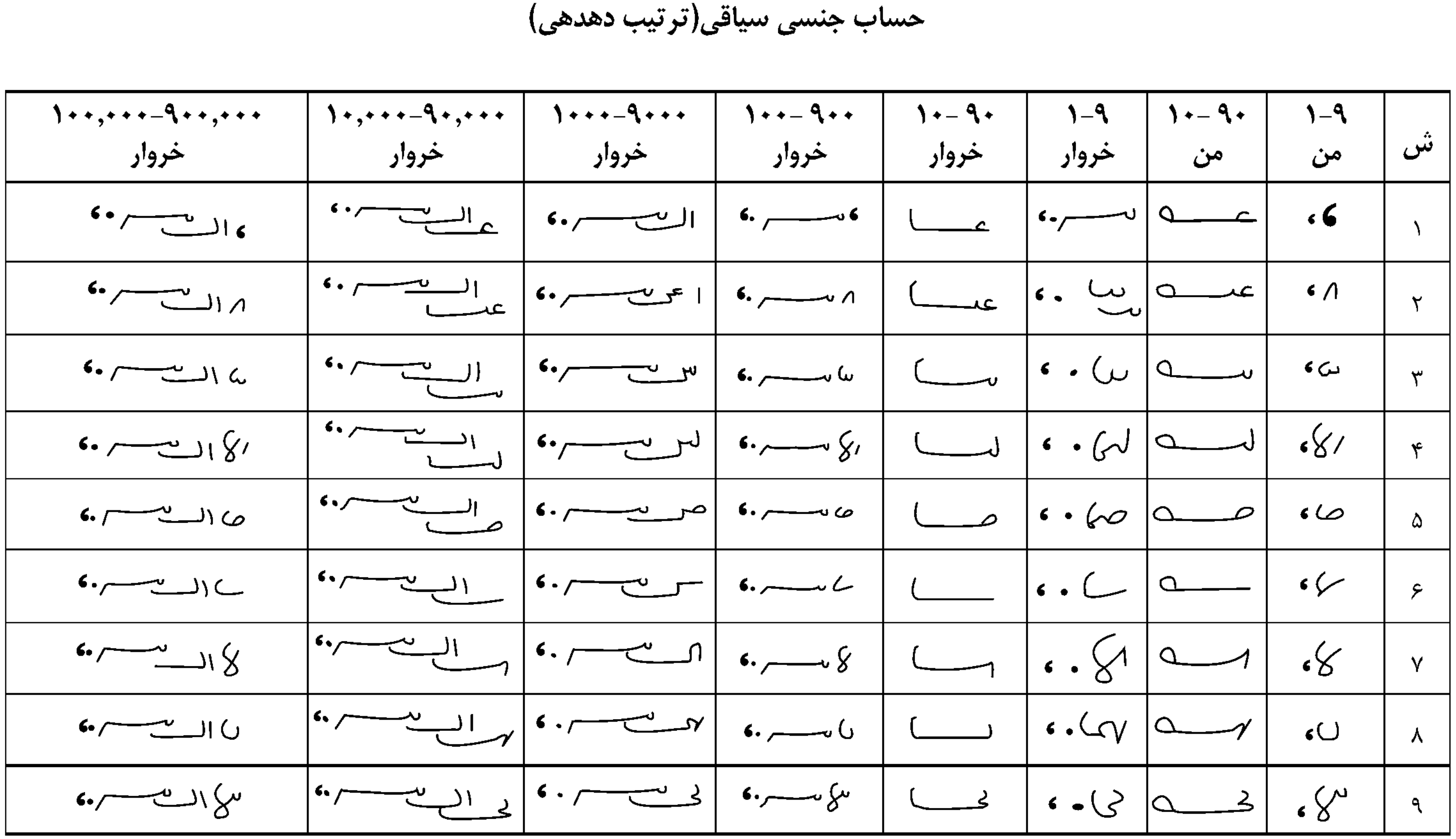
Identification

Script Encoding Process: Identification

- Existing character-encoding standards
- Request from a native user community
- Proposal from scholarly user community
- Submission from writing-systems researchers and enthusiasts
- Fieldwork and archival research

Identifying a script

Planned: Persian Siyaq

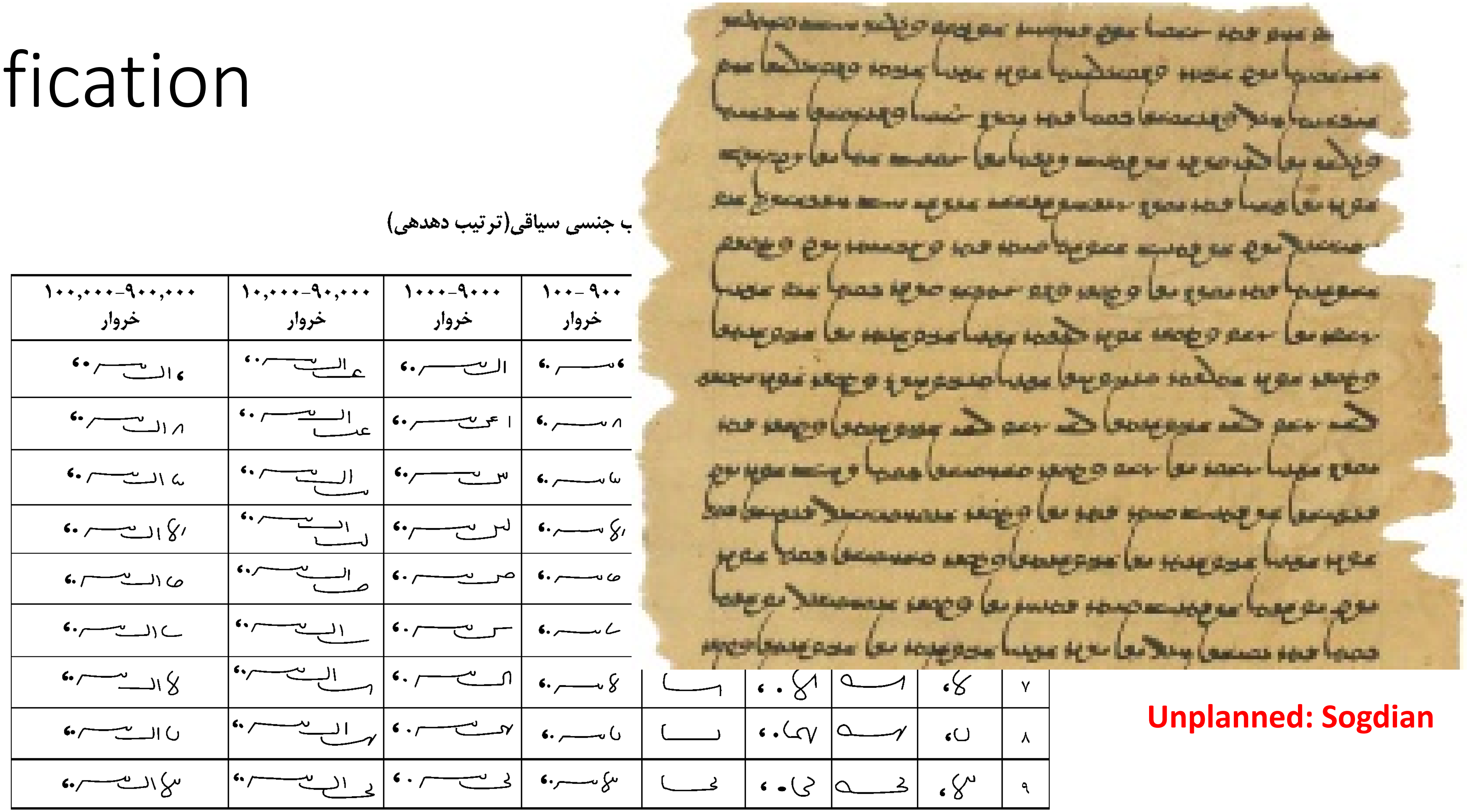


حساب جنسی سیاقی(ترتیب دهدهی)

ش	۱-۹ من	۱۰-۹۰ من	۱-۹ خروار	۱۰-۹۰ خروار	۱۰۰-۹۰۰ خروار	۱۰۰۰-۹۰۰۰ خروار	۱۰,۰۰۰-۹۰,۰۰۰ خروار	۱۰۰,۰۰۰-۹۰۰,۰۰۰ خروار
۱	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
۲	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
۳	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
۴	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
۵	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
۶	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
۷	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
۸	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
۹	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Identification

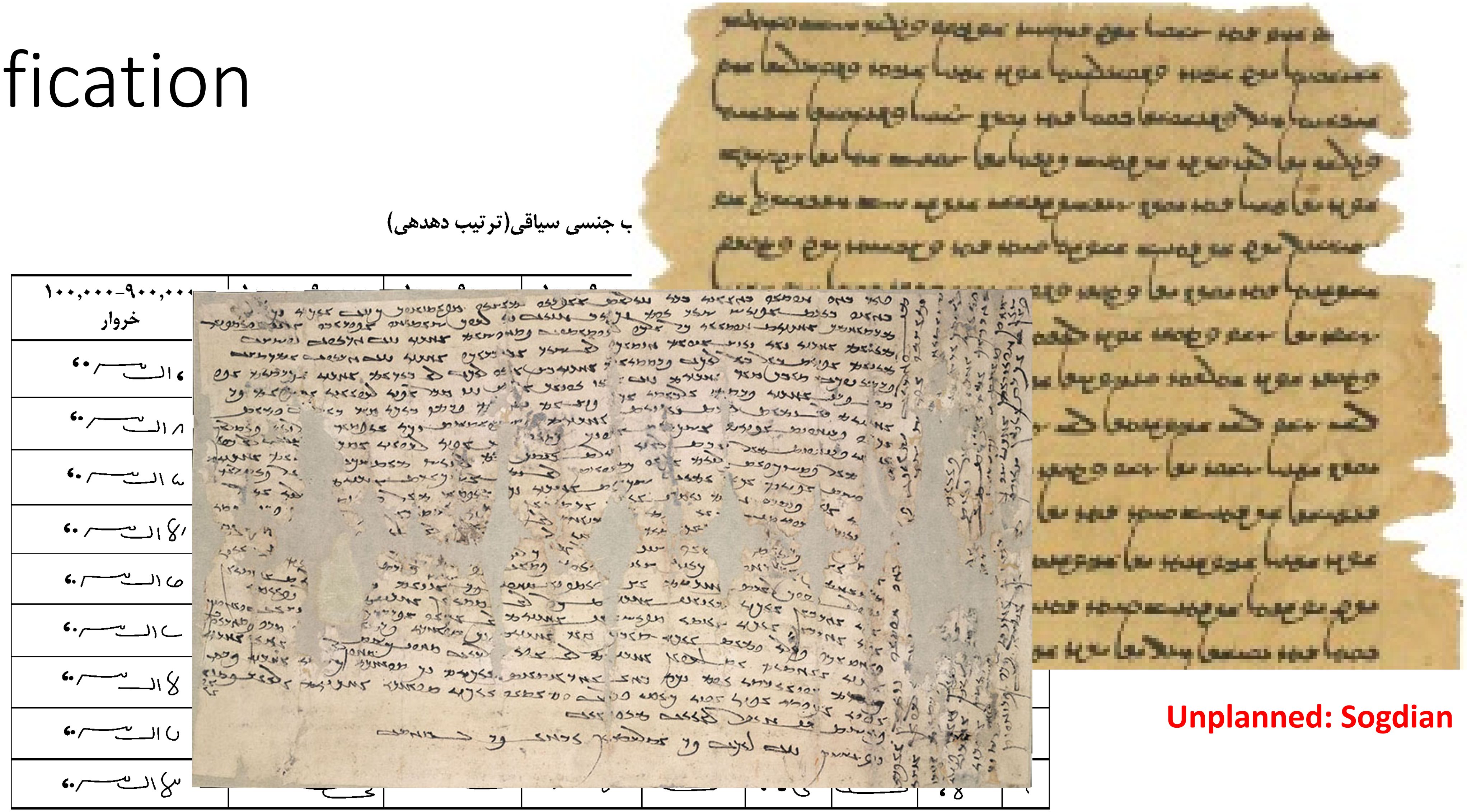
Planned: Persian Siyaq


ب جنسی سیاقی(ترتیب دهدهی)
۱۰۰,۰۰۰-۹۰۰,۰۰۰ خروار
۱۰,۰۰۰-۹۰,۰۰۰ خروار
۱۰۰۰-۹۰۰۰ خروار
۱۰۰- ۹۰۰ خروار
۷
۸
۹

Unplanned: Sogdian

Identification

Planned: Persian Siyaq



ب جنسی سیاقی(ترتیب دهدهی)

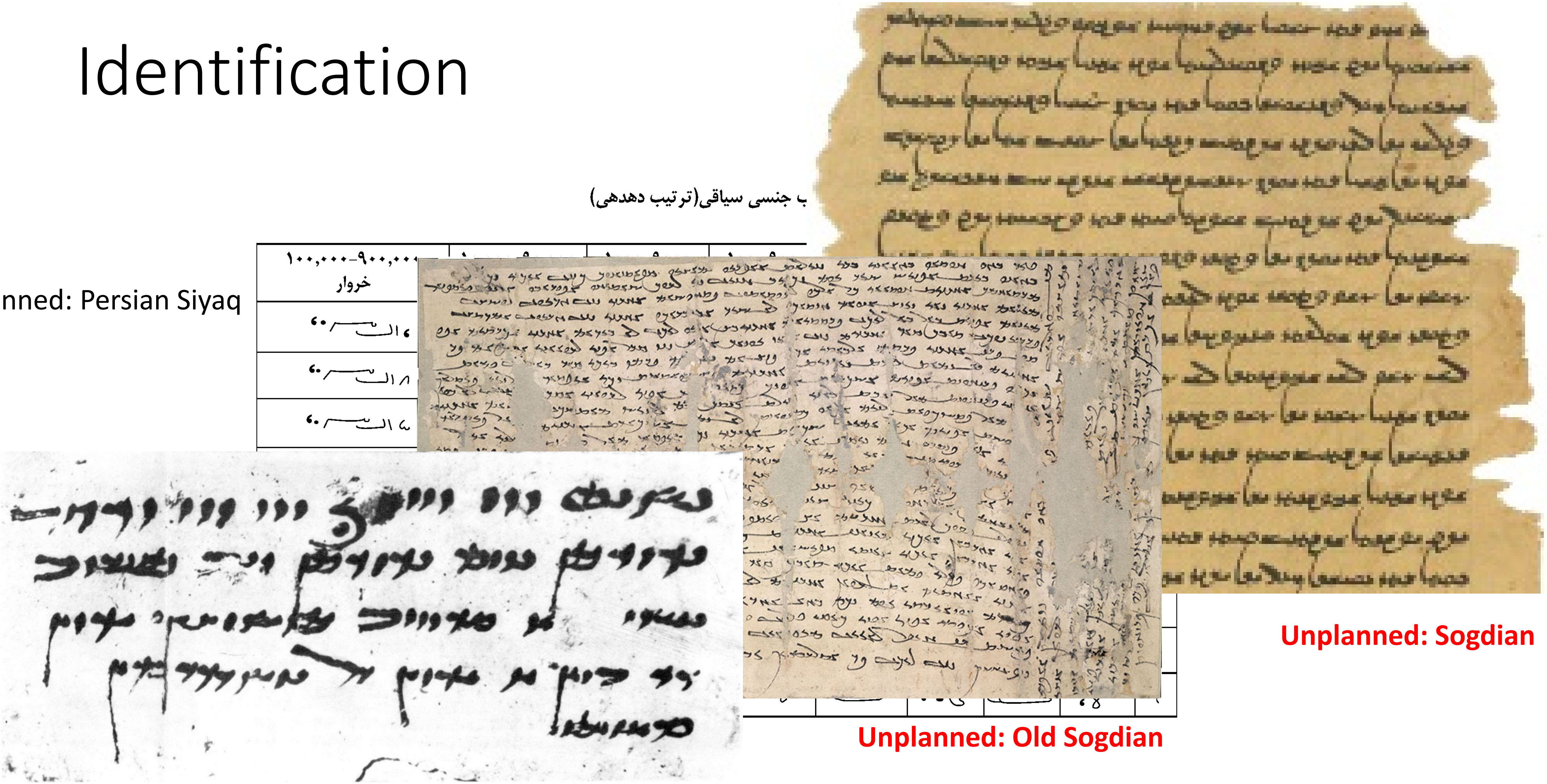
۱۰۰,۰۰۰-۹۰۰,۰۰۰ خروار

Unplanned: Sogdian

Unplanned: Old Sogdian

Identification

Planned: Persian Siyaq


ب جنسی سیاقی(ترتیب دهدهی)
۱۰۰,۰۰۰-۹۰۰,۰۰۰
خروار

Unplanned: Sogdian

Unplanned: Old Sogdian

Unplanned: Chorasmian

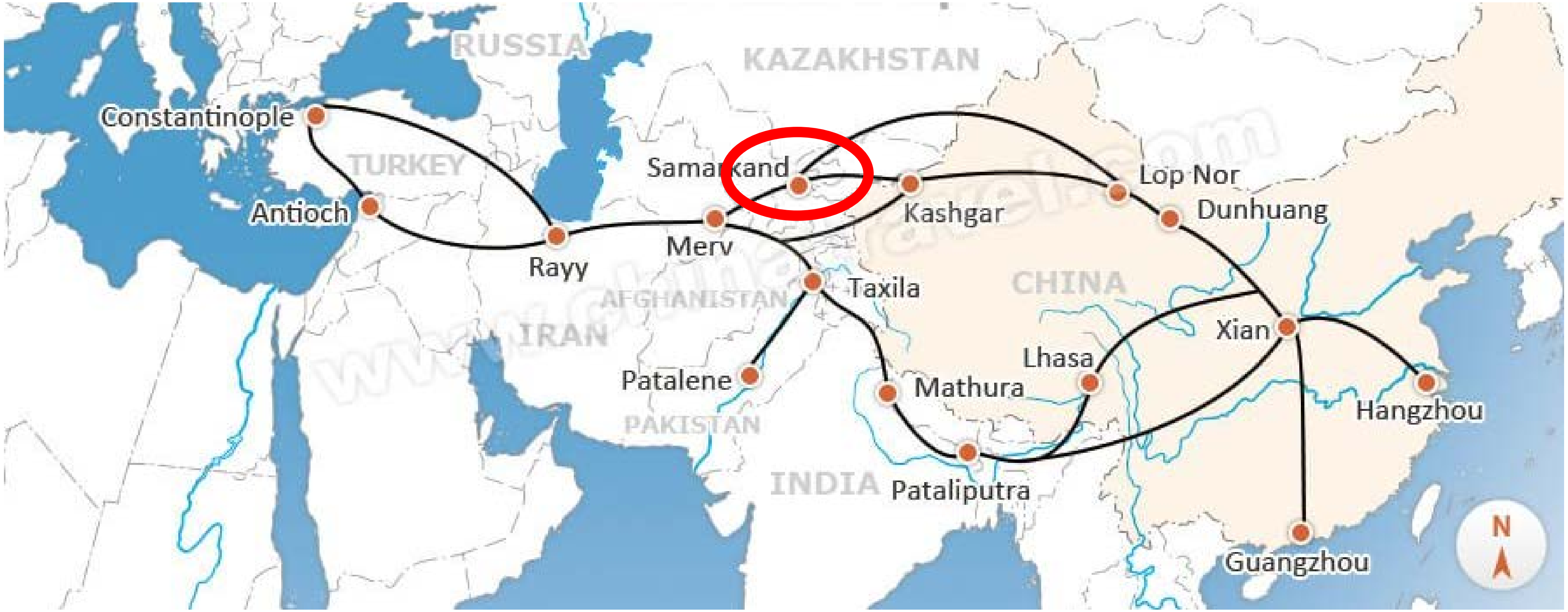
Case Study: Old Sogdian

- Consonantal alphabet
- Used for writing Sogdian
  - now-extinct eastern Iranian language
- Derived from Imperial Aramaic
  - Variation of Achaemenid administrative script
- In use from 3rd – 7th century CE

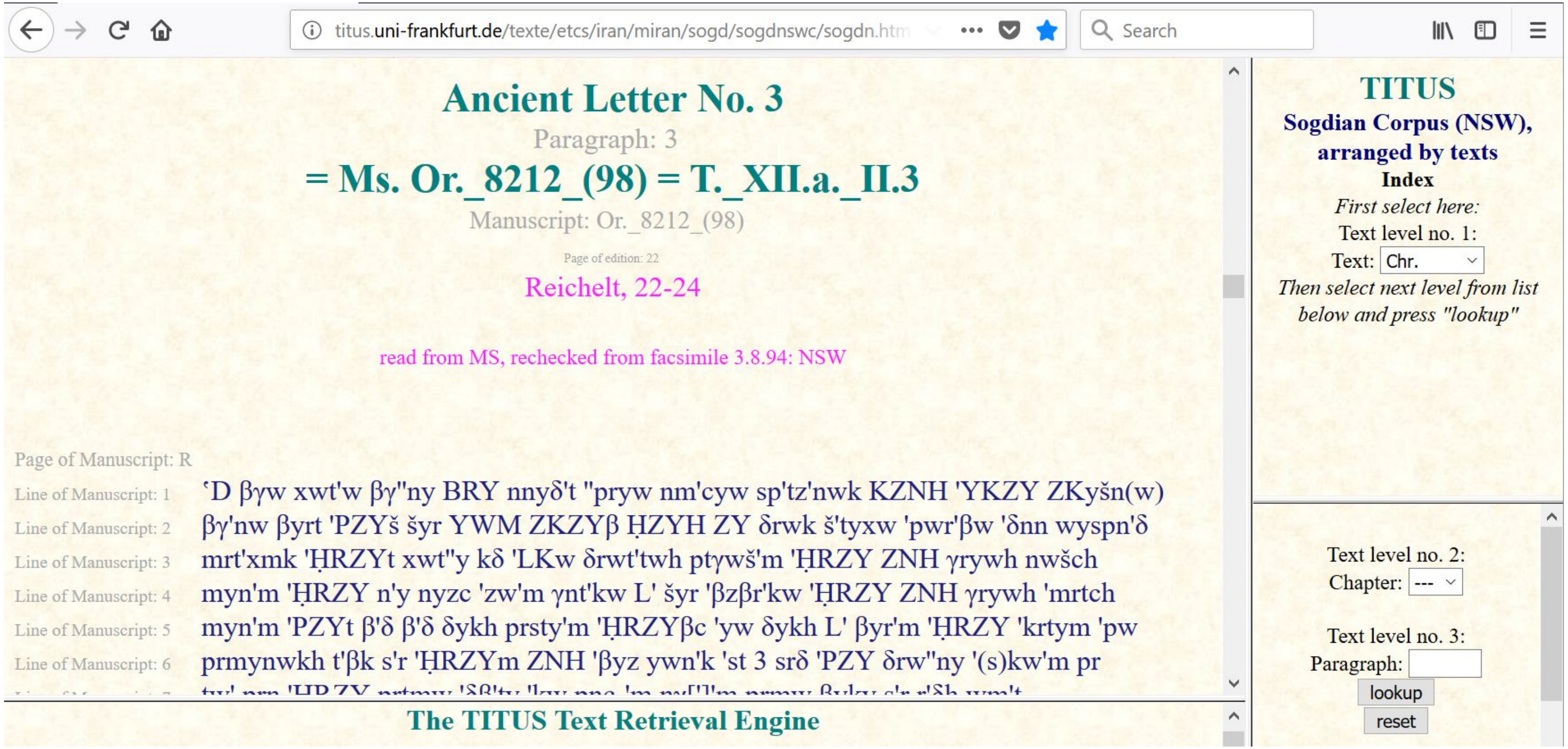
Old Sogdian: Identification

- Known to scholars
  - 19th century to present
  - active scholarship, publication trends
- Digital humanities content
  - TITUS project: transliterations
  - UW Silk Road Project: translations
- Manuscript digitization
  - International Dunhuang Project (IDP)

Situating the Sogdians



Old Sogdian: Identification



titus.uni-frankfurt.de/texte/etcs/iran/miran/sogd/sogdnswc/sogdn.htm
Search

Ancient Letter No. 3

Paragraph: 3

= Ms. Or._8212_(98) = T._XII.a._II.3

Manuscript: Or._8212_(98)

Page of edition: 22

Reichelt, 22-24

read from MS, rechecked from facsimile 3.8.94: NSW

Page of Manuscript: R

Line of Manuscript: 1 ʿD βγw xwt'w βγ''ny BRY nnyδ't ''pryw nm'cyw sp'tz'nwk KZNH 'YKZY ZKyšn(w)

Line of Manuscript: 2 βγ'nw βyrt 'PZYš šyr YWM ZKZYβ ḤZYH ZY δrwk š'tyxw 'pwr'βw 'δnn wyspn'δ

Line of Manuscript: 3 mrt'xmk 'ḤRZYt xwt''y kδ 'LKw δrwt'twh ptγwš'm 'ḤRZY ZNH γrywh nwšch

Line of Manuscript: 4 myn'm 'ḤRZY n'y nyzc 'zw'm γnt'kw L' šyr 'βzβr'kw 'ḤRZY ZNH γrywh 'mrtch

Line of Manuscript: 5 myn'm 'PZYt β'δ β'δ δykh prsty'm 'ḤRZYβc 'yw δykh L' βyr'm 'ḤRZY 'krtym 'pw

Line of Manuscript: 6 prmynwkh t'βk s'r 'ḤRZYm ZNH 'βyz ywn'k 'st 3 srδ 'PZY δrw''ny '(s)kw'm pr

The TITUS Text Retrieval Engine

TITUS

Sogdian Corpus (NSW), arranged by texts

Index

First select here:

Text level no. 1:

Text: Chr.

Then select next level from list below and press "lookup"

Text level no. 2:

Chapter: ---

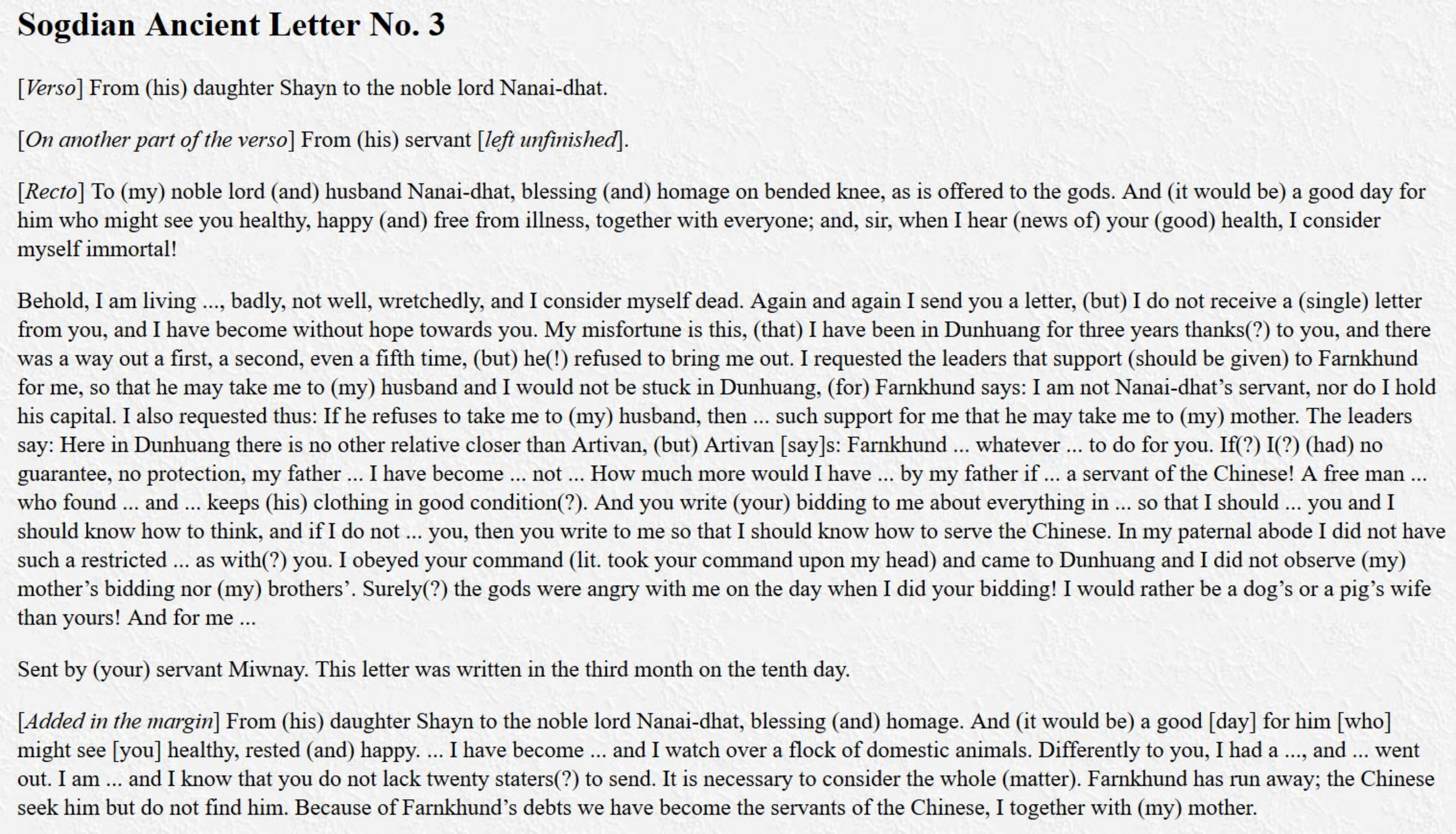
Text level no. 3:

Paragraph:

lookup

reset

Old Sogdian: Identification



Sogdian Ancient Letter No. 3

[*Verso*] From (his) daughter Shayn to the noble lord Nanai-dhat.

[*On another part of the verso*] From (his) servant [*left unfinished*].

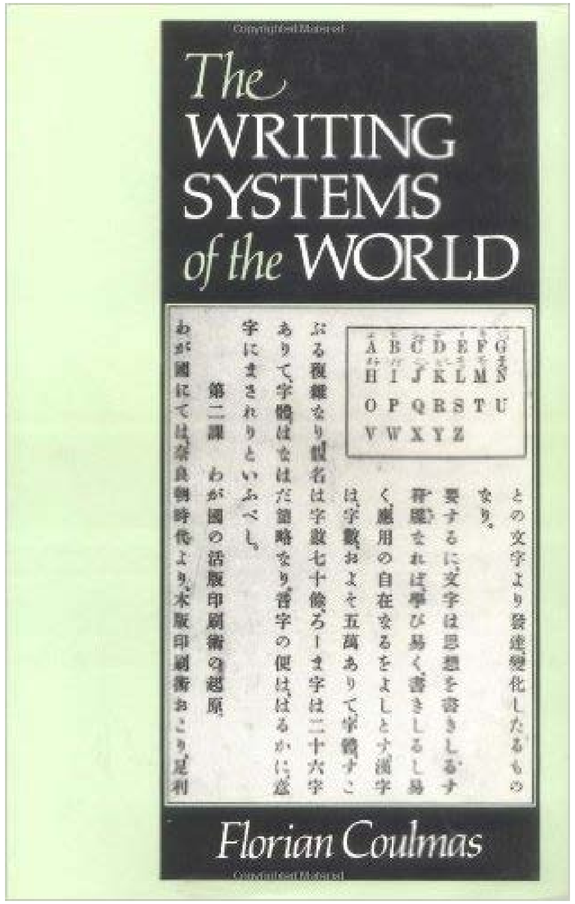
[*Recto*] To (my) noble lord (and) husband Nanai-dhat, blessing (and) homage on bended knee, as is offered to the gods. And (it would be) a good day for him who might see you healthy, happy (and) free from illness, together with everyone; and, sir, when I hear (news of) your (good) health, I consider myself immortal!

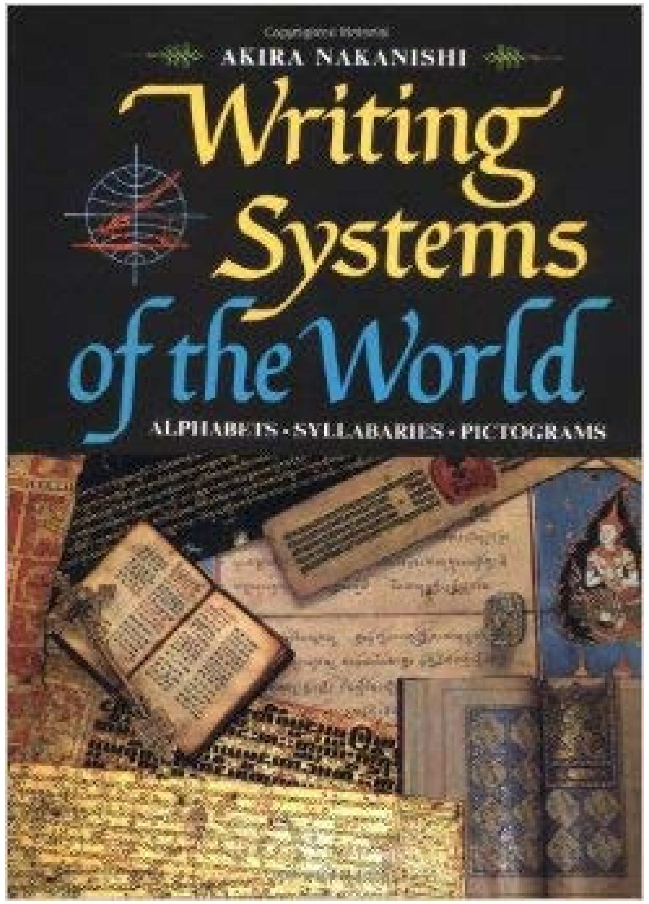
Behold, I am living ..., badly, not well, wretchedly, and I consider myself dead. Again and again I send you a letter, (but) I do not receive a (single) letter from you, and I have become without hope towards you. My misfortune is this, (that) I have been in Dunhuang for three years thanks(?) to you, and there was a way out a first, a second, even a fifth time, (but) he(!) refused to bring me out. I requested the leaders that support (should be given) to Farnkhund for me, so that he may take me to (my) husband and I would not be stuck in Dunhuang, (for) Farnkhund says: I am not Nanai-dhat's servant, nor do I hold his capital. I also requested thus: If he refuses to take me to (my) husband, then ... such support for me that he may take me to (my) mother. The leaders say: Here in Dunhuang there is no other relative closer than Artivan, (but) Artivan [say]s: Farnkhund ... whatever ... to do for you. If(?) I(?) (had) no guarantee, no protection, my father ... I have become ... not ... How much more would I have ... by my father if ... a servant of the Chinese! A free man ... who found ... and ... keeps (his) clothing in good condition(?). And you write (your) bidding to me about everything in ... so that I should ... you and I should know how to think, and if I do not ... you, then you write to me so that I should know how to serve the Chinese. In my paternal abode I did not have such a restricted ... as with(?) you. I obeyed your command (lit. took your command upon my head) and came to Dunhuang and I did not observe (my) mother's bidding nor (my) brothers'. Surely(?) the gods were angry with me on the day when I did your bidding! I would rather be a dog's or a pig's wife than yours! And for me ...

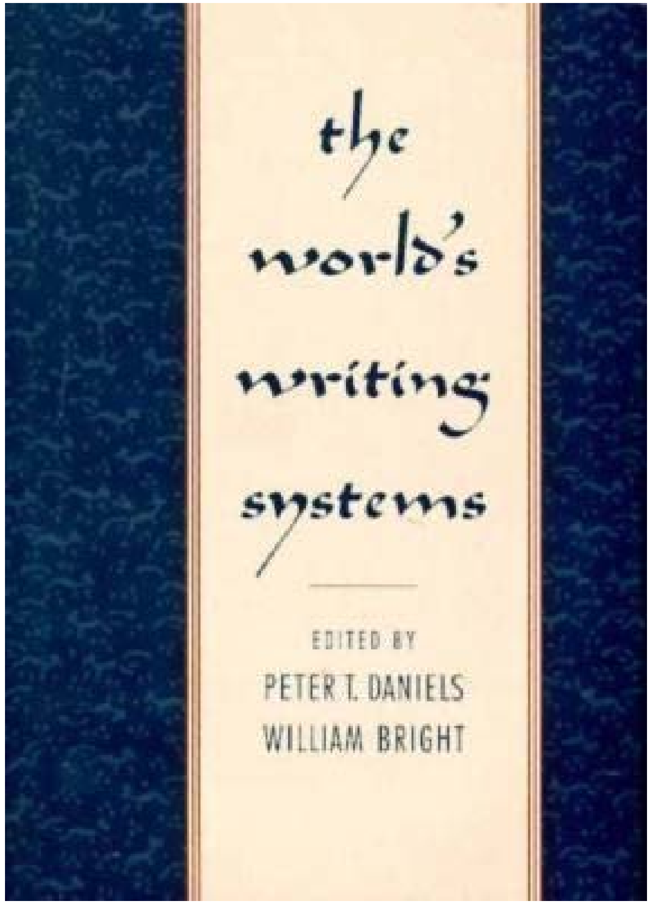
Sent by (your) servant Miwnay. This letter was written in the third month on the tenth day.

[*Added in the margin*] From (his) daughter Shayn to the noble lord Nanai-dhat, blessing (and) homage. And (it would be) a good [day] for him [who] might see [you] healthy, rested (and) happy. ... I have become ... and I watch over a flock of domestic animals. Differently to you, I had a ..., and ... went out. I am ... and I know that you do not lack twenty staters(?) to send. It is necessary to consider the whole (matter). Farnkhund has run away; the Chinese seek him but do not find him. Because of Farnkhund's debts we have become the servants of the Chinese, I together with (my) mother.

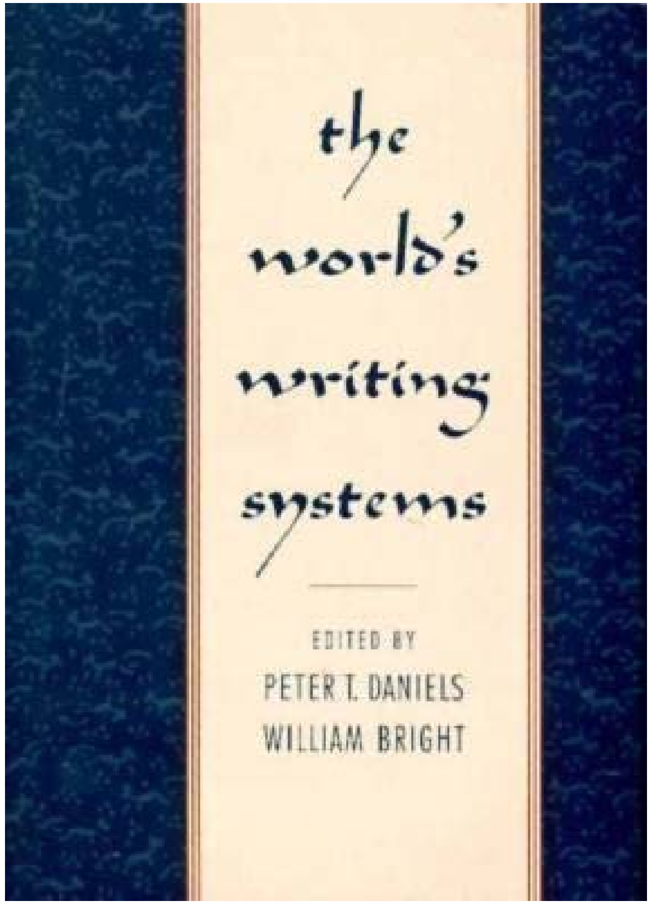
Script Encoding Process: Identification

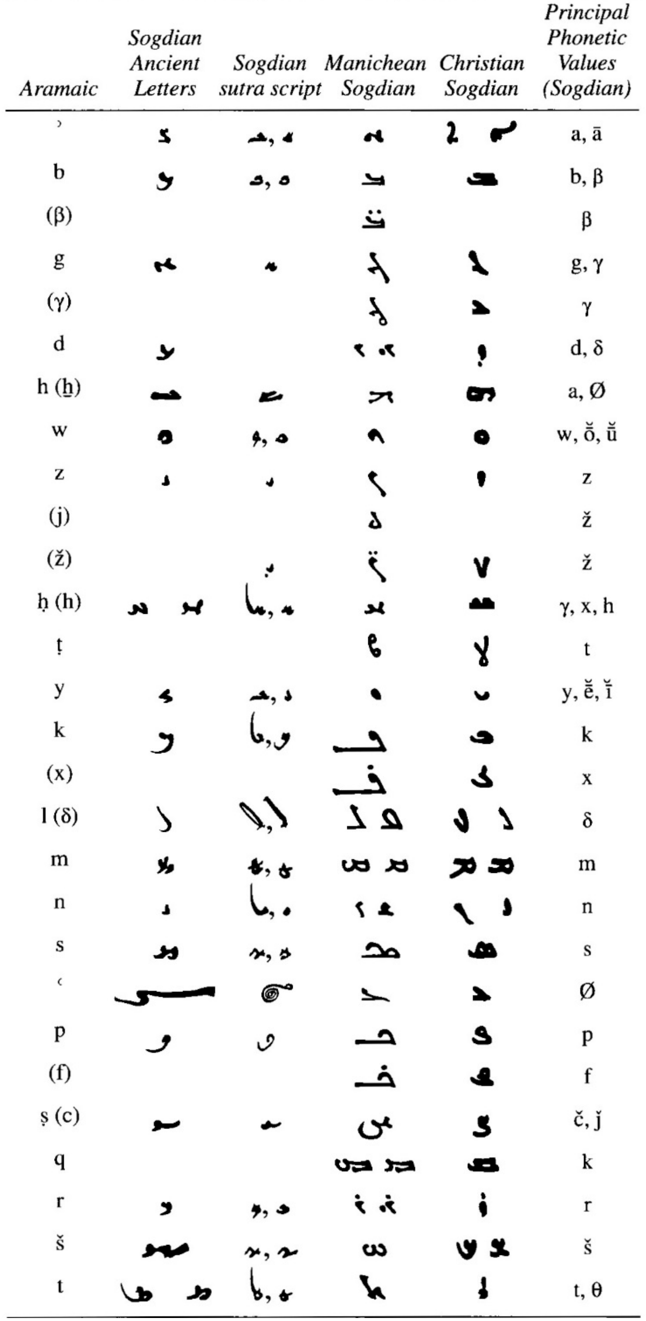



AKIRA NAKANISHI
Writing Systems of the World
ALPHABETS · SYLLABARIES · PICTOGRAMS



Old Sogdian: Identification





Aramaic	*Sogdian Ancient Letters*	*Sogdian sutra script*	*Manichean Sogdian*	*Christian Sogdian*	*Principal Phonetic Values (Sogdian)*
ʾ					a, ā
b					b, β
(β)					β
g					g, γ
(γ)					γ
d					d, δ
h (ẖ)					a, Ø
w					w, ŏ, ŭ
z					z
(j)					ž
(ž)					ž
ḥ (h)					γ, x, h
ṭ					t
y					y, ĕ, ĭ
k					k
(x)					x
l (δ)					δ
m					m
n					n
s					s
ʿ					Ø
p					p
(f)					f
ṣ (c)					č, ǰ
q					k
r					r
š					š
t					t, θ

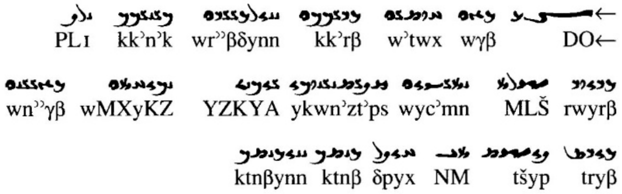
Sogdian script

In the Sogdian script used in the "Ancient Letters" (TABLE 48.2), most of the letters are distinct and do not change shape when joined. In the "formal" and "Uyghur" Sogdian scripts, most of the letters are joined and, owing to the use of a broad pen, are frequently difficult to distinguish. In the earlier form, ʾ is still distinguished from **n**; but in the later, ʾ = **n**, ʾ**n** = **n**ʾ. Some scribes distinguish **z** from **n** by not connecting **z** to the preceding letter, but others make no distinction. In the later, increasingly cursive, form, other letters tend to become indistinguishable as well: **γ/x/s/š**, **r/β/y**. Some letters are distinguished only in final position (by some scribes), e.g., **n** ~ **z**, **x** ~ **γ**.

z is sometimes distinguished from **n** or **z** from **ž** by a diacritical point, and the foreign sound *b* was noted as **ṗ**.

SAMPLES OF SOGDIAN

ANCIENT LETTERS



PL1 kkʼnʼk wrʼʼβδynn kkʼrβ wʼtwx wγβ DO←

wnʼʼγβ wMXyKZ YZKYA ykwnʼztʼps wycʼmn MLŠ rwyrβ

ktnβynn ktnβ δpyx NM tšyp tryβ

1. Transliteration:	OD	βγw	xwtʼw	βrʼkk	nnyδβʼʼrw	kʼnʼkk
2. Normalization:	at	βaγu	xutāw	βarak	nanē-θβār	kanak
3. Gloss:	to	lord.ACC	master	Barak	Nana's-gift	Kanak

1. ɪLP	βrywr	ŠLM	nmʼcyw	spʼtzʼnwky	AYKZY
2. (ēw-)zār	βrēwar	*āfrīwan	namācyu	spātzānūk	kaδ-uti
3. thousand	ten.thousand	greeting(?)	reverence.ACC	bended.knee	when-that.and

1. ZKyXMw	βγʼʼnw	βyrt	pyšt	MN	xypθ	βntk	nnyβntk
2. wēšanu	βaγān(u)	βyart	pišt	con	xēpθ	βantē	nanē-βantē
3. them.OBL	lords.OBL	received	written	from	own	servant	Nana's-servant

> 'To the Divine Master Barak(?) Nanethvar Kanak a thousand, ten thousand greetings, reverently with bended knees when received by their divinities. Written by his own servant Nanevante.'
>
> —*From the Old Sogdian "Ancient Letters" found in a mailbag in the Great Wall (AL II, Reichelt 1931: 12 and pl. 2).*

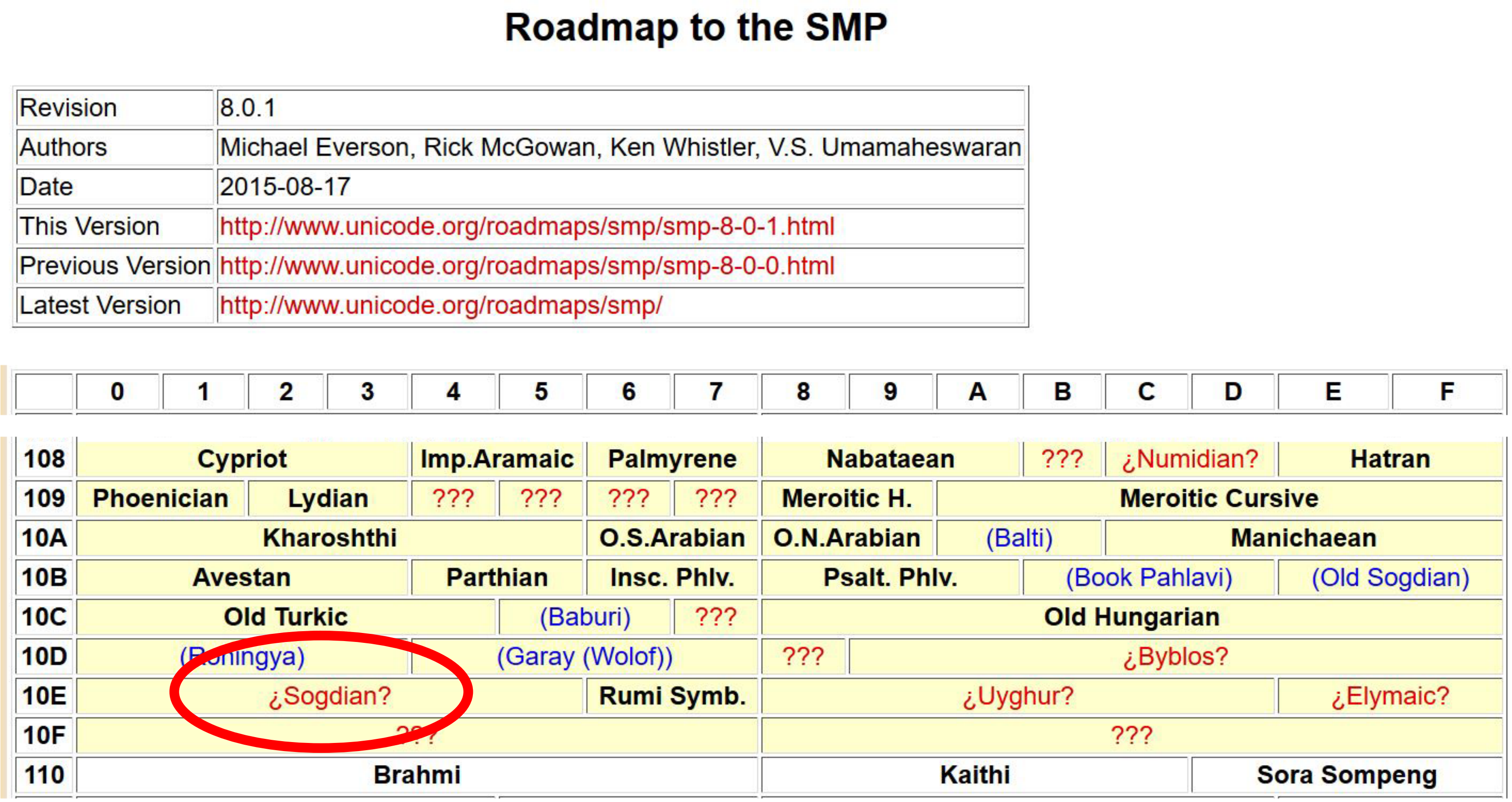
Script Encoding Process: Identification

- Challenges
  - Charts or specimens not in published in books on writing systems
  - Studies may be outdated or very limited
- Resolution
  - Conduct research: primary sources!

Script Encoding Process

Research

Old Sogdian: Research (2015)

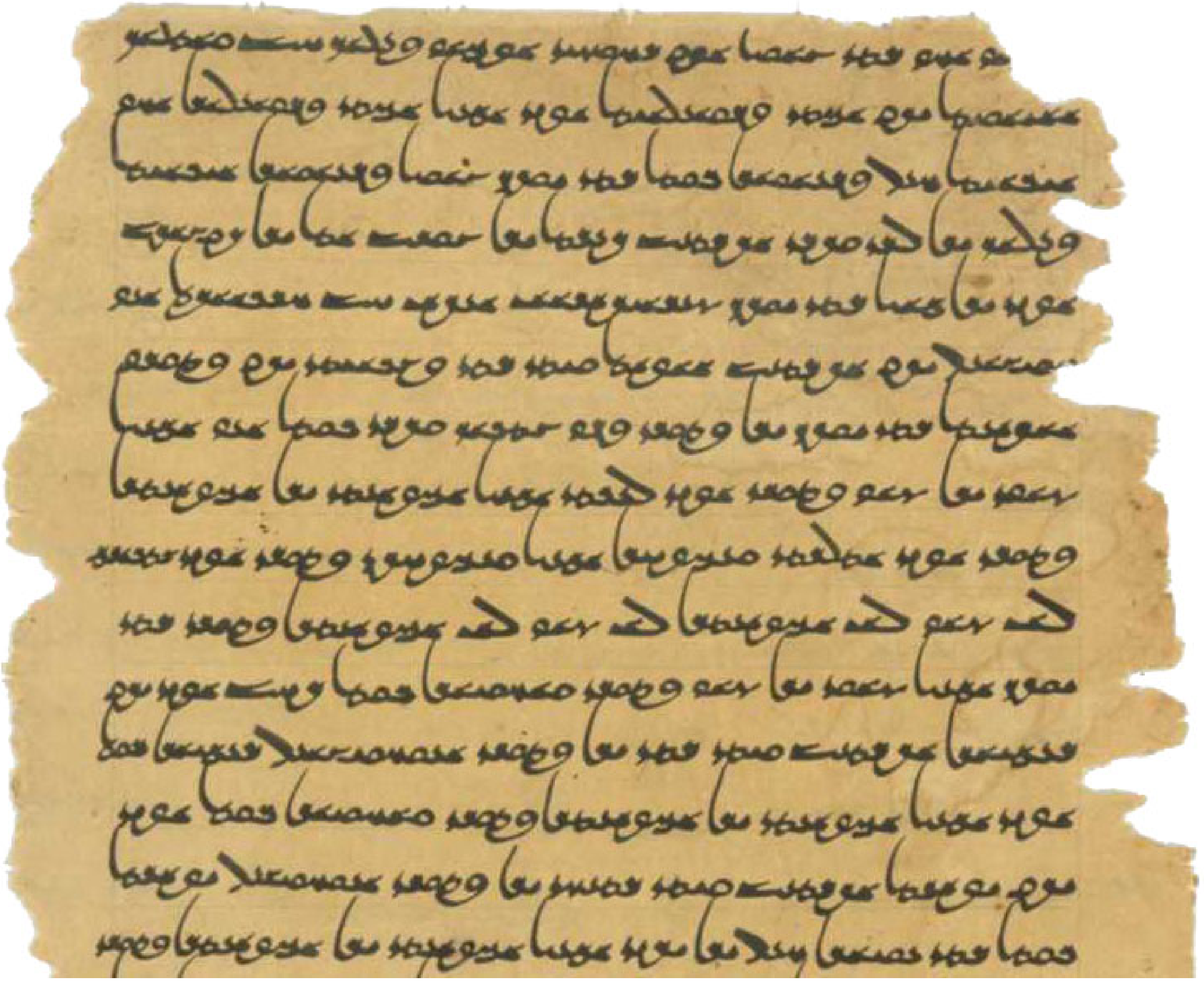


Roadmap to the SMP

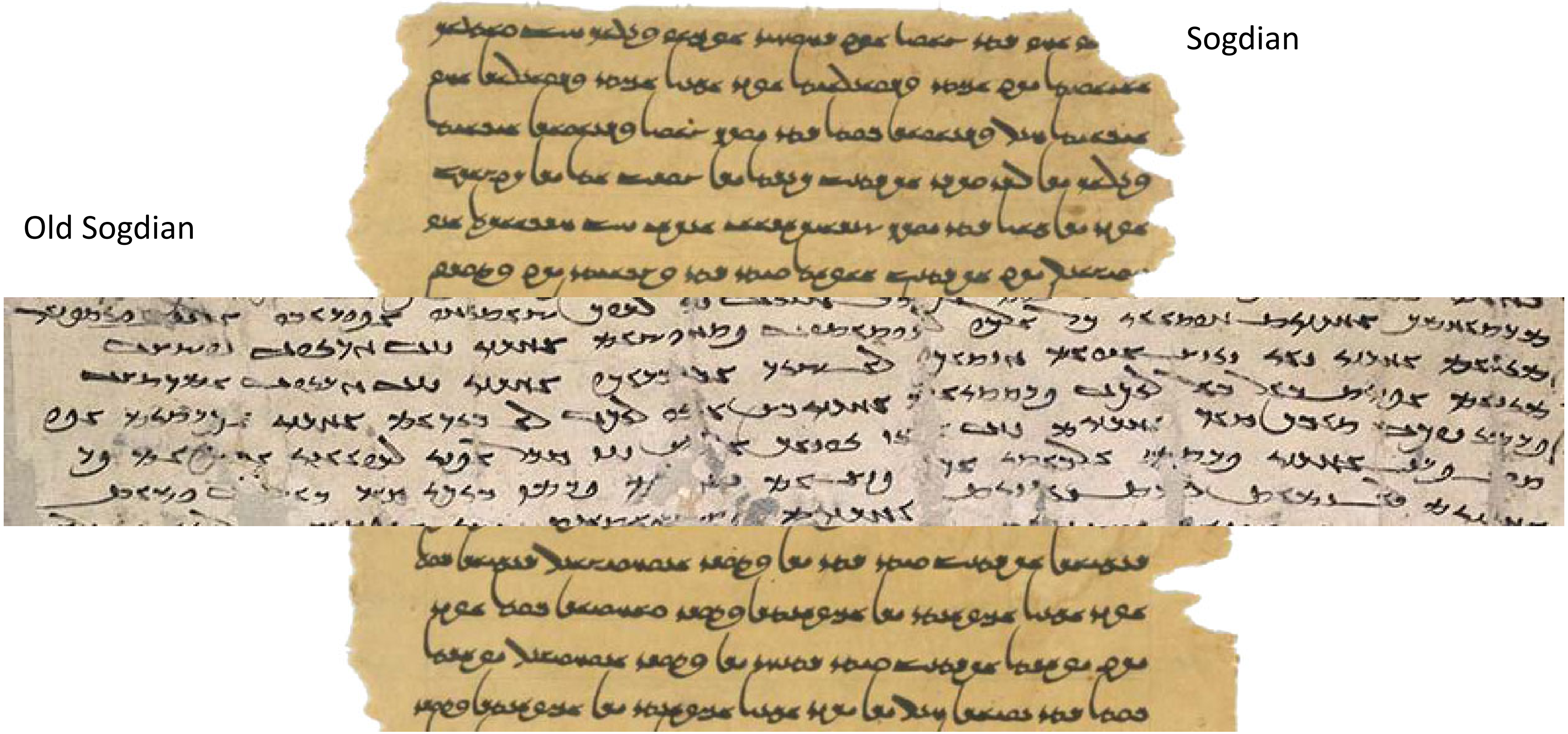
Revision	8.0.1
Authors	Michael Everson, Rick McGowan, Ken Whistler, V.S. Umamaheswaran
Date	2015-08-17
This Version	http://www.unicode.org/roadmaps/smp/smp-8-0-1.html
Previous Version	http://www.unicode.org/roadmaps/smp/smp-8-0-0.html
Latest Version	http://www.unicode.org/roadmaps/smp/

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
108	Cypriot				Imp.Aramaic		Palmyrene		Nabataean			???	¿Numidian?		Hatran	
109	Phoenician		Lydian		???	???	???	???	Meroitic H.		Meroitic Cursive					
10A	Kharoshthi						O.S.Arabian		O.N.Arabian		(Balti)		Manichaean			
10B	Avestan				Parthian		Insc. Phlv.		Psalt. Phlv.			(Book Pahlavi)			(Old Sogdian)	
10C	Old Turkic					(Baburi)		???	Old Hungarian							
10D	(Rohingya)				(Garay (Wolof))				???	¿Byblos?						
10E	¿Sogdian?						Rumi Symb.		¿Uyghur?						¿Elymaic?	
10F	???								???							
110	Brahmi								Kaithi					Sora Sompeng		

Old Sogdian: Mistaken identity



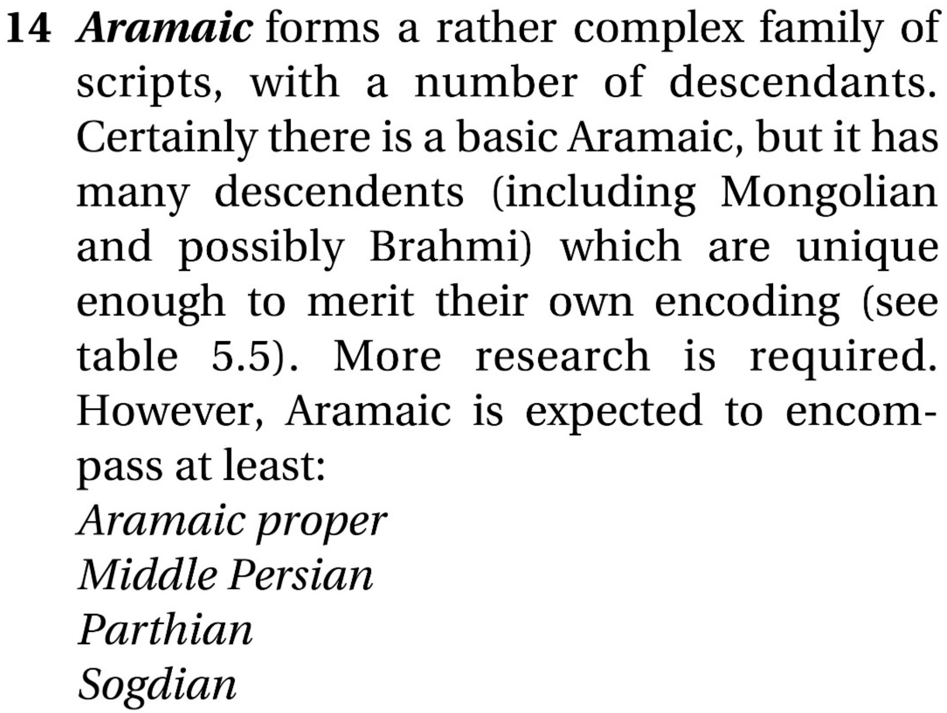
Old Sogdian: Mistaken identity



Sogdian

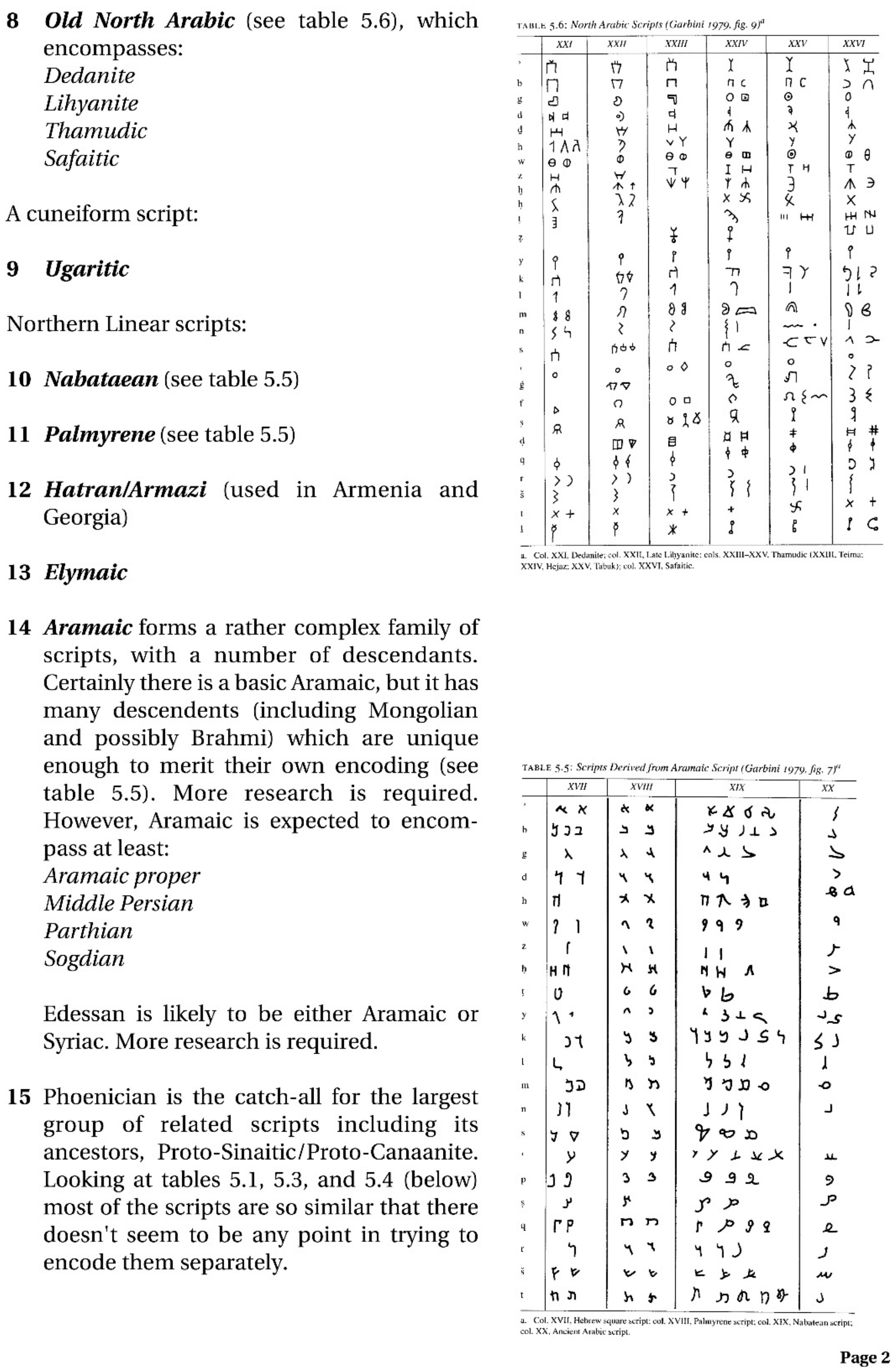
Old Sogdian

Old Sogdian: Unicode Status



14 ***Aramaic*** forms a rather complex family of scripts, with a number of descendants. Certainly there is a basic Aramaic, but it has many descendents (including Mongolian and possibly Brahmi) which are unique enough to merit their own encoding (see table 5.5). More research is required. However, Aramaic is expected to encompass at least:
Aramaic proper
Middle Persian
Parthian
Sogdian

Everson, “Roadmapping early Semitic scripts”, 2001.



8 ***Old North Arabic*** (see table 5.6), which encompasses:
Dedanite
Lihyanite
Thamudic
Safaitic

A cuneiform script:

9 ***Ugaritic***

Northern Linear scripts:

10 ***Nabataean*** (see table 5.5)

11 ***Palmyrene*** (see table 5.5)

12 ***Hatran/Armazi*** (used in Armenia and Georgia)

13 ***Elymaic***

14 ***Aramaic*** forms a rather complex family of scripts, with a number of descendants. Certainly there is a basic Aramaic, but it has many descendents (including Mongolian and possibly Brahmi) which are unique enough to merit their own encoding (see table 5.5). More research is required. However, Aramaic is expected to encompass at least:
Aramaic proper
Middle Persian
Parthian
Sogdian

Edessan is likely to be either Aramaic or Syriac. More research is required.

15 Phoenician is the catch-all for the largest group of related scripts including its ancestors, Proto-Sinaitic/Proto-Canaanite. Looking at tables 5.1, 5.3, and 5.4 (below) most of the scripts are so similar that there doesn't seem to be any point in trying to encode them separately.

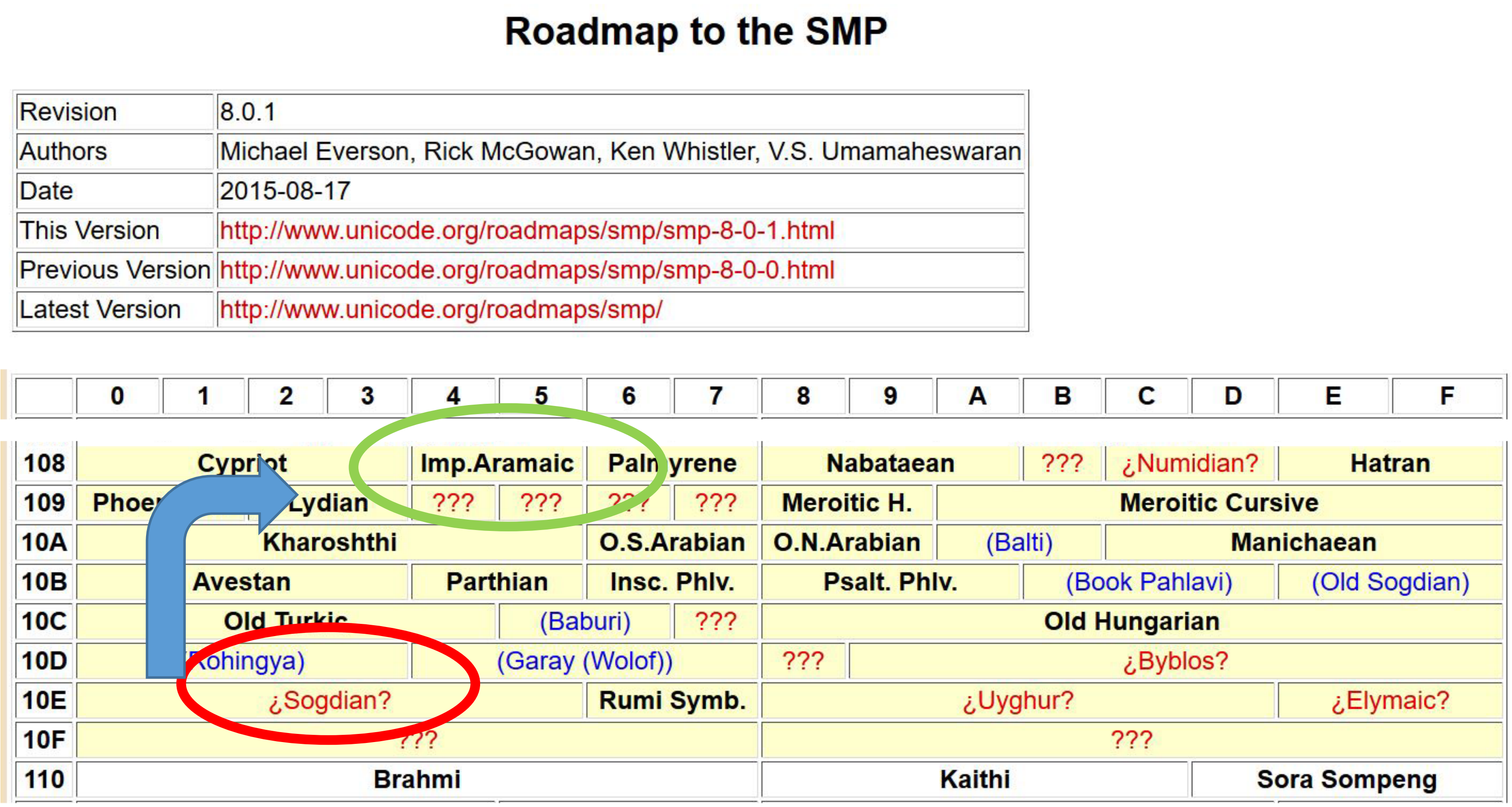
TABLE 5.6: *North Arabic Scripts (Garbini 1979, fig. 9)*[a]

[a] Col. XXI, Dedanite; col. XXII, Late Lihyanite; cols. XXIII–XXV, Thamudic (XXIII, Teima; XXIV, Hejaz; XXV, Tabuk); col. XXVI, Safaitic.

TABLE 5.5: *Scripts Derived from Aramaic Script (Garbini 1979, fig. 7)*[a]

[a] Col. XVII, Hebrew square script; col. XVIII, Palmyrene script; col. XIX, Nabatean script; col. XX, Ancient Arabic script.

Sogdian: Unicode Status – Unification?



Roadmap to the SMP

Revision	8.0.1
Authors	Michael Everson, Rick McGowan, Ken Whistler, V.S. Umamaheswaran
Date	2015-08-17
This Version	http://www.unicode.org/roadmaps/smp/smp-8-0-1.html
Previous Version	http://www.unicode.org/roadmaps/smp/smp-8-0-0.html
Latest Version	http://www.unicode.org/roadmaps/smp/

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
108	Cypriot				Imp.Aramaic		Palmyrene		Nabataean			???	¿Numidian?		Hatran	
109	Phoenician		Lydian		???	???	???	???	Meroitic H.		Meroitic Cursive					
10A	Kharoshthi						O.S.Arabian		O.N.Arabian		(Balti)		Manichaean			
10B	Avestan				Parthian		Insc. Phlv.		Psalt. Phlv.			(Book Pahlavi)			(Old Sogdian)	
10C	Old Turkic					(Baburi)		???	Old Hungarian							
10D	(Rohingya)				(Garay (Wolof))				???	¿Byblos?						
10E	¿Sogdian?						Rumi Symb.		¿Uyghur?						¿Elymaic?	
10F	???								???							
110	Brahmi								Kaithi					Sora Sompeng		

Script Encoding Process: Research

- Analysis of sources to document repertoire and script grammar
- Compare grammar and orthography of related scripts
- Outreach to user community to understand current usage

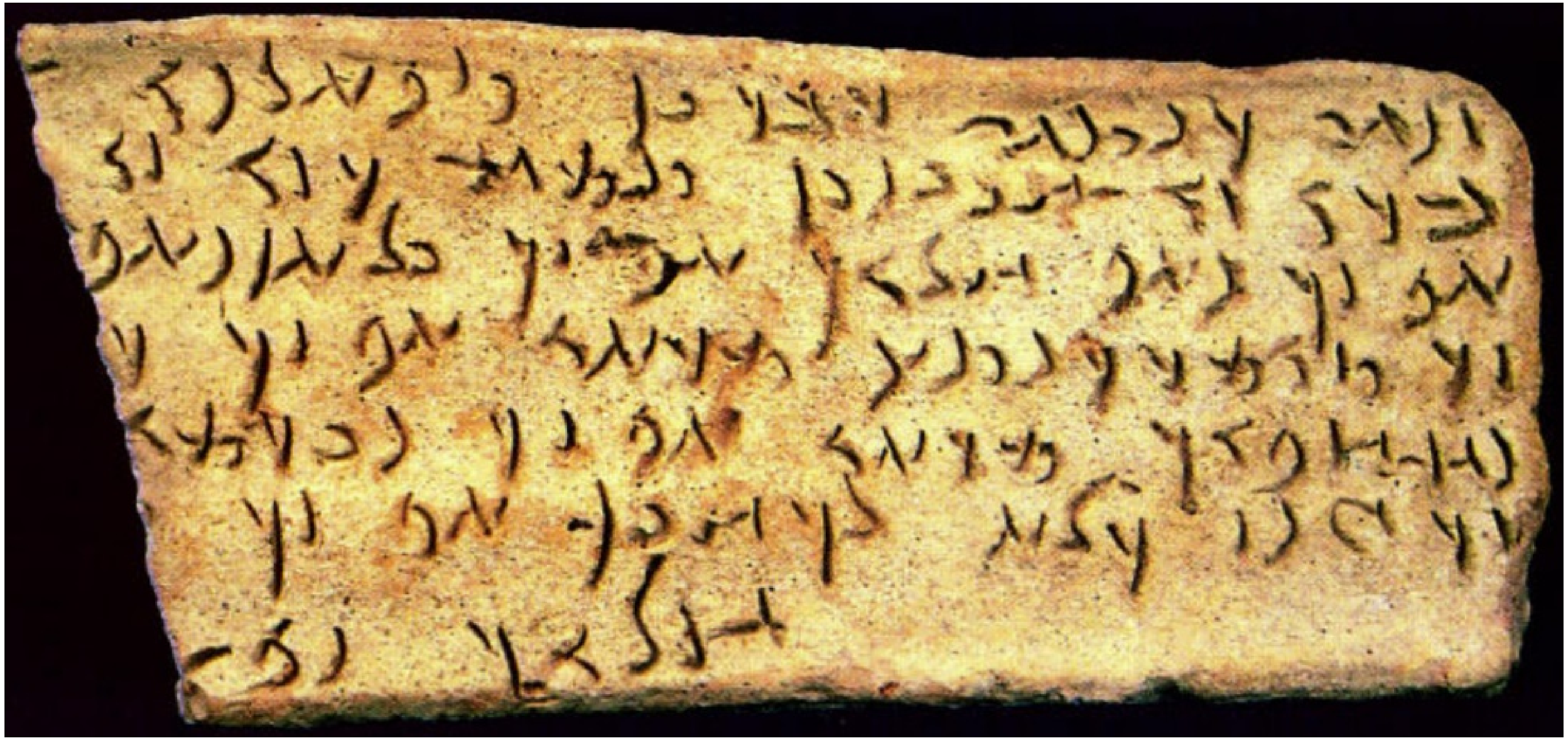
Sogdian sources

- Kultobe: the oldest Sogdian records are stone inscriptions found at Kultobe in modern Kazakhstan

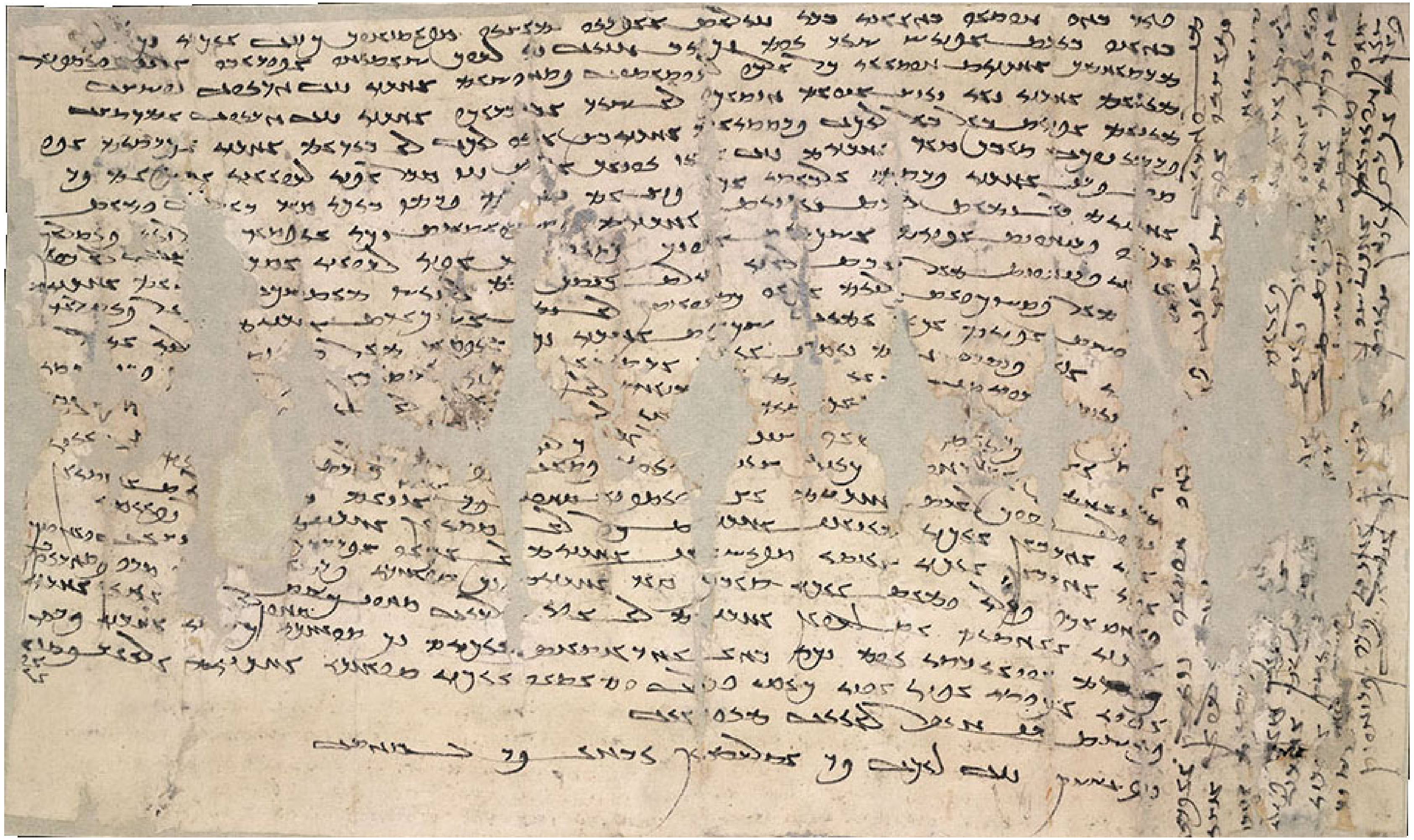
- 'Ancient Letters': the earliest attested Sogdian manuscripts are known as the 'Ancient Letters' (c.312 CE). These paper documents were found in 1907 by Aurel Stein in Dunhuang, western China

- 'Upper Indus Graffiti': appears on more than 600 rock carvings at Shatial and other sites in the Gilgit region of Pakistan (4th-7th c. CE)

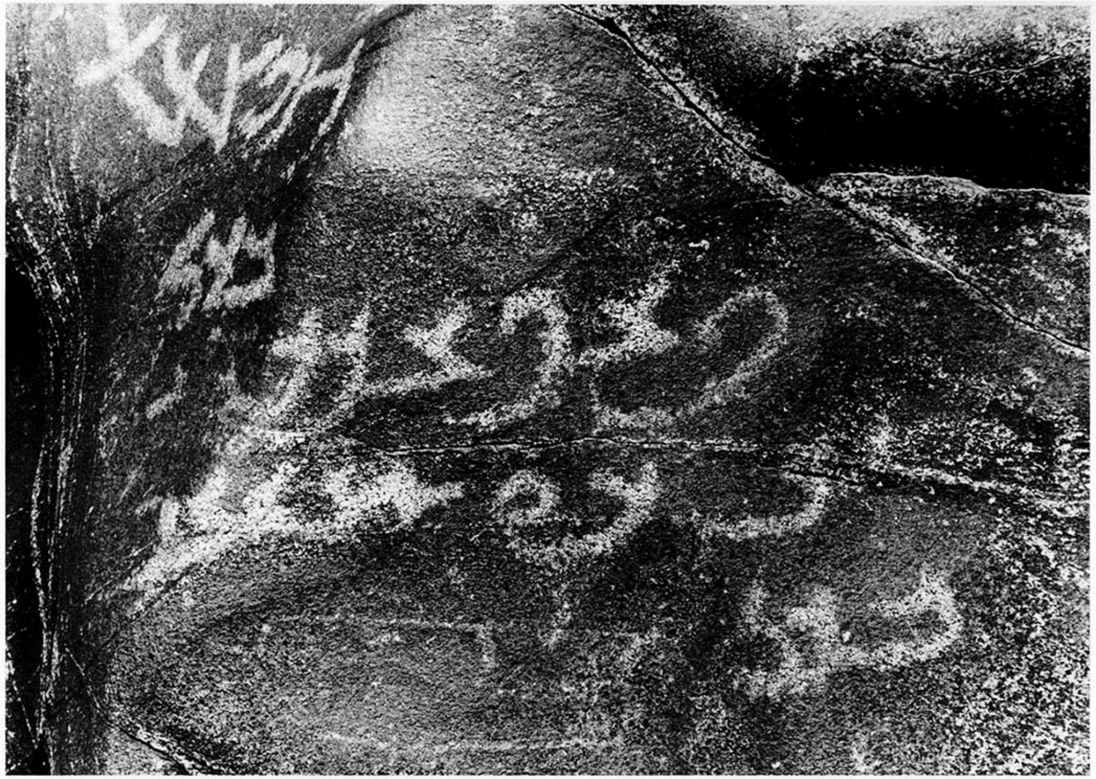
Old Sogdian: Research

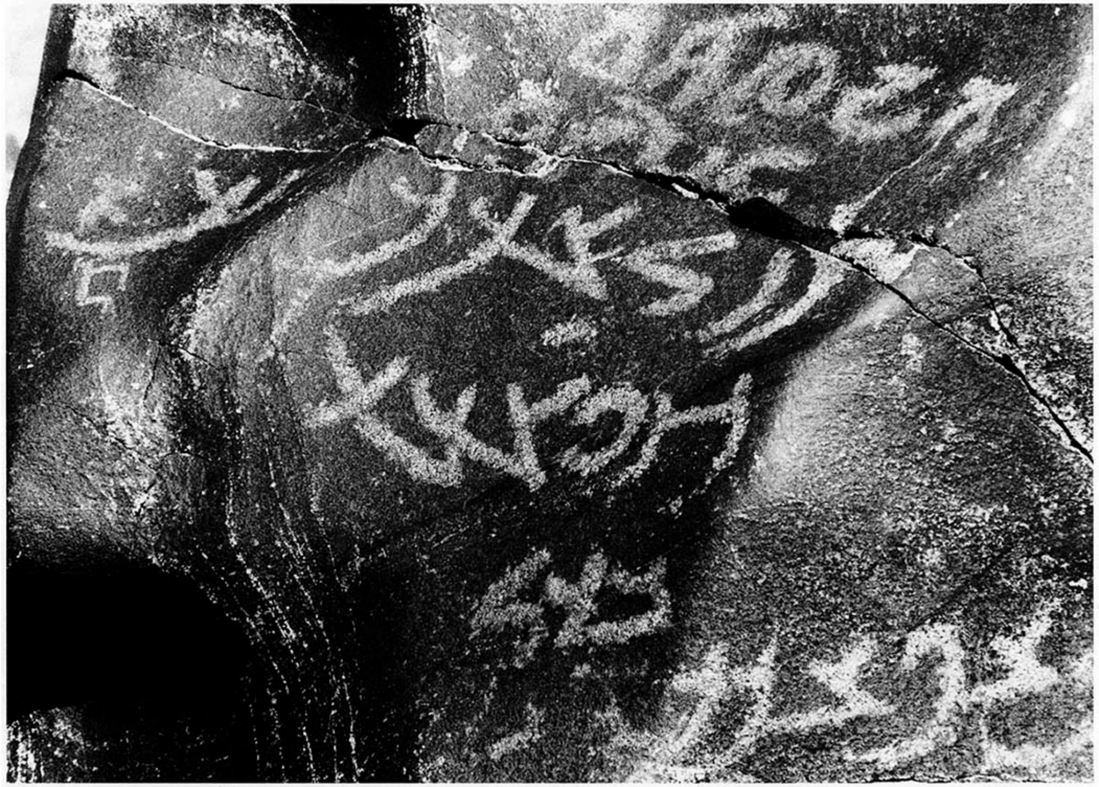


Old Sogdian: Research



Old Sogdian: Research





Script Encoding Process: ‘Script’

Define the ‘script’:

- Identify repertoire of distinctive signs and variants
- Establish representative glyphs
- Generate script grammar
- Develop encoding model
  - convert conventions for ‘writing’ into ‘typing’
  - translate qualitative data into technical metadata

Script Encoding Process: ‘Script’

Latin script for English: Distinctive signs

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Latin script for English: Variant signs

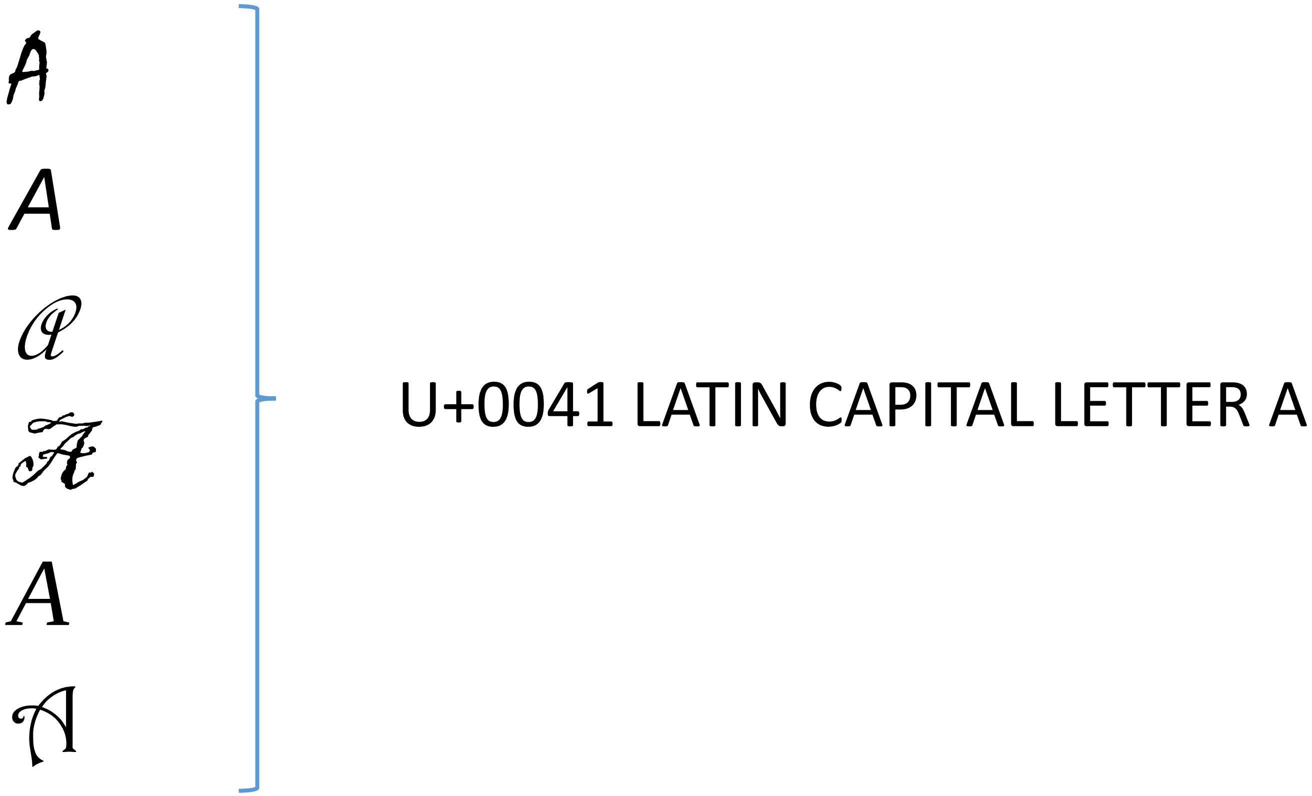
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Script Encoding: Unicode



A

A

A

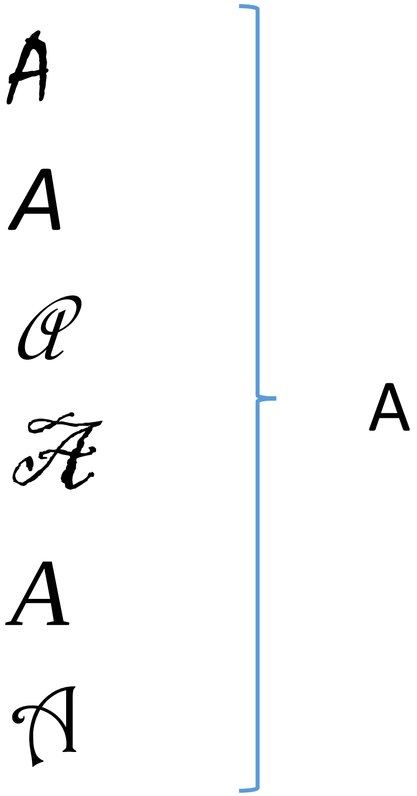
A

A

A

U+0041 LATIN CAPITAL LETTER A

Script Encoding: Unicode



Not-plain text

Plain text

Character-glyph model

Script Encoding Process: ‘Script’

- Is the ‘script’ a distinctive system or a stylistic variant?

- Character-glyph model:
  - encodes a ‘normative’ or representation form for a sign as a ‘character’
  - defines other forms as ‘glyphic’ variants
  - plain text vs. not-plain text / ‘rich text’ (font selection)

- “Unicode encodes characters, not glyphs”

Script Encoding Process: ‘Script’

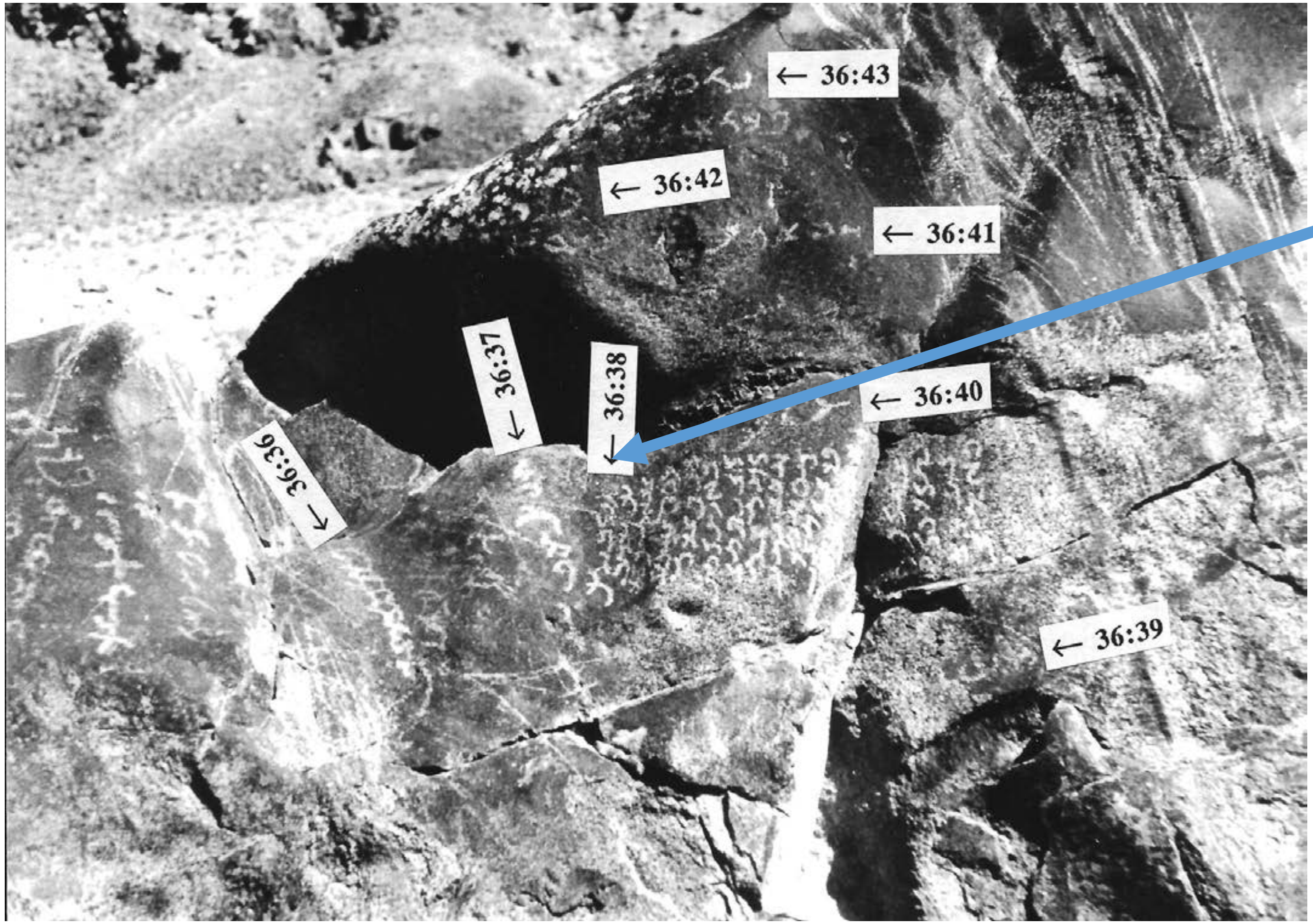
- A script may have historical, scribal, and stylistic variants

- Considerable differences may exist between early and late forms
  - Old Sogdian and Sogdian

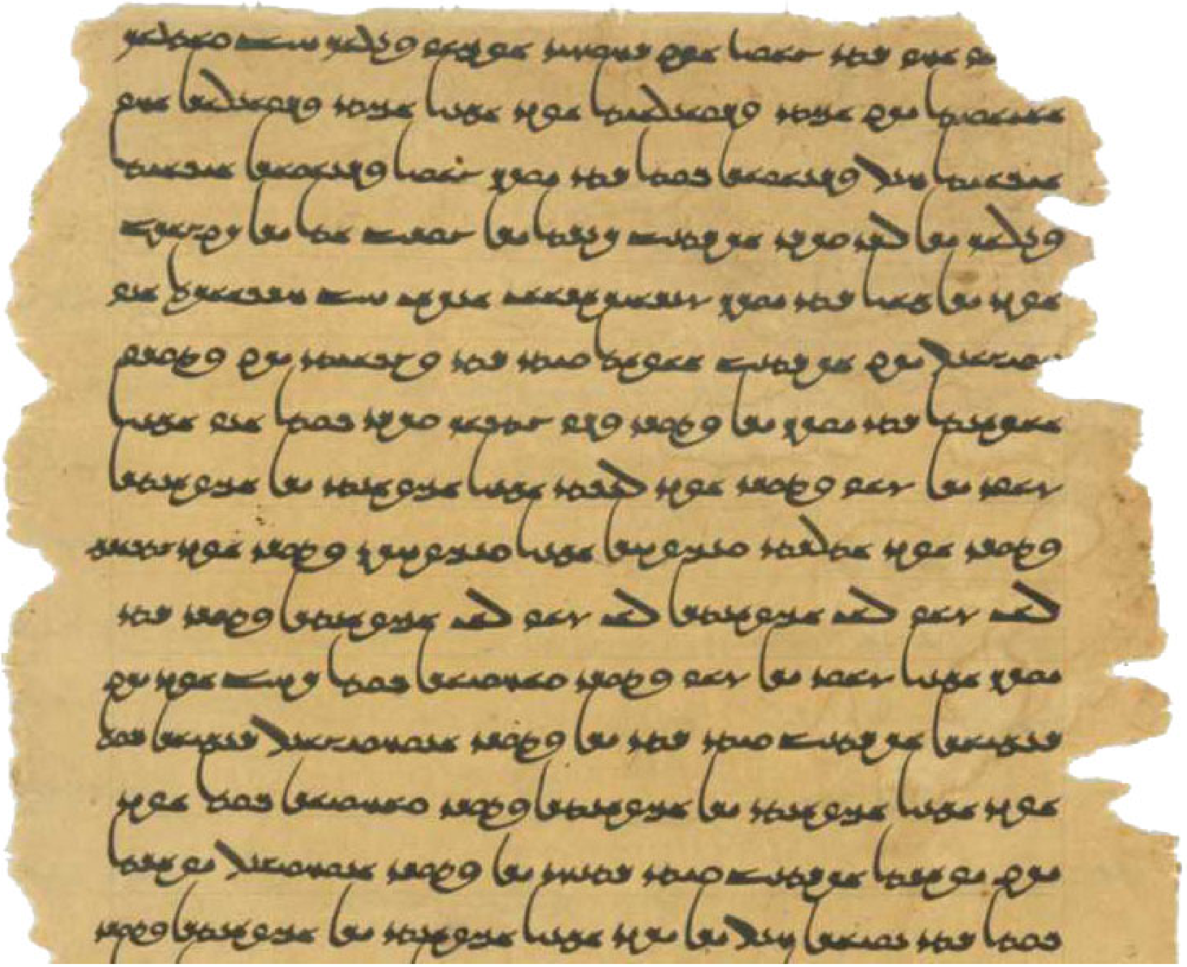
- Script boundaries unclear
  - Form of script in early phase may resemble an ancestor
  - Late form of script may resemble a descendant

- Graphically similar scripts may have structural differences

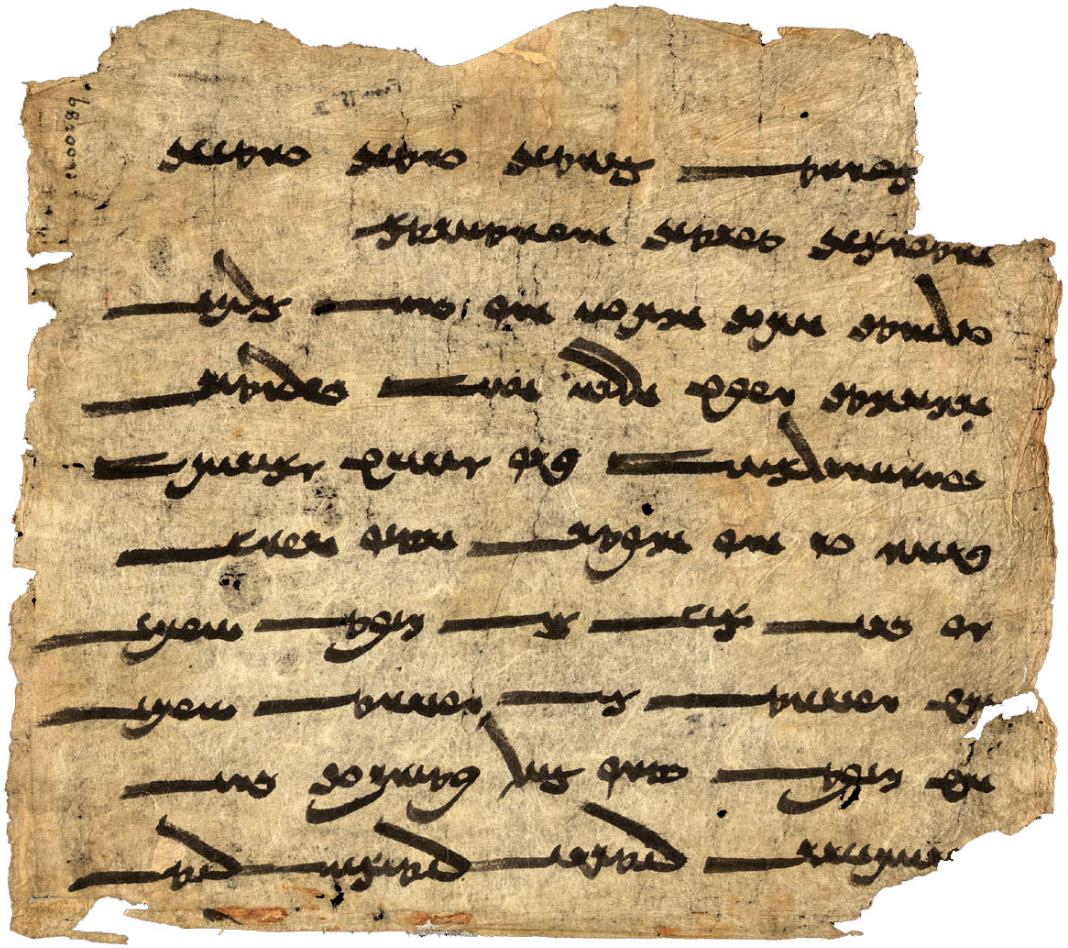
Developmental ‘Variant’: Vertical Old Sogdian



Developmental ‘Variant’: Formal Sogdian



Developmental ‘Variant’?: Cursive Sogdian



‘Script’ or ‘Variant’?: Chorasmian

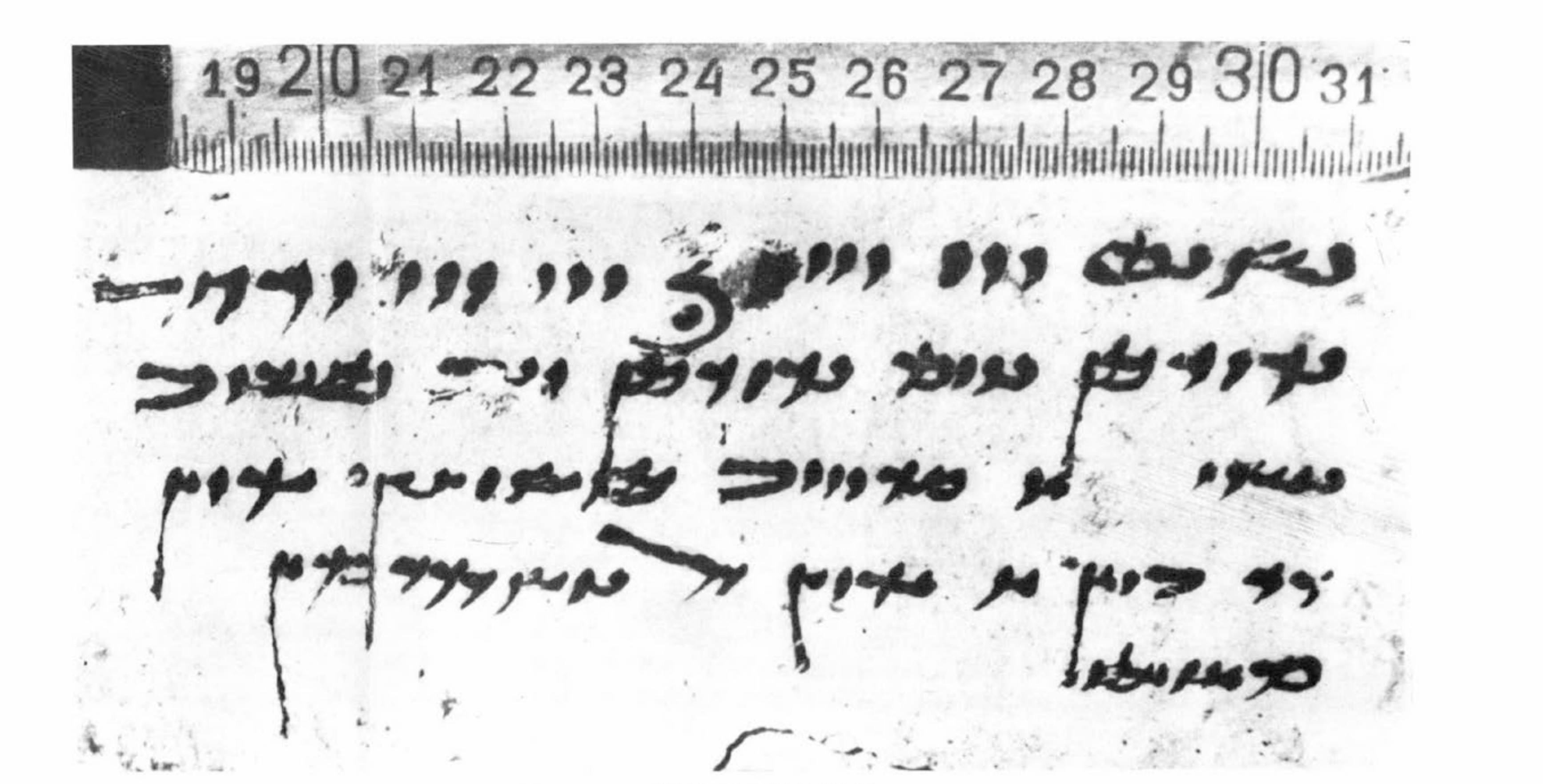
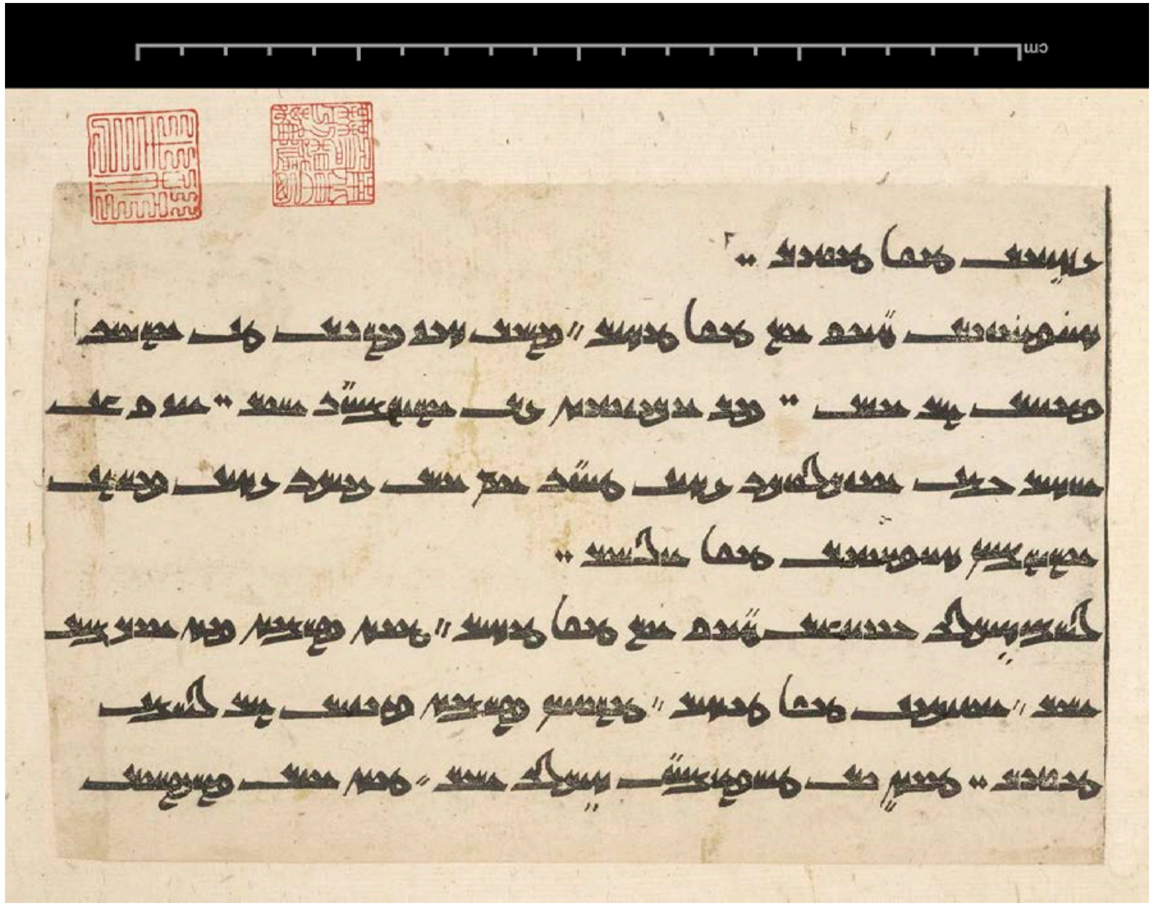
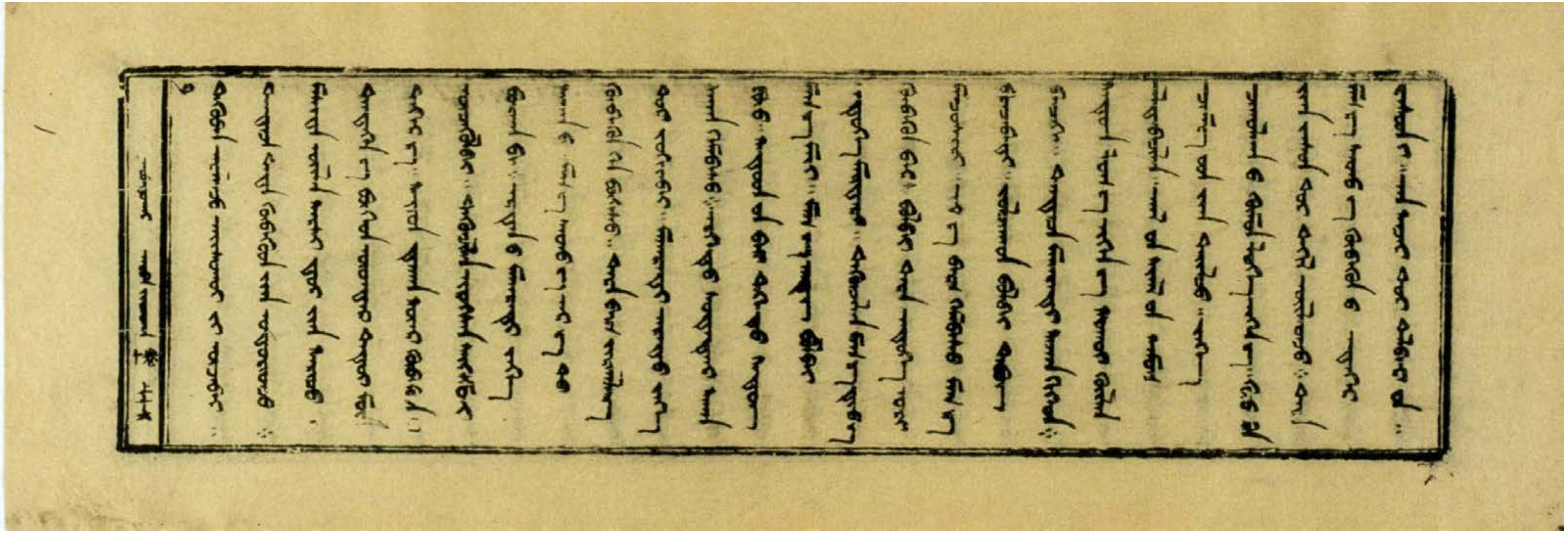


Fig. 2. Inscription No. 25.

‘Script’ or ‘Variant’?: Old Uyghur



‘Script’ or ‘Variant’: Mongolian



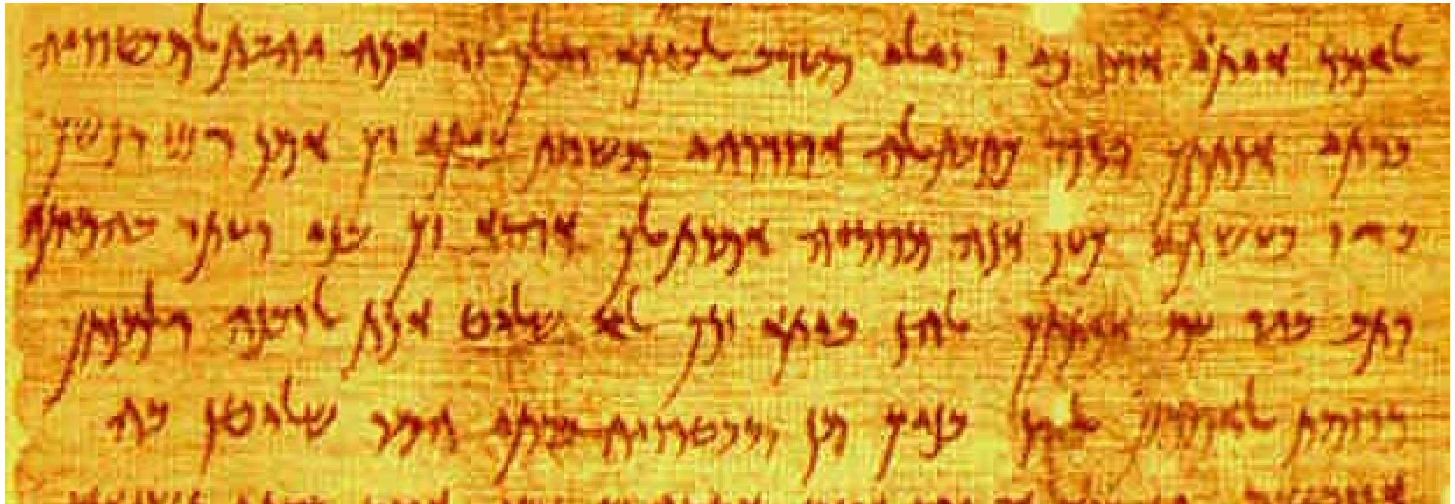
Old Sogdian: How to encode?

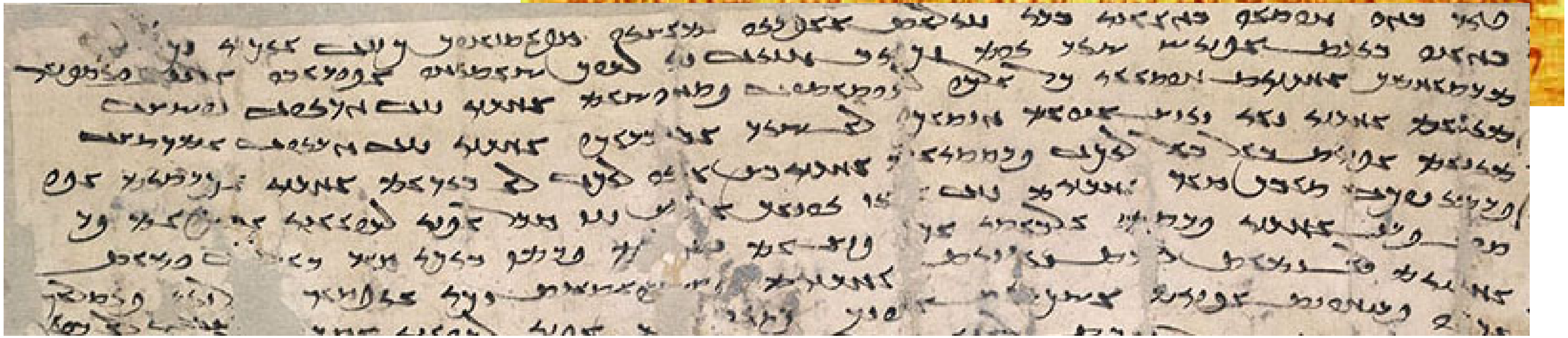
- Unification with Aramaic?
  - uses characters / code points from Imperial Aramaic block?
  - Palaeographically valid, but should Sogdian be considered a ‘style’ of Aramaic?

- Single encoding:
  - same character set for representing different forms of the script
  - usage of fonts to display different script styles?
  - but, which form is to be representative for plain text?

- Separate encoding:
  - two different character sets for historical stages of Sogdian
  - enables both scripts to be displayed simultaneously, without font changes

Aramaic & Old Sogdian

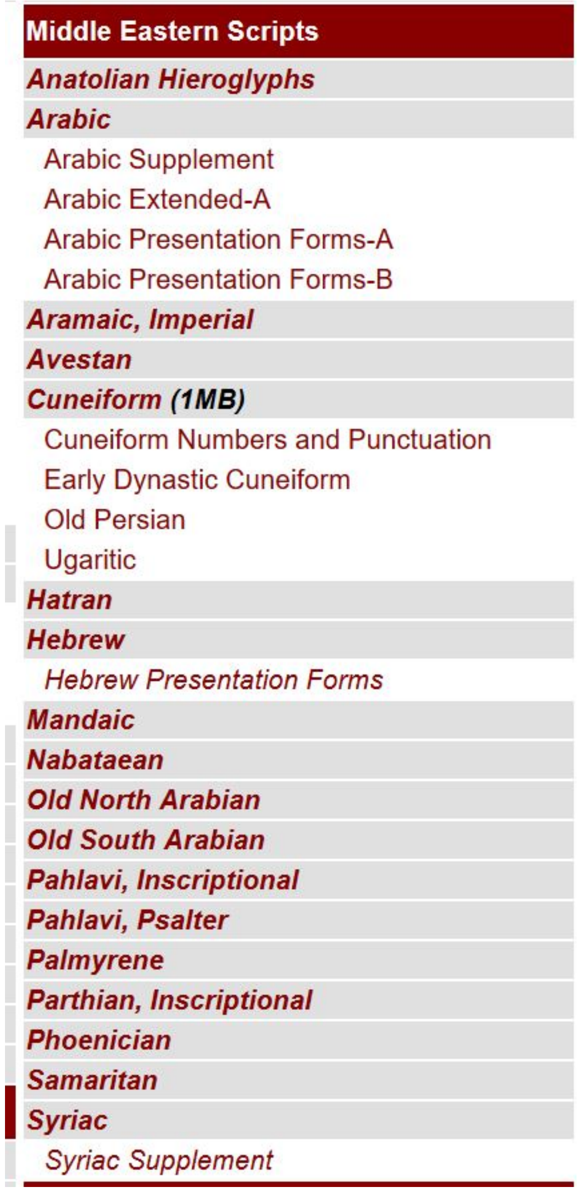




Imperial Aramaic

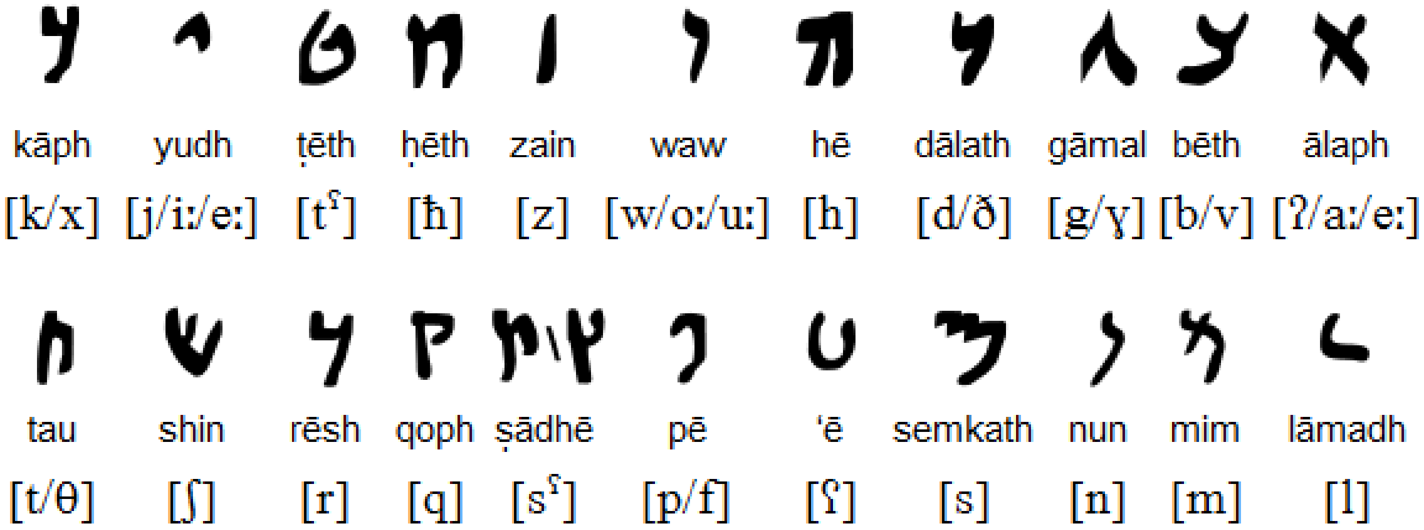
Central Asian Scripts

- Manichaean
- Marchen
- Mongolian
  - Mongolian Supplement
- Old Sogdian
- Old Turkic
- Phags-Pa
- Sogdian
- Soyombo
- Tibetan
- Zanabazar Square



Middle Eastern Scripts

- Anatolian Hieroglyphs
- Arabic
  - Arabic Supplement
  - Arabic Extended-A
  - Arabic Presentation Forms-A
  - Arabic Presentation Forms-B
- Aramaic, Imperial
- Avestan
- Cuneiform (1MB)
  - Cuneiform Numbers and Punctuation
  - Early Dynastic Cuneiform
  - Old Persian
  - Ugaritic
- Hatran
- Hebrew
  - Hebrew Presentation Forms
- Mandaic
- Nabataean
- Old North Arabian
- Old South Arabian
- Pahlavi, Inscriptional
- Pahlavi, Psalter
- Palmyrene
- Parthian, Inscriptional
- Phoenician
- Samaritan
- Syriac
  - Syriac Supplement



kāph	yudh	ṭēth	ḥēth	zain	waw	hē	dālath	gāmal	bēth	ālaph
[k/x]	[j/iː/eː]	[tˤ]	[ħ]	[z]	[w/oː/uː]	[h]	[d/ð]	[g/ɣ]	[b/v]	[ʔ/aː/eː]
tau	shin	rēsh	qoph	ṣādhē	pē	ʻē	semkath	nun	mim	lāmadh
[t/θ]	[ʃ]	[r]	[q]	[sˤ]	[p/f]	[ʕ]	[s]	[n]	[m]	[l]

Sogdian Scripts: Repertoire

	Aramaic	Old Sogdian	Sogdian
aleph	𐡀	𐼀, 𐼁	𐼰
beth	𐡁	𐼂, 𐼃	𐼱
gimel	𐡂	𐼄	𐼲
daleth	𐡃	𐼘	—
he	𐡄	𐼅, 𐼆	𐼳
waw	𐡅	𐼇	𐼴
zayin	𐡆	𐼈	𐼵
heth	𐡇	𐼉	𐼶
teth	𐡈	—	—
yodh	𐡉	𐼊	𐼷
kaph	𐡊	𐼋	𐼸

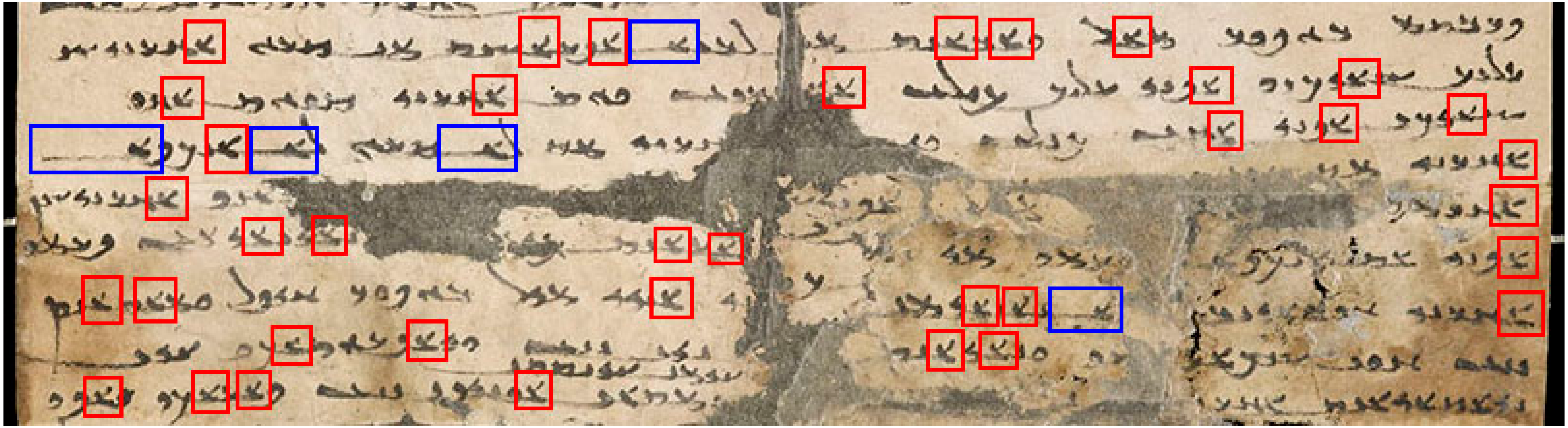
	Aramaic	Old Sogdian	Sogdian
lamedh	𐡋	𐼌	𐼹
mem	𐡌	𐼍	𐼺
nun	𐡍	𐼎, 𐼏, 𐼐	𐼻
samekh	𐡎	𐼑	𐼼
ayin	𐡏	𐼒, 𐼓, 𐼘	𐼽, 𐽀
pe	𐡐	𐼔	𐼾
sadhe	𐡑	𐼕, 𐼖, 𐼗	𐼿
qoph	𐡒	—	—
resh	𐡓	𐼘	𐽀
shin	𐡔	𐼙	𐽅, 𐽁
taw	𐡕	𐼚, 𐼛, 𐼜	𐽂

Encoding: Sogdian and Aramaic

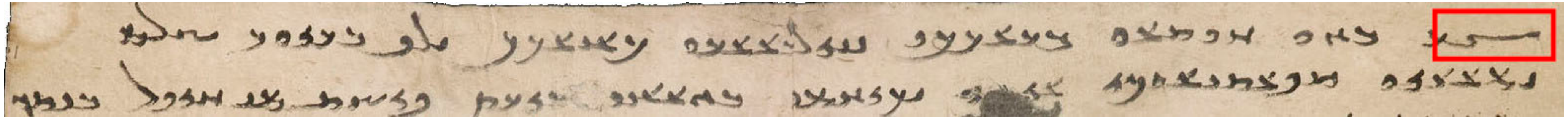
- Differences in repertoire: Sogdian
  - *Heth*, *teth*, *qoph* dropped
  - *Daleth*, *ayin*, *resh* merged
  - Special *ayin* used for saluations, eg. *`D* ‘to’
  - Scribal innovations, eg. tail extensions for final characters
  - Special numerical sign for 100

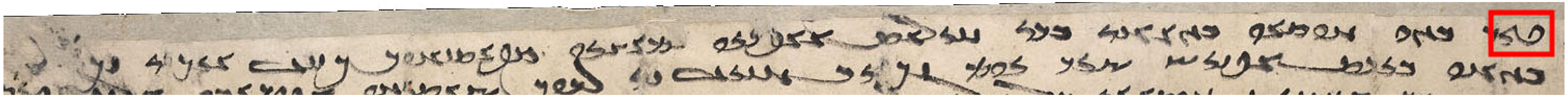
- Cultural and scribal divergences

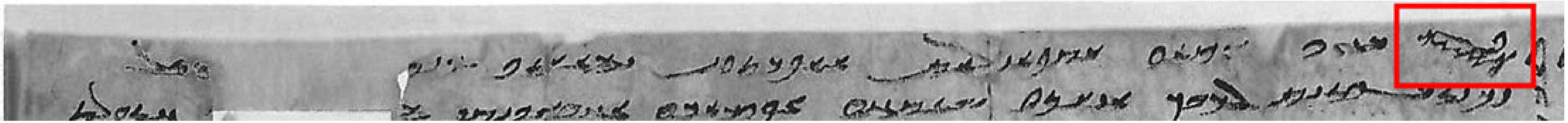
Old Sogdian: final letters: *aleph*



Old Sogdian: *ayin*







Conclusions

- Script encoding process
    - not art, not science, but a bit of both

- Skills to encode scripts:
    - interest in writing systems and willingness to learn
        - don’t need to be a linguist, typographer, or computer scientist
    - research and analytical skills
        - search, search, and search more, then dig deeper, and follow leads