Title: Preliminary Proposal to Encode the Old Sogdian Script in Unicode
Source: Script Encoding Initiative (SEI)
Author: Anshuman Pandey (anshuman.pandey@berkeley.edu)
Date: 2015-04-14

1 Introduction

This is a preliminary proposal to encode the 'Old Sogdian' script in Unicode. A discussion and typology of the various 'Sogdian' scripts and the requirements for encoding Old Sogdian and its descendants is provided in the forthcoming "Roadmap for Encoding Sogdian Scripts in Unicode".

This document provides a brief description of Old Sogdian, its character repertoire, and specimens of the script. The code points shown in the code chart and names list are based upon the current allocation for a 'Sogdian' script in the Roadmap to the SMP; they are only tentative and the assignments may change. The representative glyphs are also illustrative and are not intended to be normative or typographically aesthetic. Provisional character properties are also included.

The proposal author seeks feedback from scholars regarding the proposed characters and representative glyphs. Issues requiring further discussion are enumerated in section 6. The information presented here may be incomplete and may change as more information on the script is obtained. Research on Old Sogdian is ongoing and a formal proposal to encode it is forthcoming.

2 Background

The Old Sogdian script appears in manuscripts and inscriptions dated between the 4th to 7th centuries CE. The earliest manuscripts containing the script are known as 'Sogdian Ancient Letters' (see figures 1–5). These paper documents were found in 1907 by Aurel Stein in Dunhuang, now in Gansu province, western China. Based upon internal evidence it has been suggested that the 'Ancient Letters' were written in 312–313 CE (Sims-Williams 1985). A script similar to that used in the 'Ancient Letters' appear upon hundreds of rock carvings in the Gilgit region of Pakistan. The 'Upper Indus graffiti' have been dated to the 4th–7th centuries CE (Sims-Williams 1989, 2000; see figures 6, 7).

3 Script Details

3.1 Structure

Old Sogdian is an *abjad* that is written from right to left. Letters retain their basic shapes in different positions within a word, but a few letters have distinctive word-final shapes. As such, Old Sogdian is structurally a non-joining *abjad*, similar to Hebrew. The available sources show instances where letters are connected, but such conjunctions result from the regular flow of writing or from cursive practices rather than any intrinsic conjoining behavior of the script, as is the case with Arabic, Mongolian, etc. Similar to other *abjad* systems, vowels in Old Sogdian are represented using *aleph*, *yodh*, and *waw*.

3.2 Script name

There is no standard name for the script of the 'Ancient Letters'. The catalogue of the International Dunhuang Project at the British Library refers to it as "Sogdian" and does not differentiate between the varieties of the script grouped under this designation. Skjærvø (1996) refers to the script as "Sogdian Aramaic". The tentative identifier for the script block in Unicode is 'Old Sogdian'. This name was chosen because it differentiates the script from the 'Sogdian' script proper.

3.3 Character names

The names of Old Sogdian characters are tentatively based upon analogous Unicode names for letters of the 'Imperial Aramaic' block. The sort order is identical to the encoding order.

3.4 Character Repertoire

The repertoire of Old Sogdian letters is based upon that of Aramaic, but it contains 20 letters as opposed to the original 22 of the latter. It lacks distinctive letters that correspond to Aramaic *teth* and *qoph*. The actual number of distinctive letters may be fewer, considering that *daleth*, *ayin*, and *resh* may be represented using a single letter; the same may apply to *zayin* and *nun*.

Old Sogdian glyph shapes are quite uniform, as observed in the 'Ancient Letters'. The glyphs clearly reveal their Aramaic origins, but several changes in the shapes of letters are noticeable between the two scripts (see table 5). Some glyphs for Old Sogdian letters resemble those of Parthian letters, but they are used for different letters between the two.

Distinctive numerical signs are attested for one, ten, twenty, thirty, one hundred, and the fraction one-half. An Aramaic heterogram is used for the thousands. Ten thousand is expressed using number words.

Punctuation marks are not used in Old Sogdian.

4 Proposed Encoding

The proposed repertoire for Old Sogdian contains 31 characters: 20 basic letters, 1 logographic letter, 9 numbers, and 1 fraction.

4.1 Letters

Representative glyphs, names, and phonetic values of the proposed Sogdian letters are given below:

	Character name	Phonetic values
𼧠	OLD SOGDIAN LETTER ALEPH	a, ā
𼧡	OLD SOGDIAN LETTER BETH	β
𼧢	OLD SOGDIAN LETTER GIMEL	γ
𼧣	OLD SOGDIAN LETTER DALETH	—

𐼹	OLD SOGDIAN LETTER HE	Ø, a
𐼺	OLD SOGDIAN LETTER WAW	w, u, ū, o, ō
𐼻	OLD SOGDIAN LETTER ZAYIN	z, ž
𐼼	OLD SOGDIAN LETTER HETH	x
𐼽	OLD SOGDIAN LETTER YODH	y, i, ī, e, ē
𐼾	OLD SOGDIAN LETTER KAPH	k, g
𐼿	OLD SOGDIAN LETTER LAMEDH	δ, l
𐽀	OLD SOGDIAN LETTER MEM	m
𐽁	OLD SOGDIAN LETTER NUN	n
𐽂	OLD SOGDIAN LETTER SAMEKH	s
𐽃	OLD SOGDIAN LETTER AYIN	Ø, a
𐽄	OLD SOGDIAN LETTER PE	p, b, f
𐽅	OLD SOGDIAN LETTER SADHE	č, ǰ
𐽆	OLD SOGDIAN LETTER RESH	r
𐽇	OLD SOGDIAN LETTER SHIN	š
𐽈	OLD SOGDIAN LETTER TAW	t, d

The following logographic letter is proposed:

	Character name	Phonetic value
𐽉	OLD SOGDIAN LOGOGRAPHIC LETTER AYIN	—

The letter 𐽃 DALETH is used only in Aramaic heterograms.

The letter 𐼹 HE is used only at the end of words.

The letter 𐼼 HETH has the variant form 𐼸 (see figure 9).

The letter 𐽉 LOGOGRAPHIC AYIN is used only in the Aramaic heterogram 𐽃𐽉 *OD* <𐽉 LOGOGRAPHIC LETTER AYIN, 𐽃 DALETH>, a salutation meaning “to” that is used in correspondence.

Long vowels The long vowel *ā* is sometimes representing by two instances of 𐼰 ALEPH.

Attrition Old Sogdian lacks letters that correspond to 𐡶 U+10848 ARAMAIC LETTER TETH and 𐡸 U+10852 ARAMAIC LETTER QOPH. It is likely that Old Sogdian *teth* is not attested. Some scholars proposed that *qoph*

was retained and reassigned for the number 100 (Sims-Williams 1985); however, the general view is that *qoph* was never used in Old Sogdian.

Homoglyphic letters The following sets of letters appear to be highly similar, if not identical:

- 𐼴 DALETH, 𐼳 AYIN, and 𐼵 RESH Sims-Williams noted that these three letters are identical in the available sources. It may, then, be practical to fold them into a single character *DALETH-AYIN-RESH instead of encoding the letters separately. Such folding occurs in Inscriptional Pahlavi, a related Aramaic-based script, where *waw*, *ayin*, and *resh* have the same glyph. In the Unicode block for 'Inscriptional Pahlavi' these three are unified in the character 𐭲 U+10B65 INSCRIPTIONAL PAHLAVI LETTER WAW-AYIN-RESH. A similar case occurs with *mem* and *qoph*, which are unified in the character 𐭬 U+10B6C INSCRIPTIONAL PAHLAVI LETTER MEM-QOPH.
- 𐼶 ZAYIN and 𐼷 NUN The distinction between these two is observed at the end of a word, where NUN is written with a vertically elongated tail. If the glyphs for ZAYIN and NUN are truly undistinguished in Old Sogdian, it may be practical to fold them into a single character *ZAYIN-NUN. However, if NUN and ZAYIN are distinguished when they occur at the ends of words, then they should be retained as separately characters.

Word-final shapes In the 'Ancient Letters' certain letters are written as what appears to be distinct forms when they occur in word-final position:

	Word-final	Regular
HE	𐼨	𐼩
NUN	𐼷	𐼶
SAMEKH	𐼻	𐼺
TAW	𐽆	𐼽

The usage of these forms is not uniform. For instance, both 𐼺 and 𐼻 are observed in final position for SAMEKH in adjacent words. The strokes that distinguish word-final forms from regular letters may be interpreted as swash features. But, there is an indication that elongations of strokes occur with certain letters in final position, such as those described above, and not with others. The requirement to distinguish between regular and word-final forms at the character level, as is the case for Hebrew in Unicode, is unclear at present.

4.2 Numbers

Distinctive signs for the numbers 1, 10, and 100 are attested in Old Sogdian manuscripts. Other numbers are produced using repetitions of these signs (specimens shown in figure 8). The following numerical characters are proposed for encoding:

	Character name	Numeric value
𐼶	OLD SOGDIAN NUMBER ONE	1
𐼶𐼶	OLD SOGDIAN NUMBER TWO	2
𐼶𐼶𐼶	OLD SOGDIAN NUMBER THREE	3

𐼳	OLD SOGDIAN NUMBER FOUR	4
𐼴	OLD SOGDIAN NUMBER TEN	10
𐼵	OLD SOGDIAN NUMBER TWENTY	20
𐼶	OLD SOGDIAN NUMBER THIRTY	30
𐼷	OLD SOGDIAN NUMBER ONE HUNDRED	100
𐼸	OLD SOGDIAN NUMBER ONE THOUSAND	1000

The ordering of numbers follows the right-to-left directionality of the script. The expression of numbers is additive. Composite numbers of different decimal orders are written by placing the larger numbers first.

4.2.1 Primary units

The numbers 2–9 are produced using repetitions of the number 𐼱 ONE. These sequences are generally joined together at the left edge, eg. '3' 𐼲. Sequences for the numbers 4–9 are arranged into groups of three or four separated by a space, eg. '6' 𐼲𐼲 𐼲; '7' 𐼲𐼲 𐼲 (see figure 8 for examples).

The pattern of expressing the primary units is borrowed from Aramaic, ie. the number 3 𐡻 is a repetition of 1 𐡹. For the Aramaic block, the numbers 1–3 are encoded as separate characters: 𐡻 U+1085A IMPERIAL ARAMAIC NUMBER THREE. The Parthian encoding follows the same approach and has separate characters for numbers 1–4, eg. 𐭻 U+10B5B INSCRIPTIONAL PARTHIAN NUMBER FOUR. This encoding of separate characters for the numbers 2–3 in Aramaic and 2–4 in Parthian appears to have been motivated by the principle of grouping.

In order to facilitate the linking of sequences of 𐼱 ONE and the grouping of linked sequences, a model similar to Imperial Aramaic and Parthian is adopted for Old Sogdian. This approach requires the encoding of the following additional characters: 𐼰 TWO, 𐼲 THREE, 𐼳 FOUR.

4.2.2 Tens

The number 𐼴 TEN resembles the letter 𐼍 LAMEDH, but it is condensed. The numbers 𐼵 TWENTY and 𐼶 THIRTY are produced by stacking two and three instances, respectively, of the number 𐼴 TEN. The available sources show one instance of the number 30 represented as 𐼴𐼵 <𐼵 TWENTY, 𐼴 TEN>; it occurs in the number 𐼲𐼴𐼵 '32' (see figure 8).

4.2.3 Hundreds

The number 100 is written using 𐼷 ONE HUNDRED. The hundred units are represented using primary numbers followed by 𐼷 ONE HUNDRED, eg. 500 is expressed as 𐼷𐼰𐼲 <𐼲 THREE, 𐼰 TWO, 𐼷 ONE HUNDRED>.

4.2.4 Thousands

There number 1,000 is written using 𐼸 ONE THOUSAND. This numerical sign is a ligature consisting of the number 𐼱 ONE joined to the sequence 𐼍𐼺 <𐼍 LAMEDH, 𐼺 PE>. The word 𐼍𐼺 represents the Aramaic

heterogram *LP* "thousand" (= 𐼽 U+1085A IMPERIAL ARAMAIC NUMBER ONE THOUSAND). Although 𐼭 can be represented using sequences of letters, it is encoded as a character because of its specific numerical properities and because it occurs in conjunction with other numbers. The form 𐼭 is interpreted atomically.

The thousands units are written using the appropriate repetitions of 𐼁 '1' followed by 𐼭 ONE THOUSAND, eg. 𐼁𐼭 '2,000'. The presence of 𐼁 ONE in the glyph for ONE THOUSAND does not have a separate value.

4.2.5 Ten thousands

There is no distinctive sign for the number 10,000. In the available sources, this number is represented using the word 𐼴𐼹𐼺𐼴 *βrewar*. It occurs, for example, in 'Ancient Letter 2", line 1:

4.3 Fractions

The following fraction is attested:

	Character name	Numeric value
𐼰	OLD SOGDIAN FRACTION ONE HALF	½

The 𐼰 FRACTION ONE HALF occurs in 'Ancient Letter 5'. This character was identified as an alternate form for 𐼷 ONE HUNDRED by early scholars of Sogdian, but Grenet, et al (1998) have interpreted it as a sign for the fraction ½.

5 Character Data

Character Properties Properties in the format of `UnicodeData.txt`:

```
xx00;OLD SOGDIAN LETTER ALEPH;Lo;0;R;;;;N;;;;;
xx01;OLD SOGDIAN LETTER BETH;Lo;0;R;;;;N;;;;;
xx02;OLD SOGDIAN LETTER GIMEL;Lo;0;R;;;;N;;;;;
xx03;OLD SOGDIAN LETTER DALETH;Lo;0;R;;;;N;;;;;
xx04;OLD SOGDIAN LETTER HE;Lo;0;R;;;;N;;;;;
xx05;OLD SOGDIAN LETTER WAW;Lo;0;R;;;;N;;;;;
xx06;OLD SOGDIAN LETTER ZAYIN;Lo;0;R;;;;N;;;;;
xx07;OLD SOGDIAN LETTER HETH;Lo;0;R;;;;N;;;;;
xx08;OLD SOGDIAN LETTER YODH;Lo;0;R;;;;N;;;;;
xx09;OLD SOGDIAN LETTER KAPH;Lo;0;R;;;;N;;;;;
xx0A;OLD SOGDIAN LETTER LAMEDH;Lo;0;R;;;;N;;;;;
xx0B;OLD SOGDIAN LETTER MEM;Lo;0;R;;;;N;;;;;
xx0C;OLD SOGDIAN LETTER NUN;Lo;0;R;;;;N;;;;;
xx0D;OLD SOGDIAN LETTER SAMEKH;Lo;0;R;;;;N;;;;;
xx0E;OLD SOGDIAN LETTER AYIN;Lo;0;R;;;;N;;;;;
xx0F;OLD SOGDIAN LETTER PE;Lo;0;R;;;;N;;;;;
xx10;OLD SOGDIAN LETTER SADHE;Lo;0;R;;;;N;;;;;
xx11;OLD SOGDIAN LETTER RESH;Lo;0;R;;;;N;;;;;
xx12;OLD SOGDIAN LETTER SHIN;Lo;0;R;;;;N;;;;;
```

```
xx13;OLD SOGDIAN LETTER TAW;Lo;0;R;;;;N;;;;;
xx14;OLD SOGDIAN LOGOGRAPHIC LETTER AYIN;Lo;0;R;;;;;N;;;;;
xx15;OLD SOGDIAN NUMBER ONE;No;0;R;;;;1;N;;;;;
xx16;OLD SOGDIAN NUMBER TWO;No;0;R;;;;2;N;;;;;
xx17;OLD SOGDIAN NUMBER THREE;No;0;R;;;;3;N;;;;;
xx18;OLD SOGDIAN NUMBER FOUR;No;0;R;;;;4;N;;;;;
xx19;OLD SOGDIAN NUMBER TEN;No;0;R;;;;10;N;;;;;
xx1A;OLD SOGDIAN NUMBER TWENTY;No;0;R;;;;20;N;;;;;
xx1B;OLD SOGDIAN NUMBER THIRTY;No;0;R;;;;30;N;;;;;
xx1C;OLD SOGDIAN NUMBER ONE HUNDRED;No;0;R;;;;100;N;;;;;
xx1D;OLD SOGDIAN NUMBER ONE THOUSAND;No;0;R;;;;1000;N;;;;;
xx1E;OLD SOGDIAN FRACTION ONE HALF;;No;0;R;;;;1/2;N;;;;;
```

Linebreaking Linebreaking properties in the format of `LineBreak.txt`:

```
xx00..xx13;AL     # Lo    [20] OLD SOGDIAN LETTER ALEPH..OLD SOGDIAN LETTER TAW
xx14;AL           # Lo         OLD SOGDIAN LOGOGRAPHIC LETTER AYIN
xx15..xx1E;AL     # No    [10] OLD SOGDIAN NUMBER ONE..OLD SOGDIAN FRACTION ONE HALF
```

6 Questions

1. *Script name* Is ‘Old Sogdian’ an acceptable name for the script block in Unicode? Another option is ‘Early Sogdian’. If the script’s identity with Aramaic should be emphasized, then ‘Sogdian Aramaic’ is yet another possibility.

2. *ZAYIN and NUN* Are distinctive shapes for 𐼩 ZAYIN and 𐼨 NUN attested? If not, then should they be unified into the single character ZAYIN-NUN? The sources show NUN written as the long stroke 𐼨 in word-final position. Does ZAYIN have similar behavior?

3. *AYIN, DALETH, and RESH* Are distinctive shapes for these three letters attested? Should they be encoded separately or as a single character AYIN-DALETH-RESH?

4. *TETH* Is an Old Sogdian analogue for Aramaic TETH attested?

5. *QOPH* Is an Old Sogdian analogue for Aramaic QOPH attested?

6. *Final forms* Should final forms be encoded separately, particularly 𐼫 for 𐼪 HE and 𐼨 for 𐼨 NUN?

7 References

Everson, Michael. 2001. “Roadmapping early Semitic scripts” (L2/01-024). http://www.unicode.org/L2/L2001/01024-n2311.pdf

Grenet, Frantz; Nicholas Sims-Williams; Étienne de La Vaissière. 1998. “The Sogdian Ancient Letter V”. *Bulletin of the Asia Institute*, Alexander’s Legacy in the East: Studies in Honor of Paul Bernard, New series, vol. 12, edited by Osmund Bopearachchi, Carol Altman Bromberg, and Frantz Grenet, pp. 91–104.

Sims-Williams, Nicholas. 1985. “Ancient Letters”. *Encyclopædia Iranica*, vol. II, fasc. 1, pp. 7–9. http://www.iranicaonline.org/articles/ancient-letters

———. 1989. *Sogdian and Other Iranian Inscriptions of the Upper Indus*. Corpus Inscriptionum Iranicarum, pt. II (Inscriptions of the Seleucid and Parthian Periods and of Eastern Iran and Central Asia),

v. III (Sogdian), no. I. London: Published on behalf of Corpus Inscriptionum Iranicarum by School of Oriental and African Studies.

———. 2000. "The Iranian Inscriptions of Shatial". *Indologica Taurinensia*, Professor Gregory M. Bongard-Levin Felicitation Volume, v. 23–24, pp. 523–541.

Skjærvø, Prods Oktor. 1996. "Aramaic Scripts for Iranian Languages." *The World's Writing Systems*, edited by Peter T. Daniels and W. Bright, pp. 515–535. New York and Oxford: Oxford University Press.

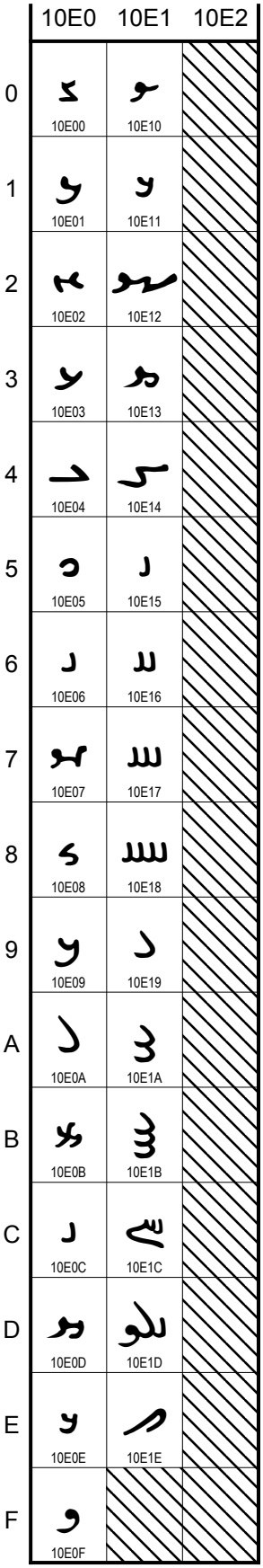
———. 2006. "Iran. VI. Iranian Languages and Scripts. (3) Writing Systems." *Encyclopædia Iranica*, vol. XIII, fasc. 4, pp. 366–370. http://www.iranicaonline.org/articles/iran-vi3-writing-systems

Waugh, Daniel C. [comp]. 2004. "The Sogdian Ancient Letters", translated by Nicholas Sims-Williams. https://depts.washington.edu/silkroad/texts/sogdlet.html

8 Acknowledgments

I am grateful to Nicholas Sims-Williams (SOAS, University of London) for providing detailed comments on am earlier draft of this document. I am also thankful to Roozbeh Pournader (Google, San Francisco) for discussing the encodings of Aramaic-based Iranian scripts in Unicode and also for his feedback on an earlier draft of this proposal.

This project was made possible in part through a Google Research Award, granted to Deborah Anderson for the Script Encoding Initiative at the University of California, Berkeley, which in turn is funding the post-doctoral research position held by the author in the Department of Linguistics at Berkeley.



Letters

10E00 OLD SOGDIAN LETTER ALEPH
10E01 OLD SOGDIAN LETTER BETH
10E02 OLD SOGDIAN LETTER GIMEL
10E03 OLD SOGDIAN LETTER DALETH
10E04 OLD SOGDIAN LETTER HE
10E05 OLD SOGDIAN LETTER WAW
10E06 OLD SOGDIAN LETTER ZAYIN
10E07 OLD SOGDIAN LETTER HETH
10E08 OLD SOGDIAN LETTER YODH
10E09 OLD SOGDIAN LETTER KAPH
10E0A OLD SOGDIAN LETTER LAMEDH
10E0B OLD SOGDIAN LETTER MEM
10E0C OLD SOGDIAN LETTER NUN
10E0D OLD SOGDIAN LETTER SAMEKH
10E0E OLD SOGDIAN LETTER AYIN
10E0F OLD SOGDIAN LETTER PE
10E10 OLD SOGDIAN LETTER SADHE
10E11 OLD SOGDIAN LETTER RESH
10E12 OLD SOGDIAN LETTER SHIN
10E13 OLD SOGDIAN LETTER TAW

Logographic letters

10E14 OLD SOGDIAN LOGOGRAPHIC LETTER AYIN

Numbers

10E15 OLD SOGDIAN NUMBER ONE
10E16 OLD SOGDIAN NUMBER TWO
10E17 OLD SOGDIAN NUMBER THREE
10E18 OLD SOGDIAN NUMBER FOUR
10E19 OLD SOGDIAN NUMBER TEN
10E1A OLD SOGDIAN NUMBER TWENTY
10E1B OLD SOGDIAN NUMBER THIRTY
10E1C OLD SOGDIAN NUMBER ONE HUNDRED
10E1D OLD SOGDIAN NUMBER ONE THOUSAND

Fraction

10E1E OLD SOGDIAN FRACTION ONE HALF

	Old Sogdian	Inscriptional Pahlavi	Inscriptional Parthian	Imperial Aramaic
ALEPH	𐼀	𐭠	𐭰	𐡀
BETH	𐼁	𐭡	𐭱	𐡁
GIMEL	𐼂	𐭢	𐭲	𐡂
DALETH	𐼃	𐭣	𐭳	𐡃
HE	𐼄	𐭤	𐭴	𐡄
WAW	𐼅	𐭥	𐭵	𐡅
ZAYIN	𐼆	𐭦	𐭶	𐡆
HETH	𐼇	𐭧	𐭷	𐡇
TETH	—	𐭨	𐭸	𐡈
YODH	𐼈	𐭩	𐭹	𐡉
KAPH	𐼉	𐭪	𐭺	𐡊
LAMEDH	𐼊	𐭫	𐭻	𐡋
MEM	𐼋	𐭬	𐭼	𐡌
NUN	𐼌	𐭭	𐭽	𐡍
SAMEKH	𐼍	𐭮	𐭾	𐡎
AYIN	𐼎	*	𐭿	𐡏
PE	𐼏	𐭯	𐮀	𐡐
SADHE	𐼐	𐭰	𐮁	𐡑
QOPH	—	*	𐮂	𐡒
RESH	𐼑	*	𐮃	𐡓
SHIN	𐼒	𐭲	𐮄	𐡔
TAW	𐼓	𐭳	𐮅	𐡕

Table 5: Comparison of Sogdian and other Iranian scripts derived from Aramaic. In Inscriptional Pahlavi, the representations of *ayin* and *resh* are identical to *waw*, and *qoph* is identical to *mem*. Parentheses indicate that a distinct character for that letter does not exist in the given script, but may be represented using other characters.

	Old Sogdian	Inscriptional Pahlavi	Inscriptional Parthian	Imperial Aramaic
ONE	𐼧	𐭠	𐭡	𐡠
TWO	𐼨	𐭡	𐭢	𐡡
THREE	𐼩	𐭢	𐭣	𐡢
FOUR	𐼪	𐭣	𐭤	—
TEN	𐼫	𐭩	𐭪	𐡩
TWENTY	𐼬	𐭪	𐭫	𐡪
THIRTY	𐼭	—	—	—
ONE HUNDRED	𐼮	𐭫	𐭬	𐡫
ONE THOUSAND	𐼯	𐭬	𐭭	𐡬
TEN THOUSAND	—	—	—	𐡭
ONE HALF	𐼰	—	—	—

Table 6: Comparison of numerical characters proposed for Old Sogdian and those encoded in Unicode blocks for Iranian scripts derived from Aramaic. Note: There is no distinctive numerical sign for 10,000 in Old Sogdian, instead this number is represented using the word 𐼴𐼯𐼺𐼴 *βrewar*.

Figure 1: Fragment of 'Ancient Letter 1' (British Library, International Dunhuang Project: Or. 8212/92.1 recto 1). "From her daughter, the free-woman Miwnay, to her d[ear] mother [Chatis]." (translation by Sims-Williams in Waugh 2004).

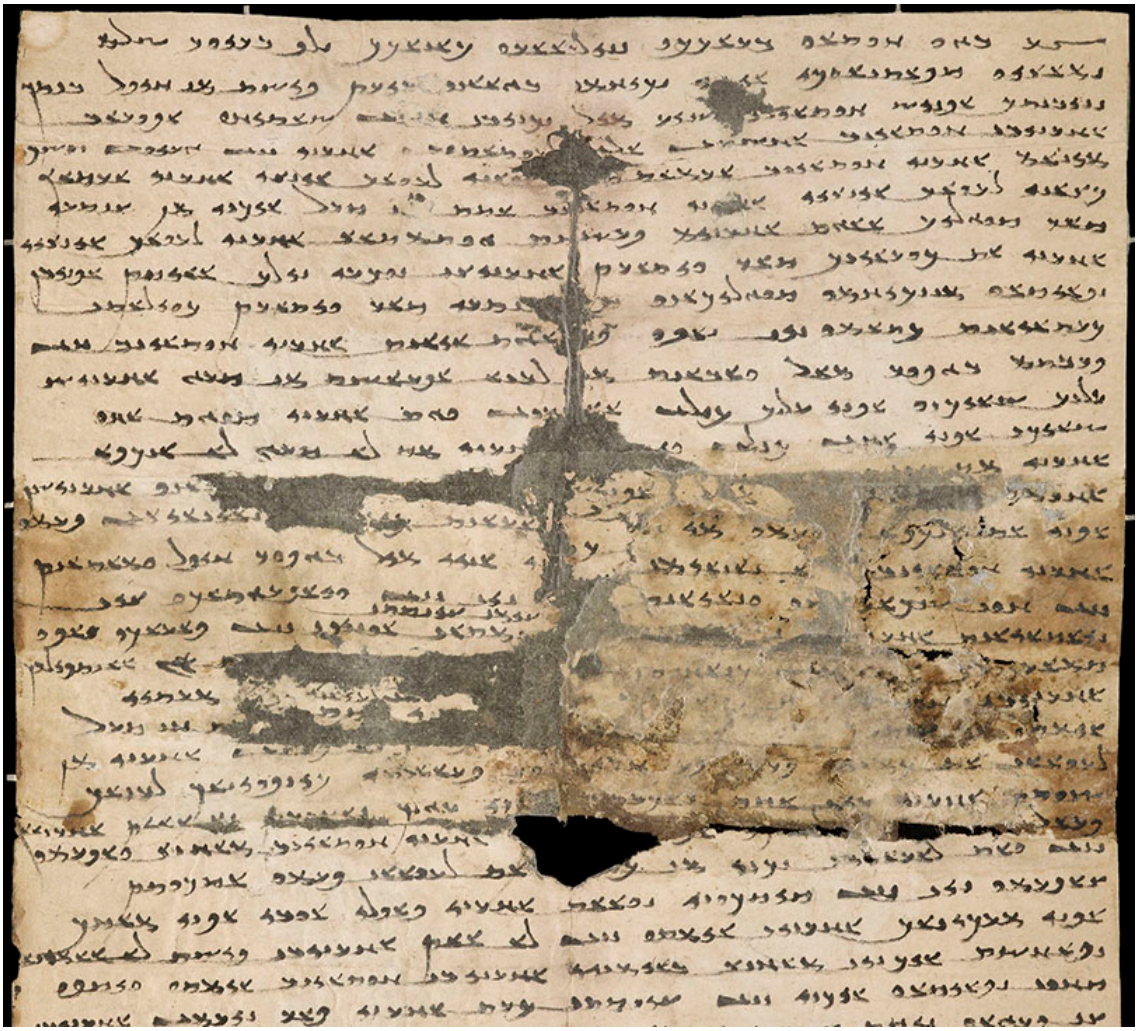


Figure 2: Fragment of 'Ancient Letter 2' (British Library, International Dunhuang Project: Or. 8212/95 side a). "To the noble lord Varzakk (son of) Nanai-thvar (of the family) Kanakk. Sent [by] his servant Nanai-vandak." (translation by Sims-Williams in Waugh 2004).

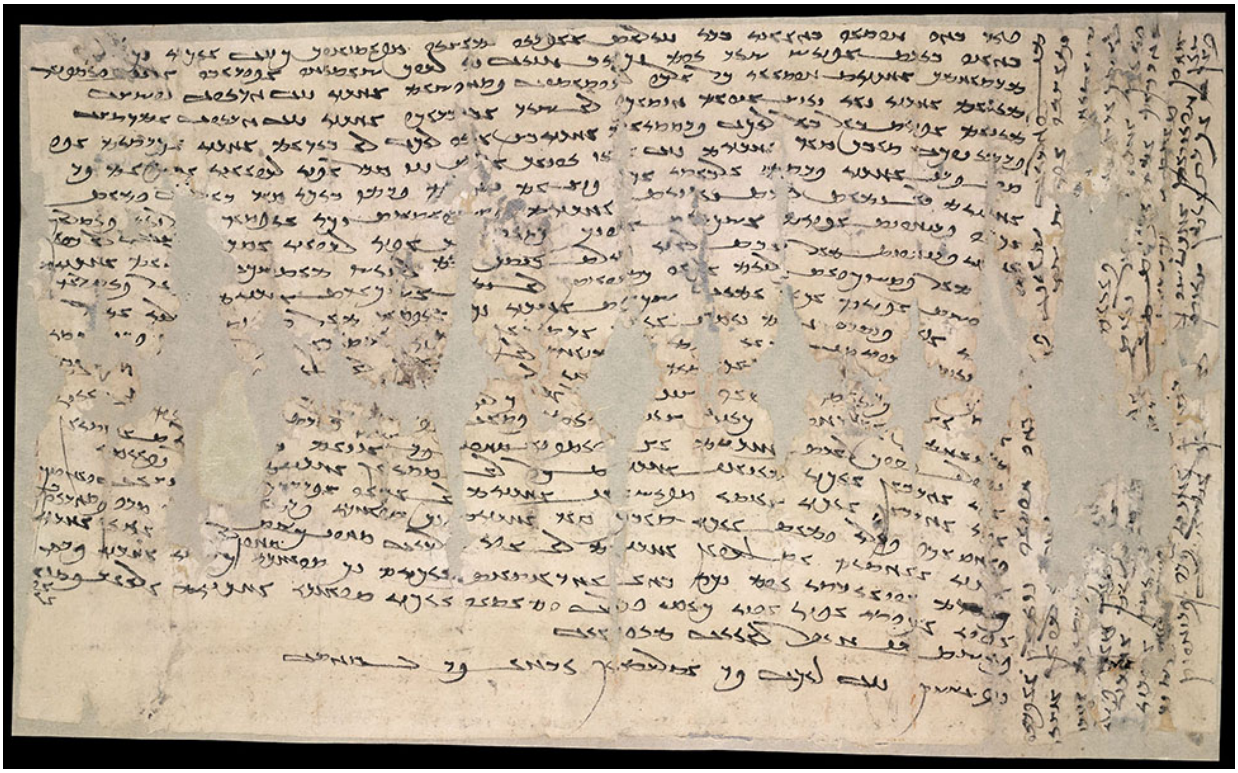


Figure 3: Fragment of 'Ancient Letter 3' (British Library, International Dunhuang Project: Or. 8212/98 recto 1). "From (his) daughter Shayn to the noble lord Nanai-dhat." (translation by Sims-Williams in Waugh 2004).

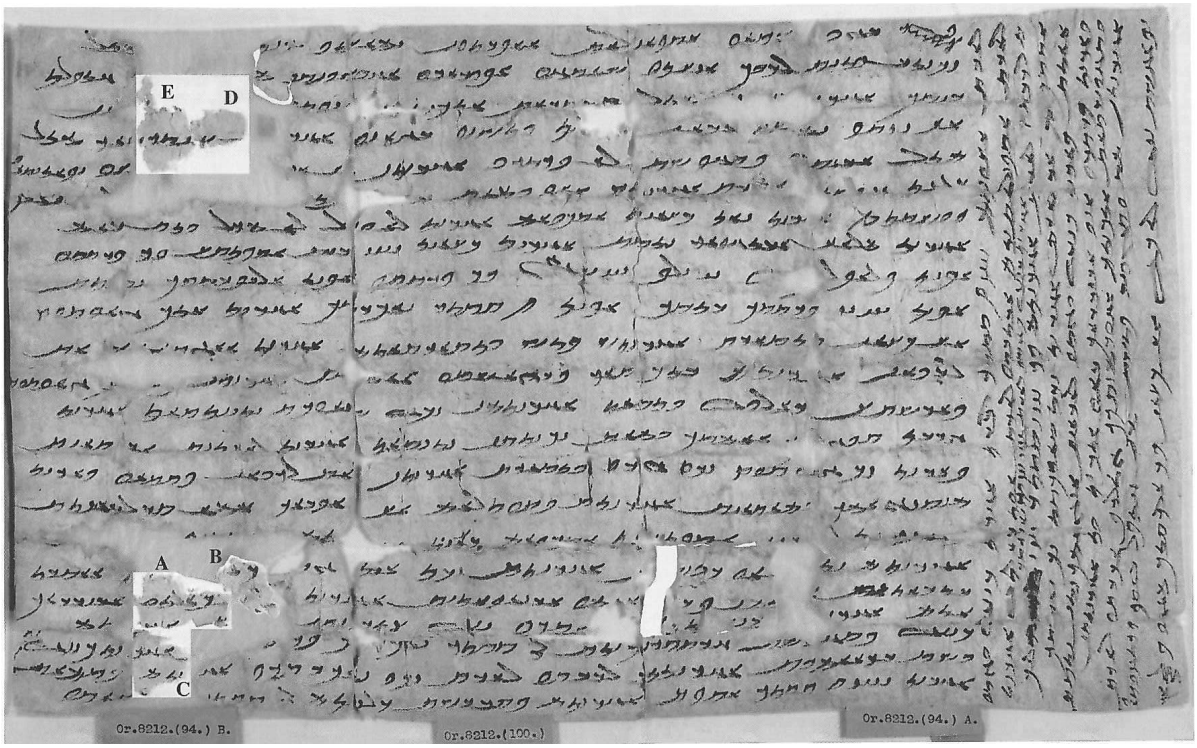


Figure 4: Fragment of 'Ancient Letter 5' (from Grenet et al. 1998: 94). "To the noble lord, the chief merchant Aspandhãt. [Sent] by your servant [Frī-khwatãw]."

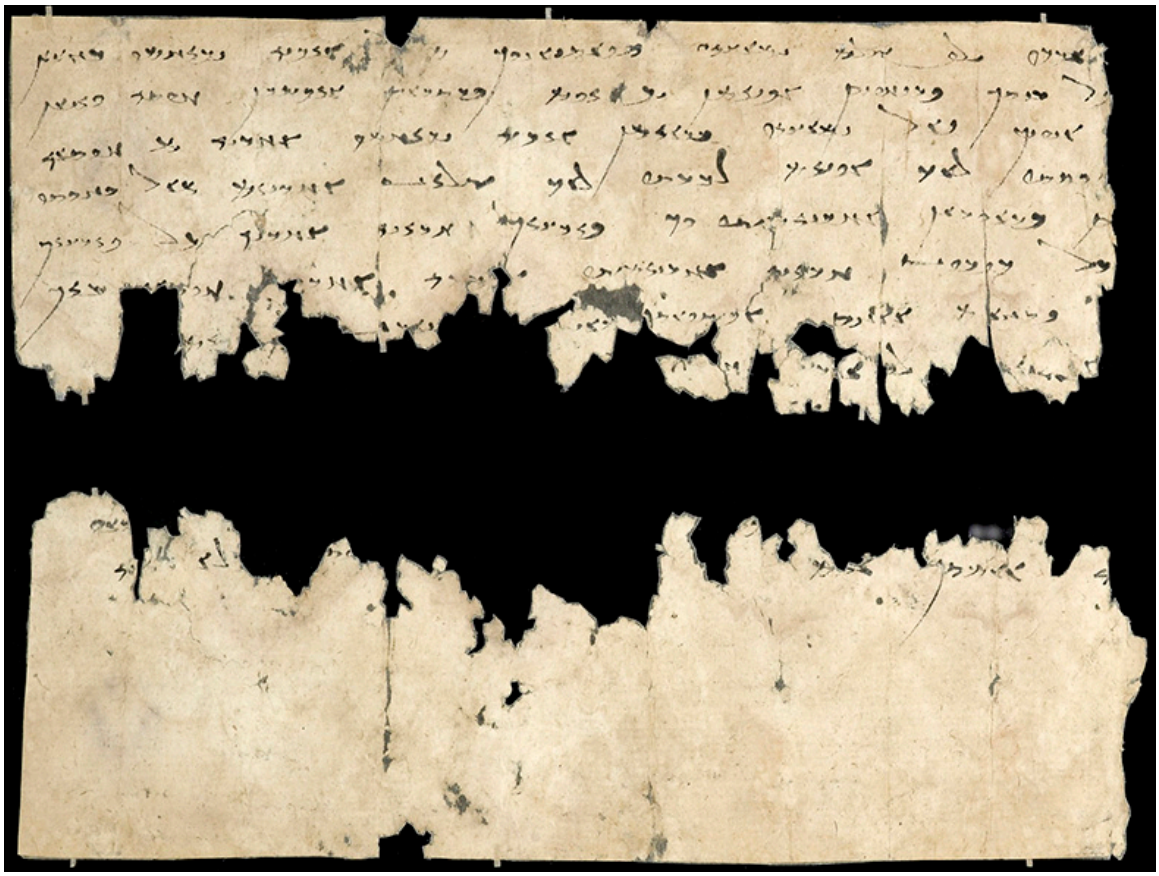


Figure 5: Fragment of Ancient Letter 6 (British Library, International Dunhuang Project: Or. 8212/97).

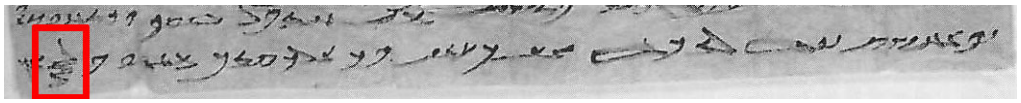
Figure 6: Sogdian rock inscription from Shatial (from Sims-Williams 1989: plate 10b) The inscription reads 𐼻𐼻𐼷𐼰𐼸𐼸 𐼵𐼸 *nny'kk ZK* (top line), 𐼼𐼴𐼰𐼾𐽀 *sw'βr* (middle), 𐼱𐽀𐼷 *BRY* (bottom). Latin transcription from *ibid*: 14. The inscription in the bottom right-hand corner is shown in detail in figure 7.

Figure 7: Sogdian rock inscription from Shatial (from Sims-Williams 1989: plate 10a). The central inscription reads 𐼾𐼰𐼾𐼰𐼸𐼸 *p'p'kk* (top line), 𐼵𐼸 𐼸𐼴𐽁𐼰𐼰𐼻 *ZK kwš''n* (middle), 𐼱𐽀𐼷 *BRY* (bottom). Latin transcription from *ibid*: 14. The inscription in the top left-hand corner is shown in detail in figure 6.

The number 4 written 𐼴𐼵 (from ‘Ancient Letter 5’, line 26).

The number 8 written 𐼵𐼵𐼵 (from ‘Ancient Letter 2’, line 26).

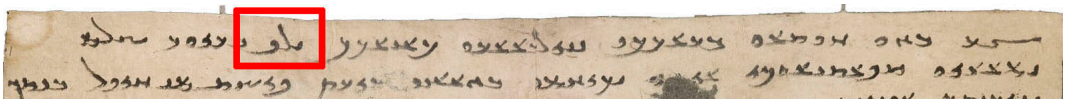
The number 20 𐼻 (from ‘Ancient Letter 2’, line 21).



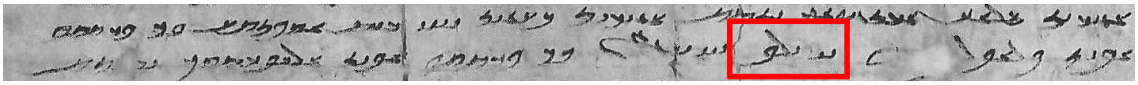
The number 30 𐼼 (from ‘Ancient Letter 2’, line 1).

The number 32 written as 𐼻𐼵𐼵 (from ‘Ancient Letter 2’, line 62).

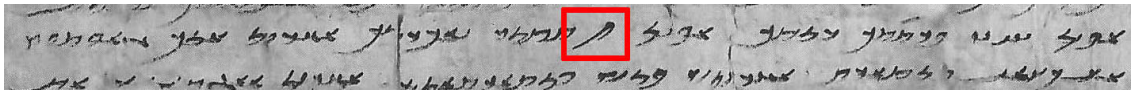
The number 500 written 𐼵𐼵 𐼴 (from ‘Ancient Letter 5’, line 9).



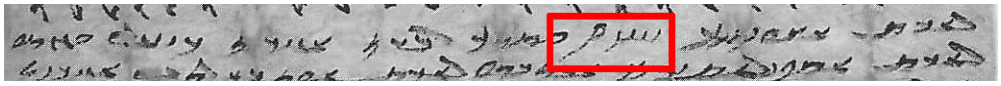
The number 1,000 written 𐼲 (from ‘Ancient Letter 2’, line 1).



The number 2,000 written 𐼴 𐼲 (from ‘Ancient Letter 5’, line 9).



The fraction ½ written 𐼽 (from ‘Ancient Letter 5’, line 10).



The fraction 4½ written 𐼵𐼵𐼽 (from ‘Ancient Letter 5’, line 24).

Figure 8: Examples of numerical notation in Old Sogdian.

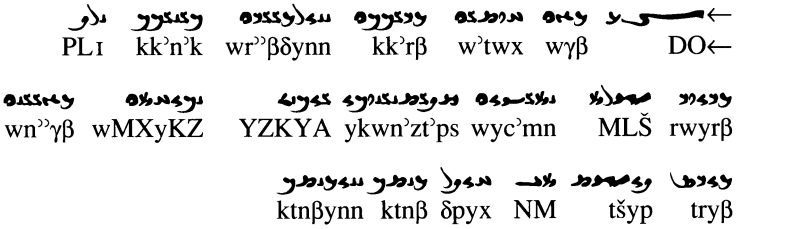
Sogdian script

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 **ẓ** **ṗ**.

SAMPLES OF SOGDIAN

ANCIENT LETTERS



1. *Transliteration:*	OD	βγw	xwtʾw	βrʾkk	nnyδβʾʾrw	kʾnʾkk
2. *Normalization:*	at	βaγu	xutāw	βarak	nanē-θβār	kanak
3. *Gloss:*	to	lord.ACC	master	Barak	Nana's-gift	Kanak

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

1. ZKyXMw	βγʾʾnw	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).*

Figure 9: Description and specimen of Old Sogdian (from Skjærvø 1996: 529).

TABLE 48.2: *Main East Iranian Scripts Developed from Aramaic*

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

Figure 10: Table of Sogdian scripts (from Skjærvø 1996: 519).

Addendum to Acknowledgments

This project was made possible in part through a Google Research Award, granted to Deborah Anderson for the Script Encoding Initiative, and a grant from the United States National Endowment for the Humanities (PR-50205-15), which funds the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of Google or the National Endowment for the Humanities.