

# Proposal to encode the Old Sogdian script in Unicode

Anshuman Pandey  
pandey@umich.edu

December 1, 2016

## 1 Introduction

This is a proposal to encode the ‘Old Sogdian’ script in Unicode. It is a significant revision and enlargement of the following document:

- L2/15-089 “Preliminary Proposal to Encode the Old Sogdian Script in Unicode”

In addition to substantial modifications, it incorporates comments regarding L2/15-089 made in:

- L2/16-037 “Recommendations to UTC #146 January 2016 on Script Proposals”

This proposal has been reviewed by Nicholas Sims-Williams and Yutaka Yoshida, who are leading scholars of Sogdian studies.

A proposal to encode the later ‘Sogdian’ script in Unicode has been submitted as:

- L2/16-371 “Revised proposal to encode the Sogdian script in Unicode”

## 2 Background



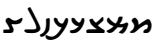

The proposed Unicode encoding for ‘Old Sogdian’ encompasses a group of related scripts used in the following records for representing Sogdian (ISO 639: sog), an ancient Eastern Iranian language:

- *Kultobe inscriptions* The oldest Sogdian records are inscriptions found at Kultobe, hereafter ‘K’, in modern Kazakhstan (see Sims-Williams & Grenet 1998; Grenet, et al 2007). Fourteen inscriptions have been discovered and studied (see figures 26, 27). They have not been dated, but the archaic features of the script and language indicate that they precede the ‘Ancient Letters’.
- ‘Ancient Letters’ The earliest attested Sogdian manuscripts are known as the ‘Ancient Letters’ (see figures 28–35), hereafter ‘AL’. These paper documents were found in 1907 by Aurel Stein in Dunhuang, western China. Based upon internal evidence the ‘AL’ may have been written during 312–314 CE (Sims-Williams 1985; Grenet, et al 1998).

- *Upper Indus inscriptions* Sogdian text appears on more than 600 rock carvings at Shatial and other sites in the Gilgit region of Pakistan (see figures 36, 37). These ‘Upper Indus inscriptions’, hereafter ‘UII’, have been dated to the 4th–7th centuries CE (Sims-Williams 1989, 2000), and some more precisely to the latter half of the 5th century (Yoshida 2013).

The scripts of these records are derived from Imperial Aramaic and exhibit the following features:

- *Repertoire* Of the 22 letters of the Aramaic alphabet, 20 are attested in the repertoires of these scripts. Analogues for *teth* and *qoph* do not exist. Of these 20, 17 have distinctive representations, while 3 share a resemblance. In AL and UII, the shapes of *daleth* and *ayin* are in general identical to *resh*, but possibly are distinctive in K. All 20 letters are exhibited in K 4 and occur collectively in AL. The AL contain letters that do not occur in K, such as final forms of *nun* and *sadhe*, special forms of *ayin*, and a new form of *he* (see § 3.2). Numerical signs are attested in AL and UII.
- *Letterforms* The shapes of letters in AL and UII are nearly identical. The letterforms of K are more archaic and reflect constraints imposed by the method and medium of inscription. The shapes of *gimel*, *he*, *yodh*, *lamedh*, *shin* in the three varieties differ from the Aramaic originals and corresponding letters in related Iranian scripts. They may be considered characteristically ‘Sogdian’. The special forms of *ayin* in AL do not occur in K or UII, or in any other script. A comparison of letters in related scripts is shown in table 1 and figure 40.
- *Structure* Each variety is a non-joining *abjad*, similar to Hebrew. Letters retain their shapes within a word. Some letters have distinctive word-final forms, but there are no formal conventions for their usage. The strokes of adjacent letters of a word may connect or overlap as the result of cursive writing. This type of conjunction differs from that of later ‘formal’ and ‘cursive’ Sogdian scripts, which possess intrinsic conjoining behaviors similar to Arabic, as shown below:

		Old Sogdian	Later Sogdian
<i>swγdyk</i>	‘Sogdian’		
<i>smʳknδc</i>	‘of Samarkand’		

- *Directionality* These old Sogdian varieties are written from right to left in lines that advance from top to bottom. Some UII are written vertically with letters rotated 90° counter-clockwise with lines that advance from left to right (see § 3.5).

These scripts may be considered typologically identical on the basis of their graphical and structural features. For purposes of character encoding they may be unified within a single Unicode script block. Using this approach texts would be represented using the same character set, but the display would be managed through the selection of fonts designed specifically for the K, AL, and UII varieties.

The proposed Unicode block is named ‘Old Sogdian’. This identifier has been selected because proper names do not exist for individual script varieties or for the family. The script of AL has been referred to as “Sogdian Aramaic” (Skjærvø 1996), which may be applied equally to the other two varieties. However, the descriptor ‘Aramaic’ is not used in Unicode names for other scripts descended from Aramaic. The bare name ‘Sogdian’ is used in the catalogue of the International Dunhuang Project for referring to both early and

later script varieties. It is, however, practical to reserve this name for a Unicode block for the more well-known ‘formal’ and ‘cursive’ styles, which have been proposed for encoding in a unified ‘Sogdian’ block (see L2/16-371). The designation ‘Old Sogdian’ suitably identifies these early varieties while emphasizing their genetic relationship with later ‘Sogdian’ script styles.

### 3 Proposed Encoding


#### 3.1 Bidirectional model

Old Sogdian may be implemented using the Unicode Bidirectional Algorithm. There are no requirements for shaping.








#### 3.2 Character repertoire

The proposed repertoire contains 40 characters: 29 letters, 10 numbers, 1 heterogram. Names for letters correspond to those of the ‘Imperial Aramaic’ block. Representative glyphs are based upon forms in the AL unless specified below.

The encoded set may differ from traditional and scholarly inventories of script varieties that occur in written and inscriptional sources. Such differences naturally arise from the requirements for digitally representing a script in plain text and for preserving the semantics of characters.

In this document, names in italics refer to scholarly names for graphemes while names in small capitals refer to proposed Unicode characters, eg.  is *aleph* and OLD SOGDIAN LETTER ALEPH. For sake of brevity, the descriptor ‘OLD SOGDIAN’ is dropped when referring to Old Sogdian characters, eg. OLD SOGDIAN LETTER ALEPH is referred to as ALEPH. Characters of other scripts are designated by their full Unicode names. Latin transliteration of Old Sogdian letters follows the scholarly convention. Aramaic heterograms are transliterated using the corresponding uppercase letters, with some exceptions as shown in the table below.

##### 3.2.1 Letters

Glyph	Character name	Latin
	OLD SOGDIAN LETTER ALEPH	ʾ
	OLD SOGDIAN LETTER FINAL ALEPH	ʾ
	OLD SOGDIAN LETTER BETH	β ; B
	OLD SOGDIAN LETTER FINAL BETH	β ; B
	OLD SOGDIAN LETTER GIMEL	γ ; G
	OLD SOGDIAN LETTER HE	h
	OLD SOGDIAN LETTER FINAL HE	h

𐰪	OLD SOGDIAN LETTER WAW	w
𐰫	OLD SOGDIAN LETTER ZAYIN	z
𐰬	OLD SOGDIAN LETTER HETH	x ; 𐰬
𐰭	OLD SOGDIAN LETTER YODH	y
𐰮	OLD SOGDIAN LETTER KAPH	k
𐰯	OLD SOGDIAN LETTER LAMEDH	δ ; L
𐰰	OLD SOGDIAN LETTER MEM	m
𐰱	OLD SOGDIAN LETTER NUN	n
𐰲	OLD SOGDIAN LETTER FINAL NUN	n
𐰳	OLD SOGDIAN LETTER VERTICAL FINAL NUN	n
𐰴	OLD SOGDIAN LETTER SAMEKH	s
𐰵	OLD SOGDIAN LETTER AYIN	ʿ
𐰶	OLD SOGDIAN LETTER ALTERNATE AYIN	ʿ
𐰷	OLD SOGDIAN LETTER PE	p
𐰸	OLD SOGDIAN LETTER SADHE	c
𐰹	OLD SOGDIAN LETTER FINAL SADHE	c ; 𐰹
𐰺	OLD SOGDIAN LETTER VERTICAL FINAL SADHE	c
𐰻	OLD SOGDIAN LETTER RESH-DALETH-AYIN	r, d, ʿ
𐰼	OLD SOGDIAN LETTER SHIN	š
𐰽	OLD SOGDIAN LETTER TAW	t
𐰾	OLD SOGDIAN LETTER FINAL TAW	t
𐰿	OLD SOGDIAN LETTER VERTICAL FINAL TAW	t

---

**aleph** In word-final positions in AL, *aleph* is written as 𐰿 FINAL ALEPH, in which the horizontal stroke at the baseline is elongated. The letter 𐰿 ALEPH has the shape 𐰿 in K. This form is a glyphic variant. See figure 1 for attestations.

**beth** In word-final positions in AL, *beth* is written as 𐰽 FINAL BETH, in which the horizontal stroke at the baseline is elongated. See figure 2 for attestations.

**gimel** See figure 3 for attestations of ג GIMEL.

**daleth** The letter ד *daleth* occurs only in Aramaic heterograms, eg. גבען 'BDt (K 4.1) and ד' (AL 2.1). In AL, the shape of *daleth* is identical to *resh*. In K, there is a possibility that *daleth* is distinguished from *resh*. See figure 4 for attestations. The issue regarding the shape of *daleth* is inherited from Aramaic, in which ד ARAMAIC LETTER DALETH and ר ARAMAIC LETTER RESH are nearly identical. Despite the possibility of a distinctive *daleth* in K, there is insufficient information at this time for defining it as a separate character. For this reason, *daleth* is unified with *resh* and is to be represented using ר RESH-DALETH-AYIN. Space has been reserved in the code block at the appropriate position in the alphabetic order in order to accommodate the encoding of *daleth* in the event that a distinctive form is identified in the future.

**he** The letter *he* is represented by both ה HE and ה FINAL HE. The ה HE occurs only in K, eg. שרחה ŠWRH and ח(ה)WH (K 2); חנה knth and חנה ZNH (K 4). The representative glyph is derived from K 2. In AL, *he* is written as ה, eg. חנה knth (AL 2.12), חנה ZNH (AL 2.10). The ה FINAL HE is not attested in K, and in contrast to ה, it occurs only at the end of a word. It has the variant shape ח, in which the terminal is truncated or merged back into the baseline stroke. See figure 5 for attestations.

**waw** See figure 6 for attestations of ו WAW.

**zayin** See figures 7 and 8 for attestations of ז ZAYIN.

**heth** See figure 9 for attestations of ח HETH.

**teth** An Old Sogdian analogue for Aramaic *teth* does not exist. In K, the *teth* in the Aramaic heterogram QTLt is represented using ת TAW: תתל KTLt (K 3.3).

**yodh** See figure 10 for attestations of י YODH.

**kaph** See figure 11 for attestations of כ KAPH.

**lamedh** The letter LAMEDH has the shape ל in K and ל in AL (see figure 12). The AL form is the representative glyph. In AL 5, *lamedh* appears as ל. Differences between ל, ל, ל are stylistic, not semantic. The forms ל and ל are to be treated as glyphic variants of ל.

**mem** See figure 13 for attestations of מ MEM.

**nun** Occurrences of *nun* are represented using נ NUN, נ FINAL NUN, ן VERTICAL FINAL NUN (see figure 14). The representative glyph נ for NUN is derived from K. The final forms occur only in AL. While *nun* has the distinctive shape נ in K, it has the shape ן in AL when non-final, which is identical to ז ZAYIN, eg. חנה ZNH (K 4.1) and חנה ZNH (AL 2.10). When word-final in AL, *nun* is written as both נ and ן, eg. מן MN (AL 2.2) and מן MN (AL 2.6). The regular and final forms are contrastive in AL (see figure 8). They are not glyphic variants. All three characters are required for fully representing *nun* in plain text.

**samekh** The letter ס SAMEKH occurs as the two-part form ס in K 4. This archaic form is to be treated as a glyphic variant. See figure 15 for attestations.

**ayin** The letter *ayin* occurs only in Aramaic heterograms. It has the regular shape ע and the special shapes פ and פ (see figure 16). The regular ע *ayin* occurs in both K and AL, eg. גבען 'BDt (K 4.1), עזק 'LZK (AL 2.12), ע 'L (AL 6.6). In AL, the shape of regular *ayin* is identical to *resh* (and *daleth*). In K,

there is a possibility that *ayin* might be a distinctive letter. The similarity between *ayin* and *resh* is inherited from Aramaic, compare 𐤀 ARAMAIC LETTER AYIN and 𐤂 ARAMAIC LETTER RESH. However, there is insufficient information for determining whether or not the differences between *ayin* and *resh* in K are semantically significant. Therefore, a separate character for regular *ayin* is not proposed at present. It is to be represented using 𐰆 RESH-DALETH-AYIN. The special forms 𐰇 AYIN and 𐰈 ALTERNATE AYIN occur only in AL for writing the heterogram *D*, eg. 𐰇𐰆 (AL 2.1), 𐰇𐰆 (AL 3 verso), 𐰇𐰆 (AL 3.1), 𐰇𐰆𐰇 (AL 5.1). Although 𐰇, 𐰈, 𐰉, 𐰆 could be considered glyphic variants of a single character AYIN, it is appropriate to define two characters on account of their graphical structures. The 𐰉 is a glyphic variant of 𐰈 with an ornate tail; the 𐰆 is a variant with a truncated tail. These three forms are unified as 𐰈 ALTERNATE AYIN, which may be used for representing these special forms in plain text. See figure 25 and § 3.2.3 for attestations.

**pe** The letter 𐰊 PE has the variant shape 𐰋, which is to be treated as a glyphic variant (see figure 17).

**sadhe** This letter is represented using 𐰌 SADHE, 𐰍 FINAL SADHE, and 𐰎 VERTICAL FINAL SADHE (see figure 18). The final forms occur only in AL. In AL 2, *sadhe* has the shape 𐰎 whenever it occurs at the margin, eg. 𐰎𐰏𐰐𐰑 𐰒HRZYnnc (AL 2.54). In other positions within a line, final *sadhe* is written using 𐰍 FINAL SADHE, eg. 𐰍𐰎𐰏𐰑 𐰒HRZYnc (AL 2.34). The stroke of 𐰍 may be curved as 𐰏.

**qoph** An Old Sogdian analogue for Aramaic *qoph* does not exist. In K, the *qoph* in the Aramaic heterogram *QTLt* is represented using 𐰐 KAPH: 𐰐𐰑𐰒 KTLt (K 3.3). It used to be believed that 𐰐 *qoph* was retained in AL as 𐰑 and reassigned for the number 100. This 𐰑 is now identified as the fraction ½ (Grenet, et al 1998).

**resh** In AL, the letter 𐰆 is used for *resh*, *daleth*, and *ayin* (see figure 19). According to the Unicode character-glyph model, letters with identical glyphic representations are considered variants and are unified as a single character. Accordingly, *daleth* and *ayin* are unified with *resh* as 𐰆 RESH-DALETH-AYIN. This approach follows the Unicode model for Inscriptional Pahlavi, in which *waw*, *ayin*, *resh* are represented by 𐭑 U+10B65 INSCRIPTIONAL PAHLAVI LETTER WAW-AYIN-RESH; and *mem* and *qoph* by 𐭒 U+10B6C INSCRIPTIONAL PAHLAVI LETTER MEM-QOPH. Despite occurring after *daleth* and *ayin* in the alphabetical order, *resh* is ordered first in the name RESH-DALETH-AYIN because it occurs more frequently in the sources.

**shin** See figure 20 for attestations of 𐰓 SHIN.

**taw** This letter is represented using 𐰔 TAW, 𐰕 FINAL TAW, 𐰖 ALTERNATE FINAL TAW (see figure 21). Usage of the two forms in word-final position is contrastive in both K and AL. In K 4, *taw* appears as 𐰕 and final *taw* as 𐰖. These archaic two-part forms are to be treated as glyphic variants of TAW and VERTICAL FINAL TAW, respectively. A curved form 𐰗 of 𐰕 appears in the sources.

**Note on final forms** Distinctive final forms of *aleph*, *beth*, *nun*, *sadhe*, and *taw* are encoded as separate characters. These forms differ from the nominal forms in the shape of their terminals, eg. elongated horizontally, curved upwards, or descending vertically. An analysis of K and AL shows that the final forms of these letters were deliberately distinguished from the nominal forms by the scribe, and are not stylistic flourishes. The available sources do not suggest that distinctive final forms exist for other letters (see figures 1–21).

### 3.2.2 Numbers

The repertoire contains 10 numerical characters. These occur in AL and UII, but not in the extant K sources.

Glyph	Character name	Numeric value
𐰀	OLD SOGDIAN NUMBER ONE	1
𐰁	OLD SOGDIAN NUMBER TWO	2
𐰂	OLD SOGDIAN NUMBER THREE	3
𐰃	OLD SOGDIAN NUMBER FOUR	4
𐰄	OLD SOGDIAN NUMBER FIVE	5
𐰅	OLD SOGDIAN NUMBER TEN	10
𐰆	OLD SOGDIAN NUMBER TWENTY	20
𐰇	OLD SOGDIAN NUMBER THIRTY	30
𐰈	OLD SOGDIAN NUMBER ONE HUNDRED	100
𐰉	OLD SOGDIAN FRACTION ONE HALF	½

**Primary units** The numbers 𐰀 ONE .. 𐰄 FOUR are encoded atomically for facilitating the expression of primary units using groups composed of repetitions of 𐰀 ONE. The numbers 5–9 are written using sequences of ONE .. FOUR arranged in groups separated by spaces. The number 5 is also written as 𐰄 FIVE. This model for ONE .. FOUR follows the Unicode encoding for Inscriptional Parthian, eg. 𐰀 U+10B58 INSCRIPTIONAL PARTHIAN NUMBER ONE .. 𐰄 U+10B5B INSCRIPTIONAL PARTHIAN NUMBER FOUR.

**Tens** The 𐰅 TEN resembles a vertically compressed 𐰁 LAMEDH. The 𐰆 TWENTY and 𐰇 THIRTY are vertical stacks of 𐰅 TEN. The number 30 is also represented as 𐰆𐰅, which is a compound of TWENTY and TEN. Multiples of 10 greater than 30 would be produced using appropriate repetitions and groupings of TEN, TWENTY, and THIRTY.

**Hundreds** The number 100 is written using 𐰈 ONE HUNDRED. The glyph resembles the letter 𐰁 GIMEL above a serpentine form, but it is an atomic character. The ONE HUNDRED also functions as a unit mark for the hundreds. Multiples of hundred are represented by prefixing the appropriate groupings of ONE .. FOUR before ONE HUNDRED.

**Thousands** The number 1000 is expressed using the Aramaic heterogram 𐰈𐰅 *ILP*, which is represented using the sequence <𐰀 ONE, 𐰁 LAMEDH, 𐰅 PE>. The sequence 𐰈𐰅 also functions as a unit mark for the thousands. The 𐰀 ONE is an inherent part of the *ILP* unit. Multiples are expressed by prefixing primary numbers before the unit, eg. 2000 is 𐰂𐰈𐰅, 3000 is 𐰃𐰈𐰅.

**Ten thousands** The number 10000 is expressed using the Sogdian word 𐰉𐰅𐰅𐰅 *brywr*. There is no distinctive numerical sign for this value.

**Fraction** The 𐰉 FRACTION ONE HALF is placed after another numerical character.

**Notation system** The ordering of numbers follows the right-to-left directionality of the script. The expression of numbers is additive. Compound numbers of different decimal orders are produced by placing larger units first.

Value	Number	Input string →
4½	𐰪𐰣𐰆	<𐰪 FOUR, 𐰆 FRACTION ONE HALF>
5	𐰪𐰣	<𐰪 THREE, [SP] SPACE, 𐰣 TWO>
5	𐰪𐰣𐰆	<𐰪𐰣 FIVE>
6	𐰪𐰣𐰣	<𐰪 THREE, [SP] SPACE, 𐰣 THREE>
7	𐰪𐰣𐰣	<𐰪𐰣 FOUR, [SP] SPACE, 𐰣 THREE>
7½	𐰪𐰣𐰆𐰆	<𐰪𐰣 FOUR, [SP] SPACE, 𐰣 THREE, 𐰆 FRACTION ONE HALF>
8	𐰪𐰣𐰣	<𐰪𐰣 FOUR, [SP] SPACE, 𐰪𐰣 FOUR>
9	𐰪𐰣𐰣𐰣	<𐰪 THREE, [SP] SPACE, 𐰣 THREE, [SP] SPACE, 𐰣 THREE>
13	𐰣𐰣	<𐰣 TEN, 𐰣 THREE>
15	𐰣𐰣𐰆	<𐰣 TEN, 𐰪𐰣 FIVE>
30	𐰣	<𐰣 THIRTY>
30	𐰣𐰣	<𐰣 TWENTY, 𐰣 TEN>
32	𐰣𐰣𐰣	<𐰣 TWENTY, 𐰣 TEN, 𐰣 TWO>
100	𐰣	<𐰣 ONE HUNDRED>
200	𐰣𐰣	<𐰣 TWO, [SP] SPACE, 𐰣 TWO, 𐰣 ONE HUNDRED>
500	𐰣𐰣𐰣	<𐰣 THREE, [SP] SPACE, 𐰣 TWO, 𐰣 ONE HUNDRED>
1000	𐰣𐰣	<𐰣 ONE, 𐰣 LAMEDH, 𐰣 PE>
2000	𐰣𐰣𐰣	<𐰣 TWO, [SP] SPACE, 𐰣 ONE, 𐰣 LAMEDH, 𐰣 PE>
10000	𐰣𐰣𐰣𐰣	<𐰣 BETH, 𐰣 RESH, 𐰣 YODH, 𐰣 WAW, 𐰣 RESH>

Attestations for the above numbers are shown in figures 22–24. The repertoire provides for the presentation of any numerical value, even if not attested. For example, the number 2453 could be represented as:



Value	Number	Input string →
2453	𐰇 𐰆 𐰅 𐰄 𐰃 𐰂 𐰁 𐰀	<𐰇 TWO, [SP] SPACE, 𐰆 ONE, 𐰄 LAMEDH, 𐰃 PE, [SP] SPACE, 𐰂 FOUR, 𐰁 ONE HUNDRED, [SP] SPACE, 𐰀 THIRTY, 𐰀 TWENTY, 𐰀 THREE>
2453	𐰇 𐰆 𐰅 𐰄 𐰃 𐰂 𐰁 𐰀	<𐰇 TWO, [SP] SPACE, 𐰆 ONE, 𐰄 LAMEDH, 𐰃 PE, [SP] SPACE, 𐰂 FOUR, 𐰁 ONE HUNDRED, [SP] SPACE, 𐰀 TWENTY, 𐰀 TWENTY, 𐰀 TEN, 𐰀 THREE>

### 3.2.3 Heterogram

The repertoire contains 1 heterogram.

Glyph	Character name	Value
𐰇	OLD SOGDIAN HETEROGRAM OD	‘D

Aramaic heterograms are represented as words spelled using conventional letters, eg. 𐤆𐤌𐤆𐤀 𐤇𐤓𐤕𐤕𐤕𐤕 *HRZY* is written <𐤆 ALEPH, 𐤌 HETH, 𐤇 RESH, 𐤕 ZAYIN, 𐤕 YODH>. The heterogram ‘D’ is the sole exception. Meaning “to”, ‘D’ occurs in the address and salutation of a letter, eg. 𐰇 𐰆𐰅𐰄𐰃𐰂𐰁𐰀 𐰇𐰆𐰅𐰄𐰃𐰂𐰁𐰀 *D βγw xwtʾw βʾrkkw* “to lord master Barak”. Morphologically, it is comprised of *ayin* and *daleth*. Yet, instead of the expected spelling \*𐰇𐰆 <𐰇 RESH-DALETH-AYIN, 𐰇 RESH-DALETH-AYIN>, the *ayin* is written using special forms: 𐰇, 𐰇, 𐰇, 𐰇 (see figure 25). An explanation for this curious orthography may be that *ayin* and *daleth* had disappeared from the script by the time of AL, and the original phonetic values of these letters never existed in Sogdian. Therefore, scribes were unaware of these letters and of the original spelling of the Aramaic word, so they stylized the writing of ‘D’ (Sims-Williams, personal correspondence, 2016).

There are two ways to analyze these representations of ‘D’. First, as a conventional word comprised of the letters *ayin* and *daleth*. These forms of *ayin*, which occur only in this heterogram, are included in the repertoire as 𐰇 AYIN and 𐰇 ALTERNATE AYIN; the 𐰇 and 𐰇 could be considered glyphic variants of ALTERNATE AYIN. Accordingly, ‘D’ may be represented as <AYIN | ALTERNATE AYIN, RESH-DALETH-AYIN>. Secondly, ‘D’ is a logographic unit comprised of a ligature or a set of two letters. This unit may be treated as an atomic character, eg. 𐰇 OLD SOGDIAN HETEROGRAM OD. These approaches are not mutually exclusive and both are practical for character encoding. Depending upon the context, ‘D’ may be spelled using a sequence of letters or represented using an atomic character.

The case of ‘D’ is similar to the Latin ‘&’ ampersand. The ‘&’ represents the Latin word *et* “and”. Morphologically, it is a ligation of the Latin letters ‘e’ and ‘t’, eg. 𐌆𐌗, 𐌆𐌗. The base letters began to be obscured as the ligature became more stylized, eg. 𐌆𐌗. The logographic nature of ‘&’ is apparent in the abbreviation “&c” for Latin *et cetera* “and so forth”, where it masks ‘et’. Latin *et* can be represented both using the sequence <e, t> and atomic characters, such as 𐌆𐌗 U+1F670 SCRIPT LIGATURE ET ORNAMENT.

The character name for HETEROGRAM OD is derived from the normalization ‘OD’ of the transliteration ‘D’ (Skjærvø 1996). The representative glyph 𐰇 is derived from AL 3 and has been selected because it is structurally a ligature. Variant forms may be managed through fonts.

### 3.3 Punctuation

Punctuation marks are not attested. Words are separated using spaces in K and AL. Inter-word spacing is inconsistent in the UII.

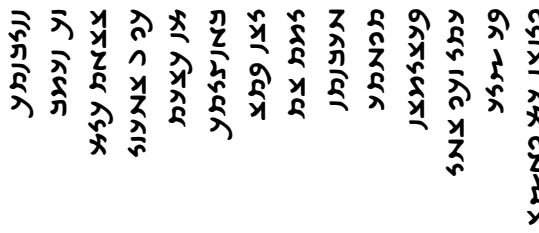
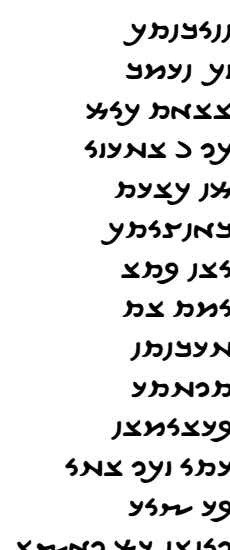
### 3.4 Line-breaking

There are no rules for line-breaking. The available sources show line-breaks after the end of a word. Word are not split across lines. Consequently, hyphens or other continuation marks are not attested. In digital layouts, line-breaks may occur after any character.

### 3.5 Vertical text

The majority of Old Sogdian records have horizontal orientations. Some UII records are inscribed vertically (Yoshida 2013). There are no formal conventions for text orientation. However, in vertical environments, Old Sogdian text is oriented from top to bottom with lines that advance from left to right. Letters are rotated 90° counter-clockwise from their regular upright shapes.

By default, Old Sogdian may be oriented horizontally in plain text representations. However, support for vertical orientations of the script is required for accurately displaying Old Sogdian text that is natively vertical. Below is a vertical text from Shatial rock 36:38 (see figure 38) and its horizontal representation:

Vertical (rotated 90° CCW)	Horizontal
	

*nnyβntk / ZK nrsβ / ʾyγt kym / kw 10 ʾHRZY / MN kʾrt / βγncytk / yʾn ptʾ[-] / [-]yst ʾt / xrβntn / twxtr / prʾysʾn / rty ZKw ʾHY / pr šyr / wyn ʾn ʾM wγšʾ*

“(I), Nanai-vandak the (son of) Narisaf have come (here) in/on the (day/year) ten and asked a boon from the spirit of the sacred place Kârt (that) I may arrive at Kharvandan (= Tashkurgan) very quickly and see (my) brother in good (health) with joy.” (Yoshida 2013: 379–380).

The “Unicode Technical Report #50: Unicode Vertical Text Layout” describes the `Vertical_Orientation` (`vo`) property for specifying the orientation of characters in vertical environments. For Old Sogdian, this property would be defined as: `Vertical_Orientation=R` or `vo=R`, where the value ‘R’ indicates that the glyphs are rotated in vertical layout.

### 3.6 Collation

The sort order for Old Sogdian is as follows:

```

𐰀 ALEPH << 𐰁 FINAL ALEPH < 𐰂 BETH << 𐰃 FINAL BETH < 𐰄 GIMEL < 𐰅 HE <<
𐰆 FINAL HE < 𐰇 WAW < 𐰈 ZAYIN < 𐰉 HETH < 𐰊 YODH < 𐰋 KAPH < 𐰌 LAMEDH <
𐰍 MEM < 𐰎 NUN << 𐰏 FINAL NUN << 𐰐 VERTICAL FINAL NUN < 𐰑 SAMEKH <
𐰒 AYIN << 𐰓 ALTERNATE AYIN < 𐰔 PE < 𐰕 SADHE << 𐰖 FINAL SADHE <<
𐰗 VERTICAL FINAL SADHE < 𐰘 RESH-DALETH-AYIN < 𐰙 SHIN < 𐰚 TAW <<
𐰛 FINAL TAW << 𐰜 VERTICAL FINAL TAW

```

### 3.7 Character Data

#### 3.7.1 Character properties

In the format of `UnicodeData.txt`:

```

10E00;OLD SOGDIAN LETTER ALEPH;Lo;0;R;;;;;N;;;;;
10E01;OLD SOGDIAN LETTER FINAL ALEPH;Lo;0;R;;;;;N;;;;;
10E02;OLD SOGDIAN LETTER BETH;Lo;0;R;;;;;N;;;;;
10E03;OLD SOGDIAN LETTER FINAL BETH;Lo;0;R;;;;;N;;;;;
10E04;OLD SOGDIAN LETTER GIMEL;Lo;0;R;;;;;N;;;;;
10E05;<reserved>
10E06;OLD SOGDIAN LETTER HE;Lo;0;R;;;;;N;;;;;
10E07;OLD SOGDIAN LETTER FINAL HE;Lo;0;R;;;;;N;;;;;
10E08;OLD SOGDIAN LETTER WAW;Lo;0;R;;;;;N;;;;;
10E09;OLD SOGDIAN LETTER ZAYIN;Lo;0;R;;;;;N;;;;;
10E0A;OLD SOGDIAN LETTER HETH;Lo;0;R;;;;;N;;;;;
10E0B;OLD SOGDIAN LETTER YODH;Lo;0;R;;;;;N;;;;;
10E0C;OLD SOGDIAN LETTER KAPH;Lo;0;R;;;;;N;;;;;
10E0D;OLD SOGDIAN LETTER LAMEDH;Lo;0;R;;;;;N;;;;;
10E0E;OLD SOGDIAN LETTER MEM;Lo;0;R;;;;;N;;;;;
10E0F;OLD SOGDIAN LETTER NUN;Lo;0;R;;;;;N;;;;;
10E10;OLD SOGDIAN LETTER FINAL NUN;Lo;0;R;;;;;N;;;;;
10E11;OLD SOGDIAN LETTER VERTICAL FINAL NUN;Lo;0;R;;;;;N;;;;;
10E12;OLD SOGDIAN LETTER SAMEKH;Lo;0;R;;;;;N;;;;;
10E13;OLD SOGDIAN LETTER AYIN;Lo;0;R;;;;;N;;;;;
10E14;OLD SOGDIAN LETTER ALTERNATE AYIN;Lo;0;R;;;;;N;;;;;
10E15;OLD SOGDIAN LETTER PE;Lo;0;R;;;;;N;;;;;
10E16;OLD SOGDIAN LETTER SADHE;Lo;0;R;;;;;N;;;;;
10E17;OLD SOGDIAN LETTER FINAL SADHE;Lo;0;R;;;;;N;;;;;
10E18;OLD SOGDIAN LETTER VERTICAL FINAL SADHE;Lo;0;R;;;;;N;;;;;
10E19;OLD SOGDIAN LETTER RESH-DALETH-AYIN;Lo;0;R;;;;;N;;;;;
10E1A;OLD SOGDIAN LETTER SHIN;Lo;0;R;;;;;N;;;;;
10E1B;OLD SOGDIAN LETTER TAW;Lo;0;R;;;;;N;;;;;
10E1C;OLD SOGDIAN LETTER FINAL TAW;Lo;0;R;;;;;N;;;;;

```

```

10E1D;OLD SOGDIAN LETTER VERTICAL FINAL TAW;Lo;0;R;;;N;;;;;
10E1E;OLD SOGDIAN NUMBER ONE;No;0;R;;;1;N;;;;;
10E1F;OLD SOGDIAN NUMBER TWO;No;0;R;;;2;N;;;;;
10E20;OLD SOGDIAN NUMBER THREE;No;0;R;;;3;N;;;;;
10E21;OLD SOGDIAN NUMBER FOUR;No;0;R;;;4;N;;;;;
10E22;OLD SOGDIAN NUMBER FIVE;No;0;R;;;5;N;;;;;
10E23;OLD SOGDIAN NUMBER TEN;No;0;R;;;10;N;;;;;
10E24;OLD SOGDIAN NUMBER TWENTY;No;0;R;;;20;N;;;;;
10E25;OLD SOGDIAN NUMBER THIRTY;No;0;R;;;30;N;;;;;
10E26;OLD SOGDIAN NUMBER ONE HUNDRED;No;0;R;;;100;N;;;;;
10E27;OLD SOGDIAN FRACTION ONE HALF;;No;0;R;;;1/2;N;;;;;
10E28;OLD SOGDIAN HETEROGRAM OD;Lo;0;R;;;N;;;;;

```

### 3.7.2 Linebreaking

In the format of `LineBreak.txt`:

```

10E00..10E04;AL # Lo [5] OLD SOGDIAN LETTER ALEPH..OLD SOGDIAN LETTER GIMEL
10E06..10E1D;AL # Lo [24] OLD SOGDIAN LETTER HE..
                          OLD SOGDIAN LETTER VERTICAL FINAL TAW
10E1E..10E27;AL # No [10] OLD SOGDIAN NUMBER ONE..OLD SOGDIAN FRACTION ONE HALF
10E28;AL # Lo OLD SOGDIAN HETEROGRAM OD

```

## 4 References

- 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, ed. Osmund Bopearachchi, Carol Altman Bromberg, and Frantz Grenet, pp. 91–104.
- Grenet, Frantz; Nicholas Sims-Williams; Aleksandr Podushkin. 2007. “Les plus anciens monuments de la langue sogdienne: les inscriptions de Kultobe au Kazakhstan”. *Comptes rendus des séances de l’Académie des Inscriptions et Belles-Lettres*, 151<sup>e</sup> année, N. 2, 2007, pp. 1005–1034.
- Pandey, Anshuman. 2016. “Revised proposal to encode the Sogdian script in Unicode” (L2/16-371). <http://www.unicode.org/L2/L2016/16371-sogdian.pdf>
- Reichelt, Hans. 1928–31. *Die soghdischen Handschriftenreste des Britischen Museums*. Heidelberg: C. Winter’s Universitätsbuchhandlung.
- Sims-Williams, Nicholas. 1981. “The Sogdian sound-system and the origins of the Uyghur script”. *Journal Asiatique*, pp. 347–360.
- . 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*, v. 23–24 (Professor Gregory M. Bongard-Levin Felicitation Volume), pp. 523–541.
- Sims-Williams, Nicholas; Frantz Grenet. 2007. “The Sogdian Inscriptions of Kultobe”. *Shygys*, 2006, vol. 1 pp. 95–111.
- 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.
- Waugh, Daniel C. [comp]. 2004. “The Sogdian Ancient Letters”, translated by Nicholas Sims-Williams. <https://depts.washington.edu/silkroad/texts/sogdlet.html>
- Yoshida, Yutaka. 2013. “When Did Sogdians Begin to Write Vertically?”. *Tokyo University Linguistic Papers*, vol. 33, pp. 375–394.

## 5 Acknowledgments

I express my sincere gratitude to Nicholas Sims-Williams (SOAS, University of London) for providing detailed comments on earlier versions of this proposal and for informative discussions regarding all facets of the script. I am also thankful to Yutaka Yoshida (University of Kyoto) for reviewing versions of this proposal and for providing valuable feedback. I thank them both for their patient responses to my numerous inquiries and for overlooking my ignorance of the script. I am grateful to Roozbeh Pournader (Google, San Francisco) for discussing Unicode encodings for Iranian scripts and for his feedback on the earliest draft proposal.

The present proposal was funded in part by the Adopt-A-Character Program of the Unicode Consortium. A previous version was made possible in part through a Google Research Award, granted to Deborah Anderson for the Script Encoding Initiative, which funded a post-doctoral research position for me in the Department of Linguistics, University of California, Berkeley during 2015–2016. Preliminary research was made possible through the Script Encoding Initiative at Berkeley. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the Unicode Consortium; the University of California, Berkeley; or Google.

	10E0	10E1	10E2
0	𐰀 10E00	𐰁 10E10	𐰂 10E20
1	𐰃 10E01	𐰄 10E11	𐰅 10E21
2	𐰆 10E02	𐰇 10E12	𐰈 10E22
3	𐰉 10E03	𐰊 10E13	𐰋 10E23
4	𐰌 10E04	𐰍 10E14	𐰎 10E24
5	𐰏 10E05	𐰐 10E15	𐰑 10E25
6	𐰒 10E06	𐰓 10E16	𐰔 10E26
7	𐰕 10E07	𐰖 10E17	𐰗 10E27
8	𐰘 10E08	𐰙 10E18	𐰚 10E28
9	𐰛 10E09	𐰜 10E19	𐰝 10E29
A	𐰞 10E0A	𐰟 10E1A	𐰠 10E2A
B	𐰡 10E0B	𐰢 10E1B	𐰣 10E2B
C	𐰤 10E0C	𐰥 10E1C	𐰦 10E2C
D	𐰧 10E0D	𐰨 10E1D	𐰩 10E2D
E	𐰪 10E0E	𐰫 10E1E	𐰬 10E2E
F	𐰭 10E0F	𐰮 10E1F	𐰯 10E2F

This block unifies the scripts used in the Ancient Letters and the Kultobe and Upper Indus inscriptions.

### Letters

10E00	𐰀	OLD SOGDIAN LETTER ALEPH
10E01	𐰃	OLD SOGDIAN LETTER FINAL ALEPH
10E02	𐰆	OLD SOGDIAN LETTER BETH
10E03	𐰉	OLD SOGDIAN LETTER FINAL BETH
10E04	𐰌	OLD SOGDIAN LETTER GIMEL
10E05	𐰏	<reserved>
10E06	𐰒	OLD SOGDIAN LETTER HE
10E07	𐰕	OLD SOGDIAN LETTER FINAL HE
10E08	𐰘	OLD SOGDIAN LETTER WAW
10E09	𐰛	OLD SOGDIAN LETTER ZAYIN
10E0A	𐰞	OLD SOGDIAN LETTER HETH
10E0B	𐰡	OLD SOGDIAN LETTER YODH
10E0C	𐰤	OLD SOGDIAN LETTER KAPH
10E0D	𐰧	OLD SOGDIAN LETTER LAMEDH
10E0E	𐰪	OLD SOGDIAN LETTER MEM
10E0F	𐰭	OLD SOGDIAN LETTER NUN
10E10	𐰛	OLD SOGDIAN LETTER FINAL NUN
10E11	𐰜	OLD SOGDIAN LETTER VERTICAL FINAL NUN
10E12	𐰟	OLD SOGDIAN LETTER SAMEKH
10E13	𐰛	OLD SOGDIAN LETTER AYIN
10E14	𐰟	OLD SOGDIAN LETTER ALTERNATE AYIN
10E15	𐰡	OLD SOGDIAN LETTER PE
10E16	𐰡	OLD SOGDIAN LETTER SADHE
10E17	𐰡	OLD SOGDIAN LETTER FINAL SADHE
10E18	𐰛	OLD SOGDIAN LETTER VERTICAL FINAL SADHE
10E19	𐰛	OLD SOGDIAN LETTER RESH-DALETH-AYIN
10E1A	𐰡	OLD SOGDIAN LETTER SHIN
10E1B	𐰡	OLD SOGDIAN LETTER TAW
10E1C	𐰡	OLD SOGDIAN LETTER FINAL TAW
10E1D	𐰡	OLD SOGDIAN LETTER VERTICAL FINAL TAW

### Numbers

10E1E	𐰛	OLD SOGDIAN NUMBER ONE
10E1F	𐰡	OLD SOGDIAN NUMBER TWO
10E20	𐰂	OLD SOGDIAN NUMBER THREE
10E21	𐰅	OLD SOGDIAN NUMBER FOUR
10E22	𐰈	OLD SOGDIAN NUMBER FIVE
10E23	𐰘	OLD SOGDIAN NUMBER TEN
10E24	𐰛	OLD SOGDIAN NUMBER TWENTY
10E25	𐰛	OLD SOGDIAN NUMBER THIRTY
10E26	𐰔	OLD SOGDIAN NUMBER ONE HUNDRED
10E27	𐰗	OLD SOGDIAN FRACTION ONE HALF

### Heterogram

10E28	𐰚	OLD SOGDIAN HETEROGRAM OD
-------	---	---------------------------

	Old Sogdian	Inscriptional Pahlavi	Inscriptional Parthian	Imperial Aramaic
<i>aleph</i>	𐰀, 𐰁	𐰀	𐰀	𐤀
<i>beth</i>	𐰂, 𐰃	𐰂	𐰂	𐤁
<i>gimel</i>	𐰄	𐰄	𐰄	𐤂
<i>daleth</i>	(𐰅)	𐰅	𐰅	𐤃
<i>he</i>	𐰆, 𐰇	𐰆	𐰆	𐤄
<i>waw</i>	𐰈	𐰈	𐰈	𐤅
<i>zayin</i>	𐰉	𐰉	𐰉	𐤆
<i>heth</i>	𐰊	𐰊	𐰊	𐤇
<i>teth</i>	—	𐰋	𐰋	𐤈
<i>yodh</i>	𐰌	𐰌	𐰌	𐤉
<i>kaph</i>	𐰍	𐰍	𐰍	𐤊
<i>lamedh</i>	𐰎	𐰎	𐰎	𐤋
<i>mem</i>	𐰏	𐰏	𐰏	𐤌
<i>nun</i>	𐰐, 𐰑, 𐰒	𐰐	𐰐	𐤍
<i>samekh</i>	𐰓	𐰓	𐰓	𐤎
<i>ayin</i>	𐰔, 𐰕, (𐰖)	(2)	𐰔	𐤏
<i>pe</i>	𐰗	𐰗	𐰗	𐤐
<i>sadhe</i>	𐰘, 𐰙, 𐰚	𐰘	𐰘	𐤑
<i>qoph</i>	—	(𐰛)	𐰛	𐤒
<i>resh</i>	𐰜	(2)	𐰜	𐤓
<i>shin</i>	𐰝	𐰝	𐰝	𐤔
<i>taw</i>	𐰞, 𐰟, 𐰠	𐰞	𐰞	𐤕

Table 1: Comparison of Old Sogdian letters with those in Unicode blocks for related Iranian scripts and Aramaic. Parenthesis indicate that a letter has been unified with another in the respective encoding. In Inscriptional Pahlavi, *ayin* and *resh* are unified with *waw*, and *qoph* with *mem*.

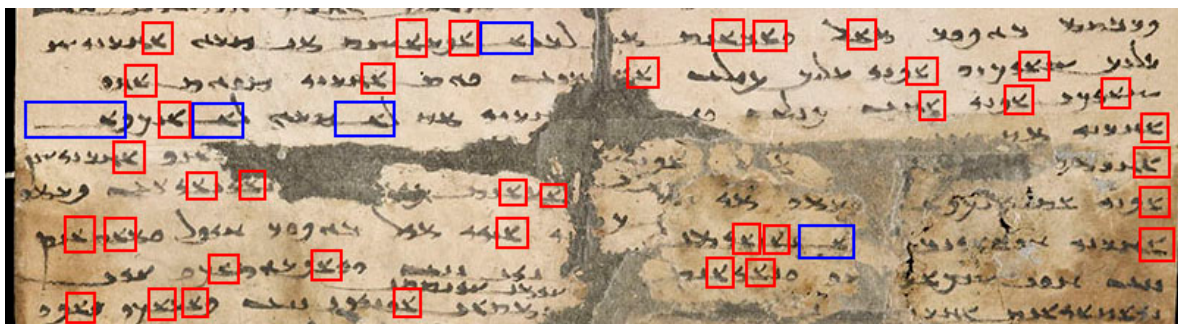
	Old Sogdian	Inscriptional Pahlavi	Inscriptional Parthian	Imperial Aramaic
ONE	𐰪	𐰬	𐰭	𐤁
TWO	𐰫	𐰮	𐰯	𐤂
THREE	𐰬	𐰱	𐰲	𐤃
FOUR	𐰭	𐰴	𐰵	—
FIVE	𐰮	—	—	—
TEN	𐰱	𐰶	𐰷	𐤄
TWENTY	𐰲	𐰹	𐰸	𐤅
THIRTY	𐰳	—	—	—
ONE HUNDRED	𐰶	𐰺	𐰹	𐤆
ONE THOUSAND	—	𐰻	𐰺	𐤇
TEN THOUSAND	—	—	—	𐤈
ONE HALF	𐰷	—	—	—

Table 2: Comparison of Old Sogdian number signs with those of related Iranian scripts and Aramaic.





Archaic form **x** of **𐰀** ALEPH (K 4.1–4).

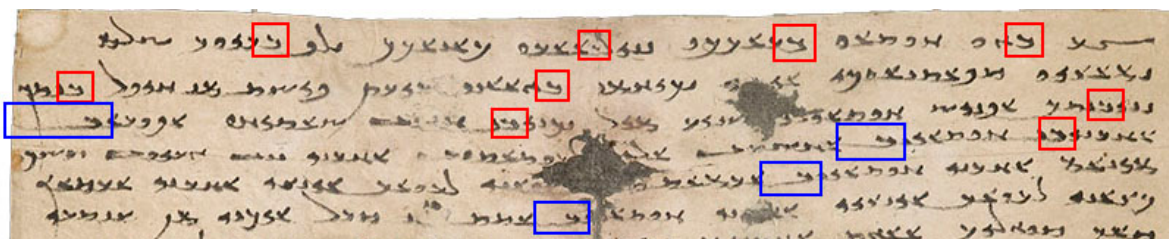


Usage of **𐰀** ALEPH (red) and **𐰀** FINAL ALEPH (blue) (AL 2.1–6).

Figure 1: Specimens of *aleph*.

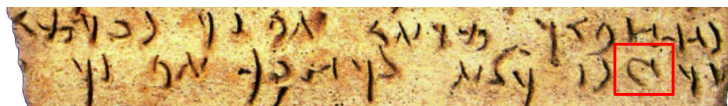


Inscriptional forms of 𐰪 BETH (K 4.1–2).

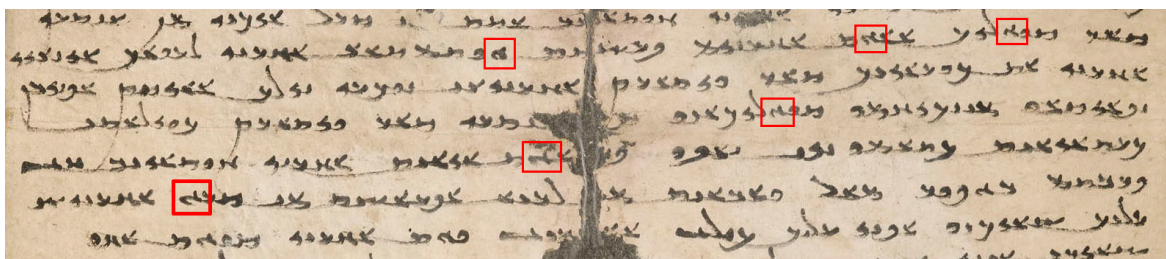


Usage of 𐰪 BETH (red) and 𐰫 FINAL BETH (blue) (AL 2.1–6).

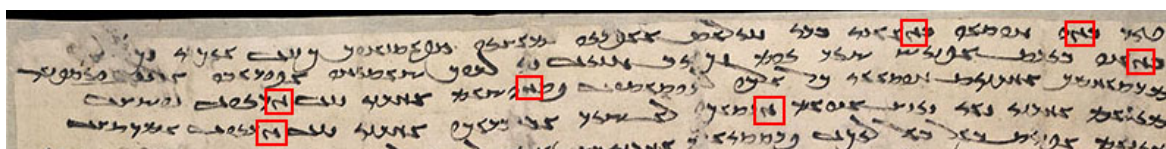
Figure 2: Specimens of *beth*.



Inscripational form of **𐰇** GIMEL (K 4.6).



Usage of **𐰇** GIMEL (AL 2.7–12).

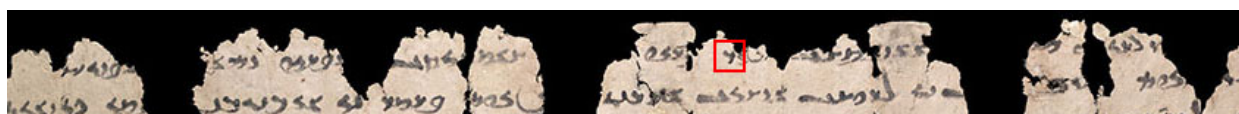


Usage of **𐰇** GIMEL (AL 3.1–4).

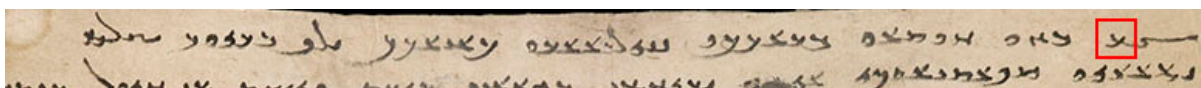
Figure 3: Specimens of *gimel*



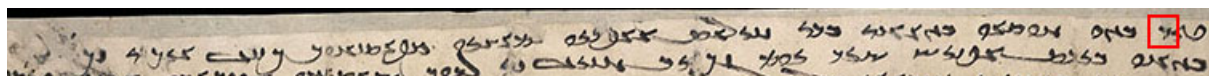
The letter *daleth* in 𐰽𐰺𐰽 'BDt' written as 𐰽 (= RESH-DALETH-AYIN) (K 4.1).



The letter *daleth* in 𐰽𐰺𐰽 'D' written as 𐰽 (= RESH-DALETH-AYIN) (AL 1.1).



The letter *daleth* in 𐰽𐰺 'D' written as 𐰽 (= RESH-DALETH-AYIN) (AL 2.1).



The letter *daleth* in 𐰽𐰺 'D' written as 𐰽 (= RESH-DALETH-AYIN) (AL 3.1).

Figure 4: Specimens of *daleth*.



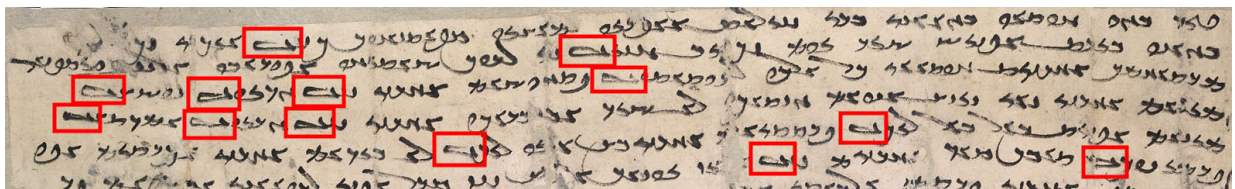
The letter **𐰪** HE in **𐰪𐰸𐰸𐰸** *ṢWRH* and **𐰪𐰸** (*H*)*WH* (K 2.3–4).



The letter **𐰪** HE in **𐰪𐰸** *ZNH*, **𐰪𐰸𐰸** *knth*, **𐰪𐰸𐰸** *TMH* (K 4.1–2).

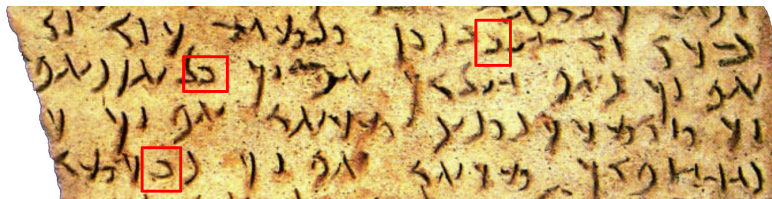


The letter **𐰪** FINAL HE in **𐰪𐰸** *ZNH* and **𐰪𐰸𐰸** *knḥ* (AL 2.10, 12).

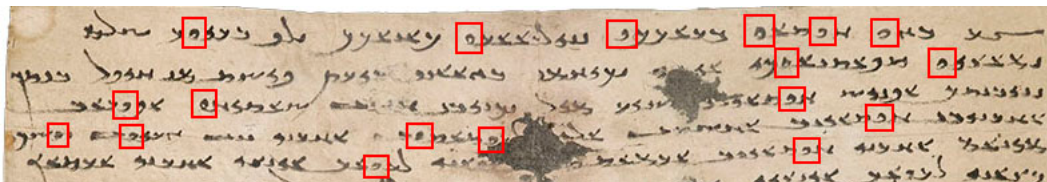


Ubiquitous usage of **𐰪** FINAL HE in AL 3.1–6.

Figure 5: Specimens of *he*.



Inscriptional forms of 𐰪 waw (K 4.2-4).

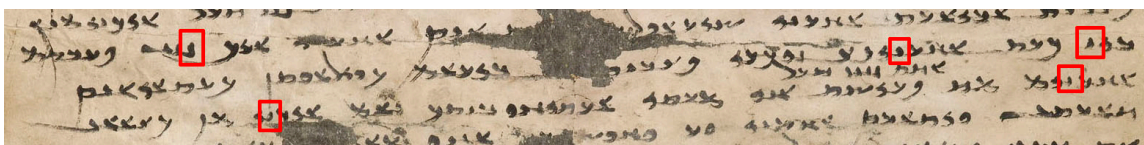


Usage of 𐰪 waw (AL 2.1-5).

Figure 6: Specimens of *waw*.

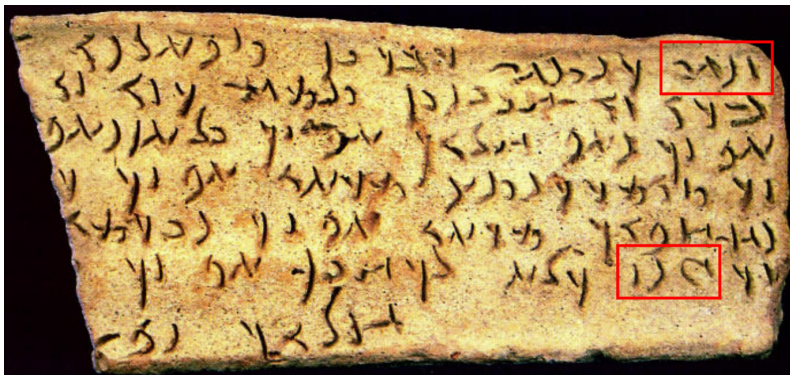


Inscripational form of 𐰽 ZAYIN (K 4).

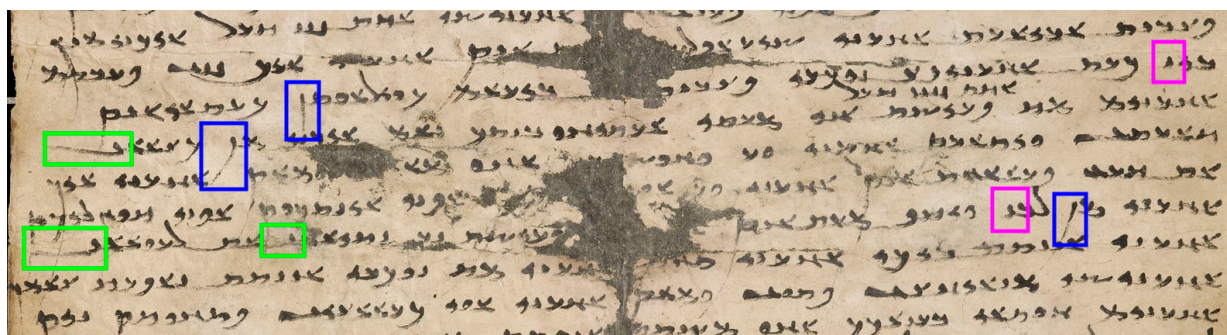


Usage of 𐰽 ZAYIN (AL 2.34–36).

Figure 7: Specimens of *zayin*. See also figure 8.



Distinctive representations of 𐰇 ZAYIN and 𐰆 NUN in K 4: 𐰇𐰆 ZNH (line 1) and 𐰆𐰇 GNZ (line 6).



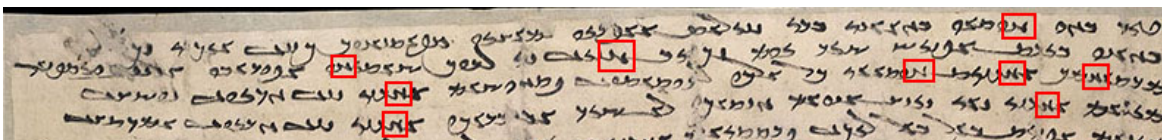
Representations of 𐰇 *zayin* (magenta) and *nun* at the end of word (AL 2.33–41). Final *nun* is written using both 𐰆 VERTICAL FINAL NUN (blue), and 𐰇 FINAL NUN (green).

Figure 8: Comparison of *zayin* and *nun*. See also figure 14.





Inscripational forms of **𐰇** HETH (K 4.3–7).

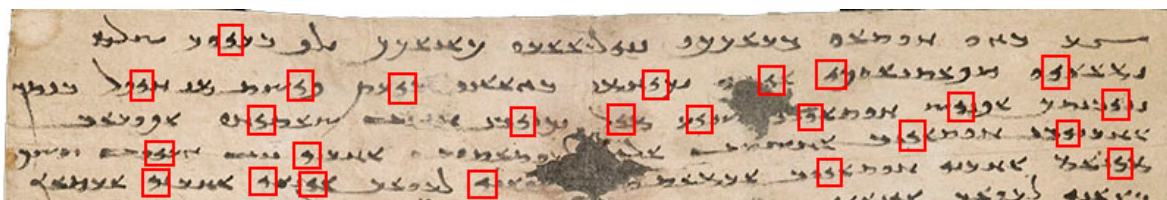


Usage of **𐰇** HETH (AL 3.1–4).

Figure 9: Specimens of *heth*.



Inscriptional forms of 𐰽 YODH (K 4.1–3).

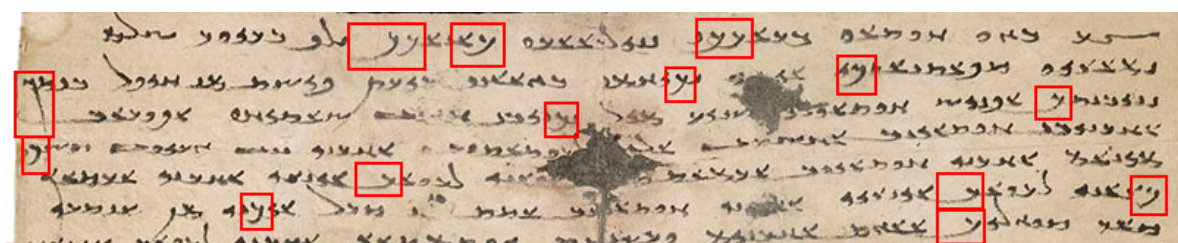


Usage of 𐰽 YODH (AL 2.1–5).

Figure 10: Specimens of *yodh*.



Inscriptional forms of  $\gamma$  KAPH (K 4.1–3).

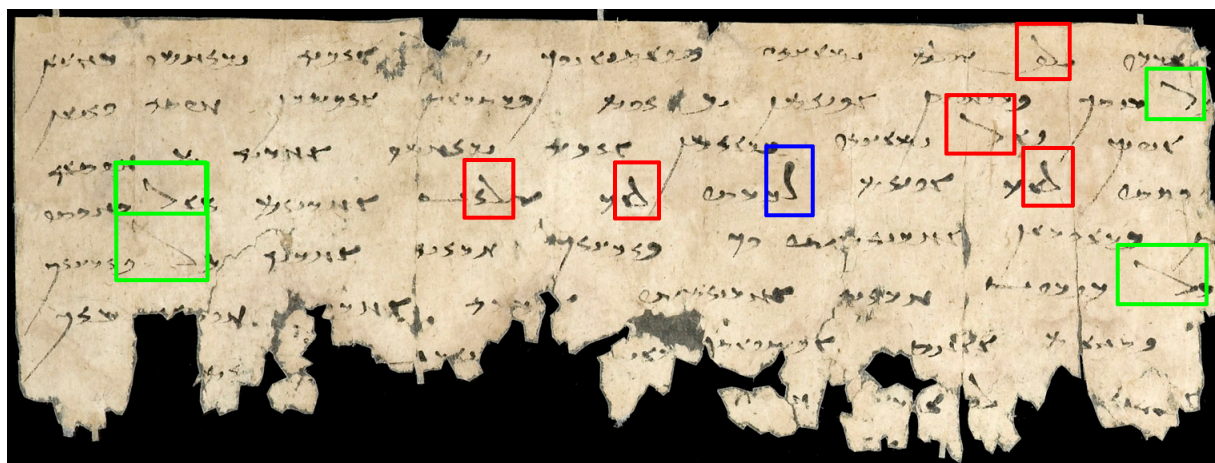


Usage of  $\gamma$  KAPH (AL 2.1–4).

Figure 11: Specimens of *kaph*.

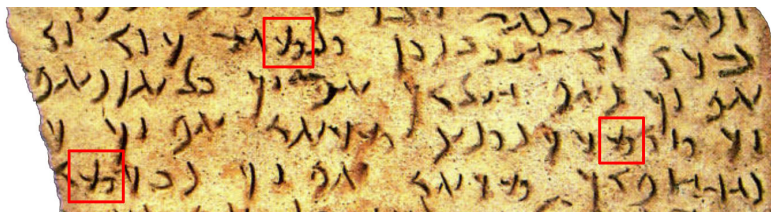


Archaic form  $\text{𐰇}$  of  $\text{𐰆}$  LAMEDH (K 4.1).

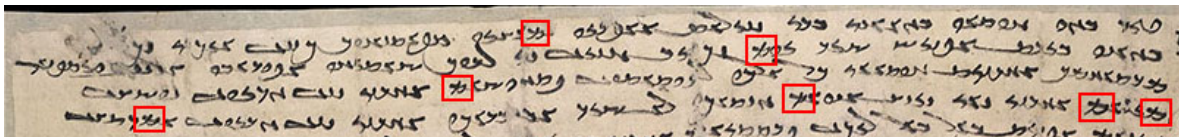


Occurrences of  $\text{𐰆}$  LAMEDH (red) and its variant forms  $\text{𐰇}$  (green) and  $\text{𐰈}$  (blue) (AL 6.1–8).

Figure 12: Specimens of *lamedh*.



Inscriptural forms of **𐰜** MEM (K 4.1–3).

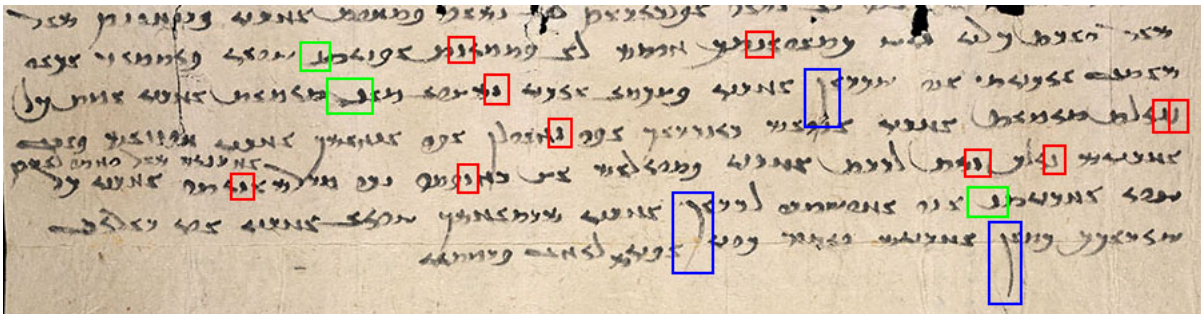


Usage of **𐰜** MEM (AL 3.1–4).

Figure 13: Specimens of *mem*.



Inscripational form of 𐰽 NUN (K 4).



Usage of 𐰽 NUN (red), 𐰾 FINAL NUN (green), 𐰿 VERTICAL FINAL NUN (blue) (AL 1.7–12).

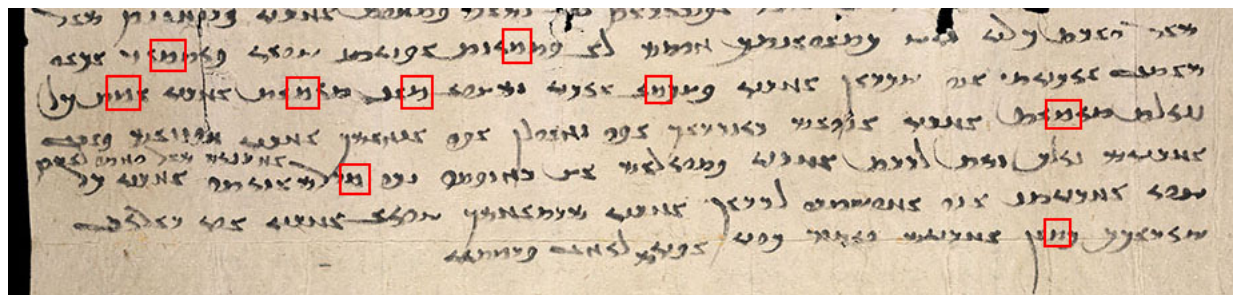


Contrastive usage of 𐰾 FINAL NUN (red) and 𐰿 VERTICAL FINAL NUN (blue) in the word MN: 𐰾𐰽 and 𐰿𐰽 (AL 2.2–7).

Figure 14: Specimens of *nun*. See also figure 8.

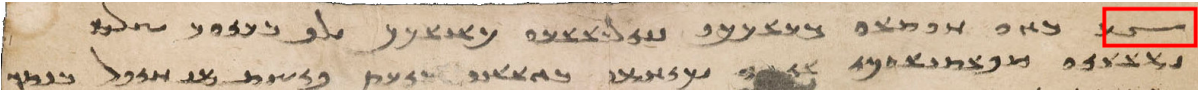


Archaic form כִּי of 𐰽 SAMEKH (K 4.1-4).



Form of 𐰽 SAMEKH (AL 1.7-12).

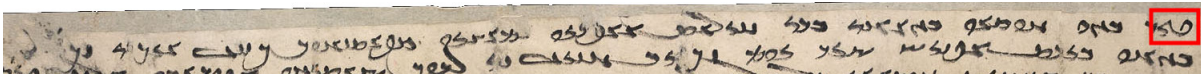
Figure 15: Specimens of *samekh*.



The letter *ayin* in  $\text{𐰇𐰆}$  'D' written using  $\text{𐰇}$  AYIN (AL 2.1).



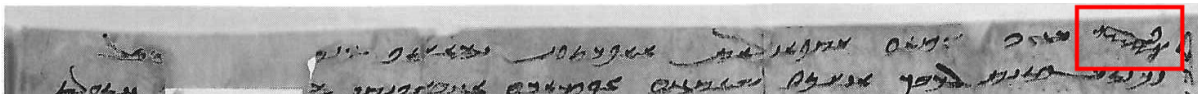
The *ayin* in  $\text{𐰇𐰆}$  'D' written using  $\text{𐰆}$  ALTERNATE AYIN (AL 1.1).



The *ayin* in  $\text{𐰇𐰆}$  'D' written using  $\text{𐰆}$  ALTERNATE AYIN (AL 3.1).



The *ayin* in  $\text{𐰇𐰆}$  'D' written using the glyphic variant  $\text{𐰆}$  of  $\text{𐰆}$  ALTERNATE AYIN (AL 3 verso).



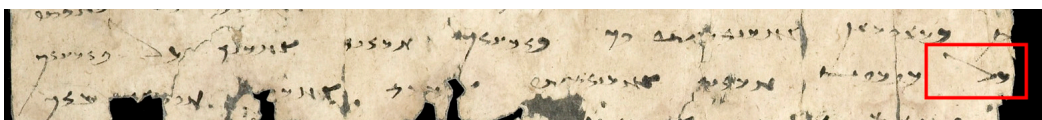
The *ayin* in  $\text{𐰇𐰆}$  'D' written using the glyphic variant  $\text{𐰆}$  of  $\text{𐰆}$  ALTERNATE AYIN (AL 5.1).



The letter *ayin* in  $\text{𐰇𐰆𐰇}$  'BDt' written as  $\text{𐰇}$  (= RESH-DALETH-AYIN) (K 4.1).



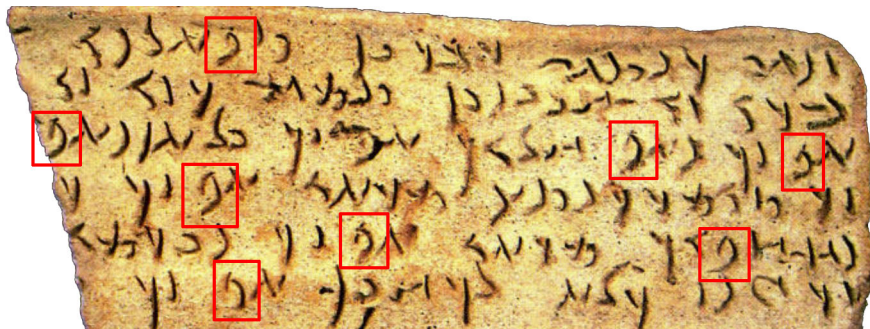
The letter *ayin* in  $\text{𐰇𐰆𐰇}$  'LZK' written using  $\text{𐰇}$  RESH-DALETH-AYIN (AL 2.12).



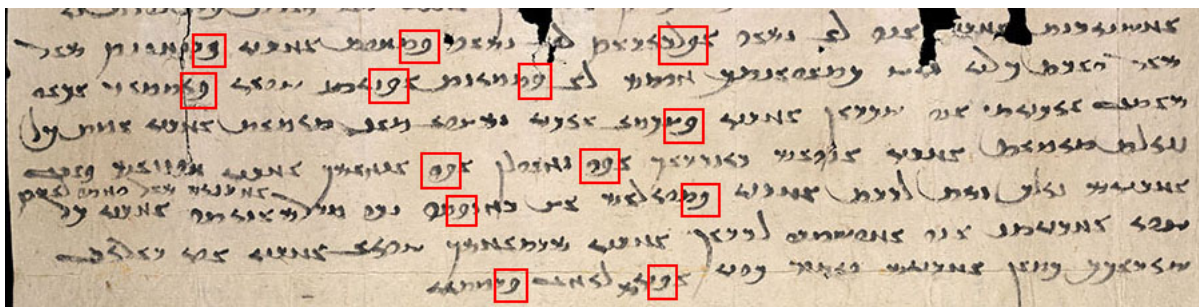
The letter *ayin* in  $\text{𐰇𐰆}$  'L' written using  $\text{𐰇}$  RESH-DALETH-AYIN (AL 6.6).

Figure 16: Specimens of *ayin*.



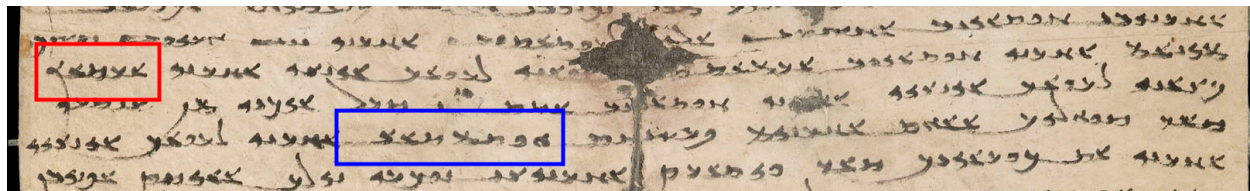


Glyphic variant 9 of 9 PE (K 4.1–6).

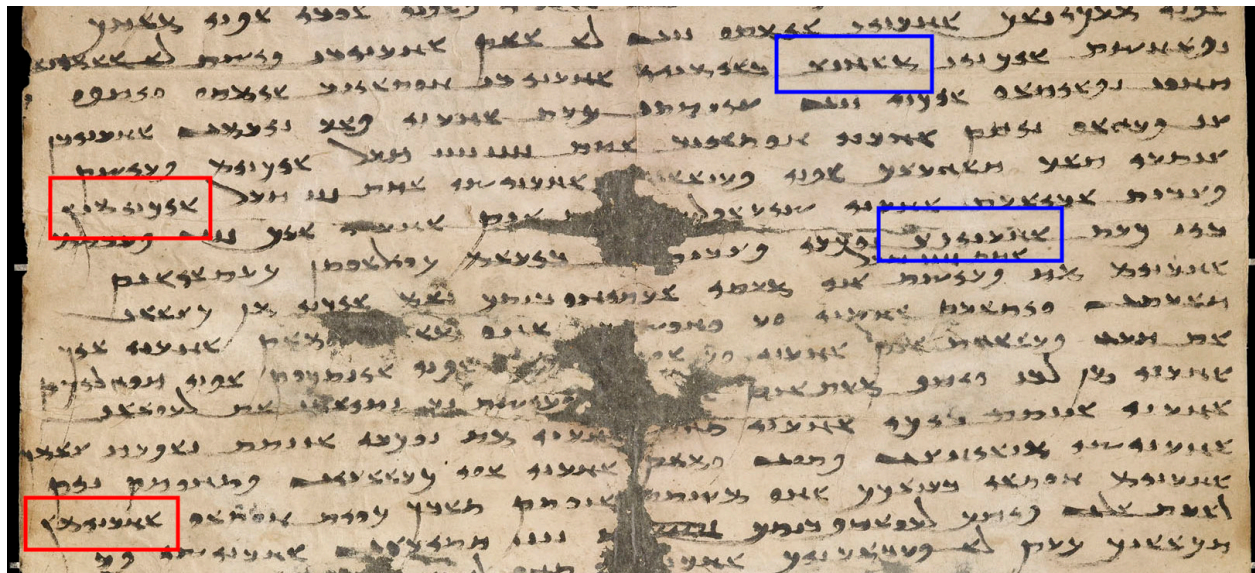


The letter 9 PE (AL 1.6–12).

Figure 17: Specimens of *pe*.



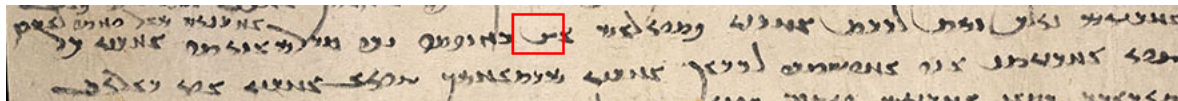
[...]



[...]

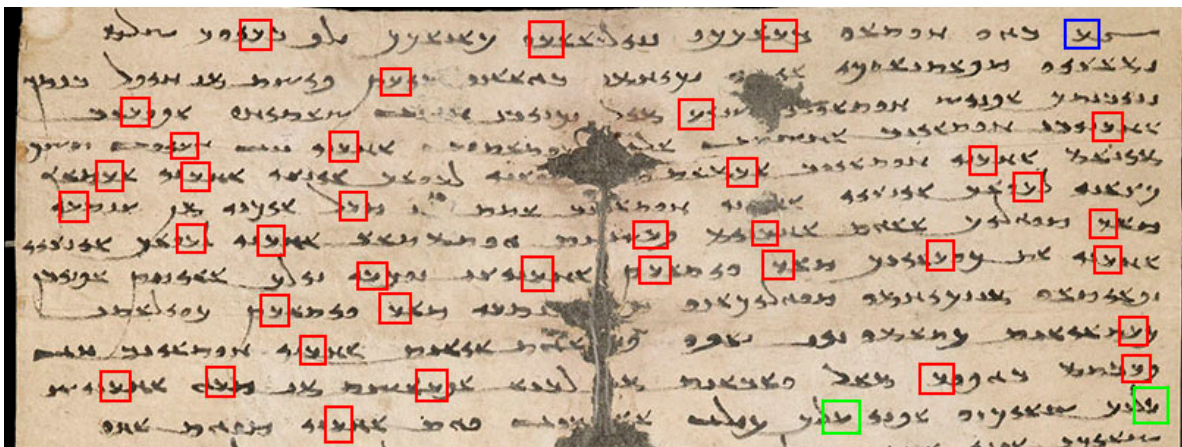


Usage of 𐰽 FINAL SADHE (blue): 𐰽𐰺𐰸𐰸𐰽 *wtš'c* (AL 2.7) 𐰽𐰺𐰸𐰸𐰽 *m'xnc* (line 29), 𐰽𐰺𐰸𐰸𐰽 *'HRZYnc* (line 34), 𐰽𐰺𐰸𐰸𐰽 *pnc* (line 59). The use of 𐰽 VERTICAL FINAL SADHE at the end of line / margin (red): 𐰽𐰺𐰸𐰸𐰽 *rs'c* (line 5), 𐰽𐰺𐰸𐰸𐰽 *'YKZYmnc* (line 32), 𐰽𐰺𐰸𐰸𐰽 *'HRZYmc* (line 41), 𐰽𐰺𐰸𐰸𐰽 *'HRZYnnc* (line 54).



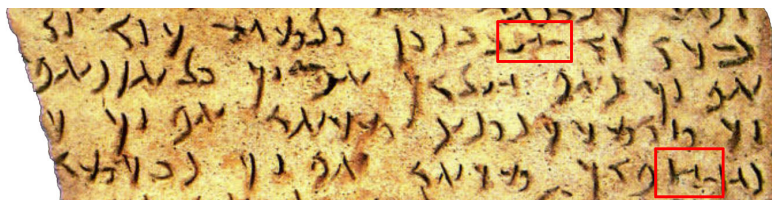
Curved variant 𐰽 of 𐰽 FINAL SADHE (AL 1.10).

Figure 18: Specimens of *sadhe*.

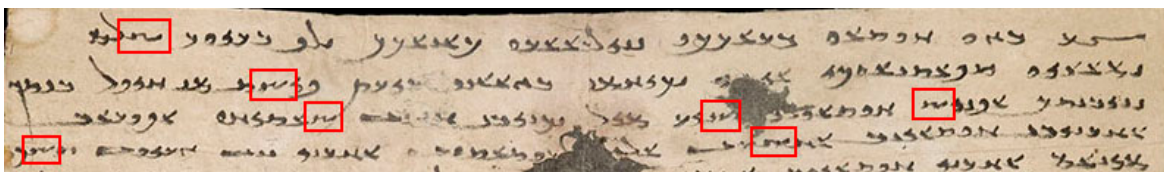


Usage of  $\text{𐰽}$  for representing *resh* (red), *daleth* (blue), and *ayin* (green) (AL 2.1–12). As shown,  $\text{𐰽}$  is most commonly used for *resh*. The letter  $\text{𐰽}$  is proposed for encoding as the unified character RESH-DALETH-AYIN.

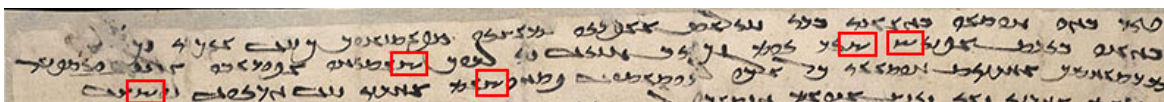
Figure 19: Comparison of *daleth*, *ayin*, and *resh*.



Inscriptional forms of 𐰽 SHIN (K 4.1–3).



Usage of 𐰽 SHIN (AL 2.1–4).

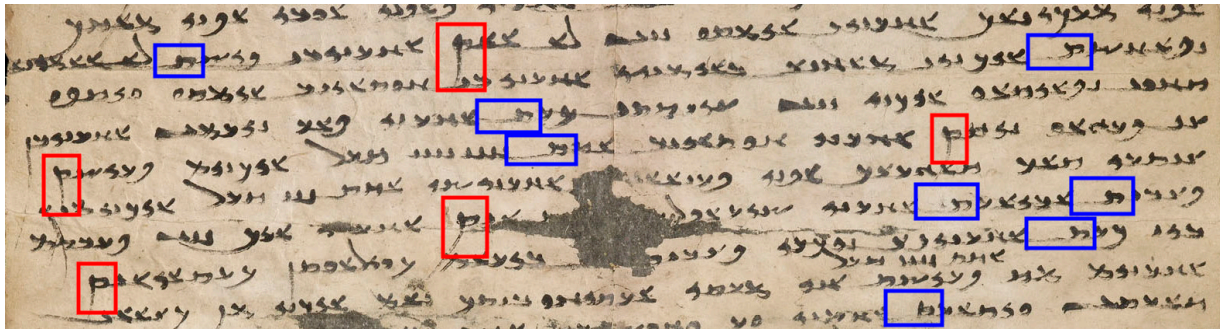


Usage of 𐰽 SHIN (AL 3.1–3).

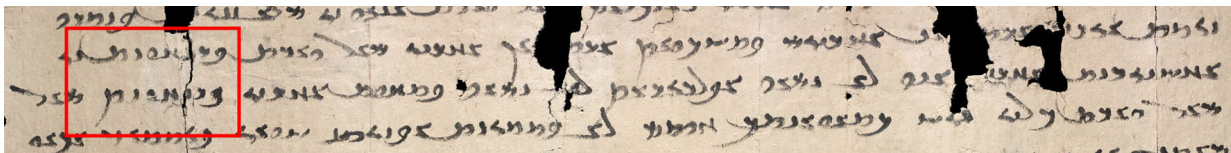
Figure 20: Specimens of *shin*.



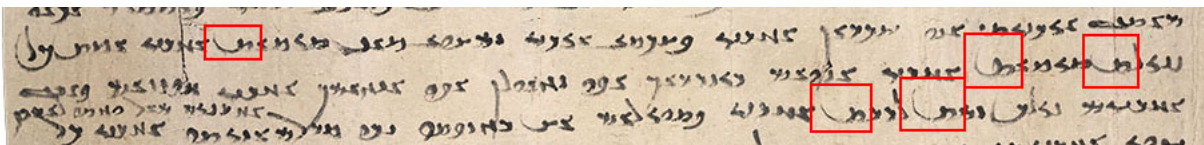
Archaic forms 𐰪 (red) and 𐰫 (blue) of 𐰪 TAW and 𐰫 VERTICAL FINAL TAW (K 4.1–2). The distinction is apparent in 𐰪𐰫𐰪𐰫 *šwt* (line 2).



Usage of 𐰬 FINAL TAW (blue) and 𐰫 VERTICAL FINAL TAW (red) at the end of word (AL 2.28–36).

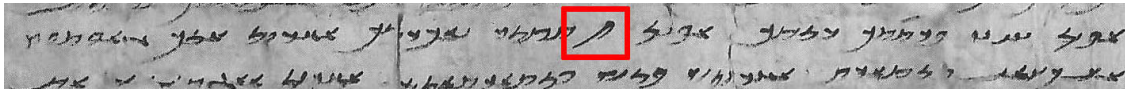


Contrastive usage of 𐰬 FINAL TAW and 𐰫 VERTICAL FINAL TAW in two instances of the word *prnxwnt*: 𐰬𐰪𐰫𐰪 and 𐰫𐰪𐰫𐰪 (AL 1.5–6).



Curved variant 𐰭 of 𐰬 FINAL TAW (AL 1.8–10).

Figure 21: Specimens of *taw*.



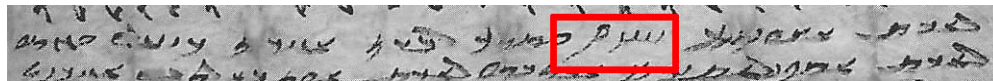
The fraction  $\frac{1}{2}$  ρ (AL 5.10).



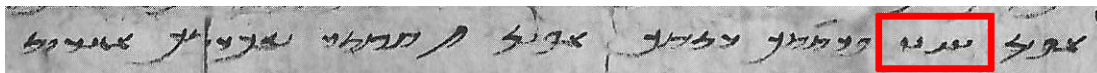
The number 3 𐰇 (AL 2.32).



The number 4 𐰇𐰇 (AL 5.26).



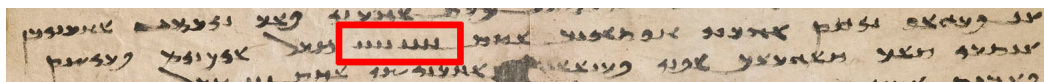
The number  $4\frac{1}{2}$  𐰇𐰇𐰇 (AL 5.24).



The number 5 𐰇𐰇𐰇 (AL 5.10).

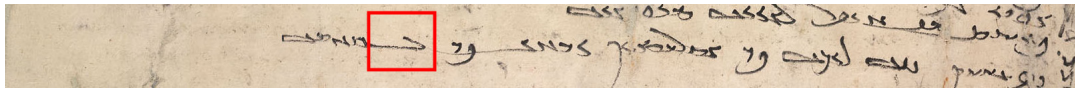


The number  $7\frac{1}{2}$  𐰇𐰇𐰇𐰇 (AL 5.26).

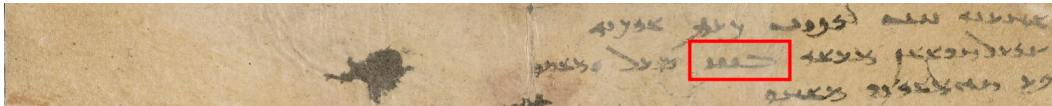


The number 8 𐰇𐰇𐰇𐰇 (AL 2.31).

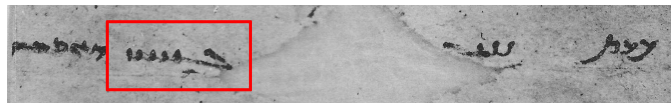
Figure 22: Examples of numbers in the ‘Ancient Letters’. See also figures 23 and 24.



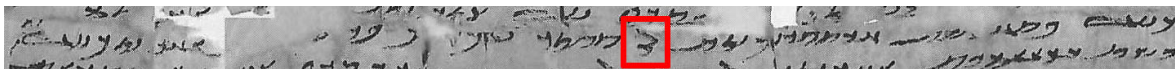
The number 10 𐰽 (AL 3.26).



The number 13 𐰽𐰽 (AL 2.62).



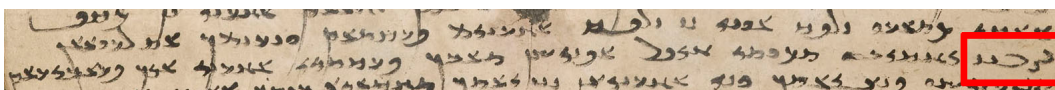
The number 15 𐰽𐰽𐰽 (AL 7.8).



The number 20 𐰽𐰽𐰽𐰽 (AL 5.21).



The number 30 𐰽𐰽𐰽𐰽𐰽 (AL 5.32).



The number 32 𐰽𐰽𐰽𐰽𐰽𐰽 (AL 2.62).

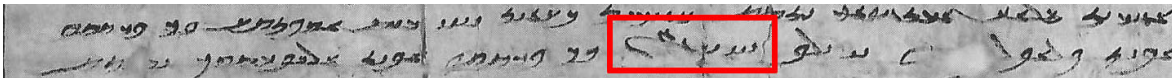
Figure 23: Additional examples of numbers in the ‘Ancient Letters’. See also figures 22 and 24.



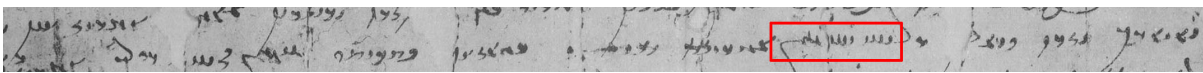
The number 100 **𐰪** (AL 2.19).



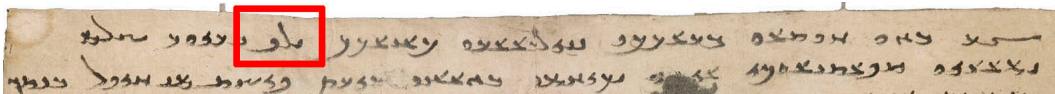
The number 200 **𐰪 𐰪** (AL 7.3).



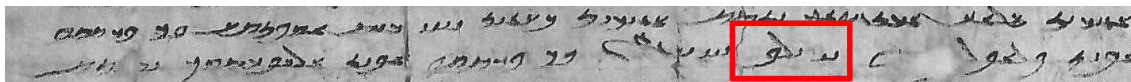
The number 500 **𐰪 𐰪 𐰪** (AL 5.9).



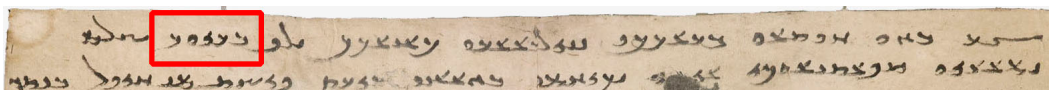
The number 800 **𐰪 𐰪 𐰪 𐰪** (AL 4.3).



The number 1000 **𐰪 𐰪 𐰪 𐰪** (AL 2.1).



The number 2000 **𐰪 𐰪 𐰪 𐰪 𐰪** (AL 5.9).



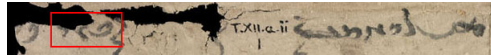
The number 10000 represented using the word **𐰪 𐰪 𐰪 𐰪 𐰪 𐰪 𐰪 𐰪 𐰪 𐰪** *brywr* (AL 2.1).

Figure 24: Further examples of numbers in the ‘Ancient Letters’. See also figures 22 and 23.

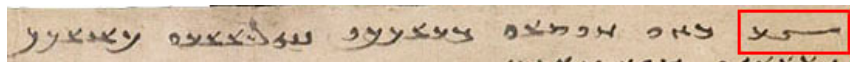




The heterogram 'D' written as 𐰽𐰺 <𐰺 ALTERNATE AYIN, 𐰽 RESH-DALETH-AYIN> (AL 1.1).



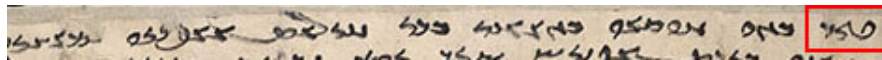
The heterogram 'D' written as (𐰽)𐰺 <𐰺 ALTERNATE AYIN, (𐰽 RESH-DALETH-AYIN)> (AL 1 verso).



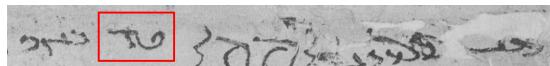
The heterogram 'D' written as 𐰽𐰾 <𐰾 AYIN, 𐰽 RESH-DALETH-AYIN> (AL 2.1).



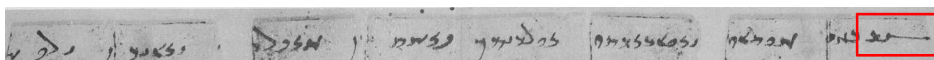
The heterogram 'D' written as 𐰽𐰾 <𐰾 AYIN, 𐰽 RESH-DALETH-AYIN> (AL 2 verso).



The heterogram 'D' written as 𐰽𐰺 <𐰺 ALTERNATE AYIN, 𐰽 RESH-DALETH-AYIN> (AL 3.1).



The heterogram 'D' written as the ligature 𐰽𐰺 HETEROGRAM OD (AL 3 verso).



The heterogram 'D' written as 𐰽𐰾 <𐰾 AYIN, 𐰽 RESH-DALETH-AYIN> (AL 4.1).



The heterogram 'D' written as 𐰽𐰺 <𐰺 ALTERNATE AYIN, 𐰽 RESH-DALETH-AYIN> using the glyphic variant 𐰽 of ALTERNATE AYIN (AL 5.1).

Figure 25: Specimens of the heterogram 'D'.

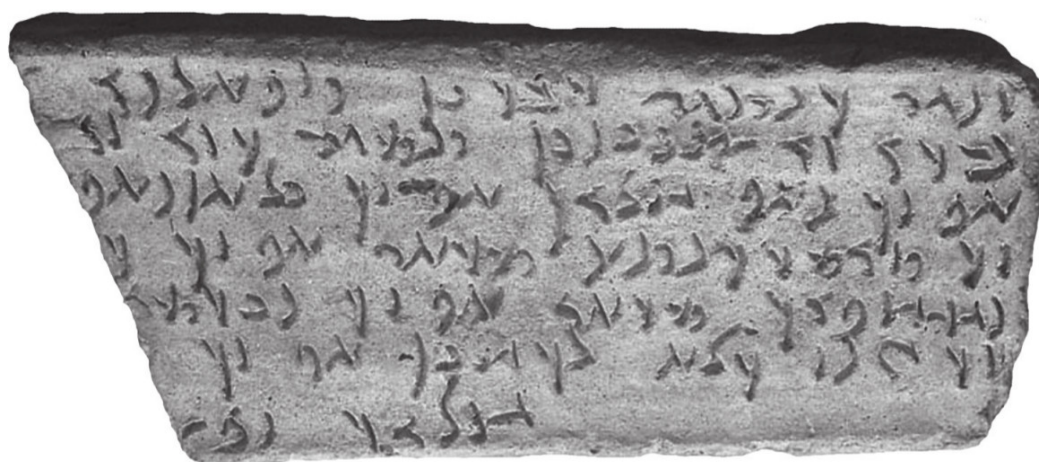
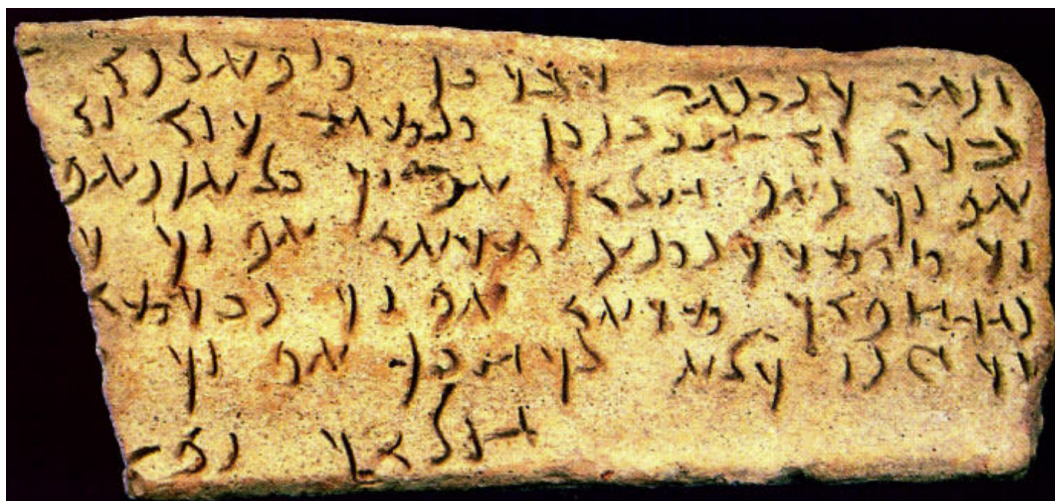


Figure 26: Two images of Kultobe inscription 4 (KII 26859/1). Top from Sims-Williams 2007; bottom from Grenet, et al 2007.



Figure 27: Kultobe inscriptions 2, 1, 3, 5, 10 (from Grenet, et al 2007).

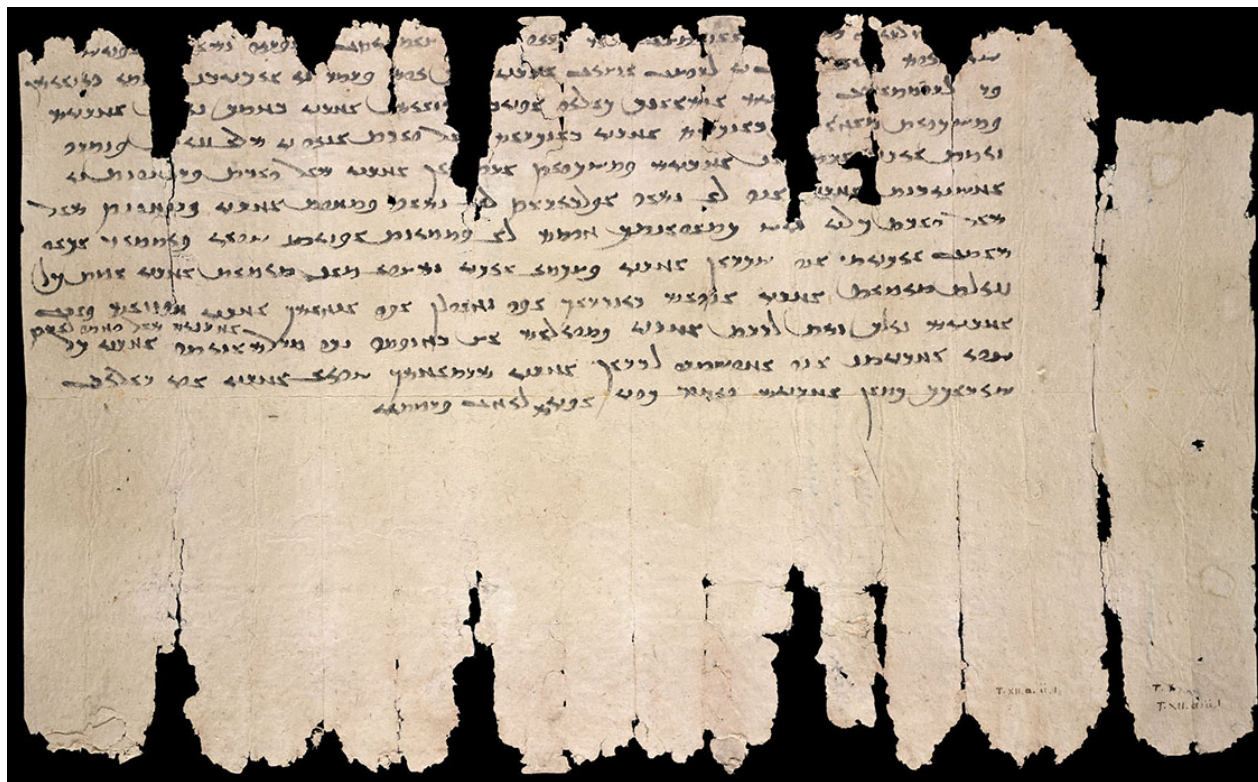


Figure 28: The ‘Ancient Letter 1’ (British Library, International Dunhuang Project: Or. 8212/92.1 recto 1). “From her daughter, the free-woman Miwnay, to her d[ear] mother [Chatis].” (translation by Sims-Williams in Waugh 2004).

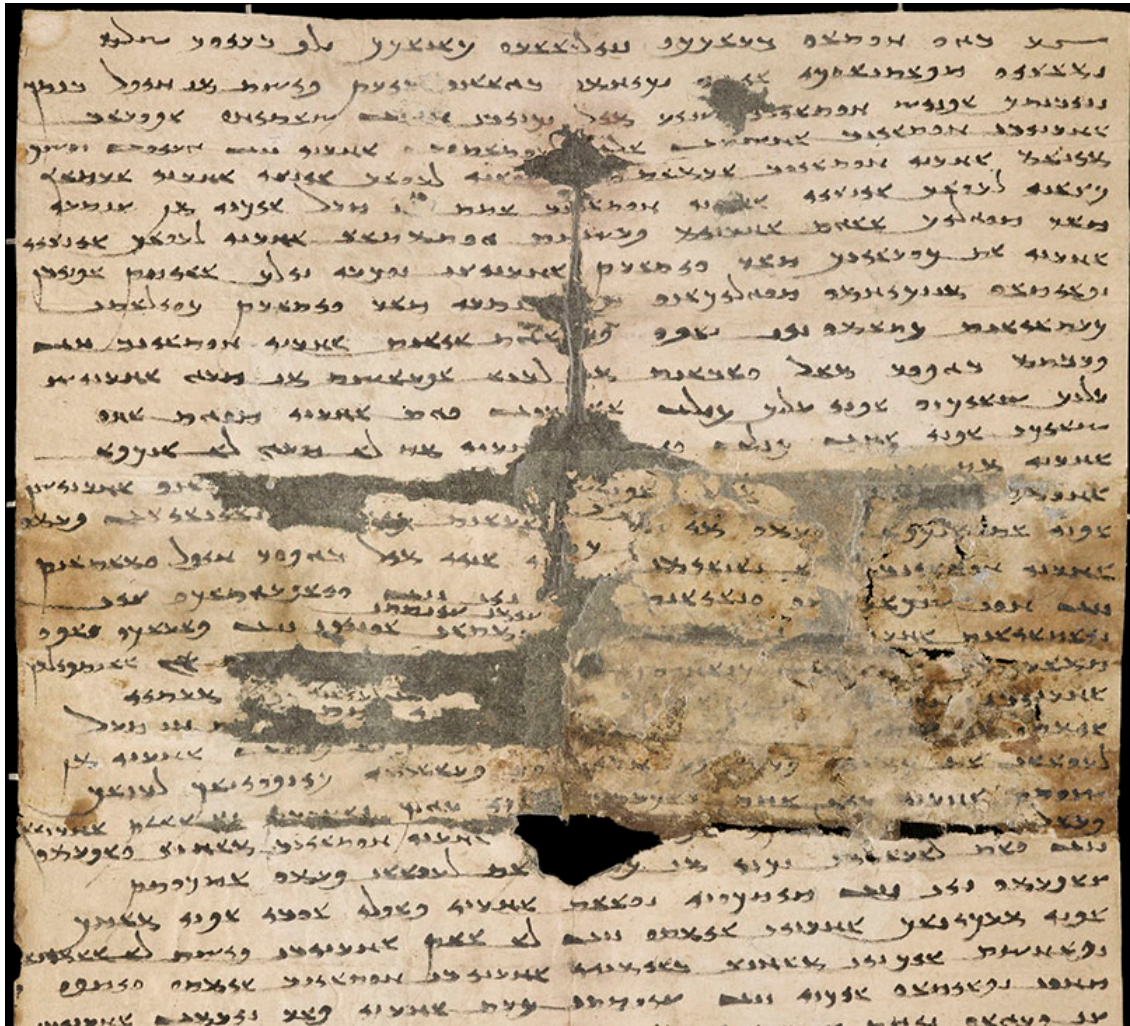


Figure 29: Top portion of ‘Ancient Letter 2’ (British Library, International Dunhuang Project: Or. 8212/95 side a). “To the noble lord Varzakk (son of) Nanai-thvar (of the family) Kanakk. Sent [by] his servant Nanai-vandak.” (translation by Sims-Williams in Waugh 2004). Continued in figure 30.



Figure 30: Bottom portion of ‘Ancient Letter 2’ (British Library, International Dunhuang Project: Or. 8212/95 side a). Continued from figure 29.

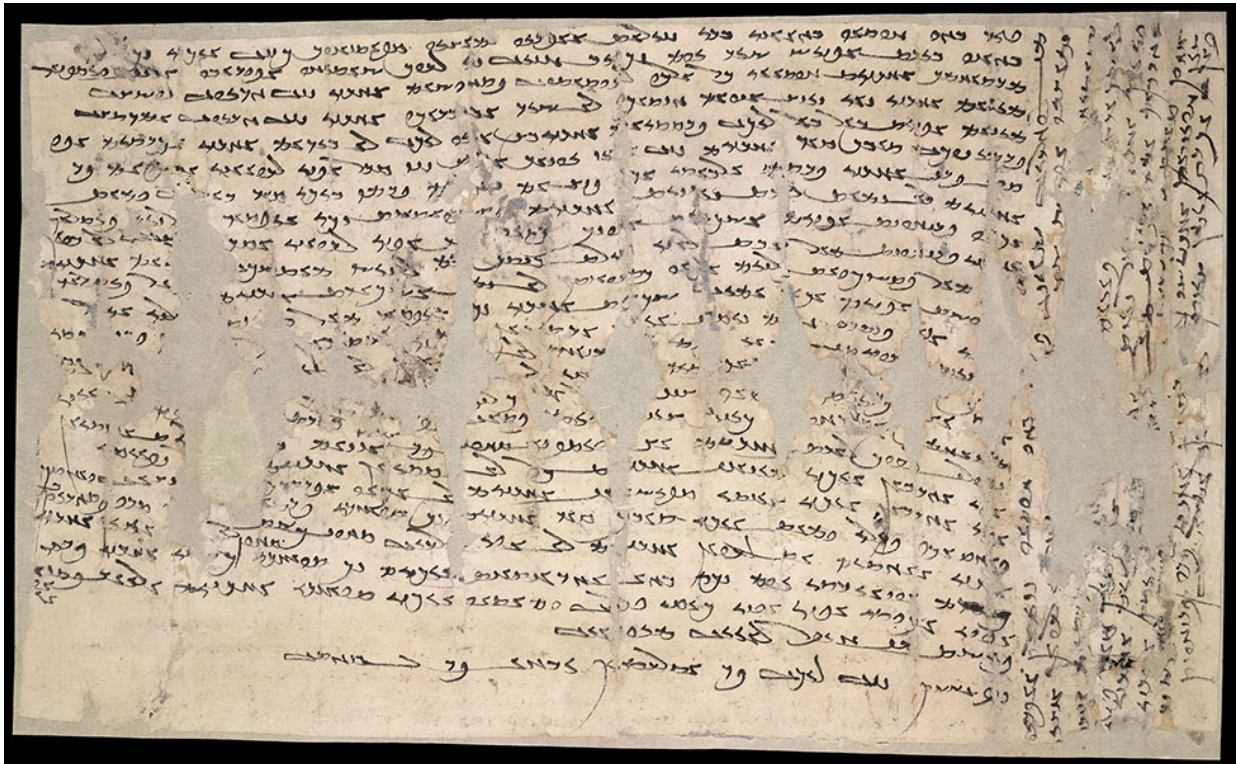


Figure 31: The ‘Ancient Letter 3’ (British Library, International Dunhuang Project: Or. 8212/98 recto 1). “From (his) daughter Shayn to the noble lord Nanai-dhat.” (translation by Sims-Williams in Waugh 2004).

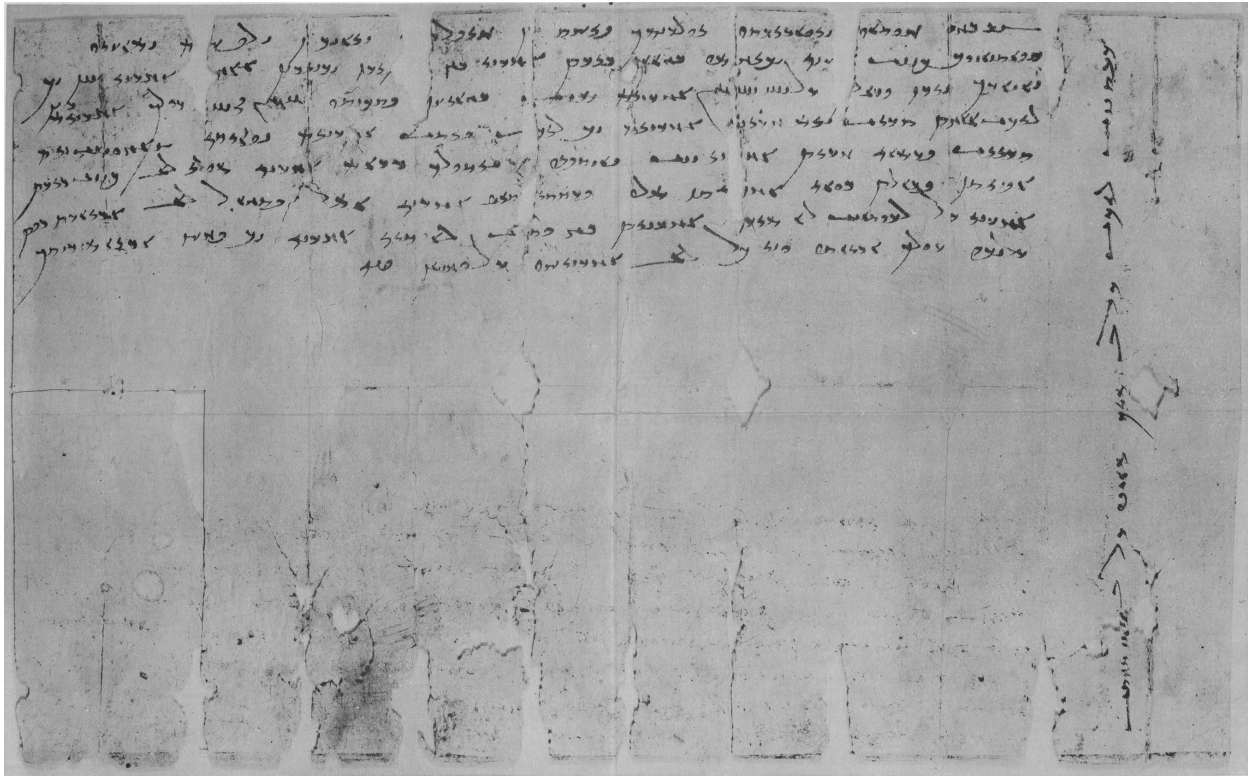


Figure 32: The ‘Ancient Letter 4’ (British Library: Or. 8212/93 recto; reproduced in Reichelt 1928: plate IV).



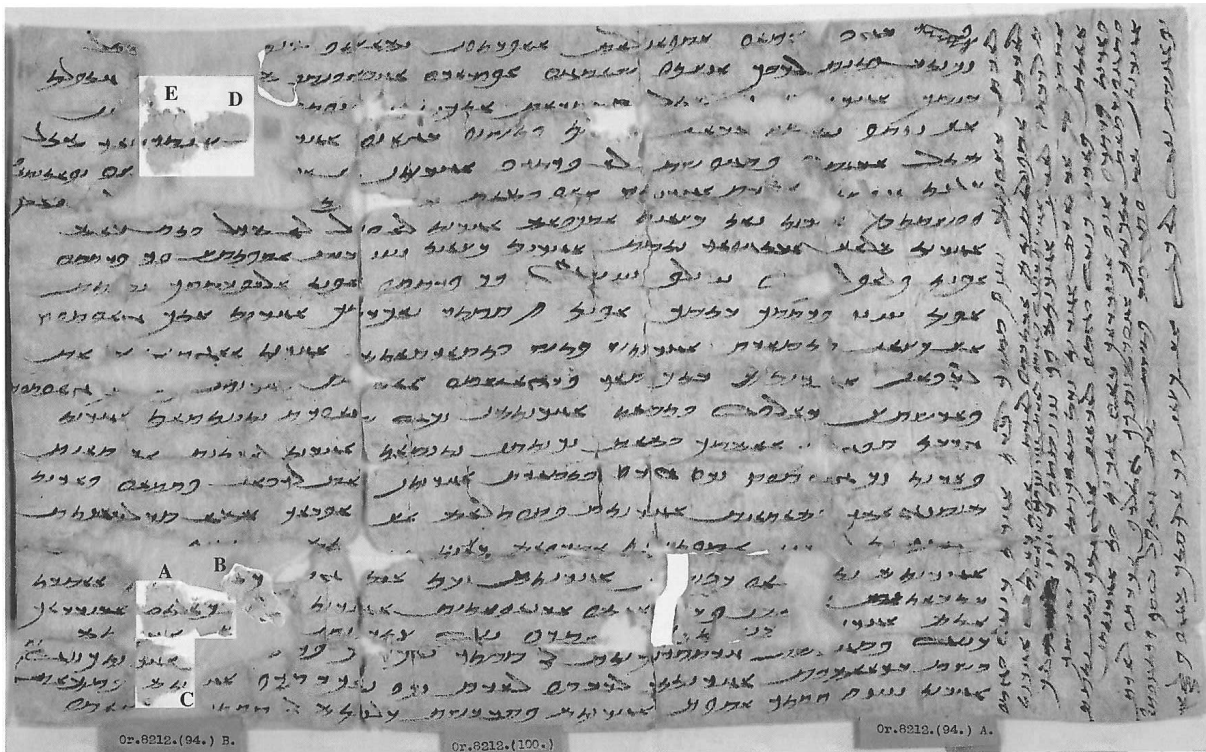


Figure 33: The ‘Ancient Letter 5’ (from Grenet, et al. 1998: 94). “To the noble lord, the chief merchant Aspandhāt. [Sent] by your servant [Frī-khwatāw].”

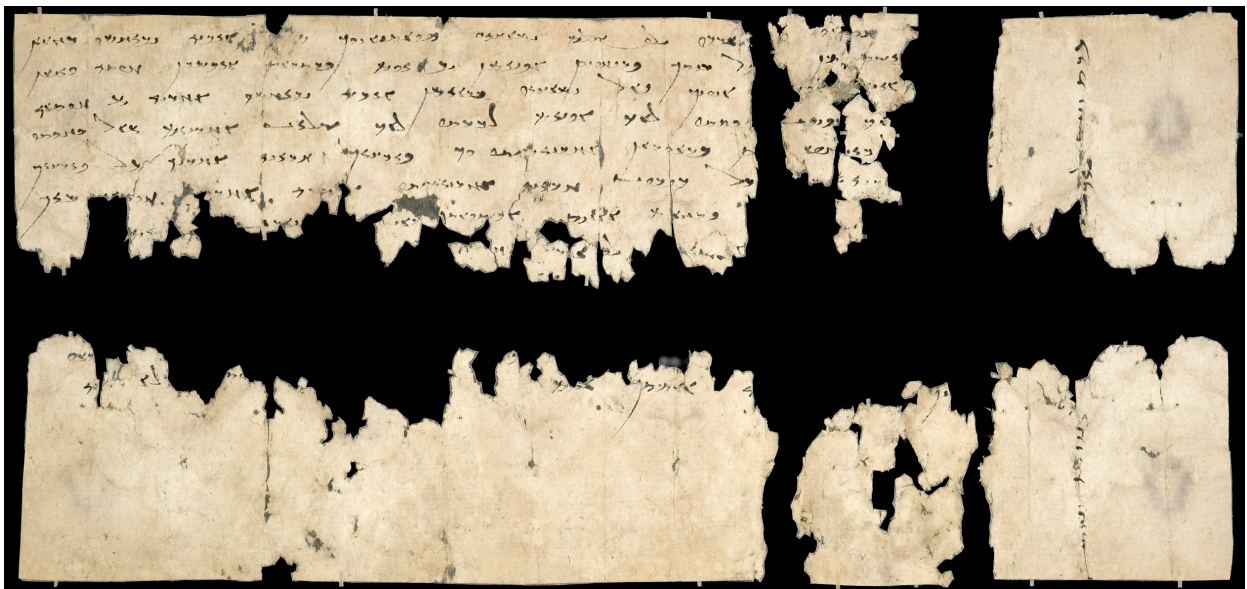


Figure 34: The ‘Ancient Letter 6’ (British Library, International Dunhuang Project: Or. 8212/97).

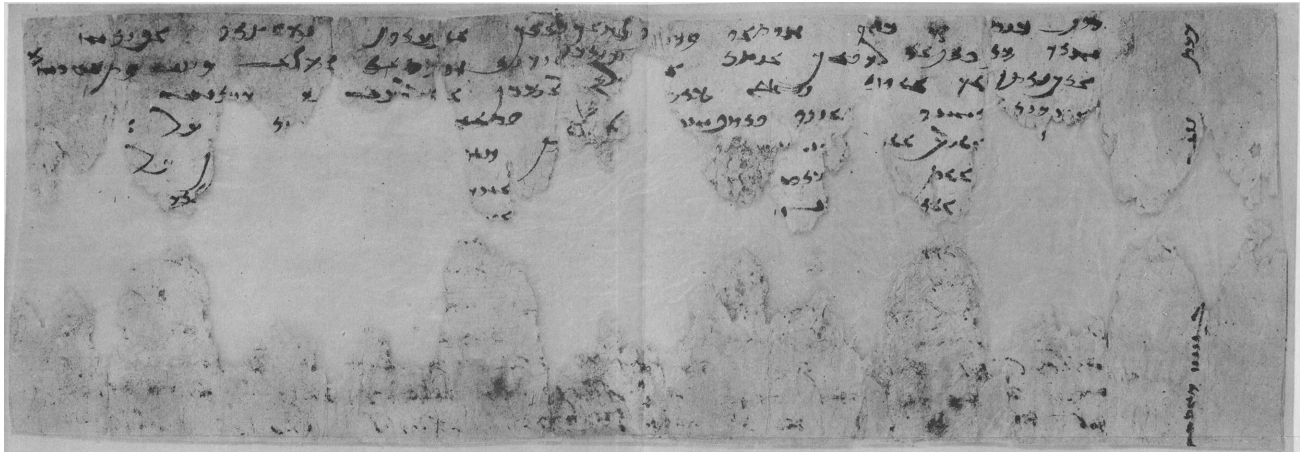


Figure 35: The ‘Ancient Letter 7’ (British Library: Or. 8212/96 recto; reproduced in Reichelt 1928: plate VII).



Figure 36: Sogdian rock inscription from Shatial (from Sims-Williams 1989: plate 10b) The inscription reads **𐰽𐰺𐰍𐰏𐰤** *nny'kk ZK* (top line), **𐰽𐰺𐰍𐰏** *sw'br* (middle), **𐰽𐰺** *BRY* (bottom). Latin transcription from *ibid*: 14. The inscription in the bottom right-hand corner is shown in detail in figure 37.



Figure 37: Sogdian rock inscription from Shatial (from Sims-Williams 1989: plate 10a). The central inscription reads **𐰽𐰺𐰍𐰏𐰤** *p'p'kk* (top line), **𐰽𐰺𐰍𐰏𐰤** *ZK kwš'n* (middle), **𐰽𐰺** *BRY* (bottom). Latin transcription from *ibid*: 14. The inscription in the top left-hand corner is shown in detail in figure 36.

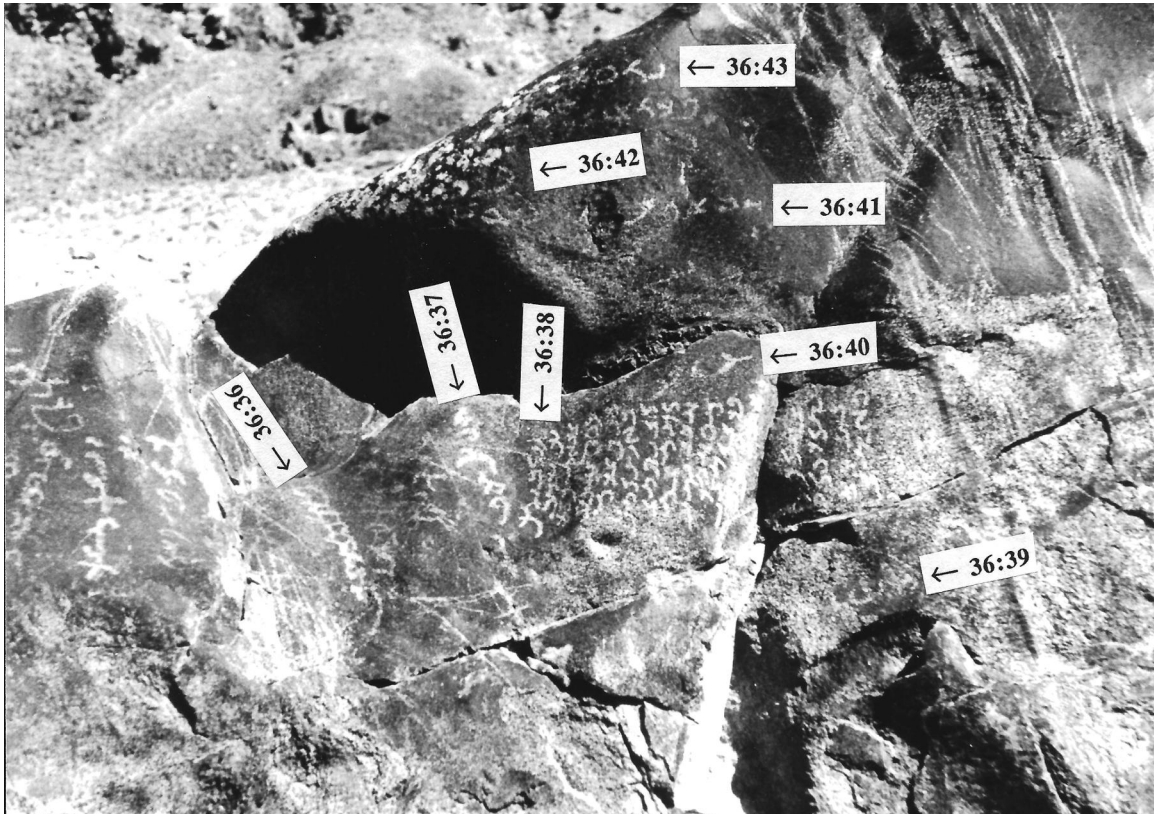


Figure 38: Rock at Shatial containing horizontal and vertical inscriptions in the Old Sogdian script (from Sims-Williams 1989: plate 109b). The text of 36:38 is shown in section 3.5.

### 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, ’ is still distinguished from **n**; but in the later, ’ = **n**, ’**n** = **n**’. 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:  $\gamma/x/s/\check{s}$ ,  $r/\beta/y$ . Some letters are distinguished only in final position (by some scribes), e.g., **n** ~ **z**, **x** ~  $\gamma$ .

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

#### SAMPLES OF SOGDIAN

##### ANCIENT LETTERS

PLI	kk'n'k	wr''βδynn	kk'rβ	w'twx	wγβ	DO	
wn''γβ	wMXyKZ	YZKYA	ykwn'zt'ps	wyc'mn	MLŠ	rwyrfβ	
ktnβynn	ktnβ	δpyx	NM	tšyp	tryβ		
1. Transliteration:	OD βγw	xwt'w	βr'kk	nnyδβ''rw	k'n'kk		
2. Normalization:	at	βayu	xutāw	βarak	nanē-θβār	kanak	
3. Gloss:	to	lord.ACC	master	Barak	Nana's-gift	Kanak	
1. ILP	βrywr	ŠLM	nm'cyw	sp'tz'nwky	AYKZY		
2. (ēw-)zār	βrēwar	*āfrīwan	namācyu	spätzānūk	kaδ-uti		
3. thousand	ten.thousand	greeting(?)	reverence.ACC	bended.knee	when-that.and		
1. ZKyXMw	βγ''nw	βyrt	pyšt	MN	xypθ	βntk	nnyβntk
2. wēšanu	βayān(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 39: Description of the Sogdian script of the ‘Ancient Letters’ (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	𐭪	𐭪, 𐭫	𐭪	𐭪 𐭫	a, ā
b	𐭬	𐭬, 𐭭	𐭬	𐭬	b, β
(β)			𐭬		β
g	𐭮	𐭮	𐭮	𐭮	g, γ
(γ)			𐭮	𐭮	γ
d	𐭰		𐭰	𐭰	d, δ
h (h)	𐭲	𐭲	𐭲	𐭲	a, Ø
w	𐭴	𐭴, 𐭵	𐭴	𐭴	w, ō, ũ
z	𐭶	𐭶	𐭶	𐭶	z
(j)			𐭶		ž
(ž)			𐭶	𐭶	ž
ḥ (h)	𐭸 𐭹	𐭸, 𐭹	𐭸	𐭸	γ, x, h
ṭ			𐭺	𐭺	t
y	𐭼	𐭼, 𐭽	𐭼	𐭼	y, ē, ĭ
k	𐭾	𐭾, 𐭿	𐭾	𐭾	k
(x)			𐭾	𐭾	x
l (δ)	𐭽	𐭽, 𐭾	𐭽	𐭽	δ
m	𐭿	𐭿, 𐭻	𐭿	𐭿	m
n	𐭻	𐭻, 𐭼	𐭻	𐭻	n
s	𐭽	𐭽, 𐭾	𐭽	𐭽	s
ʿ	𐭿	𐭿	𐭿	𐭿	Ø
p	𐭽	𐭽	𐭽	𐭽	p
(f)			𐭽	𐭽	f
š (c)	𐭽	𐭽	𐭽	𐭽	č, j
q			𐭽	𐭽	k
r	𐭽	𐭽, 𐭾	𐭽	𐭽	r
š	𐭽	𐭽, 𐭾	𐭽	𐭽	š
t	𐭽	𐭽, 𐭾	𐭽	𐭽	t, θ

Figure 40: Table showing various scripts for writing Sogdian (from Skjærvø 1996: 519).