

# Revised proposal to encode the Sogdian script in Unicode

Anshuman Pandey  
pandey@umich.edu

October 20, 2016

## 1 Introduction

This is a proposal to encode the Sogdian script in Unicode. It replaces the following document:

- L2/16-158 “Proposal to encode Sogdian in Unicode”

It incorporates comments regarding previous proposals made in:

- L2/16-216 “Recommendations to UTC #148 August 2016 on Script Proposals”

Previous versions of this proposal have been reviewed by Nicholas Sims-Williams and Yutaka Yoshida, who are leading scholars of Sogdian studies.

## 2 Background

The proposed Unicode encoding for ‘Sogdian’ encompasses a group of related script styles used for representing Sogdian (ISO 639: *sog*), an ancient Eastern Iranian language. The Sogdian script is derived from ‘Old Sogdian’ (see L2/16-312). There are two major styles of Sogdian: formal and cursive. These developed separately from Old Sogdian, with the cursive style likely emerging first. It is difficult to specify a clear division between the ‘formal’ and ‘cursive’ styles. It is appropriate to consider the Sogdian varieties as belonging to a continuum, with the most formal styles and the simplified, ambiguous cursive styles as the two terminals. Manuscript folios written in the various styles are shown in figures 21–30. The styles possess the same character repertoire and structural features:

- *Repertoire* Of the 20 letters of the Old Sogdian alphabet, 19 were retained in Sogdian (see comparison in table 1). The letter *daleth* is no longer represented. Additional letters were introduced, such as *feth* and *lesh* (‘hooked *resh*’). Diacritic marks were also added. An ostrakon found at Panjikant shows 23 Sogdian letters that correspond to the Aramaic repertoire. Of these *daleth*, *ayin*, *qoph*, *teth* are represented using non-letter signs, while *lamedh* occurs twice. The numerical signs of Old Sogdian are attested. Various marks of punctuation were also used.

- *Structure* The script is a conjoining *abjad*, similar to Arabic, Syriac, and Mongolian. Letters connect and change shape based upon their position within a word. In later styles, some letters, eg. *zayin* and *heth*, do not join to a following letter in order to distinguish themselves from other letters with identical shapes. Words are separated using spaces. The conjoining behavior of Sogdian contrasts with the non-joining Old Sogdian:

		Old Sogdian	Sogdian
<i>swydyk</i>	‘Sogdian’		
<i>smʾrknδc</i>	‘of Samarkand’		

- *Directionality* The script is written both horizontally and vertically. In horizontal mode, the writing direction is right to left and lines proceed from top to bottom. When vertical, glyphs are rotated 90° counter-clockwise and are written from top to bottom in lines that advance from the left edge of the writing surface towards the right.

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 for each script variety.

### 3 Proposed repertoire

The proposed repertoire for Sogdian contains 42 characters: 21 letters, 9 diacritic signs, 4 numbers, and 12 punctuation signs. Representative glyphs are based upon the ‘sutra’ / ‘formal’ variety. The glyphs are normalizations of diverse handwritten styles.

#### 3.1 Letters

Letters included in the proposed repertoire are shown below in their independent and positional forms:


	Character name	Roman	Final	Medial	Initial
	SOGDIAN LETTER ALEPH	ʾ			
	SOGDIAN LETTER BETH	β			
	SOGDIAN LETTER GIMEL	γ			
	SOGDIAN LETTER HE	h		—	—
	SOGDIAN LETTER WAW	w			





▲	SOGDIAN LETTER ZAYIN	z	▲	▲	▲
𐰪	SOGDIAN LETTER HETH	x	𐰪	𐰪	𐰪
𐰫	SOGDIAN LETTER YODH	y	𐰫	𐰫	𐰫
𐰬	SOGDIAN LETTER KAPH	k	𐰬	𐰬	𐰬
𐰭	SOGDIAN LETTER LAMEDH	δ	𐰭	𐰭	𐰭
𐰮	SOGDIAN LETTER MEM	m	𐰮	𐰮	𐰮
𐰯	SOGDIAN LETTER NUN	n	𐰯	𐰯	𐰯
𐰰	SOGDIAN LETTER SAMEKH	s	𐰰	𐰰	𐰰
𐰱	SOGDIAN LETTER AYIN	◌	𐰱	𐰱	𐰱
𐰲	SOGDIAN LETTER PE	p	𐰲	𐰲	𐰲
𐰳	SOGDIAN LETTER SADHE	c	𐰳	𐰳	𐰳
𐰴	SOGDIAN LETTER RESH	r	𐰴	𐰴	𐰴
𐰵	SOGDIAN LETTER SHIN	š	𐰵	𐰵	𐰵
𐰶	SOGDIAN LETTER TAW	t	𐰶	𐰶	𐰶
𐰷	SOGDIAN LETTER FETH	f	𐰷	𐰷	𐰷
𐰸	SOGDIAN LETTER LESH	l	𐰸	𐰸	𐰸



Only independent forms are included in the proposed repertoire. Positional forms are to be maintained in a font and substituted by the shaping engine. (see § 4). Some positional forms may not be palaeographically distinctive, but are differentiated typographically in order to illustrate the joining features of glyphs.



**beth** and **yodh** The forms of 𐰮 BETH and 𐰫 YODH converged graphically in later styles. In these styles YODH does not join to the left when initial and medial in order to differentiate it from BETH, which maintains its joining behavior. (see § 4.4). In some cases 𐰮 BETH is distinguished using a diacritic, eg. 𐰮̇.



**gimel** and **heth** Initial and medial forms of 𐰯 GIMEL and 𐰪 HETH are identical. In these positions ligation between GIMEL and a following letter may be broken in order to distinguish it from HETH, which maintains its regular joining behavior (see § 4.4). The letter HETH has a variant final form 𐰪.











**he** The letter  HE is used for marking a long vowel. It occurs only in final position.




**zayin** and **nun** The initial and medial forms of  ZAYIN and  NUN are identical. In these positions the ligation between ZAYIN and a following letter may be broken to distinguish it from NUN (see § 4.4). In various texts ZAYIN is marked explicitly as  or  (see § 3.2).



**kaph** The letter  KAPH has the variant final shape  (see § 4.5).





**lamedh** The letter  LAMEDH appears in several sources as the ‘hooked’ form . This form is not a distinct letter, but a glyphic variant.



**nun** The letter  NUN has the variant final shape  (see § 4.5).



**ayin** The letter *ayin* occurs in Aramaic heterograms and has two attested shapes:  and . The first form occurs less commonly and is attested in the spelling of ‘NY’W “he said”, eg. . It also occurs in a special shortened form of ‘NY’W, which is expressed as  ‘‘. See figure 1 for attestations. Initial and final forms are attested, but there is no dual-joining medial form, eg. \* . The form  is the more commonly used shape for *ayin*. As shown above, it also occurs in the spelling of ‘NY’W, in word-initial position. It also occurs in the heterogram  ‘M’ ‘with’. It is identical to  RESH. The merger of the shape of the regular letter *ayin* and *resh* occurred in Old Sogdian. Therefore, for purposes of the encoding,  is encoded as AYIN, and  is unified with RESH.

**sadhe** The letter  SADHE has the variant final forms  and  (see § 4.5).

**shin** The manuscript So 14830 shows an independent  SHIN written as  (see fig. 26). Usage of this form is rare. It may be possible to represent it using a variation selector (see § 4.5).







**taw** The letter  TAW has the variant final forms , , and  (see § 4.5).

**feth** The letter  FETH is an extension of  BETH that contains an extra hook at the left edge of the head. See figure 2 for attestations.

**lesh** The  LESH or ‘hooked *resh*’ is an extension of  RESH with a below-base hook. See figure 3 for attestations.

### 3.2 Modifier signs

The following modifier signs are used for disambiguation and transcription:

	Character name	Example
	SOGDIAN SIGN DOT BELOW	 ẓ
	SOGDIAN SIGN TWO DOTS BELOW	 ẓ̣
	SOGDIAN SIGN DOT ABOVE	 ḃ



◌̈	SOGDIAN SIGN TWO DOTS ABOVE	𐰪 ǰ
◌̇	SOGDIAN SIGN CURVE ABOVE	𐰫 ǧ
◌̆	SOGDIAN SIGN CURVE BELOW	𐰬 ǧ, 𐰭 ǧ
◌̅	SOGDIAN SIGN HOOK ABOVE	𐰮 ǧ, 𐰯 ǧ, 𐰰 ǧ
◌̄	SOGDIAN SIGN HOOK BELOW	𐰱 ǧ, 𐰲 ǧ, 𐰳 ǧ, 𐰴 ǧ, 𐰵 ǧ, 𐰶 ǧ, 𐰷 ǧ
◌̃	SOGDIAN SIGN SUBSCRIPT RESH	𐰸 ǧ

---

The shapes of dots in these combining signs may vary according to script style. In some styles, the dots may appear rounder or more square:

	Normative	Variant
SIGN DOT BELOW	◌̇	◌̇◌̇
SIGN TWO DOTS BELOW	◌̈	◌̈◌̈
SIGN DOT ABOVE	◌̅	◌̅◌̅
SIGN TWO DOTS ABOVE	◌̆	◌̆◌̆

---

### 3.3 Numbers

The following signs are used for numerical notation (see figures 13–14):

	Character name	Value	Final	Medial	Initial
𐰀	SOGDIAN NUMBER ONE	1	𐰀	𐰀	𐰀
𐰁	SOGDIAN NUMBER TEN	10	𐰁	𐰁	𐰁
𐰂	SOGDIAN NUMBER TWENTY	20	𐰂	𐰂	𐰂
𐰃	SOGDIAN NUMBER ONE HUNDRED	100	—	—	—

---

Numbers are dual joining except for ONE HUNDRED, which is non-joining.

### 3.4 Punctuation

The following punctuation signs are used for delimiting sections of text (see figures 15–17).

Character name	
	SOGDIAN PUNCTUATION TWO VERTICAL BARS
ḡ	SOGDIAN PUNCTUATION TWO VERTICAL BARS WITH DOTS ABOVE AND BELOW
⊙	SOGDIAN PUNCTUATION CIRCLE WITH MIDDLE DOT
◐	SOGDIAN PUNCTUATION HALF CIRCLE WITH MIDDLE DOT
..	SOGDIAN PUNCTUATION TWO DOTS
❖	SOGDIAN PUNCTUATION FOUR DOTS
⋄	SOGDIAN PUNCTUATION FIVE DOTS
⊕	SOGDIAN PUNCTUATION CROSS WITH EIGHT DOTS
⊖	SOGDIAN PUNCTUATION FLOWER WITH THREE PETALS
⊗	SOGDIAN PUNCTUATION FLOWER WITH FOUR PETALS
⊘	SOGDIAN PUNCTUATION FOUR HALF CIRCLES WITH DOTS
⊙	SOGDIAN PUNCTUATION THREE CIRCLES WITH RAYS

The following marks of punctuation have variant forms:

	Normative	Variant
TWO VERTICAL BARS		
TWO VERTICAL BARS WITH DOTS ABOVE AND BELOW	ḡ	ḡ ḡ

Some punctuation signs occur in combination (see fig. 25).

## 4 Encoding model

### 4.1 General principles

The encoding model is based upon that used for Arabic in Unicode. A word is produced using a sequence of letters, i.e. the following input string would be used for representing the word *rxwšnʾyrδmn* ‘paradise’:

𐰽	𐰪	𐰬	𐰺	𐰮	𐰀	𐰽	𐰶	𐰸	𐰷	𐰺	←
NUN	MEM	LAMEDH	RESH	GIMEL	ALEPH	NUN	SHIN	WAW	HETH	RESH	

The shaping engine would replace independent shapes with the appropriate positional forms:

𐰽	𐰪	𐰬	𐰺	𐰮	𐰀	𐰽	𐰶	𐰸	𐰷	𐰺	←
NUN	MEM	LAMEDH	RESH	GIMEL	ALEPH	NUN	SHIN	WAW	HETH	RESH	
<i>fin</i>	<i>med</i>	<i>med</i>	<i>med</i>	<i>med</i>	<i>med</i>	<i>med</i>	<i>med</i>	<i>med</i>	<i>med</i>	<i>init</i>	

The rendered output would be:

*rxwšnʾyrδmn*

### 4.2 Combining signs

Combining signs are placed after the base letter in the input sequence:

Output	Input string →
<i>cr</i>	𐰺𐰪 𐰽 SADHE, 𐰺 RESH, 𐰸 SIGN CURVE BELOW
<i>rʾy</i>	𐰺𐰬 𐰺 RESH, 𐰵 ZAYIN, 𐰸 SIGN DOT BELOW, 𐰬 YODH
<i>ry</i>	𐰺𐰬 𐰺 RESH, 𐰸 SIGN SUBSCRIPT RESH, 𐰬 YODH

### 4.3 Numerical notation

Numerical notation is similar to that of the ‘Old Sogdian’ script. The numbers 1–9 are expressed using repetitions of 𐰽 ONE. The numbers 5–9 are arranged in groups of three or four instances of ONE separated by spaces. The arrangements are shown below:




	Output	Input string →
1	𐰀	𐰀 ONE
2	𐰁	𐰀 ONE, 𐰀 ONE
3	𐰂	𐰀 ONE, 𐰀 ONE, 𐰀 ONE
4	𐰃	𐰀 ONE, 𐰀 ONE, 𐰀 ONE, 𐰀 ONE
5	𐰁 𐰁	𐰀 ONE, 𐰀 ONE, 𐰀 ONE, <span style="border: 1px dashed black; padding: 0 2px;">SP</span> SPACE, 𐰀 ONE, 𐰀 ONE
6	𐰂 𐰂	𐰀 ONE, 𐰀 ONE, 𐰀 ONE, <span style="border: 1px dashed black; padding: 0 2px;">SP</span> SPACE, 𐰀 ONE, 𐰀 ONE, 𐰀 ONE
7	𐰃 𐰃	𐰀 ONE, 𐰀 ONE, 𐰀 ONE, 𐰀 ONE, <span style="border: 1px dashed black; padding: 0 2px;">SP</span> SPACE, 𐰀 ONE, 𐰀 ONE, 𐰀 ONE
8	𐰄 𐰄	𐰀 ONE, 𐰀 ONE, 𐰀 ONE, 𐰀 ONE, <span style="border: 1px dashed black; padding: 0 2px;">SP</span> SPACE, 𐰀 ONE, 𐰀 ONE, 𐰀 ONE, 𐰀 ONE
9	𐰃 𐰁 𐰁	𐰀 ONE, 𐰀 ONE, 𐰀 ONE, <span style="border: 1px dashed black; padding: 0 2px;">SP</span> SPACE, 𐰀 ONE, 𐰀 ONE, 𐰀 ONE, <span style="border: 1px dashed black; padding: 0 2px;">SP</span> SPACE, 𐰀 ONE, 𐰀 ONE, 𐰀 ONE













The tens are written using sequences of 𐰄 TEN and 𐰅 TWENTY. Even multiples are expressed with repetitions of TWENTY, not TEN. Odd multiples are produced by attaching TEN at the end.

	Output	Input string →
10	𐰄	𐰄 TEN
20	𐰅	𐰅 TWENTY
30	𐰄𐰅	𐰅 TWENTY, 𐰄 TEN
40	𐰅𐰅	𐰅 TWENTY, 𐰅 TWENTY
50	𐰄𐰅𐰅	𐰅 TWENTY, 𐰅 TWENTY, 𐰄 TEN
60	𐰅𐰅𐰅	𐰅 TWENTY, 𐰅 TWENTY, 𐰅 TWENTY
70	𐰄𐰅𐰅𐰅	𐰅 TWENTY, 𐰅 TWENTY, 𐰅 TWENTY, 𐰄 TEN
80	𐰅𐰅𐰅𐰅	𐰅 TWENTY, 𐰅 TWENTY, 𐰅 TWENTY, 𐰅 TWENTY






90         TWENTY,  TWENTY,  TWENTY,  TWENTY,  TEN

---
















Hundreds are represented using  ONE HUNDRED. This sign is an atomic character. It represents the value 100, but also functions as a unit mark for multiples of hundred. Such multiples are expressed using Sogdian names for cardinal numbers followed by  or , eg. 200 is *δwy* 100; 500 is *pnc* 100:

	Output	Input string →
100		 ONE HUNDRED
200		 LAMEDH,  WAW,  YODH, <span style="border: 1px dashed black; padding: 0 2px;">SP</span> SPACE,  ONE HUNDRED
500		 PE,  NUN,  SADHE, <span style="border: 1px dashed black; padding: 0 2px;">SP</span> SPACE,  ONE HUNDRED




The thousands are represented using the Aramaic heterogram  *LP*, produced as  LAMEDH,  PE>:

	Output	Input string →
1000		 ONE,  LAMEDH,  PE,  WAW

The primary units are generally written after the tens in compound numbers:

	Output	Input string →
81		 TWENTY,  TWENTY,  TWENTY,  TWENTY,  ONE
155		 ONE HUNDRED, <span style="border: 1px dashed black; padding: 0 2px;">SP</span> SPACE,  TWENTY,  TEN,  ONE,  ONE,  ONE, <span style="border: 1px dashed black; padding: 0 2px;">SP</span> SPACE,  ONE,  ONE

#### 4.4 Ligature breaking

The letters  GIMEL,  ZAYIN,  YODH may remain disconnected from a following letter when initial or medial. The breaking of regular joining behavior can be managed using the generic Unicode control character ZWJ U+200C ZERO WIDTH NON-JOINER (abbreviated as ZWNJ). This character is placed after the letter that should remain unligated.

	Output	Input string →
$\beta\gamma\delta r$		𐰪 BETH, 𐰪 GIMEL, 𐰪 PE, 𐰪 LAMEDH, 𐰪 RESH
$\beta\gamma-p\delta r$		𐰪 BETH, 𐰪 GIMEL, 𐰪 ZWNJ, 𐰪 PE, 𐰪 LAMEDH, 𐰪 RESH
$\text{ʾ}zwn$		𐰪 ALEPH, 𐰪 ALEPH, 𐰪 ZAYIN, 𐰪 WAW, 𐰪 NUN
$\text{ʾ}z-wn$		𐰪 ALEPH, 𐰪 ALEPH, 𐰪 ZAYIN, 𐰪 ZWNJ, 𐰪 WAW, 𐰪 NUN

In such cases the letter before ZWNJ is displayed using its final form although it is in medial position.

#### 4.5 Variant forms

Alternate forms may be requested using control characters, eg.  $\text{VS1}$  U+FE00 VARIATION SELECTOR-1 (abbreviated as VS1),  $\text{VS2}$  U+FE01 VARIATION SELECTOR-2 (abbreviated as VS2), etc. There are 16 such selectors defined in Unicode. This mechanism can be used for displaying variants of independent letters:

	Output	Input string →
$\check{s}$		𐰪 SHIN
$\check{s}$		𐰪 SHIN, $\text{VS2}$ VS2

The terminals of some letters have variant orientations (see, for example, figure 27). Terminal variation occurs most often at the end of a line. It is a scribal technique for filling space or for compensating for lack of space at a margin. There is no semantic difference between normative and alternate forms of finals. Terminals occur in free variation and a letter may have different terminals within the same document. Variation selectors may be used for requesting alternate forms.

Alternate styles of final forms may be selected as follows:

	Regular final	Alternate final	Input string →
HETH	𐰇	𐰇	𐰇 HETH, <span style="border: 1px dashed black; padding: 0 2px;">VS</span> <sub>1</sub> VS1
KAPH	𐰈	𐰈	𐰈 KAPH, <span style="border: 1px dashed black; padding: 0 2px;">VS</span> <sub>1</sub> VS1
NUN	𐰉	𐰉	𐰉 NUN, <span style="border: 1px dashed black; padding: 0 2px;">VS</span> <sub>1</sub> VS1
SADHE	𐰊	𐰊	𐰊 SADHE, <span style="border: 1px dashed black; padding: 0 2px;">VS</span> <sub>1</sub> VS1
		𐰊	𐰊 SADHE, <span style="border: 1px dashed black; padding: 0 2px;">VS</span> <sub>2</sub> VS2
TAW	𐰋	𐰋	𐰋 TAW, <span style="border: 1px dashed black; padding: 0 2px;">VS</span> <sub>1</sub> VS1
		𐰋	𐰋 TAW, <span style="border: 1px dashed black; padding: 0 2px;">VS</span> <sub>2</sub> VS2

For example, alternate forms of final 𐰋 TAW would be represented in encoded input as follows:

	Output	Input string →
»pʻt	𐰇𐰇𐰇𐰇	𐰇 ALEPH, 𐰇 ALEPH, 𐰇 PE, 𐰇 ALEPH, 𐰇 TAW
»pʻt	𐰇𐰇𐰇𐰇	𐰇 ALEPH, 𐰇 ALEPH, 𐰇 PE, 𐰇 ALEPH, 𐰇 TAW, <span style="border: 1px dashed black; padding: 0 2px;">VS</span> <sub>1</sub> VS1
»pʻt	𐰇𐰇𐰇𐰇	𐰇 ALEPH, 𐰇 ALEPH, 𐰇 PE, 𐰇 ALEPH, 𐰇 TAW, <span style="border: 1px dashed black; padding: 0 2px;">VS</span> <sub>2</sub> VS2

## 5 Character Data

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

```

10E30; SOGDIAN LETTER ALEPH; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E31; SOGDIAN LETTER BETH; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E32; SOGDIAN LETTER GIMEL; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E33; SOGDIAN LETTER HE; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E34; SOGDIAN LETTER WAW; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E35; SOGDIAN LETTER ZAYIN; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E36; SOGDIAN LETTER HETH; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E37; SOGDIAN LETTER YODH; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E38; SOGDIAN LETTER KAPH; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E39; SOGDIAN LETTER LAMEDH; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E3A; SOGDIAN LETTER MEM; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E3B; SOGDIAN LETTER NUN; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E3C; SOGDIAN LETTER SAMEKH; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E3D; SOGDIAN LETTER AYIN; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E3E; SOGDIAN LETTER PE; Lo; 0; AL; ; ; ; ; N; ; ; ;
10E3F; SOGDIAN LETTER SADHE; Lo; 0; AL; ; ; ; ; N; ; ; ;
    
```

```

10E40; SOGDIAN LETTER RESH; Lo; 0; AL; ; ; ; ; N; ; ; ; ;
10E41; SOGDIAN LETTER SHIN; Lo; 0; AL; ; ; ; ; N; ; ; ; ;
10E42; SOGDIAN LETTER TAW; Lo; 0; AL; ; ; ; ; N; ; ; ; ;
10E43; SOGDIAN LETTER FETH; Lo; 0; AL; ; ; ; ; N; ; ; ; ;
10E44; SOGDIAN LETTER LESH; Lo; 0; AL; ; ; ; ; N; ; ; ; ;
10E45; SOGDIAN SIGN DOT BELOW; Mn; 220; NSM; ; ; ; ; N; ; ; ; ;
10E46; SOGDIAN SIGN TWO DOTS BELOW; Mn; 220; NSM; ; ; ; ; N; ; ; ; ;
10E47; SOGDIAN SIGN DOT ABOVE; Mn; 230; NSM; ; ; ; ; N; ; ; ; ;
10E48; SOGDIAN SIGN TWO DOTS ABOVE; Mn; 230; NSM; ; ; ; ; N; ; ; ; ;
10E49; SOGDIAN SIGN CURVE ABOVE; Mn; 230; NSM; ; ; ; ; N; ; ; ; ;
10E4A; SOGDIAN SIGN CURVE BELOW; Mn; 220; NSM; ; ; ; ; N; ; ; ; ;
10E4B; SOGDIAN SIGN HOOK ABOVE; Mn; 230; NSM; ; ; ; ; N; ; ; ; ;
10E4C; SOGDIAN SIGN HOOK BELOW; Mn; 220; NSM; ; ; ; ; N; ; ; ; ;
10E4D; SOGDIAN SIGN SUBSCRIPT RESH; Mn; 220; NSM; ; ; ; ; N; ; ; ; ;
10E4E; SOGDIAN NUMBER ONE; No; 0; AL; ; ; ; ; 1; N; ; ; ; ;
10E4F; SOGDIAN NUMBER TEN; No; 0; AL; ; ; ; ; 10; N; ; ; ; ;
10E50; SOGDIAN NUMBER TWENTY; No; 0; AL; ; ; ; ; 20; N; ; ; ; ;
10E51; SOGDIAN NUMBER ONE HUNDRED; No; 0; AL; ; ; ; ; 100; N; ; ; ; ;
10E52; SOGDIAN PUNCTUATION TWO VERTICAL BARS; Po; 0; AL; ; ; ; ; N; ; ; ; ;
10E53; SOGDIAN PUNCTUATION TWO VERTICAL BARS WITH DOTS ABOVE AND BELOW; Po; 0; AL; ; ; ; ; N; ; ; ; ;
10E54; SOGDIAN PUNCTUATION CIRCLE WITH MIDDLE DOT; Po; 0; AL; ; ; ; ; N; ; ; ; ;
10E55; SOGDIAN PUNCTUATION HALF CIRCLE WITH MIDDLE DOT; Po; 0; AL; ; ; ; ; N; ; ; ; ;
10E56; SOGDIAN PUNCTUATION MARK-5; Po; 0; AL; ; ; ; ; N; ; ; ; ;
10E57; SOGDIAN PUNCTUATION MARK-6; Po; 0; AL; ; ; ; ; N; ; ; ; ;
10E58; SOGDIAN PUNCTUATION MARK-7; Po; 0; AL; ; ; ; ; N; ; ; ; ;
10E59; SOGDIAN PUNCTUATION MARK-8; Po; 0; AL; ; ; ; ; N; ; ; ; ;
10E5A; SOGDIAN PUNCTUATION MARK-9; Po; 0; AL; ; ; ; ; N; ; ; ; ;
10E5B; SOGDIAN PUNCTUATION MARK-10; Po; 0; AL; ; ; ; ; N; ; ; ; ;
10E5C; SOGDIAN PUNCTUATION MARK-11; Po; 0; AL; ; ; ; ; N; ; ; ; ;

```

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

```

10E30..10E44;AL # Lo [21] SOGDIAN LETTER ALEPH..SOGDIAN LETTER LESH
10E45..10E4D;CM # Mn [9] SOGDIAN SIGN DOT BELOW..SOGDIAN SIGN SUBSCRIPT RESH
10E4E..10E51;AL # No [4] SOGDIAN NUMBER ONE..SOGDIAN NUMBER ONE HUNDRED
10E52..10E5C;AL # Po [11] SOGDIAN PUNCTUATION TWO VERTICAL BARS..
SOGDIAN PUNCTUATION MARK-11

```

### Shaping Properties Shaping properties in the format of ArabicShaping.txt:

```

# Sogdian Characters
10E30; SOGDIAN LETTER ALEPH; D; No_Joining_Group
10E31; SOGDIAN LETTER BETH; D; No_Joining_Group
10E32; SOGDIAN LETTER GIMEL; D; No_Joining_Group
10E33; SOGDIAN LETTER HE; D; No_Joining_Group
10E34; SOGDIAN LETTER WAW; D; No_Joining_Group
10E35; SOGDIAN LETTER ZAYIN; D; No_Joining_Group
10E36; SOGDIAN LETTER HETH; D; No_Joining_Group
10E37; SOGDIAN LETTER YODH; D; No_Joining_Group
10E38; SOGDIAN LETTER KAPH; D; No_Joining_Group
10E39; SOGDIAN LETTER LAMEDH; D; No_Joining_Group
10E3A; SOGDIAN LETTER MEM; D; No_Joining_Group
10E3B; SOGDIAN LETTER NUN; D; No_Joining_Group
10E3C; SOGDIAN LETTER SAMEKH; D; No_Joining_Group
10E3D; SOGDIAN LETTER AYIN; D; No_Joining_Group
10E3E; SOGDIAN LETTER PE; D; No_Joining_Group
10E3F; SOGDIAN LETTER SADHE; D; No_Joining_Group
10E40; SOGDIAN LETTER RESH; D; No_Joining_Group

```



















































10E41; SOGDIAN LETTER SHIN; D; No\_Joining\_Group  
 10E42; SOGDIAN LETTER TAW; D; No\_Joining\_Group  
 10E43; SOGDIAN LETTER FETH; D; No\_Joining\_Group  
 10E44; SOGDIAN LETTER LESH; D; No\_Joining\_Group

## 6 References

- Livshits, V. A. 2015. “A Sogdian alphabet from Penjikent”. *Sogdian epigraphy of Central Asia and Semirech’e*. Corpus inscriptionum Iranicarum. Part II. Inscriptions of the Seleucid and Parthian periods of Eastern Iran and Central Asia. Vol. III. Sogdian, pp. 227–232. Translated by Tom Stableford, edited by Nicholas Sims-Williams. London: School of Oriental and African Studies.
- Pandey, Anshuman. 2016a. “Proposal to encode the Old Sogdian script in Unicode”. L2/16-312.  
<http://www.unicode.org/L2/L2016/16312-old-sogdian.pdf>
- . 2016b. “Proposal to encode Sogdian in Unicode”. L2/16-158.
- Sims-Williams, Nicholas. 1981. “The Sogdian sound-system and the origins of the Uyghur script”. *Journal Asiatique*, pp. 347–360.
- 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>
- Yakubovich, Ilya, and Yutaka Yoshida. 2005. “The Sogdian Fragments of *Samghātasūtra* in the German Turfan Collection”. *Languages of Iran: Past and Present*, ed. Dieter Weber, pp. 239–268. Weisbaden: Harrassowitz Verlag.






















## 7 Acknowledgments

I express my gratitude to Nicholas Sims-Williams (SOAS, University of London) and Yutaka Yoshida (University of Kyoto) for providing detailed comments on an earlier draft of this document and for their valuable and patient discussions with me on numerous aspects of the script.

	10E3	10E4	10E5
0			
1			
2			
3			
4			
5			
6			
7			
8			
9			
A			
B			
C			
D			
E			
F			










This block unifies the 'formal' and 'cursive' scripts.  
Representative glyphs are based upon the 'formal' style.

### Letters





10E30		SOGDIAN LETTER ALEPH
10E31		SOGDIAN LETTER BETH
10E32		SOGDIAN LETTER GIMEL
10E33		SOGDIAN LETTER HE
10E34		SOGDIAN LETTER WAW
10E35		SOGDIAN LETTER ZAYIN
10E36		SOGDIAN LETTER HETH
10E37		SOGDIAN LETTER YODH
10E38		SOGDIAN LETTER KAPH
10E39		SOGDIAN LETTER LAMEDH
10E3A		SOGDIAN LETTER MEM
10E3B		SOGDIAN LETTER NUN
10E3C		SOGDIAN LETTER SAMEKH
10E3D		SOGDIAN LETTER AYIN
10E3E		SOGDIAN LETTER PE
10E3F		SOGDIAN LETTER SADHE
10E40		SOGDIAN LETTER RESH
10E41		SOGDIAN LETTER SHIN
10E42		SOGDIAN LETTER TAW
10E43		SOGDIAN LETTER FETH
10E44		SOGDIAN LETTER LESH

= hooked resh









### Modifier signs

10E45		SOGDIAN SIGN DOT BELOW
10E46		SOGDIAN SIGN TWO DOTS BELOW
10E47		SOGDIAN SIGN DOT ABOVE
10E48		SOGDIAN SIGN TWO DOTS ABOVE
10E49		SOGDIAN SIGN CURVE ABOVE
10E4A		SOGDIAN SIGN CURVE BELOW
10E4B		SOGDIAN SIGN HOOK ABOVE
10E4C		SOGDIAN SIGN HOOK BELOW
10E4D		SOGDIAN SIGN SUBSCRIPT RESH





### Numbers

10E4E		SOGDIAN NUMBER ONE
10E4F		SOGDIAN NUMBER TEN
10E50		SOGDIAN NUMBER TWENTY
10E51		SOGDIAN NUMBER ONE HUNDRED

### Punctuation

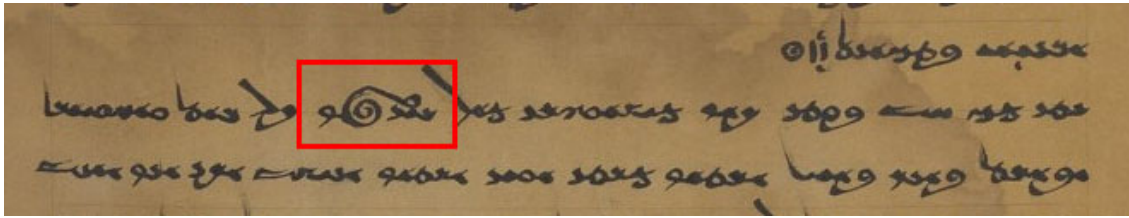
10E52		SOGDIAN PUNCTUATION TWO VERTICAL BARS
10E53		SOGDIAN PUNCTUATION TWO VERTICAL BARS WITH DOTS ABOVE AND BELOW
10E54		SOGDIAN PUNCTUATION CIRCLE WITH MIDDLE DOT
10E55		SOGDIAN PUNCTUATION HALF CIRCLE WITH MIDDLE DOT
10E56		SOGDIAN PUNCTUATION TWO DOTS
10E57		SOGDIAN PUNCTUATION FOUR DOTS
10E58		SOGDIAN PUNCTUATION FIVE DOTS
10E59		SOGDIAN PUNCTUATION CROSS WITH EIGHT DOTS

### Text ornaments

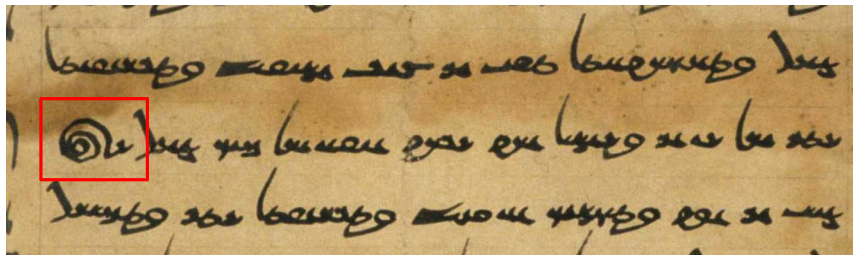
10E5A		SOGDIAN PUNCTUATION FLOWER WITH THREE PETALS
10E5B		SOGDIAN PUNCTUATION FLOWER WITH FOUR PETALS
10E5C		SOGDIAN PUNCTUATION FOUR HALF CIRCLES WITH DOTS
10E5D		SOGDIAN PUNCTUATION THREE CIRCLES WITH RAYS

	Old Sogdian	Sogdian
<i>aleph</i>	𐰀, 𐰁	𐰀
<i>beth</i>	𐰂, 𐰃	𐰂
<i>gimel</i>	𐰄	𐰄
<i>daleth</i>	(𐰅)	—
<i>he</i>	𐰆, 𐰇	𐰆
<i>waw</i>	𐰈	𐰈
<i>zayin</i>	𐰉	𐰉
<i>heth</i>	𐰊	𐰊
<i>teth</i>	—	—
<i>yodh</i>	𐰋	𐰋
<i>kaph</i>	𐰌	𐰌
<i>lamedh</i>	𐰍	𐰍
<i>mem</i>	𐰎	𐰎
<i>nun</i>	𐰏, 𐰐, 𐰑	𐰏
<i>samekh</i>	𐰒	𐰒
<i>ayin</i>	𐰓, 𐰔, (𐰕)	𐰓, (𐰕)
<i>pe</i>	𐰖	𐰖
<i>sadhe</i>	𐰗, 𐰘, 𐰙	𐰗
<i>qoph</i>	—	—
<i>resh</i>	𐰚	𐰚
<i>shin</i>	𐰛	𐰛
<i>taw</i>	𐰜, 𐰝, 𐰞	𐰜

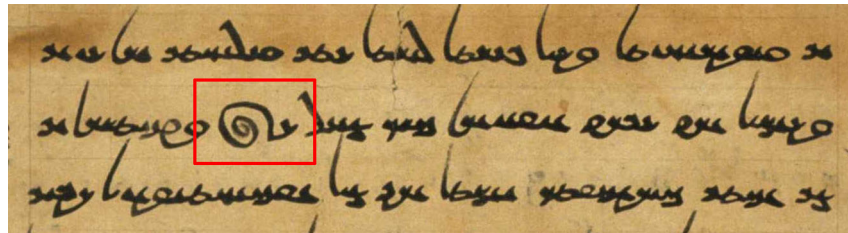
Table 1: Comparison of letters of Old Sogdian with related Iranian scripts and Aramaic. Parenthesis indicate that a letter is unified with another.



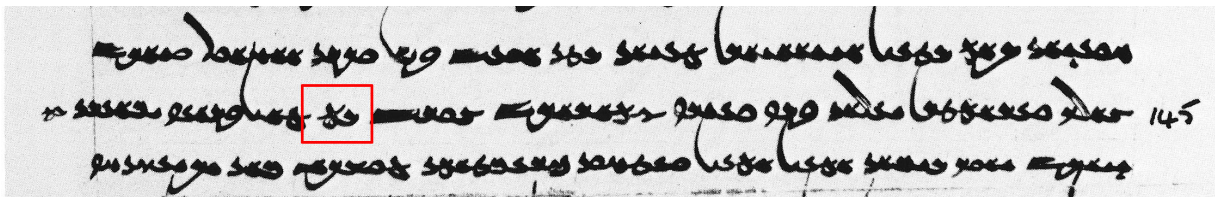
The letter *ayin* represented in the word 𐰽𐰺𐰍 𐰪𐰆𐰏, using 𐰽 RESH and 𐰺 AYIN. The encoded representation of the word is <𐰽 RESH, 𐰺 NUN, 𐰍 YODH, 𐰺 AYIN, 𐰪 WAW>. See full folio in figure 25.



[...]

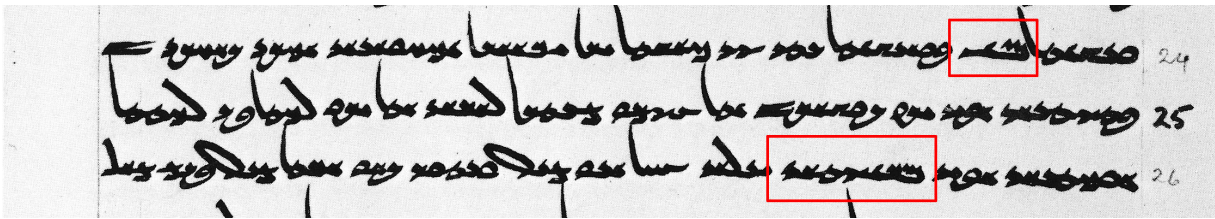


Two instances of 𐰺𐰽, which represents a shortened form of the Aramaic heterogram 𐤍𐤏𐤃. These instances may be interpreted as an isolated 𐰺 AYIN or a sequence of <𐰽 RESH, 𐰺 AYIN>. In the latter interpretation, the letter 𐰽 represents the regular *ayin*. From So 20165, lines 8, 11.



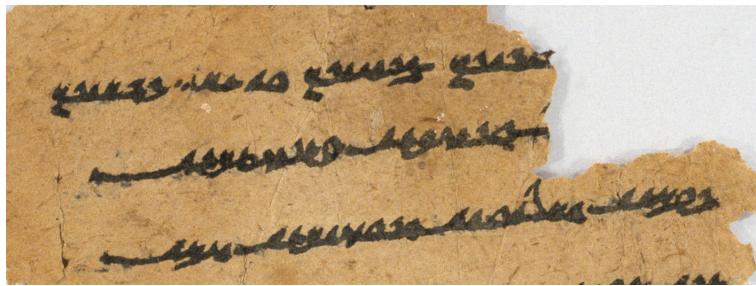
The letter *ayin* in the Aramaic heterogram 𐰽𐰪 'M' 'with'. Here *ayin* is represented using the shape 𐰽. The word would be encoded as <𐰽 RESH, 𐰪 MEM>. From Dhyana text, line 145.

Figure 1: Specimens of *ayin*.

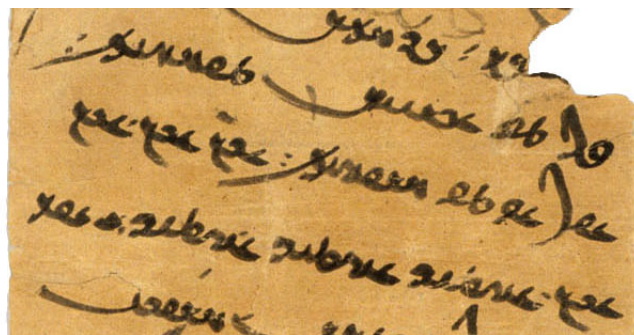


Usage of 𐰢 FETH for representing [f] (Dhyana text, lines 24, 26.).

Figure 2: Specimens of *feth*.

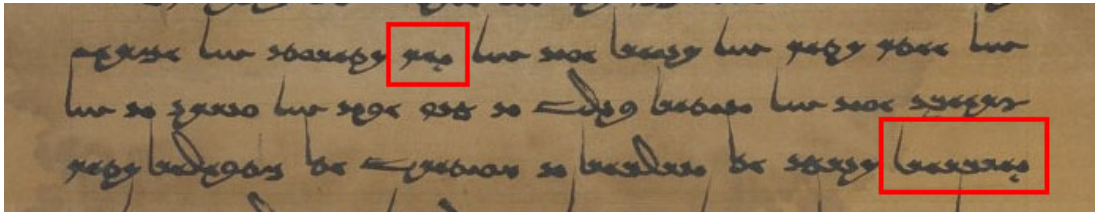


Usage of 𐰣 LESH for representing word final [l] (So 10678 r).



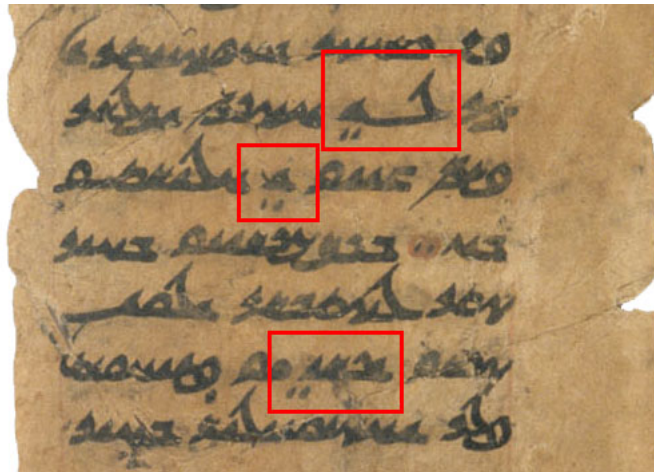
Usage of 𐰣 LESH for representing word final [l] (Ch/So 20135 v).

Figure 3: Specimens of *lesh*.

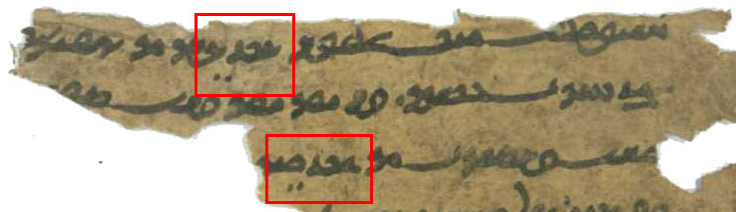


The sign ◌ used with ▲ ZAYIN for transcribing [z].

Figure 4: Usage of ◌ SOGDIAN SIGN DOT BELOW.



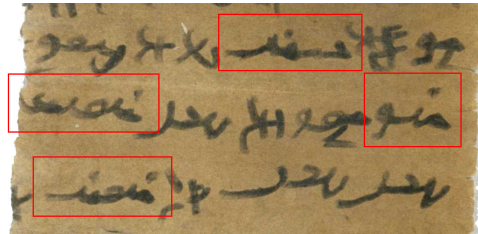
The sign ◌ used with ▲ ZAYIN for transcribing [z]. From So 18196 verso.



The sign ◌ used with ▲ ZAYIN for transcribing [z]. From So 20226 recto.

Figure 5: Usage of ◌ SOGDIAN SIGN TWO DOTS BELOW.





The sign 𐰆 used with 𐰃 HETH. From So 14800 verso.

Figure 6: Usage of 𐰆 SOGDIAN SIGN DOT ABOVE.

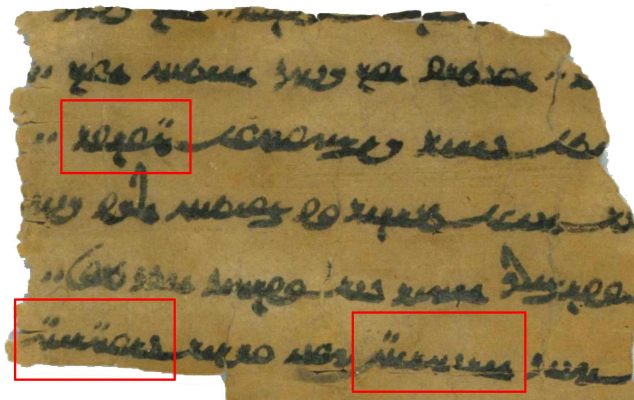
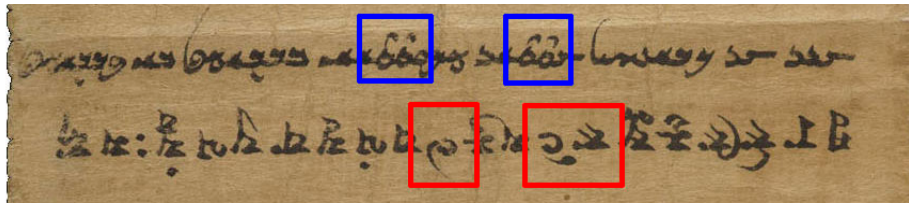
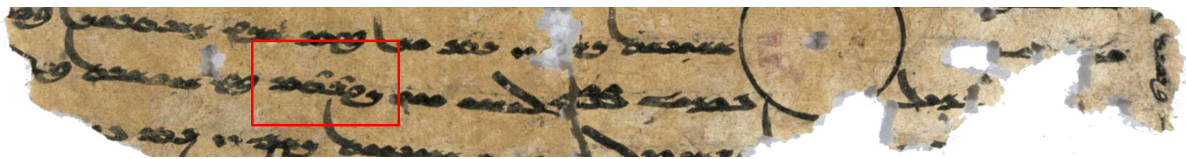


Figure 7: Usage of 𐰇 SOGDIAN SIGN TWO DOTS ABOVE. The sign is used for distinguishing 𐰃 HETH. From So 10026 verso.



The sign 𐰇 used with 𐰣 TAW for transcribing Sanskrit ट *ṭa* and त *ta* eg. 𐰣𐰇𐰣𐰣 = °जटे, 𐰣𐰇𐰣𐰣𐰣 = मकुटा.  
From BL Or.8212/175.



The sign 𐰇 used with 𐰣 TAW for transcribing Sanskrit ट *ṭa*, eg. 𐰣𐰇𐰣𐰣 = कोटि (So 14680 verso).

Figure 8: Usage of 𐰇 SOGDIAN SIGN CURVE ABOVE.

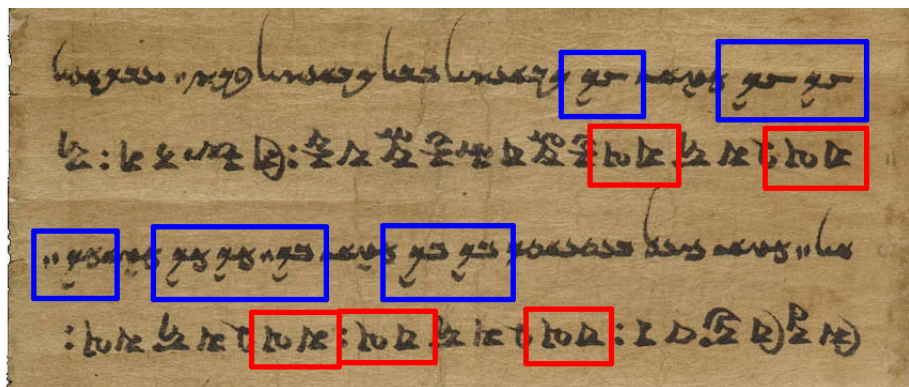


Figure 9: Usage of 𐰇 SOGDIAN SIGN CURVE BELOW. The sign is used in these instances with 𐰣 RESH for transcribing Sanskrit ल *la*, eg. 𐰣𐰇 = चल, 𐰣𐰇 = बल, 𐰣𐰇 = मल.



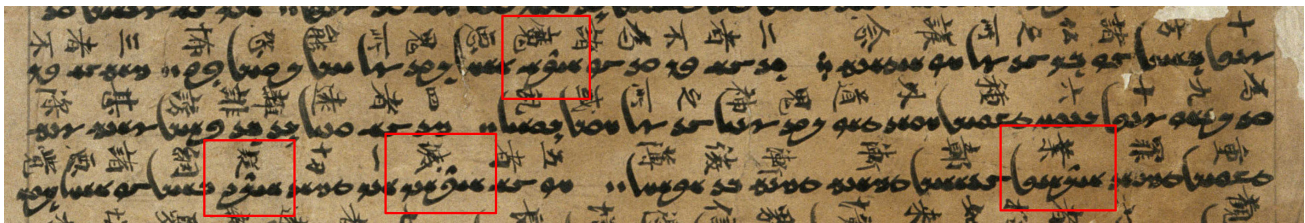
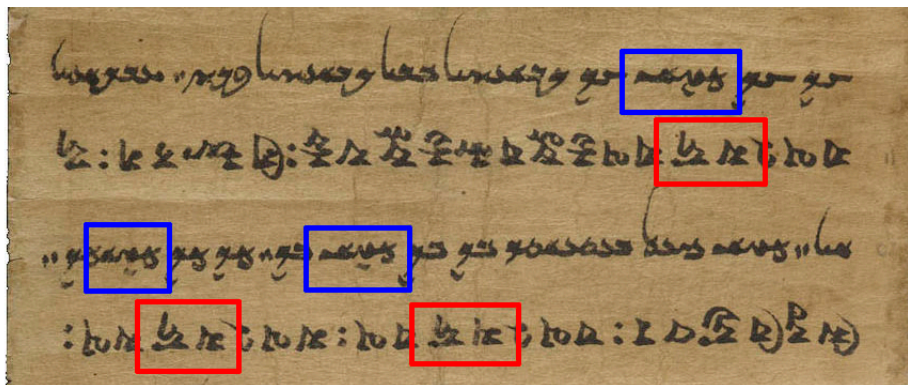
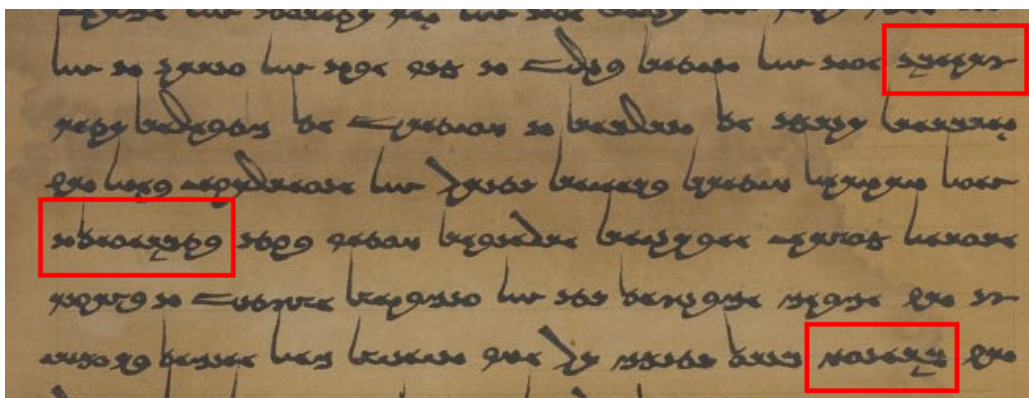


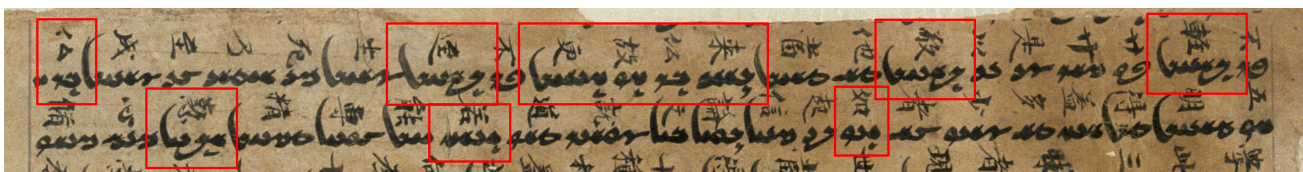
Figure 10: Usage of <sup>1</sup> SOGDIAN SIGN HOOK ABOVE. The sign is used in these instances with <sup>6</sup> KAPH and <sup>6</sup> PE for possibly transcribing voiced Chinese consonants, ie. [g], [b].



The sign ◌ used with 𐰇 GIMEL for transcribing Sanskrit ह ha, eg. 𐰇𐰃 = महा (highlighted red).



The sign ◌ used with 𐰃 BETH for transcribing [f].



The sign ◌ used for transcribing Chinese consonants.

Figure 11: Usage of ◌ SOGDIAN SIGN HOOK BELOW.

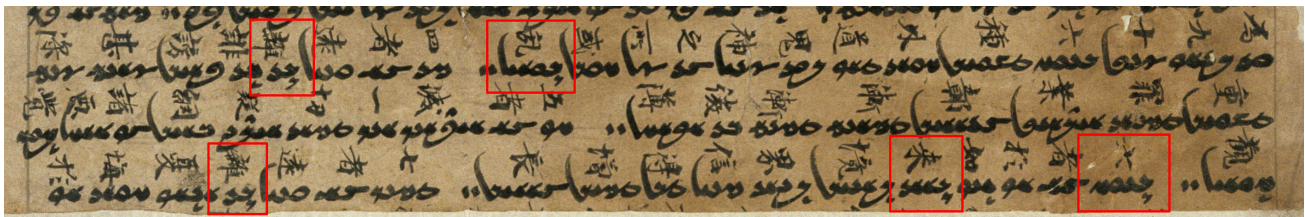
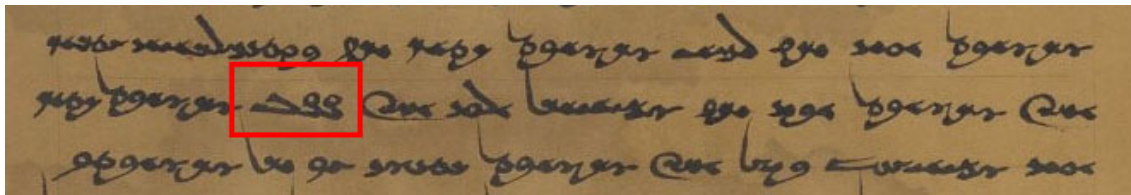
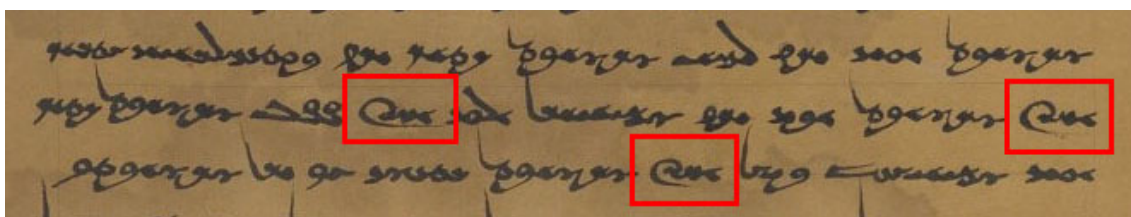


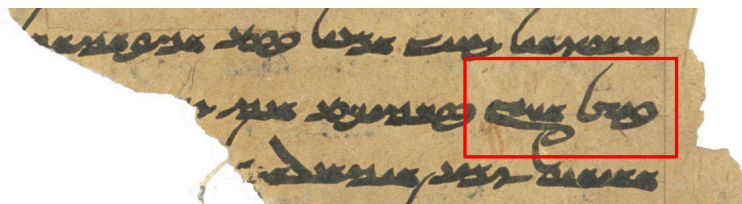
Figure 12: Usage of ◌ SOGDIAN SIGN SUBSCRIPT RESH. The sign is used in these instances with ◌ RESH for transcribing Chinese [1].



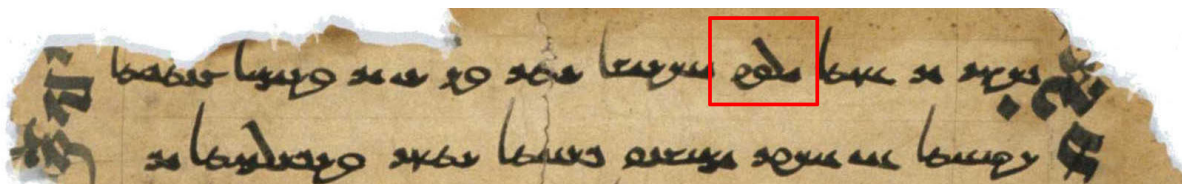
The number 50 𐰽𐰺𐰾 (Pelliot 6).



The number 100 𐰺𐰾 (Pelliot 6).



The number 500 𐰺𐰾𐰺𐰾 (So 14485).



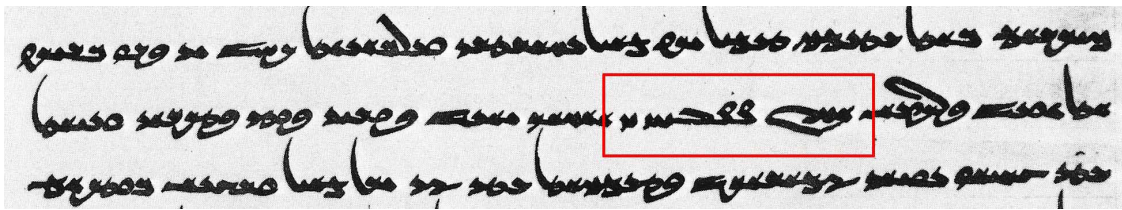
The number 1000 𐰺𐰾𐰺𐰾𐰺𐰾 (So 20165).

Figure 13: Specimens of numbers.



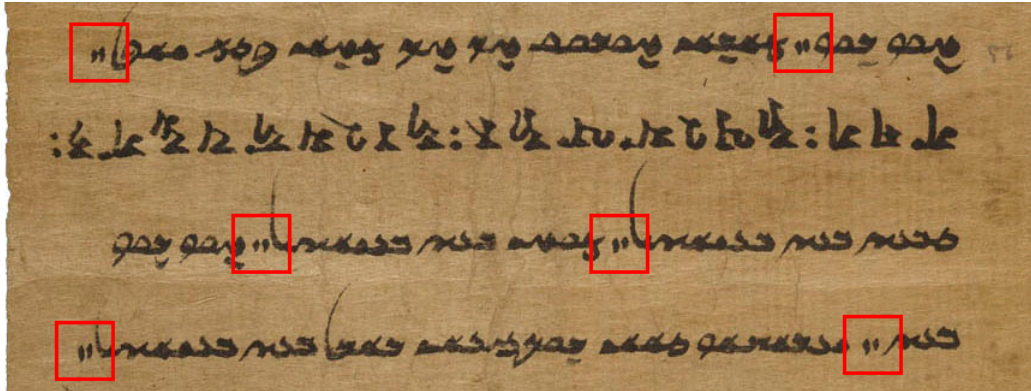


The number 81  $\text{𐰇𐰆𐰆𐰆}$  (So 18160 verso).

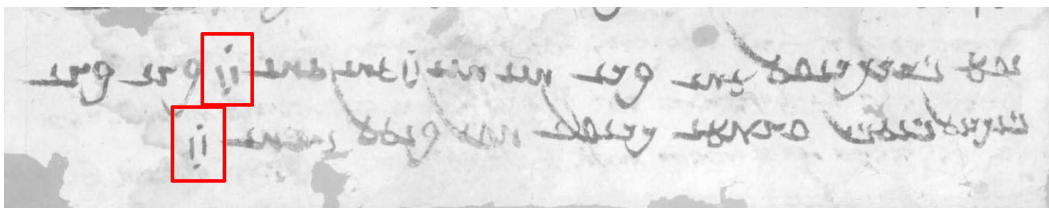


The number 155  $\text{𐰇 𐰆𐰆𐰆 𐰇𐰆}$  (from Dhyana text, line 93).

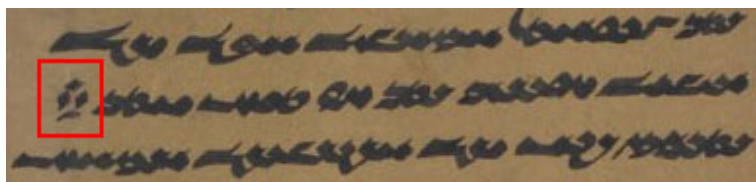
Figure 14: Additional specimens of numbers.



Usage of || SOGDIAN PUNCTUATION TWO VERTICAL BARS. From BL Or8212/174.

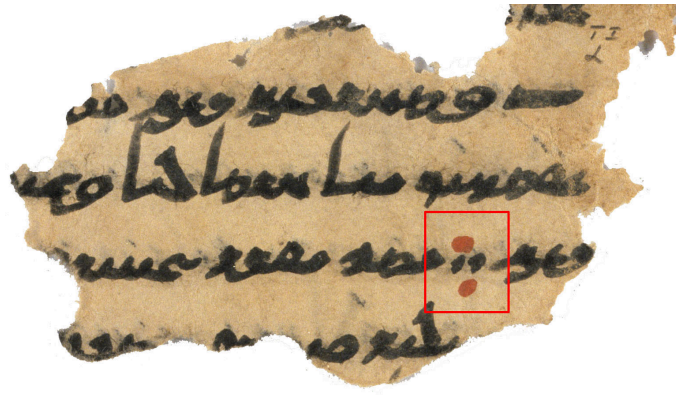


Usage of ꞥ SOGDIAN PUNCTUATION TWO VERTICAL BARS WITH DOTS ABOVE AND BELOW. From So 10006 verso.

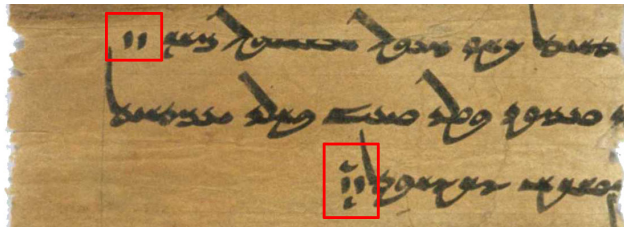


Representation of ꞥ SOGDIAN PUNCTUATION TWO VERTICAL BARS WITH DOTS ABOVE AND BELOW using the variant form ꞥ. From Pelliot 3.

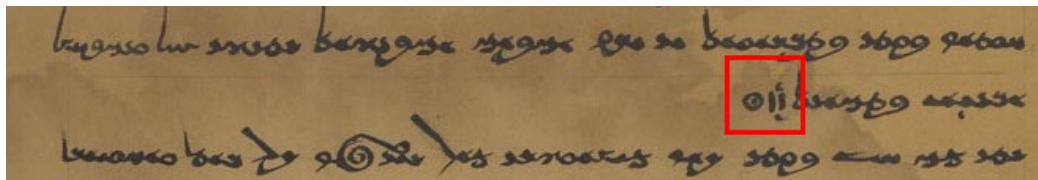
Figure 15: Specimens of punctuation signs. See figures 16 and 17 for additional specimens.



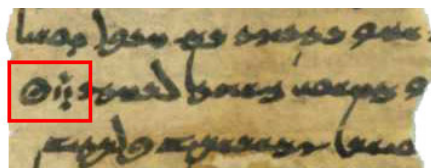
Representation of **ἢ** SOGDIAN PUNCTUATION TWO VERTICAL BARS WITH DOTS ABOVE AND BELOW in which the dots are marked using colored ink. From So 10100 (d) verso.



Usage of **||** SOGDIAN PUNCTUATION TWO VERTICAL BARS and **ἢ** SOGDIAN PUNCTUATION TWO VERTICAL BARS WITH DOTS ABOVE AND BELOW. The **ἢ** is represented using a variant form **ἢ̣̣**. From So 14800 recto.

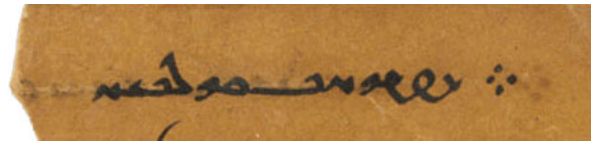


A sequence of **ἢ** SOGDIAN PUNCTUATION TWO VERTICAL BARS WITH DOTS ABOVE AND BELOW and **⊙** SOGDIAN PUNCTUATION CIRCLE WITH MIDDLE DOT. From Pelliot 6, line 17.

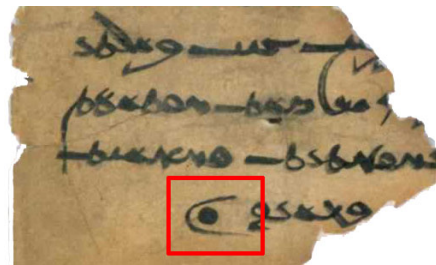


A sequence of **ἢ** SOGDIAN PUNCTUATION TWO VERTICAL BARS WITH DOTS ABOVE AND BELOW and **⊙** SOGDIAN PUNCTUATION CIRCLE WITH MIDDLE DOT. From So 10100 (i) verso.

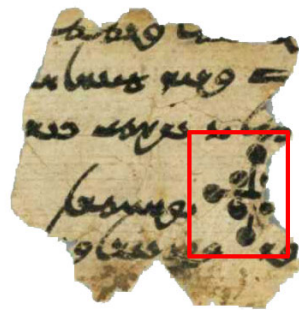
Figure 16: Additional specimens of punctuation signs. See figures 15 and 17 for other signs.



Usage of ❖ SOGDIAN PUNCTUATION FOUR DOTS (So 18224).



Usage of ◐ SOGDIAN PUNCTUATION HALF CIRCLE WITH MIDDLE DOT. From So 14700 (16a) recto.



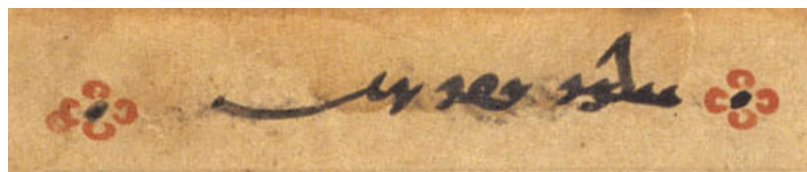
Usage of ✚ SOGDIAN PUNCTUATION CROSS WITH EIGHT DOTS. From So 10000 (4) recto.

Figure 17: Further specimens of punctuation signs. See figures 15 and 16 for other signs.

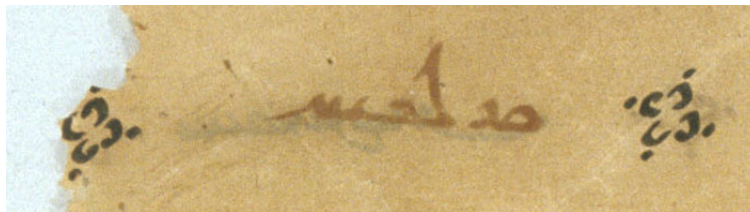




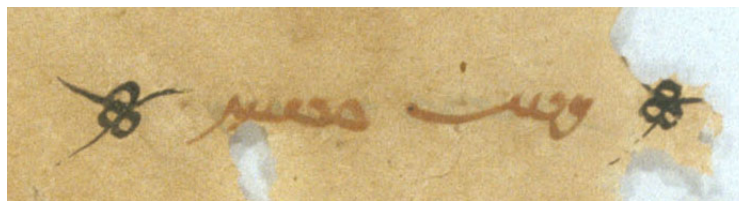
Usage of ☸ SOGDIAN PUNCTUATION FLOWER WITH THREE PETALS (So 20208).



Usage of ☸ SOGDIAN PUNCTUATION FLOWER WITH FOUR PETALS (So 14570).

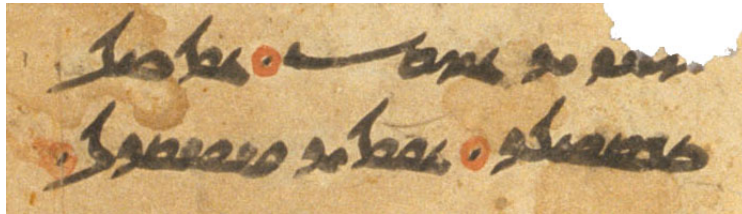


Usage of ☸ SOGDIAN PUNCTUATION FOUR HALF CIRCLES WITH DOTS (So 18248).

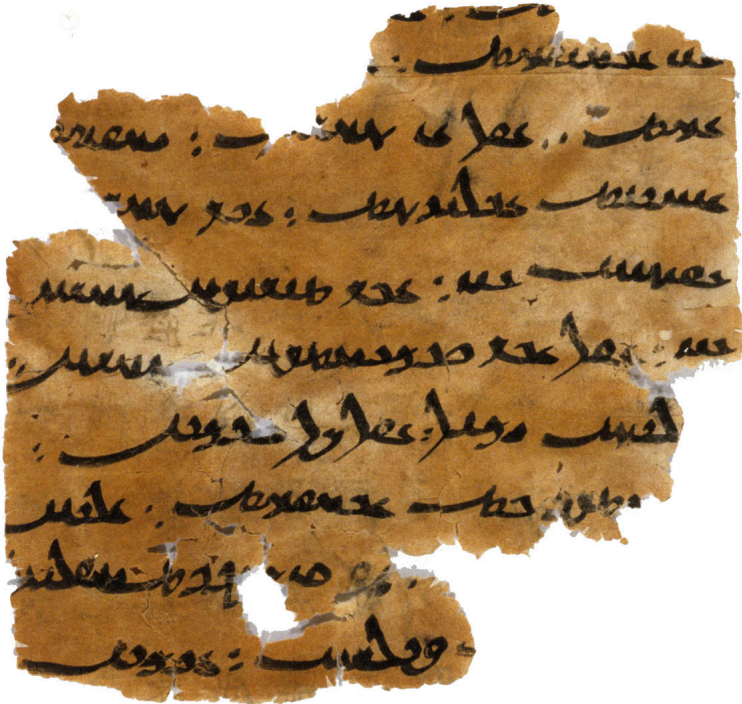


Usage of ☸ SOGDIAN PUNCTUATION THREECIRCLESWITH RAYS (So 18248).

Figure 18: Usage of text ornaments.



Usage of full-stop-like characters as punctuation (Ch/So 20208)



Usage of colon-like characters as punctuation (Ch/So 20501)

Figure 19: Usage of general punctuation.



Figure 20: An ostracon from Penjikent dated to the end of the 7th or first half of the 8th century listing the letters of the Sogdian script (Livshits 2015: 228). It contains 23 letters: “aleph, beth, gimel, daleth, he, waw, zayin, heth, teth, yodh, kaph, lamedh, mem, nun, samek, ayin, pe, sade, qoph, res, shin, taw, lamedh”.

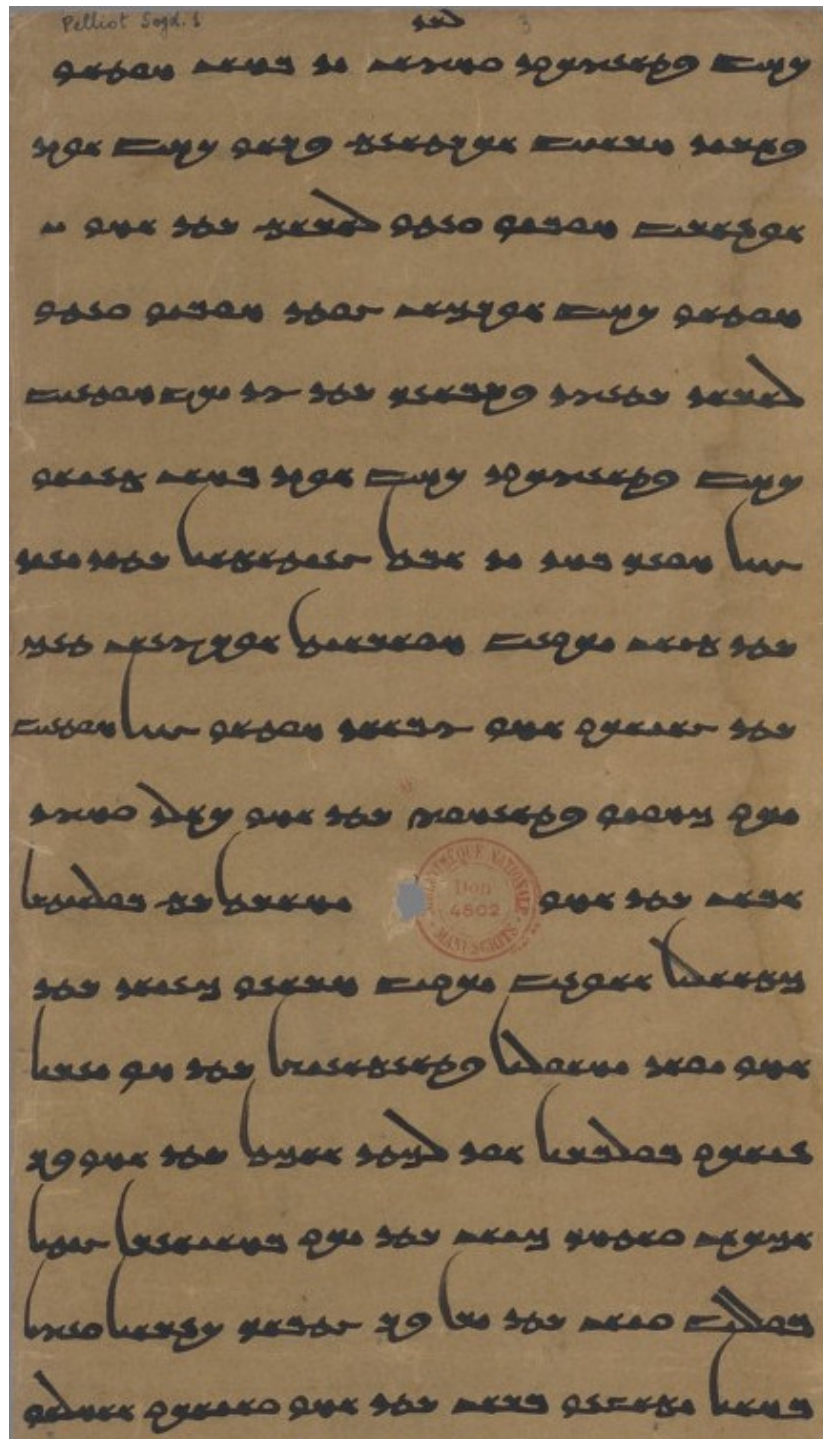


Figure 21: Excerpt from the Sogdian *Vessantara Jātaka* (Pelliot Sogdien 1).



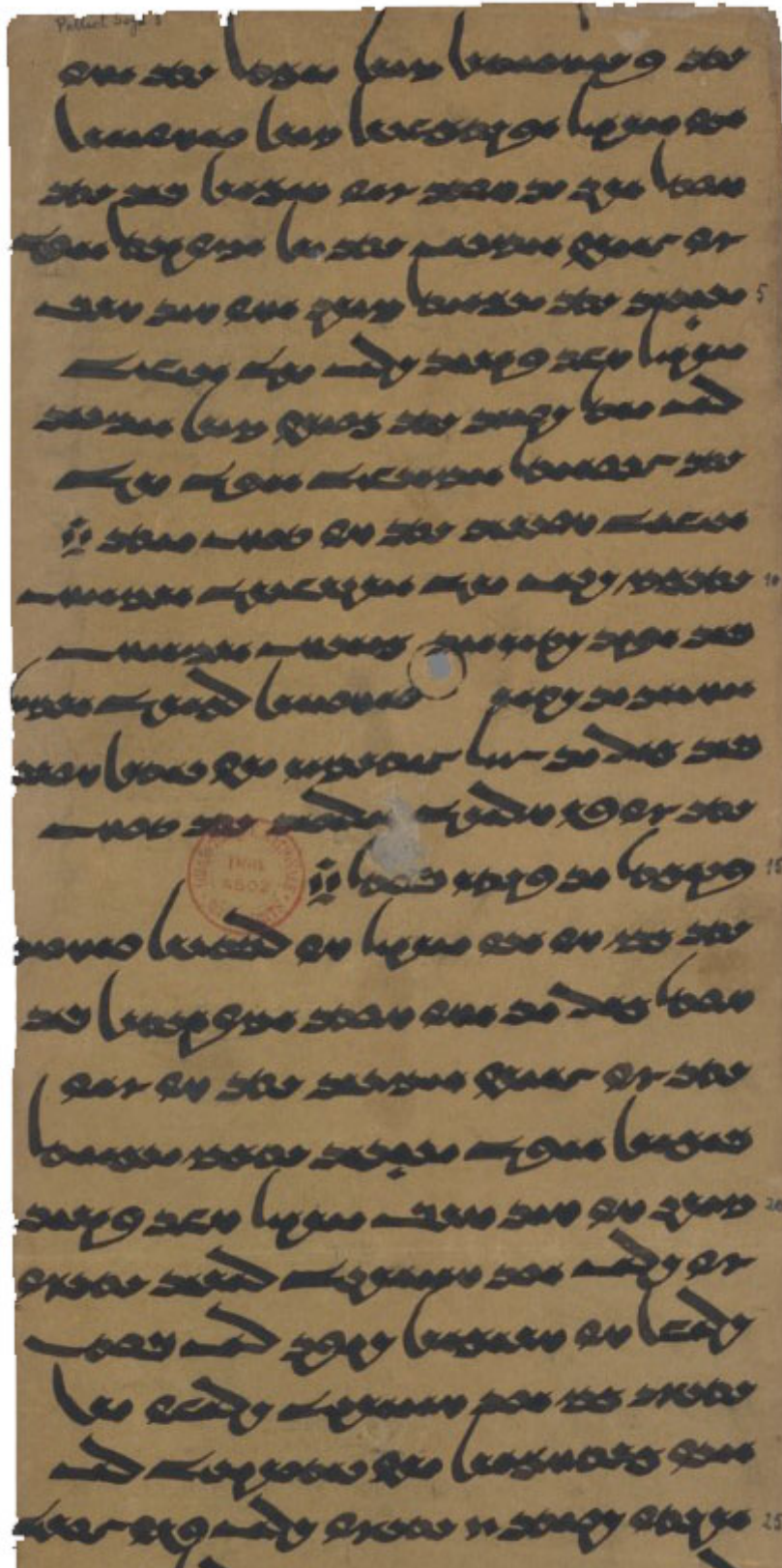


Figure 22: Excerpt from Pelliot Sogdien 3.

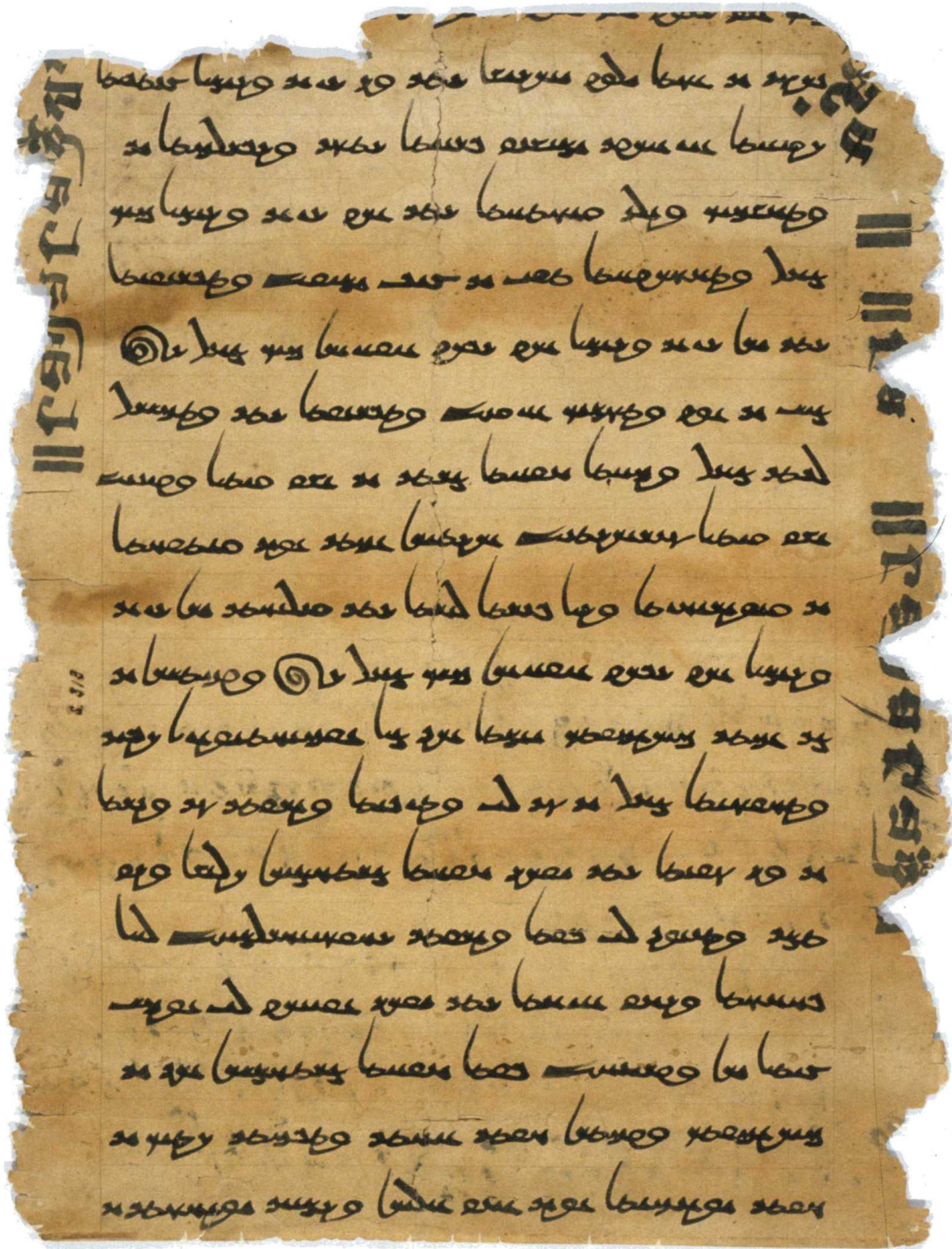


Figure 23: Fragment of the Sogdian *Samghāta Sūtra* (So 20165 r). Sutra script style. Turkestani Brahmi in the margins.



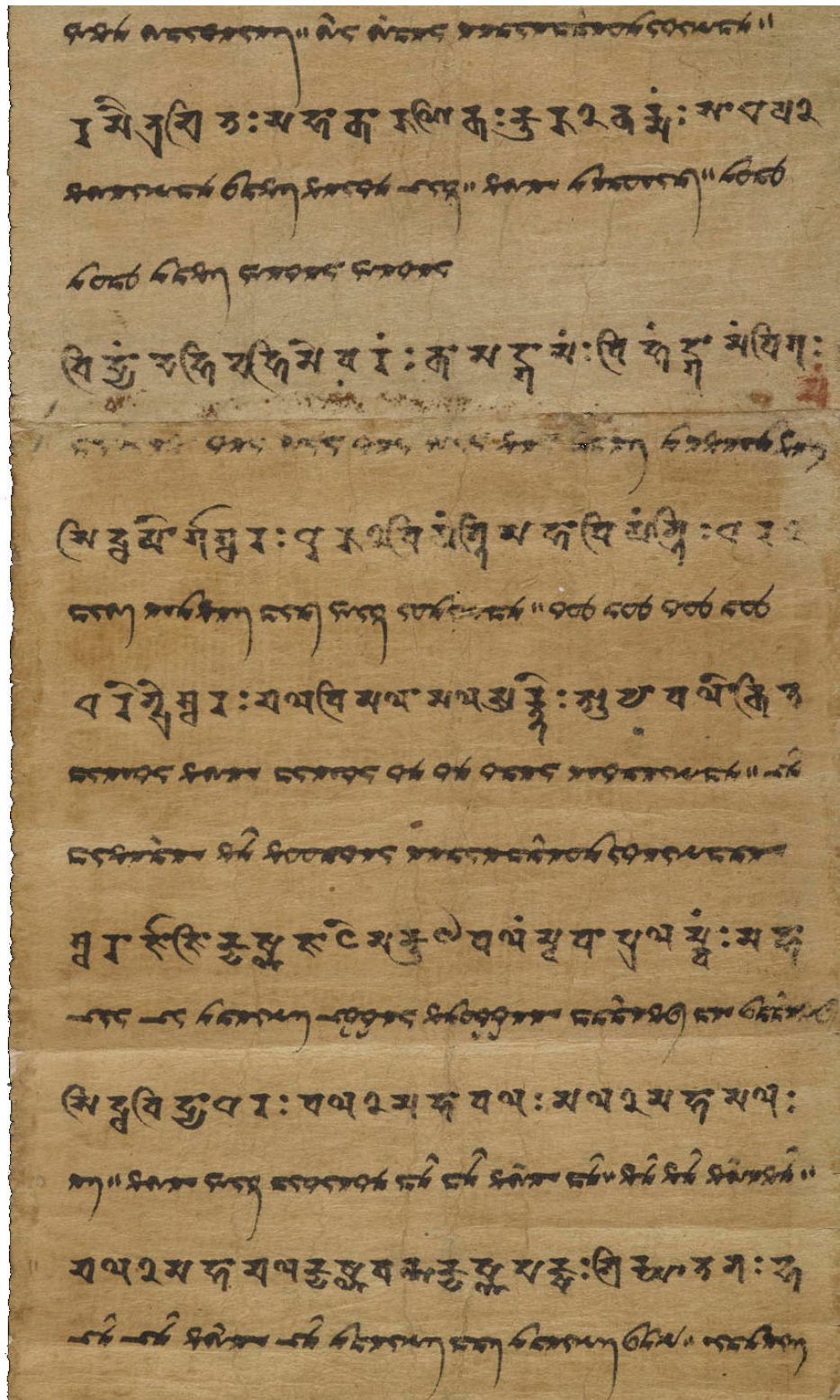


Figure 24: A folio of the *Nīlakaṇṭha Dhāraṇī* in sutra Sogdian and Siddham scripts (BL Or.8212/175). The Sogdian text is to be read by rotating the image 90° clockwise.



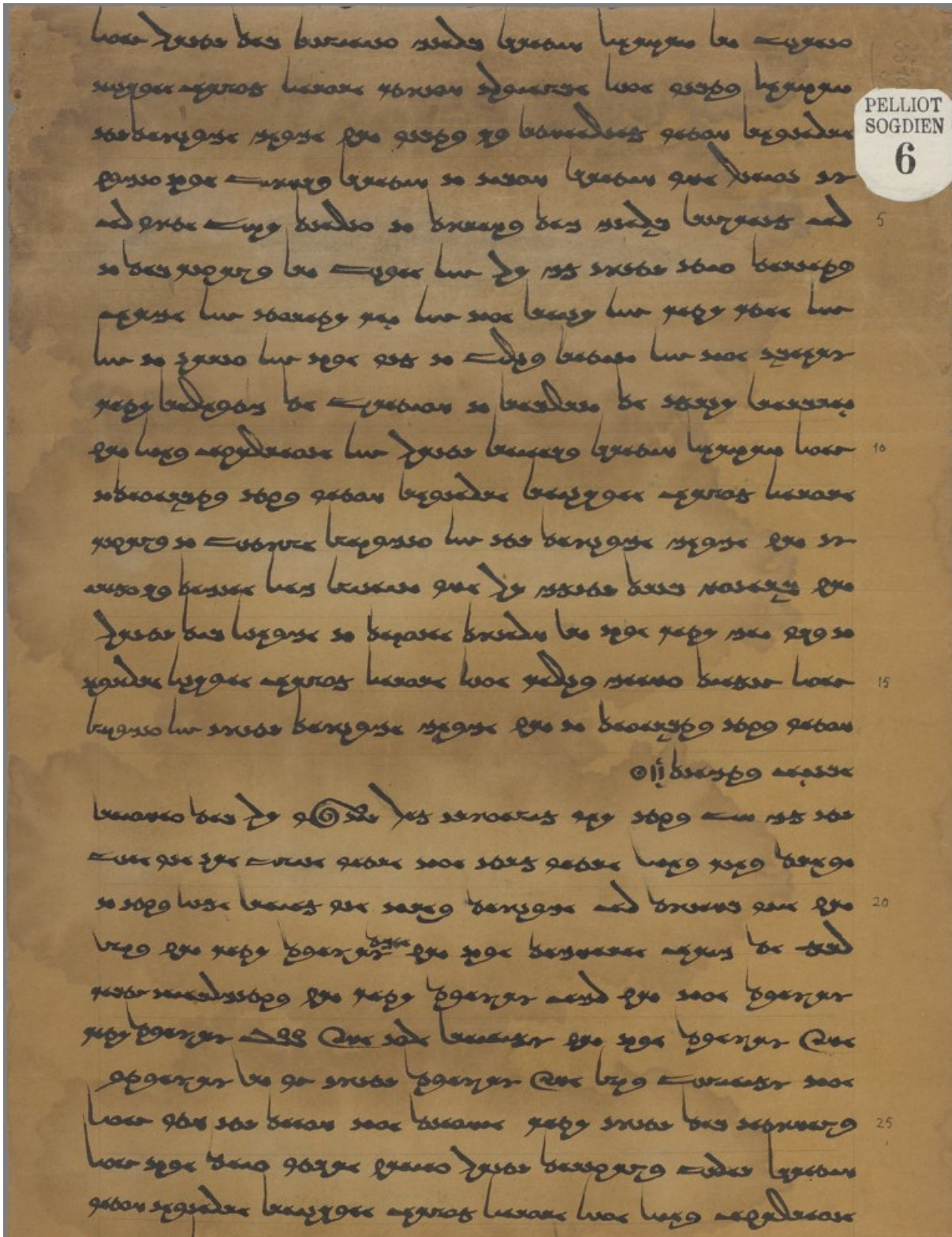


Figure 25: Excerpt from Pelliot Sogdien 6. Formal script.



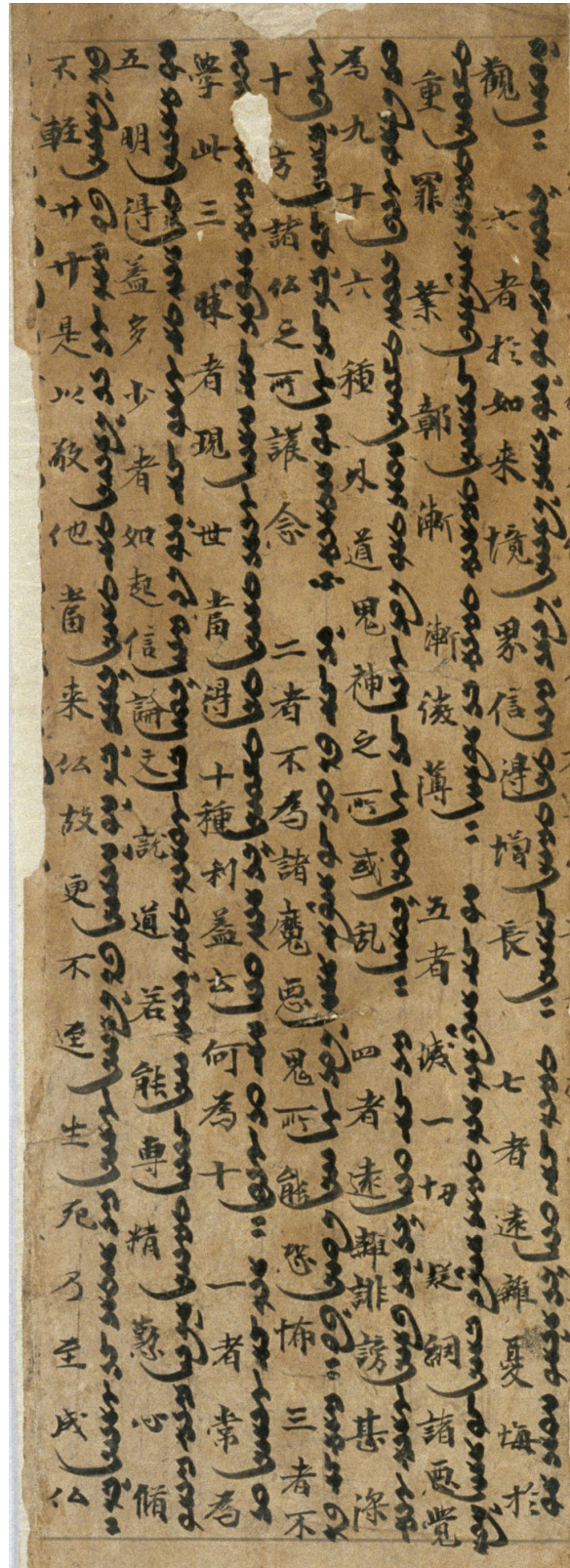


Figure 26: Transcription of a Chinese text in Sogdian (So 14830). Formal script.

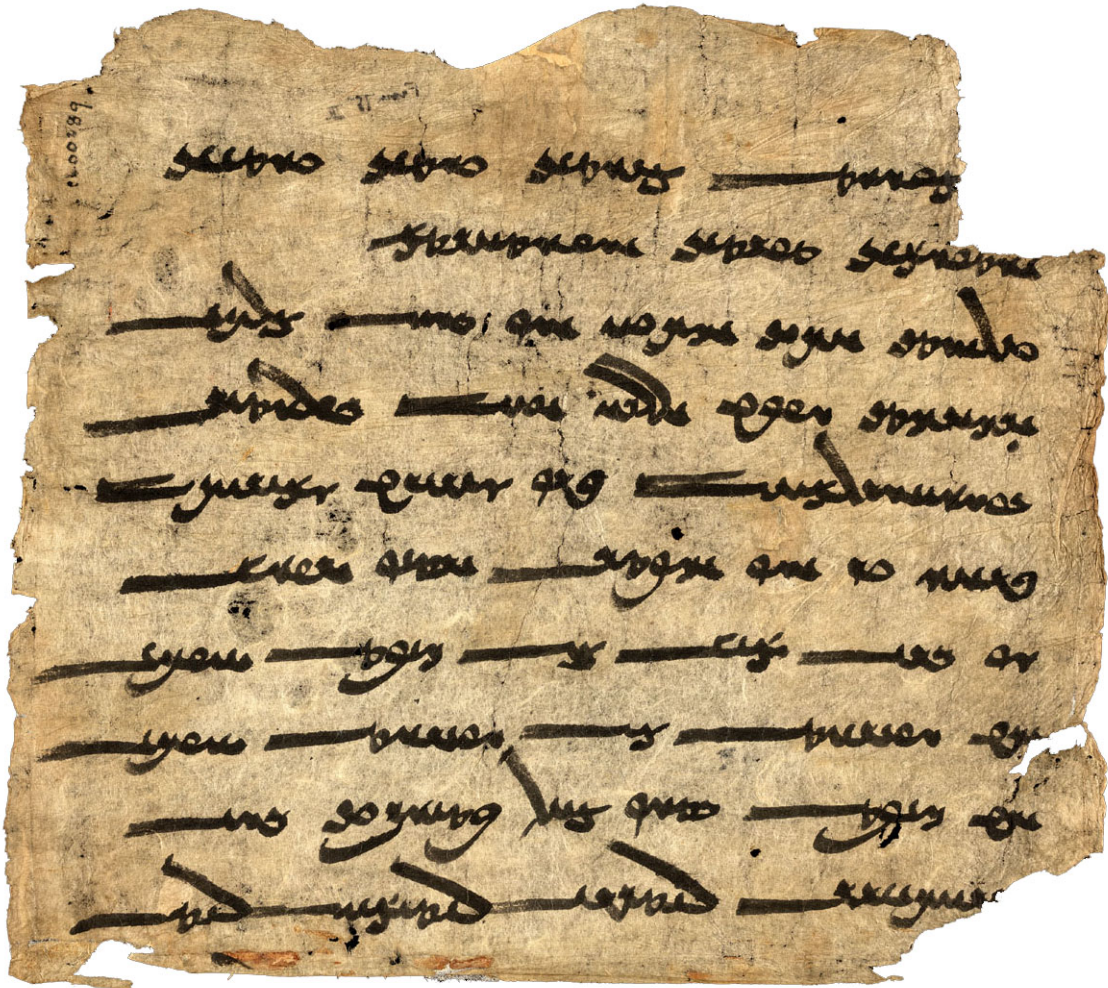


Figure 27: The Zoroastrian prayer, *ashem vohu* (*ašəm vohū*), in cursive Sogdian script, 10–11th c (Or.8212/84 recto).





Figure 28: Cursive Sogdian text (So 20165 v). Turkestani Brahmi in lower section.

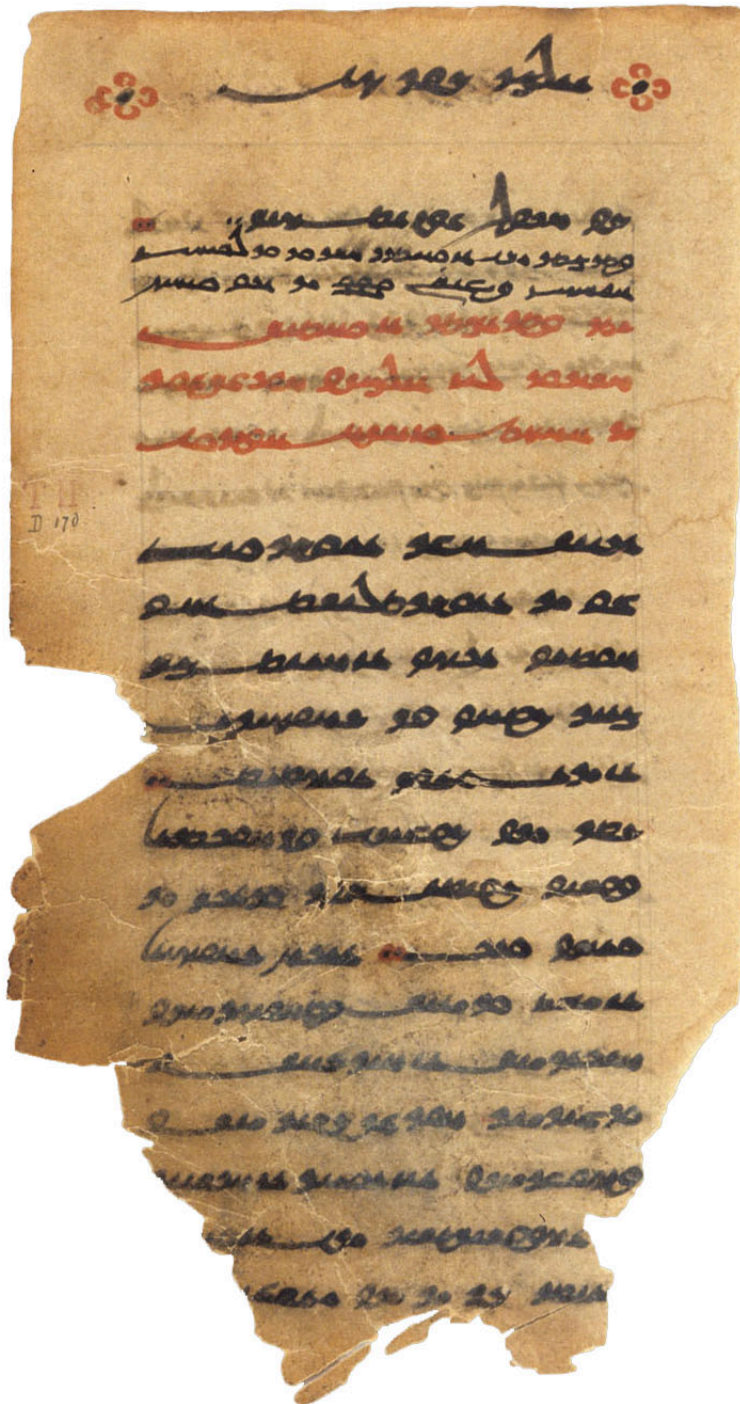


Figure 29: Cursive Sogdian manuscript (So 14570 recto).



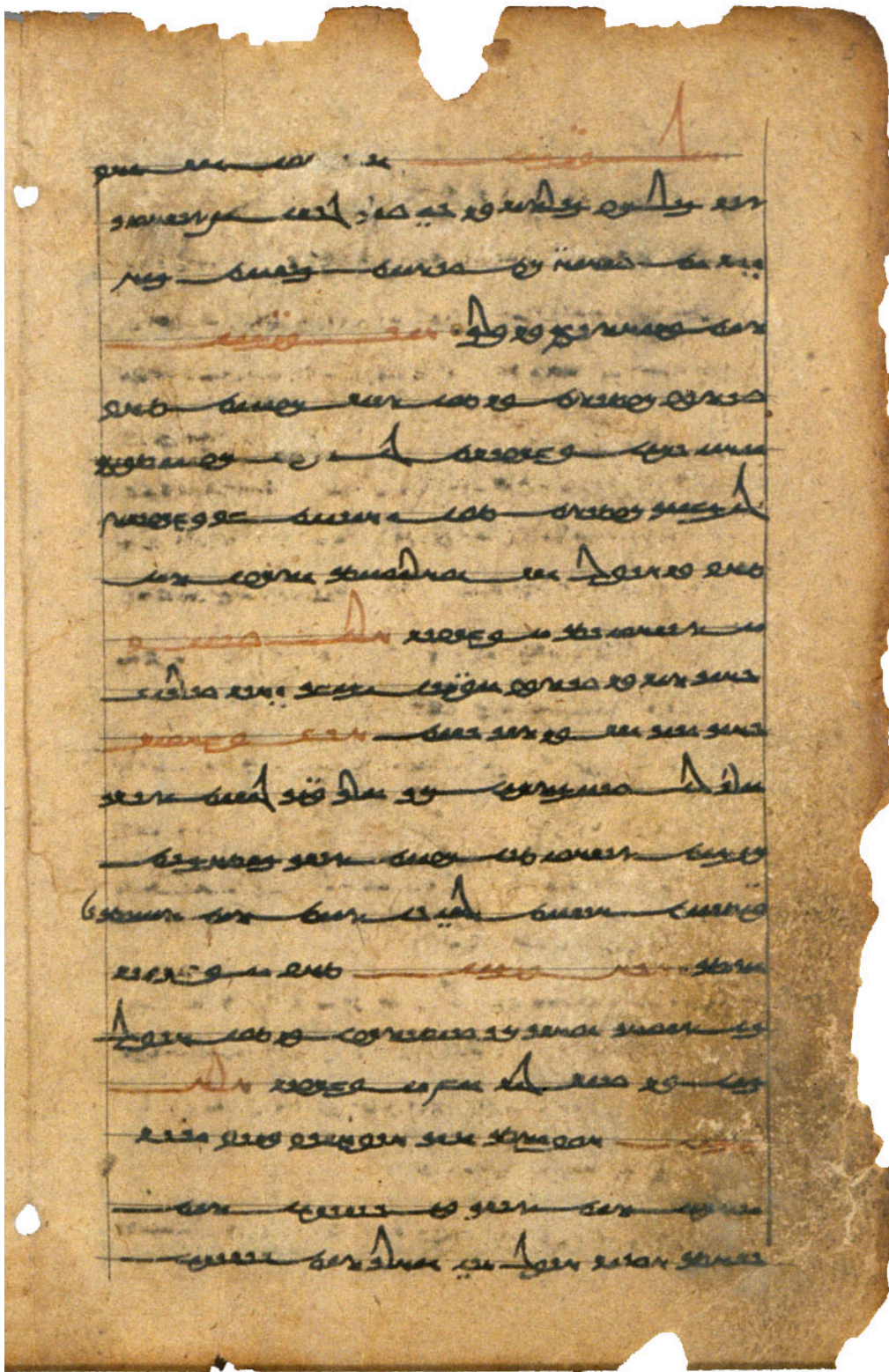


Figure 30: Excerpt from a cursive Sogdian manuscript (So 14410 verso).

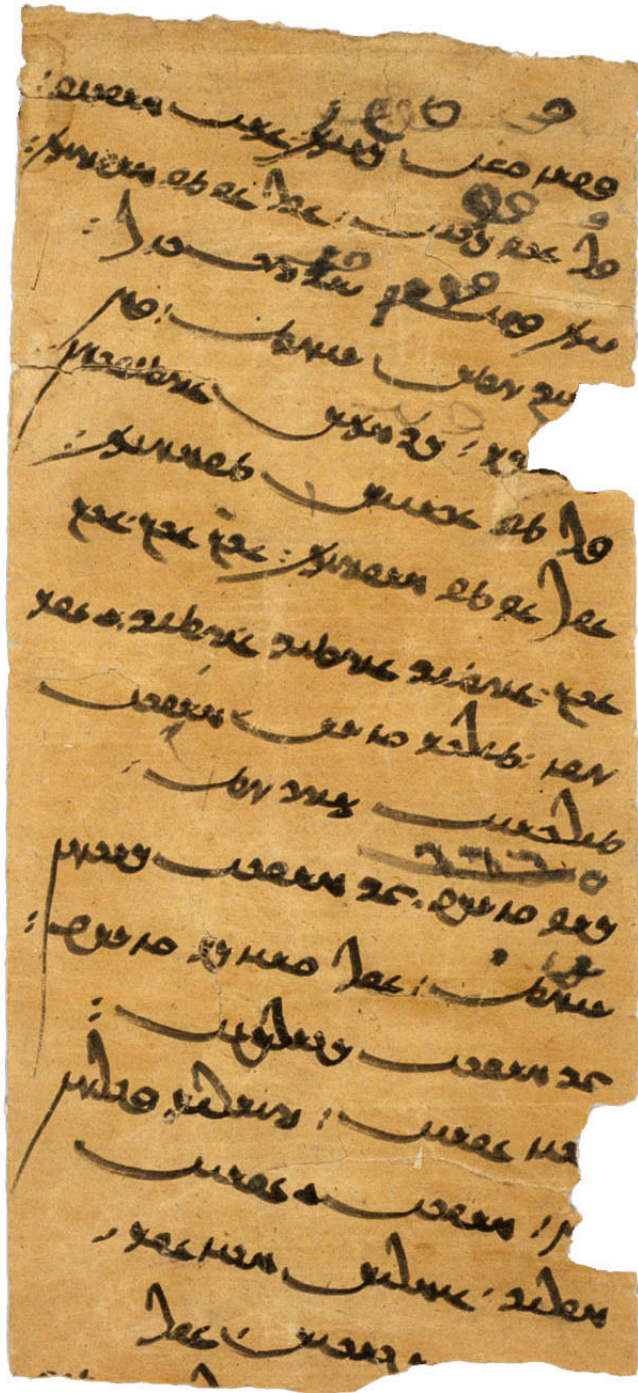


Figure 31: Cursive Sogdian manuscript (Ch/So 20135 verso).



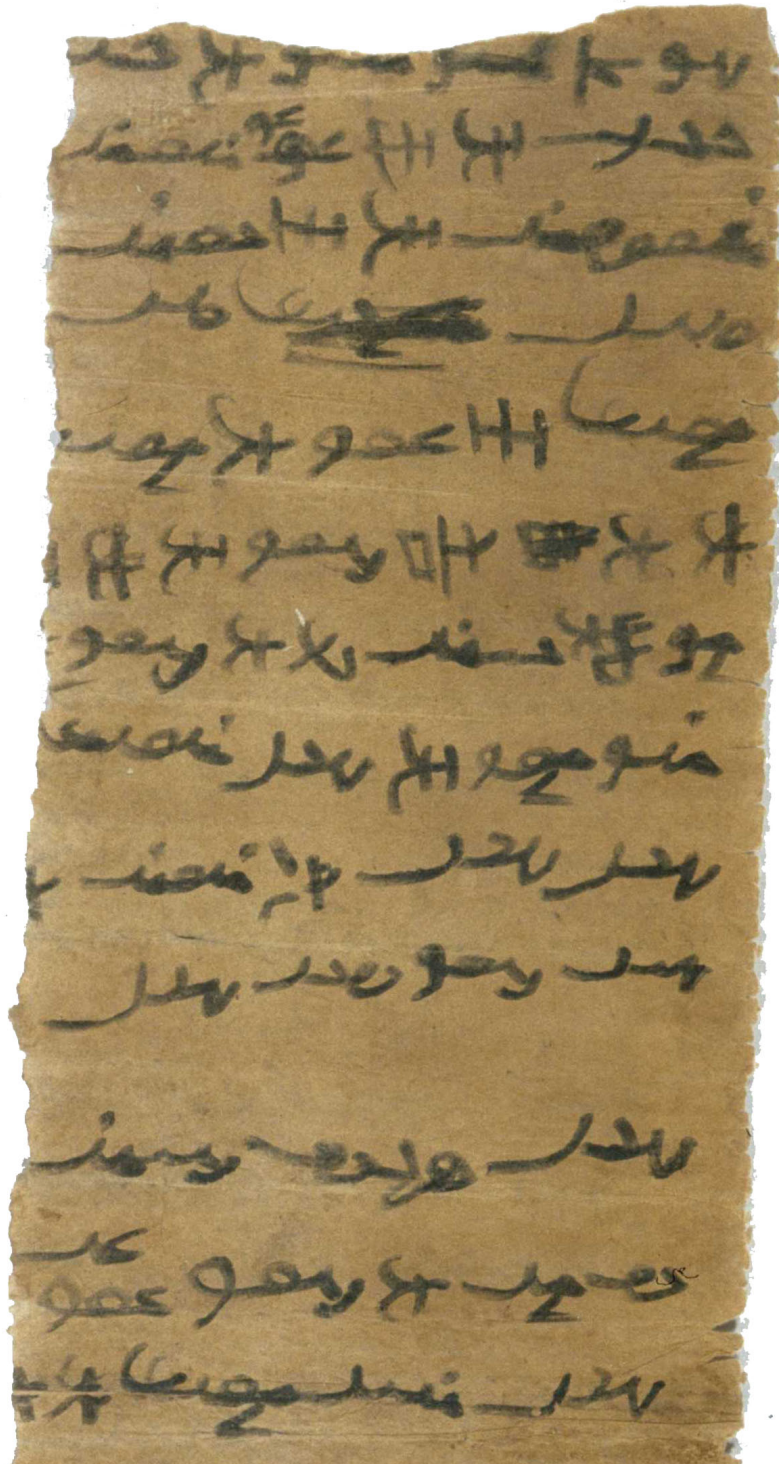


Figure 32: Cursive Sogdian with Chinese interspersed (excerpt of Ch/So 14800 verso).

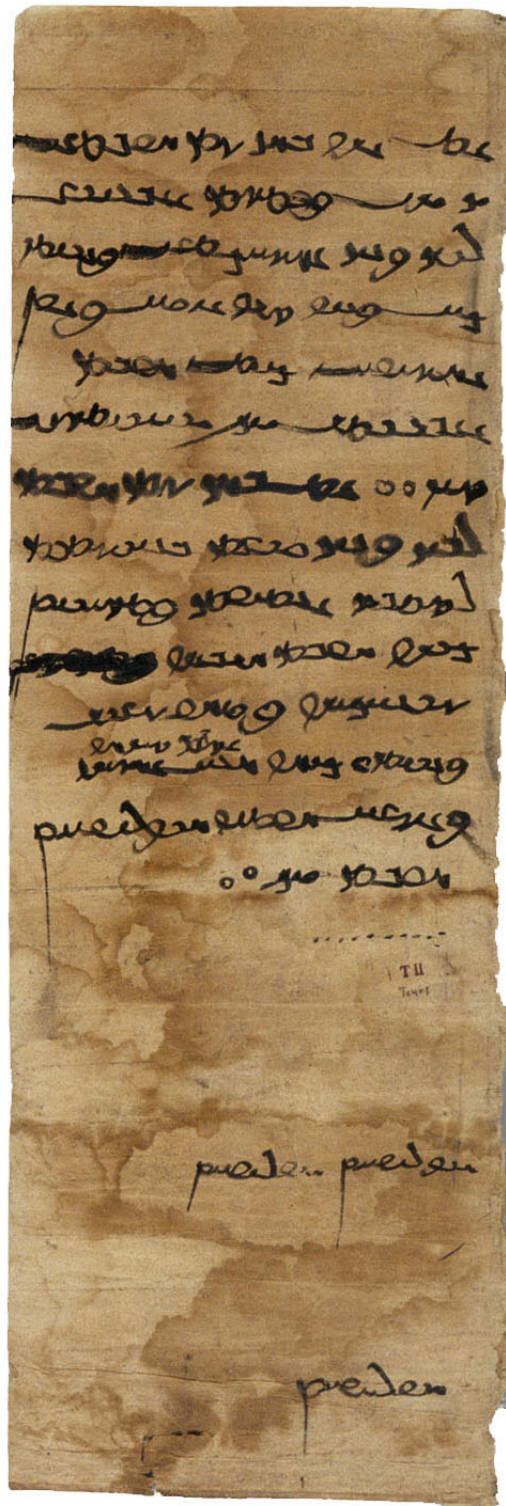
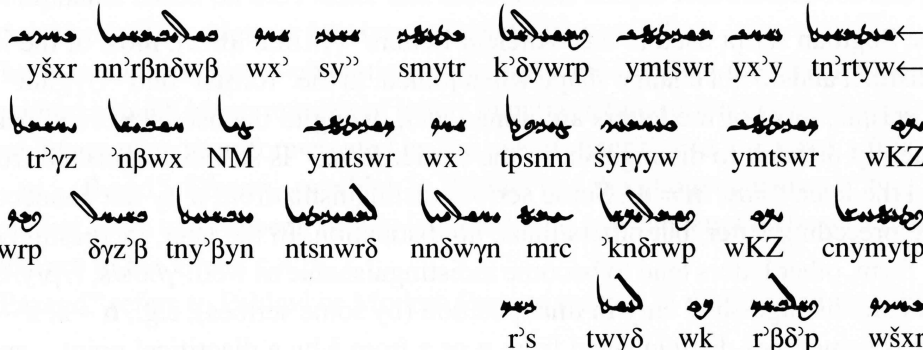


Figure 33: Excerpt from a cursive Sogdian manuscript (Ch/So 14730 verso).



## STORY OF RUSTAM


  
 yšxr nn'rβnδwβ wx' sy'' smytr k'dywrp ymtswr yx'y tn'rtyw←  
 tr'γz 'nβwx NM ymtswr wx' tpsnm šyryyw ymtswr wKZ  
 wrp δγz'β tny'βyn ntsnwrδ nnδwγn mrc 'knδrwp wKZ cnymytp  
 r's twyδ wk r'βδ'p wšxr

1. *Transliteration:* wytr'nt                      y'xy      rwstmy              prwyδ'k      rtyms
2. *Normalization:* wītarand                      yaxī      Rustami              parwēdē      rti-mas
3. *Gloss:*                      IMPF.they.departed brave      Rustam.GEN      to.seek      and-then

1. 'ys      'xw              βwδnβr'nn              rxšy              ZKw      rwstmy
2. āyas      axu              βōdan-βarān              Raxši              awu      Rustami
3. came      the.NOM      perception-bearing      Raxš.NOM      the.ACC      Rustam.ACC

1. wyryš              mnspt              'xw              rwstmy              MN      xwβn'      zy'rt
2. wīyrēš              manspat              axu              Rustami              čon      xuβna      žγart
3. IMPF.he.woke      IMPF.arose      the.NOM      Rustam.NOM      from      sleep.ABL      quickly

1. ptymync              ZKw      pwrδnk' crm      nγwδnn      δrwnstn              nyβ'ynt
2. ptīmēnč              awu      pu'δang-čarm      nγōdan      δrūn-stan              nīβēnd
3. IMPF.he.donned      the.ACC      leopard-skin      garment      bow-container      IMPF.he.tied

1. β'zγδ              prw              rxšw              p'δβ'r              kw      δywt              s'r
2. βāžγaδ              par-ō              Raxšu              pāθfār              kū      δēwt              sār
3. IMPF.mounted      on-the.ACC      Raxš.ACC      IMPF.hurried      to-      demon.PL      -ward

'They (the demons) departed in search of the brave Rustam. Then came the perceptive(?) Rakhsh (his horse) and woke Rustam. Rustam arose out of his sleep, quickly donned (his) leopard-skin garment, tied on his bow-case, mounted Rakhsh, and hurried toward the demons.'

—From a (Manichean?) version of the story of Rustam (Benveniste 1940A, pls. 193–94, 1940B: 135; Sims-Williams 1976: 54–57).

Figure 34: Example of printed Sogdian (from Skjærvø 1996: 530).

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, ū
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	𐭺	𐭺, 𐭻	𐭺	𐭺	n
s	𐭼	𐭼, 𐭽	𐭼	𐭼	s
ʿ	𐭽	𐭽	𐭽	𐭽	Ø
p	𐭿	𐭿	𐭿	𐭿	p
(f)			𐭿	𐭿	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 35: Table showing various scripts for writing Sogdian (from Skjærvø 1996: 519).