L2/16-158
2016-05-09

Preliminary proposal to encode Sogdian in Unicode

Anshuman Pandey
pandey@umich.edu

May 9, 2016

1 Introduction

This document presents a draft encoding for the Sogdian script in Unicode. The Sogdian script discussed here is the ‘cursive’ or ‘sutra’ script. Although related to ‘Old Sogdian’ it is proposed for encoding as a separate block on account of its structure (see L2/15-089R). The ‘Sogdian Uyghur’ script may be a candidate for unification with the present script, but further research is required regarding this matter.

The preliminary encoding for Sogdian is intended solely to provide a foundation for discussion. It is not a formal proposal. The character repertoire and character names are tentative. The representative glyphs are merely illustrative and are not intended to be typographically aesthetic; they are normalizations of diverse handwritten styles. Feedback is from the scholarly community is requested.

2 Script details

Sogdian is an *abjad* that is written from right to left. It is a structurally conjoining script similar to Arabic and Mongolian. As in these *abjad* systems, letters connect and change shape based upon their position within a word. Words are separated using spaces.

The script consists of 17 basic letters. Diacritic marks are used for indicating different values of base letters. A set of numbers are also attested. Signs for delimiting various sections of text are used.

The script is written both horizontally and vertically. In horizontal mode, the script is written in lines that proceed from top to bottom. In vertical mode, the writing direction is rotated 90° and lines proceed from top to bottom from the right edge of the writing surface towards the left.

2.1 Letters

Each letter is shown below in its isolated and positional forms. The alphabetic order is based upon Aramaic. The final form is tentatively selected as the isolated form because it is the most distinctive of all the forms. As is the case for structurally similar scripts such as Arabic, only the isolated form would be encoded. The the positional forms would exist only as glyphs in the font and the rendering engine would perform the necessary shaping. Although initial and medial forms may not be palaeographically distinctive, these are differentiated typographically in order to support the joining features of the script.

	Character name	Value	Final	Medial	Initial
𐼰	SOGDIAN LETTER ALEPH	ʾ, a, ā	𐼰	𐼰	𐼰
𐼱	SOGDIAN LETTER BETH	β	𐼱	𐼱	𐼱
𐼹	SOGDIAN LETTER DALETH	δ	𐼹	𐼹	𐼹
𐼳	SOGDIAN LETTER HE	h	𐼳	—	—
𐼴	SOGDIAN LETTER WAW	w, u, ū, o	𐼴	𐼴	𐼴
𐼵	SOGDIAN LETTER ZAYIN	z, ẓ	𐼵	𐼵	𐼵
𐼲	SOGDIAN LETTER HETH	γ, x	𐼲	𐼲	𐼲
𐼷	SOGDIAN LETTER YODH	y, i, ī, e	𐼷	𐼷	𐼷
𐼸	SOGDIAN LETTER KAPH	k	𐼸	𐼸	𐼸
𐼺	SOGDIAN LETTER MEM	m	𐼺	𐼺	𐼺
𐼻	SOGDIAN LETTER NUN	n	𐼻	𐼻	𐼻
𐼼	SOGDIAN LETTER SAMEKH	s	𐼼	𐼼	𐼼
𐼾	SOGDIAN LETTER PE	p	𐼾	𐼾	𐼾
𐼿	SOGDIAN LETTER SADHE	c, j	𐼿	𐼿	𐼿
𐽀	SOGDIAN LETTER RESH	r	𐽀	𐽀	𐽀
𐽁	SOGDIAN LETTER SHIN	š	𐽁	𐽁	𐽁
𐽂	SOGDIAN LETTER TAW	t	𐽂	𐽂	𐽂

All letters except for the following join to the left and right edges of adjacent letters:

- 𐼳 HE This letter is used only in final position. It does not have initial or medial forms.
- 𐼸 KAPH This letter joins to the right in medial and final positions. The top of the curve joins to adjacent letters in initial and medial positions.
- 𐼾 PE This letter joins to the right in medial and final positions. The top of the curve joins to adjacent letters in initial and medial positions.

2.2 Phonetic modifier signs

The following signs are used for indicating different pronunciations or phonetic values of letters:

	Character name
◌𐽊	SOGDIAN MODIFIER MARK-1
◌𐽋	SOGDIAN MODIFIER MARK-2

For example, in the Sogdian transcription of the *Nīlakaṇṭha Dhāraṇī*, these marks are used with base letters to represent non-Sogdian sounds:

𐼶𐽋	h	𐼶 HETH + ◌𐽋
𐽀𐽋	l	𐽀 RESH + ◌𐽋
𐽂𐽊	ṭ	𐽂 TAW + ◌𐽊

2.3 Punctuation

The following marks of punctuation are used in various sources for marking phrases and end of text sections (see figures 3, 4, 5).

	Character name
𐽕	SOGDIAN PUNCTUATION MARK-1
𐽖	SOGDIAN PUNCTUATION MARK-2
𐽗	SOGDIAN PUNCTUATION MARK-3

Some punctuation signs occur in combination (see figure 5).

2.4 Abbreviation

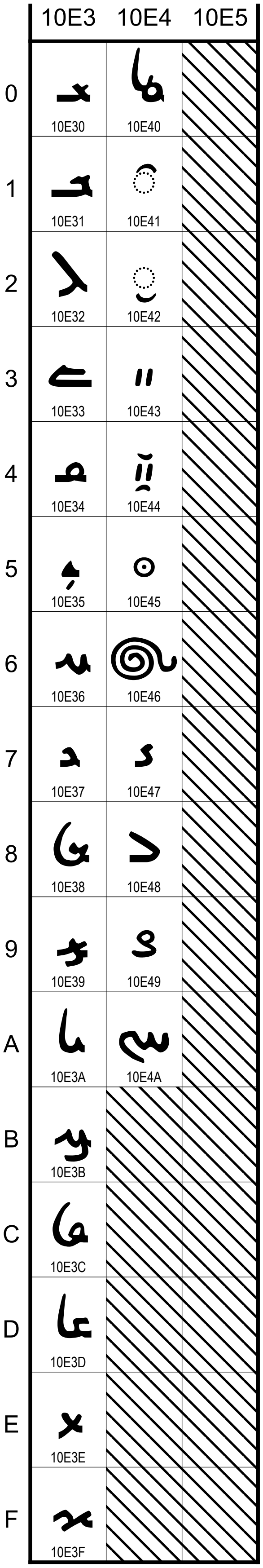
The sign 𐼽 occurs in a few sources as an abbreviation for an Aramaic heterogram (see figures 2 and 5). It may be an embellished form of the letter *ayin*, which is not distinctively represented in Sogdian. It is a connecting character, as shown in figure 5. It may be appropriate to encode this sign as a letter.

2.5 Numbers

Numbers are attested in Sogdian manuscripts:

	Character name
𐽑	SOGDIAN NUMBER ONE
𐽒	SOGDIAN NUMBER TEN
𐽓	SOGDIAN NUMBER TWENTY
𐽔	SOGDIAN NUMBER HUNDRED

See figures 2 and 5 for examples. Numbers in Sogdian follow the same pattern of numeration as in the ‘Old Sogdian’ script (see L2/15-089R).


10E3
10E4
10E5
0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
10E30
10E31
10E32
10E33
10E34
10E35
10E36
10E37
10E38
10E39
10E3A
10E3B
10E3C
10E3D
10E3E
10E3F
10E40
10E41
10E42
10E43
10E44
10E45
10E46
10E47
10E48
10E49
10E4A

Representative glyphs are based upon the 'sutra' style.

Letters

10E30	SOGDIAN LETTER ALEPH
10E31	SOGDIAN LETTER BETH
10E32	SOGDIAN LETTER DALETH
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 MEM
10E3A	SOGDIAN LETTER NUN
10E3B	SOGDIAN LETTER SAMEKH
10E3C	SOGDIAN LETTER PE
10E3D	SOGDIAN LETTER SADHE
10E3E	SOGDIAN LETTER RESH
10E3F	SOGDIAN LETTER SHIN
10E40	SOGDIAN LETTER TAW

Various signs

10E41	SOGDIAN MODIFIER SIGN ABOVE
10E42	SOGDIAN MODIFIER SIGN BELOW

Punctuation

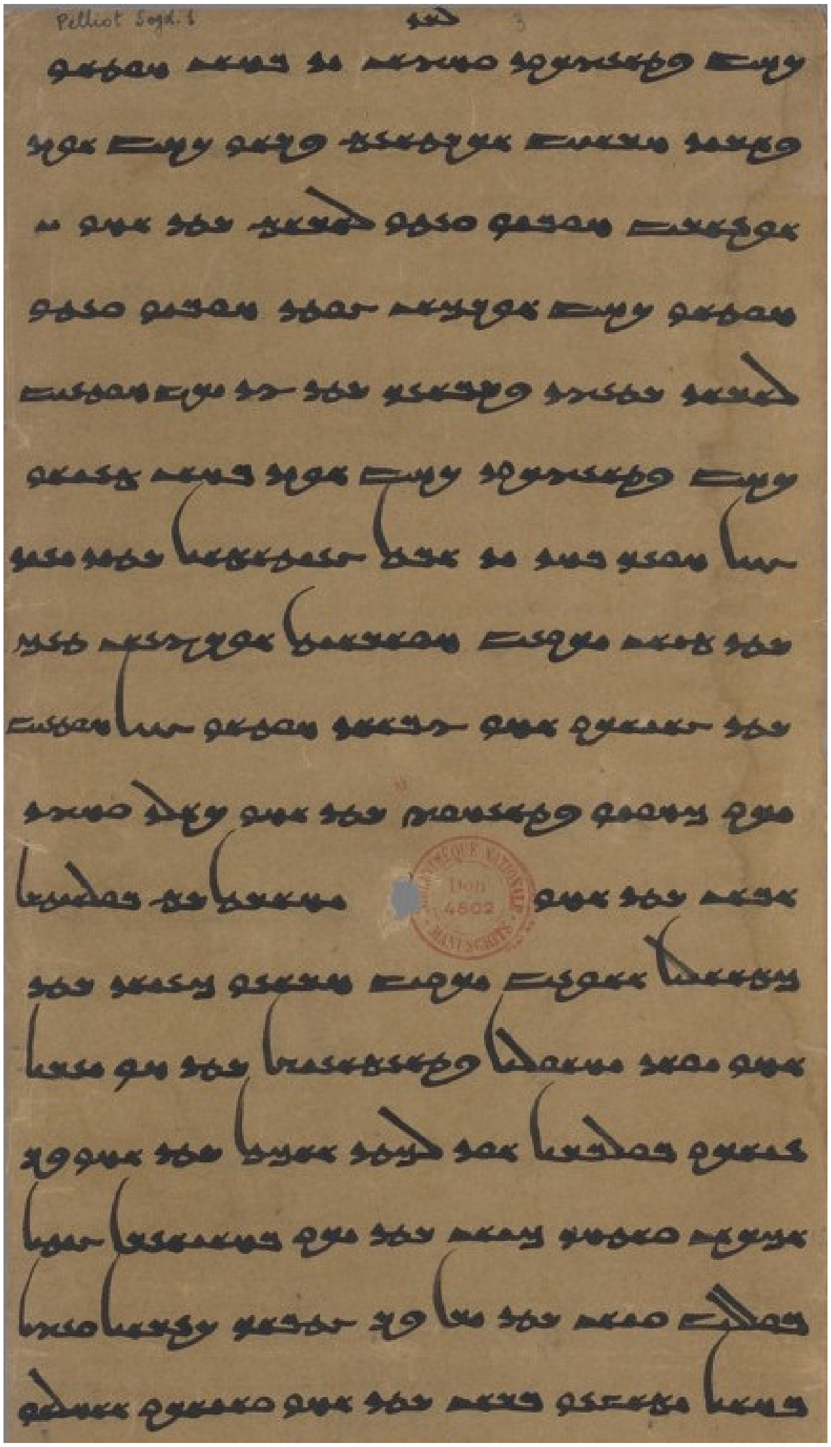
10E43	SOGDIAN PUNCTUATION MARK-1
10E44	SOGDIAN PUNCTUATION MARK-2
10E45	SOGDIAN PUNCTUATION MARK-3

Abbreviation

10E46	SOGDIAN ABBREVIATION AYIN

Numbers

10E47	SOGDIAN NUMBER ONE
10E48	SOGDIAN NUMBER TEN
10E49	SOGDIAN NUMBER TWENTY
10E4A	SOGDIAN NUMBER HUNDRED



Pelliot Sogd. 1

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

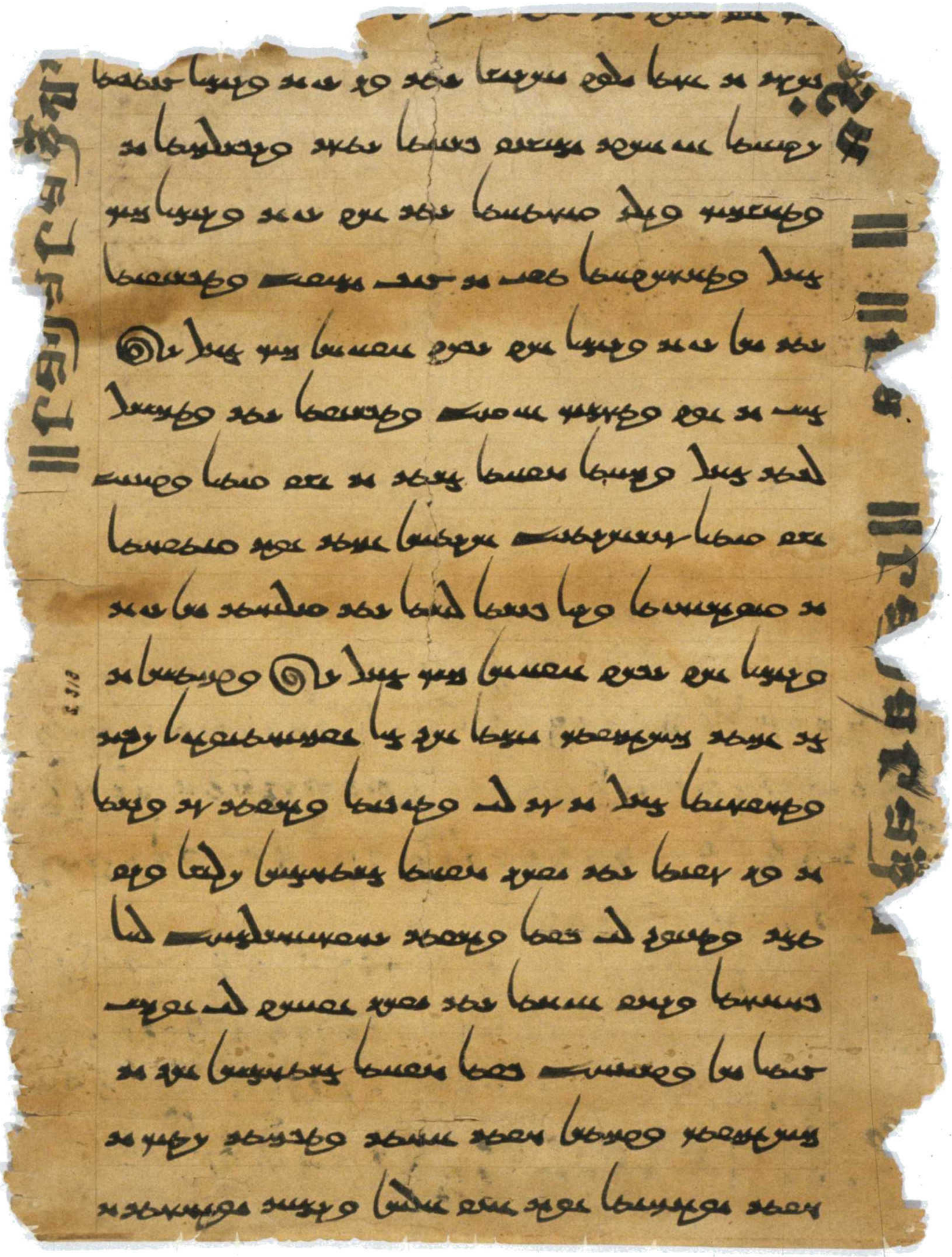


Figure 2: Fragment of the Sogdian *Saṃghaṭa Sūtra* (So 20165 r). The abbreviation ◎ occurs in lines 6, 11. The number 𐽑𐼹𐼾 ‘1000’ occurs in the word 𐽑𐼹𐼾𐼴 *1LPw* in line 2.

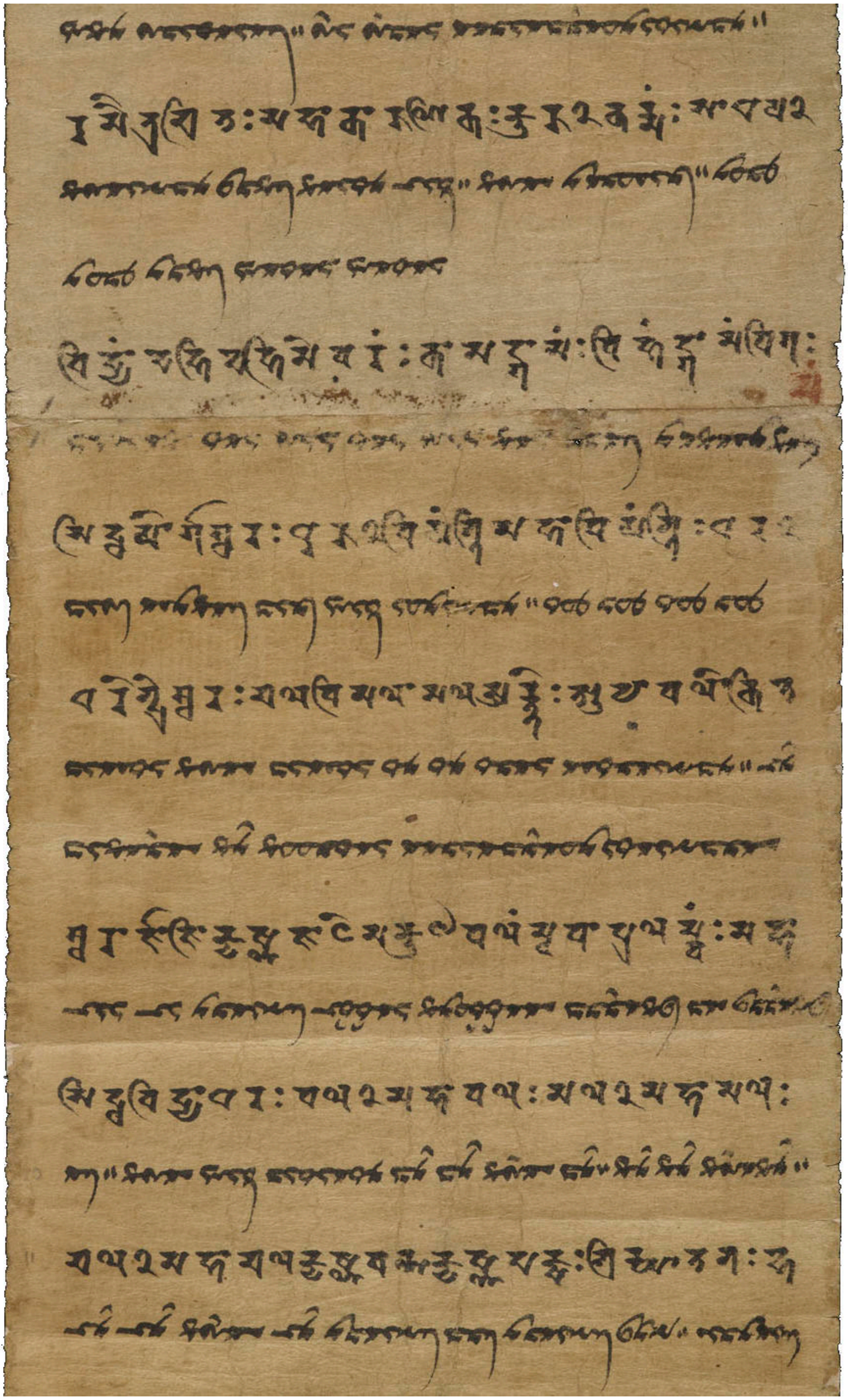


Figure 3: Excerpt (3/9) of the *Nīlakaṇṭha Dhāraṇī* (BL Or.8212/175). The Sogdian text is to be read by rotating the 90° clockwise. The consonant modifier signs ◌̑ occurs in line 13 and ◌̮ occurs in lines 1, 10, 11, 13, 15, 17.

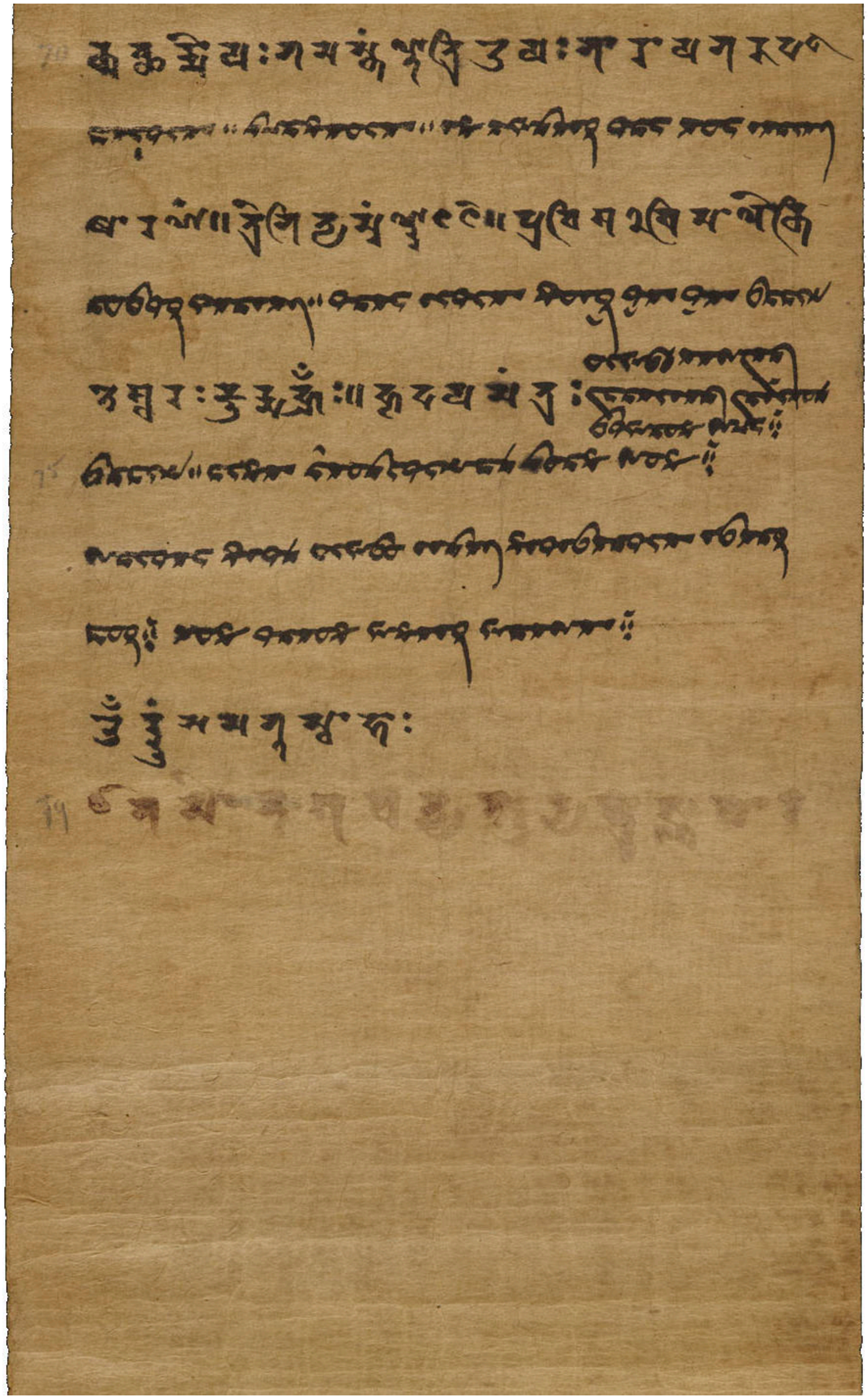


Figure 4: Excerpt (8/9) of the *Nīlakaṇṭha Dhāraṇī* (BL Or.8212/175). The Sogdian text is to be read by rotating the 90° clockwise. The punctuation sign 𐽕 occurs in lines 2, 4, 8; the sign 𐽖 occurs in lines 7, 8, 10.

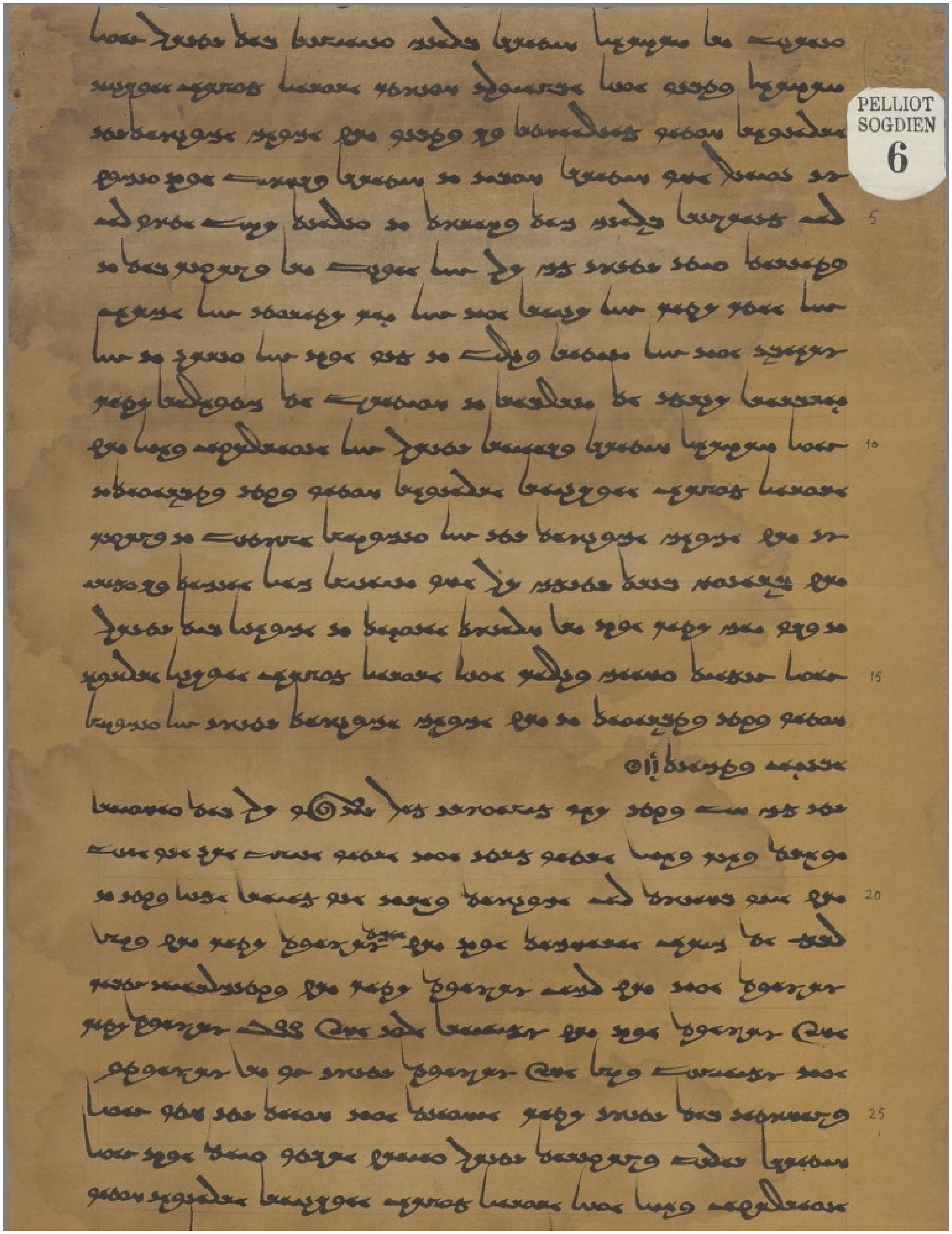


Figure 5: Excerpt from Pelliot Sogdien 6. The abbreviation 𐽐 occurs in line 18. A sequence of the punctuation signs 𐽖 and 𐽗 occurs in line 17. The number 𐽔 ‘100’ occurs in lines 23 and 24; the number 𐽓𐽓𐽒 ‘50’ occurs in line 23.