

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document

Title: Proposal for encoding the Psalter Pahlavi script in the SMP of the UCS

Source: UC Berkeley Script Encoding Initiative (Universal Scripts Project)

Authors: Michael Everson and Roozbeh Pournader

Status: Liaison Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2011-05-06

Replaces: N3286R2 (2007-09-18)

1. Introduction. The Psalter Pahlavi script was used to write Middle Persian. A conservative variety of Pahlavi used by Christians in Iran, it is so named because of its use in a fragmentary manuscript of the Psalms of David.

2. Processing. Psalter Pahlavi is an alphabetic script written right-to-left, usually with spaces between words. The script has fully-developed joining behaviour. The table below shows the joining forms.

Dual-joining Psalter Pahlavi Characters

Character	X _n	X _r	X _m	X _l
ALEPH	ⱪ	ⱪ	ⱪ	ⱪ
GIMEL	ⱪ	ⱪ	ⱪ	ⱪ
ZAYIN	ⱪ	ⱪ	ⱪ	ⱪ
HETH	ⱪ	ⱪ	ⱪ	ⱪ
YODH	ⱪ	ⱪ	ⱪ	ⱪ
LAMEDH	ⱪ	ⱪ	ⱪ	ⱪ
MEM-QOPH	ⱪ	ⱪ	ⱪ	ⱪ
SAMEKH	ⱪ	ⱪ	ⱪ	ⱪ
SHIN	ⱪ	ⱪ	ⱪ	ⱪ

Right-joining Psalter Pahlavi Characters

Character	X _n	X _r
BETH	ⱪ	ⱪ
DALETH	ⱪ	ⱪ
HE	ⱪ	ⱪ
WAW-AYIN-RESH	ⱪ	ⱪ
KAPH	ⱪ	ⱪ
NUN	ⱪ	ⱪ
PE	ⱪ	ⱪ
SADHE	ⱪ	ⱪ
TAW	ⱪ	ⱪ

3. Names and ordering. The names used for the Psalter Pahlavi characters are based on their Imperial Aramaic analogues. The order of the characters in the code charts is their alphabetical order. The historical characters WAW, AYIN, and RESH and MEM and QOPH fell together into two characters, named here WAW-AYIN-RESH and MEM-QOPH respectively.

4. Numbers. Psalter Pahlavi has its own numbers, which have right-to-left directionality. Numbers are built up out of 1, 2, 3, 4, 10, 20, and 100; the number 1000 is not attested in Psalter Pahlavi. The following is an exhaustive list of numbers attested in Psalter Pahlavi. The third column is displayed in visual order.

3	س	3 ←	<i>Hérat Cross l. 3r</i>
96	سوسوسوس	3 + 3 + 10 + 20 + 20 + 20 + 20 ←	<i>f.2r.13</i>
98	سوسوسوس	4 + 4 + 10 + 20 + 20 + 20 + 20 ←	<i>f.3r.1</i>
99	[سوسوسوس]	[3 + 3 + 3 + 10] + 20 + 20 + 20 + 20 ←	<i>f.3v.5</i>
122	وس ١ س	2 + 20 + WAW + 100 ←	<i>f.5r.10</i>
123	وس ١ س	3 + 20 + WAW + 100 ←	<i>f.5v.2</i>
124	وس ١ س	4 + 20 + WAW + 100 ←	<i>f.5v.18</i>
125	وس ١ س	2 + 3 + 20 + WAW + 100 ←	<i>f.6r.11</i>
126	وس ١ س	3 + 3 + 20 + WAW + 100 ←	<i>f.6v.6</i>
127	وس ١ س	3 + 4 + 20 + WAW + 100 ←	<i>f.7r.7</i>
128	وس ١ س	4 + 4 + 20 + WAW + 100 ←	<i>f.7v.4</i>
129	وس ١ صص	ALEPH + SHIN + TAW + 20 + WAW + 100 ←	<i>f.8r.4</i>
130	وس ١ ك	10 + 20 + WAW + 100 ←	<i>f.8v.2</i>
131	[وس ١ ك]	[1] + 10 + 20 + WAW + 100 ←	<i>f.8v.12</i>
132	[وس ١ ك]	[2 + 10] + 20 + WAW + 100 ←	<i>f.9v.14</i>
134	[وس ١ ك]	[4 + 10 + 20 + WAW] + 100 ←	<i>f.10r.3</i>
135	وس ١ ك	2 + 3 + 10 + 20 + WAW + 100 ←	<i>f.11v.2</i>
136	[وس ١ كصص]	ALEPH + TAW + SHIN + [10 + 20 + WAW + 100] ←	<i>f.12v.10</i>
507	سوسوس	3 + 4 + 100 + 2 + 3 ←	<i>Hérat Cross l. 1v</i>

Note that “129” is written “100 and 20 nine” (١ W means *ud* ‘and’ صص TŠA means *nō* ‘nine’); “136” is written “100 and 30 six” (صص ŠTA means *šāš* ‘six’).

Some Psalter Pahlavi numbers have joining behaviour. Note that they can join with letters as well as numbers (see 20-TAW in 129 above). The numbers 10 and 20 are dual joining, and the numbers 1, 2, 3, and 4 are right-joining, although these change their form in some combinations. The nominal forms of the numbers 2, 3, and 4 have a long final leg (μ, ρ, ρ), but in combination when used in building up other numbers they ligate with a short form in the first (س 5, س 6, س 7, س 8, س 9).

Dual-joining Psalter Pahlavi Numbers

Character	X _n	X _r	X _m	X _l
TEN	١	١	١	١
TWENTY	و	و	و	و

Right-joining Psalter Pahlavi Numbers

Character	X _n	X _r	X _{rm}	X _{nm}
ONE	١	١	١	١
TWO	μ	μ	μ	μ
THREE	ρ	ρ	ρ	ρ
FOUR	ρ	ρ	ρ	ρ

The following rules, R1 and R2, will need to be applied after the contextual shaping process in order to determine if the form with a short final leg should be used. These rules may be implemented either in rendering engines and/or by using font technology such as OpenType’s GSUB.

R1 A number 1, 2, 3, or 4 that has a right join-causing character to its right and a number 1, 2, 3, or 4 to its left will take the form of X_{rm} .

$$\mu 3 + \mu 4 + \omega 20 \rightarrow \mu + \mu\mu + \omega \rightarrow \mu\mu\mu\omega 27$$

R2 A number 1, 2, 3, or 4 that has no right join-causing character to its right and a number 1, 2, 3, or 4 to its left will take the form of X_{nm} .

$$\mu 3 + \mu 4 + 100 \rightarrow \mu + \mu\mu + \omega\omega \rightarrow \mu\mu\mu\omega\omega 107$$

5.1 Minor punctuation. Psalter Pahlavi punctuation appears to be related to Syriac punctuation, though the usage of the latter is probably more exactly established. Relevant related characters in Syriac are U+0701 SYRIAC SUPRALINEAR FULL STOP, U+0702 SYRIAC SUBLINEAR FULL STOP, U+0703 SYRIAC SUBLINEAR COLON, and U+0708 SYRIAC SUPRALINEAR COLON SKEWED LEFT. Psalter Pahlavi uses four characters just like these, as well as three combining characters which are used alone with letters alone, and with letters followed by one of the spacing punctuation characters. The following configurations have been observed (using the letter YODH as the base character:

- ◌Რ YODH + PSALTER PAHLAVI HIGH DOT
- ◌Ს YODH + PSALTER PAHLAVI LOW DOT
- ◌Ტ YODH + PSALTER PAHLAVI COLON
- ◌Უ YODH + PSALTER PAHLAVI COLON SKEWED RIGHT
- ◌Ფ YODH + PSALTER PAHLAVI DOT ABOVE
- ◌Ქ YODH + PSALTER PAHLAVI DOT BELOW
- ◌Ღ YODH + PSALTER PAHLAVI DOUBLE DOT ABOVE
- ◌Ყ YODH + PSALTER PAHLAVI DOT ABOVE + PSALTER PAHLAVI HIGH DOT
- ◌Შ YODH + PSALTER PAHLAVI DOT ABOVE + PSALTER PAHLAVI LOW DOT
- ◌Ჩ YODH + PSALTER PAHLAVI DOT BELOW + PSALTER PAHLAVI LOW DOT
- ◌Ც YODH + PSALTER PAHLAVI DOUBLE DOT ABOVE + PSALTER PAHLAVI LOW DOT
- ◌Ძ YODH + PSALTER PAHLAVI DOT ABOVE + PSALTER PAHLAVI HIGH DOT + PSALTER PAHLAVI LOW DOT

5.2 Major punctuation. Four types of large section-ending punctuation have been identified. The most common is PSALTER PAHLAVI SECTION MARK Წ which is written with red dots in the vertical position and black dots in the horizontal position; the red dots are often written as rings, which is reflected in the reference glyph here. Less common but found together with this is the PSALTER PAHLAVI TURNED SECTION MARK Ჭ which is written with black dots in the vertical position and red dots in the horizontal position. More rare is the PSALTER PAHLAVI FOUR DOT PUNCTUATION WITH CROSS Ხ (which is sometimes found immediately following the SECTION MARK), and the PSALTER PAHLAVI FOUR DOT PUNCTUATION WITH DOT Ჯ (which is slightly reminiscent of the MANICHAEAN PUNCTUATION FLEURON Ჰ).

5.3. Issues. Some of these configurations look very similar indeed, and may be related not to structural differences in a punctuation scheme, but rather to penmanship. This will be discussed with experts in the script so that the scheme can be better understood. Attempts were made to find ways to use Syriac characters or generic characters to build up this set, but no configuration seemed to be suitable, and in any case at least the PSALTER PAHLAVI COLON SKEWED RIGHT has no analogue elsewhere in the UCS. In particular it would be difficult to use ASCII FULL STOP and COLON because this would effectively prevent a simple monofont containing Psalter Pahlavi characters alongside Latin characters used for transliteration and discussion text. Since Psalter Pahlavi punctuation is a coherent set of characters, historically related to but not identical to what we have in Syriac, and historically unrelated to Western punctuation, it is proposed to encode them as a set of script-specific characters.

5.4. Psalter Pahlavi makes use of a *kashida* to extend a word. The character U+ 0640 ARABIC TATWEEL is proposed to be used for this function. Mandaic and Manichaean also have a similar requirement. The data file ScriptExtensions.txt would need to be changed to say:

```
0640 ; Arab Mand Mani Phlp Syrc # Lm ARABIC TATWEEL
```

6. **Line-Breaking.** The letters and digits behave like letters, and will have the line breaking class AL (Alphabetic). The stops will have the line breaking class CM (Combining Mark). It is probably safe to assume that lines could be broken both before and after the section mark and the dotted cross, but not between consecutive ones, so they should have the line breaking class B2 (Break Opportunity Before and After). The other punctuation marks happen at the end of words and should perhaps forbid a break before them. The best line breaking class for them should be EX (Exclamation/Interrogation).

7. Unicode Character Properties

```
10B80;PSALTER PAHLAVI LETTER ALEPH;Lo;0;R;;;N;;;
10B81;PSALTER PAHLAVI LETTER BETH;Lo;0;R;;;N;;;
10B82;PSALTER PAHLAVI LETTER GIMEL;Lo;0;R;;;N;;;
10B83;PSALTER PAHLAVI LETTER DALETH;Lo;0;R;;;N;;;
10B84;PSALTER PAHLAVI LETTER HE;Lo;0;R;;;N;;;
10B85;PSALTER PAHLAVI LETTER WAW-AYIN-RESH;Lo;0;R;;;N;;;
10B86;PSALTER PAHLAVI LETTER ZAYIN;Lo;0;R;;;N;;;
10B87;PSALTER PAHLAVI LETTER HETH;Lo;0;R;;;N;;;
10B88;PSALTER PAHLAVI LETTER YODH;Lo;0;R;;;N;;;
10B89;PSALTER PAHLAVI LETTER KAPH;Lo;0;R;;;N;;;
10B8A;PSALTER PAHLAVI LETTER LAMEDH;Lo;0;R;;;N;;;
10B8B;PSALTER PAHLAVI LETTER MEM-QOPH;Lo;0;R;;;N;;;
10B8C;PSALTER PAHLAVI LETTER NUN;Lo;0;R;;;N;;;
10B8D;PSALTER PAHLAVI LETTER SAMEKH;Lo;0;R;;;N;;;
10B8E;PSALTER PAHLAVI LETTER PE;Lo;0;R;;;N;;;
10B8F;PSALTER PAHLAVI LETTER SADHE;Lo;0;R;;;N;;;
10B90;PSALTER PAHLAVI LETTER SHIN;Lo;0;R;;;N;;;
10B91;PSALTER PAHLAVI LETTER TAW;Lo;0;R;;;N;;;
10B92;PSALTER PAHLAVI HIGH DOT;Po;0;R;;;N;;;
10B93;PSALTER PAHLAVI LOW DOT;Po;0;R;;;N;;;
10B94;PSALTER PAHLAVI COLON;Po;0;R;;;N;;;
10B95;PSALTER PAHLAVI COLON SKEWED RIGHT;Po;0;R;;;N;;;
10B96;PSALTER PAHLAVI DOT ABOVE;Mn;230;NSM;;;N;;;
10B97;PSALTER PAHLAVI DOT BELOW;Mn;220;NSM;;;N;;;
10B98;PSALTER PAHLAVI DOUBLE DOTS ABOVE;Mn;230;NSM;;;N;;;
10B99;PSALTER PAHLAVI SECTION MARK;Po;0;R;;;N;;;
10B9A;PSALTER PAHLAVI TURNED SECTION MARK;Po;0;R;;;N;;;
10B9B;PSALTER PAHLAVI FOUOD DOT PUNCTUATION WITH CROSS;Po;0;R;;;N;;;
10B9C;PSALTER PAHLAVI FOUOD DOT PUNCTUATION WITH DOT;Po;0;R;;;N;;;
10BA9;PSALTER PAHLAVI NUMBER ONE;No;0;R;;;1;N;;;
10BAA;PSALTER PAHLAVI NUMBER TWO;No;0;R;;;2;N;;;
10BAB;PSALTER PAHLAVI NUMBER THREE;No;0;R;;;3;N;;;
10BAC;PSALTER PAHLAVI NUMBER FOUR;No;0;R;;;4;N;;;
10BAD;PSALTER PAHLAVI NUMBER TEN;No;0;R;;;10;N;;;
10BAE;PSALTER PAHLAVI NUMBER TWENTY;No;0;R;;;20;N;;;
10BAF;PSALTER PAHLAVI NUMBER ONE HUNDRED;No;0;R;;;100;N;;;;
```

8. Additions for the file ArabicShaping.txt

```
# Psalter Pahlavi
10B80; PSALTER PAHLAVI ALEPH; D; No_Joining_Group
10B81; PSALTER PAHLAVI BETH; R; No_Joining_Group
10B82; PSALTER PAHLAVI GIMEL; D; No_Joining_Group
10B83; PSALTER PAHLAVI DALETH; R; No_Joining_Group
10B84; PSALTER PAHLAVI HE; R; No_Joining_Group
10B85; PSALTER PAHLAVI WAW-AYIN-RESH; R; No_Joining_Group
10B86; PSALTER PAHLAVI ZAYIN; D; No_Joining_Group
10B87; PSALTER PAHLAVI HETH; D; No_Joining_Group
10B88; PSALTER PAHLAVI YODH; D; No_Joining_Group
10B89; PSALTER PAHLAVI KAPH; R; No_Joining_Group
10B8A; PSALTER PAHLAVI LAMEDH; D; No_Joining_Group
10B8B; PSALTER PAHLAVI MEM-QOPH; D; No_Joining_Group
10B8C; PSALTER PAHLAVI NUN; R; No_Joining_Group
10B8D; PSALTER PAHLAVI SAMEKH; D; No_Joining_Group
```








































10B8E; PSALTER PAHLAVI PE; R; No_Joining_Group
10B8F; PSALTER PAHLAVI SADHE; R; No_Joining_Group
10B90; PSALTER PAHLAVI SHIN; D; No_Joining_Group
10B91; PSALTER PAHLAVI TAW; R; No_Joining_Group
10BA9; PSALTER PAHLAVI ONE; R; No_Joining_Group
10BAA; PSALTER PAHLAVI TWO; R; No_Joining_Group
10BAB; PSALTER PAHLAVI THREE; R; No_Joining_Group
10BAC; PSALTER PAHLAVI FOUR; R; No_Joining_Group
10BAD; PSALTER PAHLAVI TEN; D; No_Joining_Group
10BAE; PSALTER PAHLAVI TWENTY; D; No_Joining_Group
10BAF; PSALTER PAHLAVI HUNDRED; U; No_Joining_Group

9. Bibliography.



















- Andreas, F. C. 1933. *Bruchstücke einer Pehlevi-Übersetzung der Psalmen*. Aus dem Nachlass herausgegeben von Kaj Barr. Sonderausgabe aus den Sitzungsberichten der Preussischen Akademie der Wissenschaften Phil.-Hist. Klasse, 1933.I. Berlin: Verlag der Akademie der Wissenschaften in Kommission bei Walter de Gruyter u. Co.
- Faulmann, Carl. 1990 (1880). *Das Buch der Schrift*. Frankfurt am Main: Eichborn. ISBN 3-8218-1720-8
- Gignoux, Philippe. 2001. “Une croix de procession de Hérat inscrite en pehlevi”, in *Le Muséon* 114, fasc. 3-4, pp. 291-304.
- MacKenzie, D. N. 1971. *A concise Pahlavi dictionary*. London: Oxford University Press.
- Nyberg, Henrik Samuel. 1964 *A manual of Pahlavi*. Wiedbaden: Otto Harrassowitz. Reprinted 2003 Tehran: Asatir. ISBN 964-331-131-7, 964-331-132-5
- Oryan, Saeed. 2003b (1382 AP). *Zabur-e Pahlavi: matn-e Pahlavi, harf-nevisi, āvā-nevisi, tarojme-ye Fārsi va yaddāšt-hā = Pahlavi translation of the Psalms: text, transliteration, transcription, Persian translation, and notes*. Tehran: Iranian Cultural Heritage Organization. ISBN 964-7483-75-9
- Skjærvø, P. Oktor. 1996. “Aramaic scripts for Iranian languages” in *The World’s Writing Systems*, ed. Peter T. Daniels & William Bright. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0
- Taylor, Isaac. 1883. *The alphabet: an account of the origin and development of letters*. Vol. 1: Semitic alphabets; Vol. 2: Aryan alphabets. London: Kegan Paul.

10. Acknowledgements












This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley), as well as by support from the FarsiWeb Project, Sharif FarsiWeb, Tehran. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

	10B8	10B9	10BA
0	 10B80	 10B90	
1	 10B81	 10B91	
2	 10B82	 10B92	
3	 10B83	 10B93	
4	 10B84	 10B94	
5	 10B85	 10B95	
6	 10B86	 10B96	
7	 10B87	 10B97	
8	 10B88	 10B98	
9	 10B89	 10B99	
A	 10B8A	 10B9A	 10BAA
B	 10B8B	 10B9B	 10BAB
C	 10B8C	 10B9C	 10BAC
D	 10B8D		 10BAD
E	 10B8E		 10BAE
F	 10B8F		 10BAF








Letters

- 10B80  PSALTER PAHLAVI LETTER ALEPH
- 10B81  PSALTER PAHLAVI LETTER BETH
- 10B82  PSALTER PAHLAVI LETTER GIMEL
- 10B83  PSALTER PAHLAVI LETTER DALETH
- 10B84  PSALTER PAHLAVI LETTER HE
- 10B85  PSALTER PAHLAVI LETTER WAW-AYIN-RESH
- 10B86  PSALTER PAHLAVI LETTER ZAYIN
- 10B87  PSALTER PAHLAVI LETTER HETH
- 10B88  PSALTER PAHLAVI LETTER YODH
- 10B89  PSALTER PAHLAVI LETTER KAPH
- 10B8A  PSALTER PAHLAVI LETTER LAMEDH
- 10B8B  PSALTER PAHLAVI LETTER MEM-QOPH
- 10B8C  PSALTER PAHLAVI LETTER NUN
- 10B8D  PSALTER PAHLAVI LETTER SAMEKH
- 10B8E  PSALTER PAHLAVI LETTER PE
- 10B8F  PSALTER PAHLAVI LETTER SADHE
- 10B90  PSALTER PAHLAVI LETTER SHIN
- 10B91  PSALTER PAHLAVI LETTER TAW

Punctuation

- 10B92  PSALTER PAHLAVI HIGH DOT
- 10B93  PSALTER PAHLAVI LOW DOT
- 10B94  PSALTER PAHLAVI COLON
- 10B95  PSALTER PAHLAVI COLON SKEWED RIGHT
- 10B96  PSALTER PAHLAVI DOT ABOVE
- 10B97  PSALTER PAHLAVI DOT BELOW
- 10B98  PSALTER PAHLAVI DOUBLE DOTS ABOVE
- 10B99  PSALTER PAHLAVI SECTION MARK
- 10B9A  PSALTER PAHLAVI TURNED SECTION MARK
- 10B9B  PSALTER PAHLAVI FOUR DOTS WITH CROSS
- 10B9C  PSALTER PAHLAVI FOUR DOTS WITH DOT

Numbers

- 10BA9  PSALTER PAHLAVI NUMBER ONE
- 10BAA  PSALTER PAHLAVI NUMBER TWO
- 10BAB  PSALTER PAHLAVI NUMBER THREE
- 10BAC  PSALTER PAHLAVI NUMBER FOUR
- 10BAD  PSALTER PAHLAVI NUMBER TEN
- 10BAE  PSALTER PAHLAVI NUMBER TWENTY
- 10BAF  PSALTER PAHLAVI NUMBER ONE HUNDRED