

# Revised proposal to encode Book Pahlavi in Unicode

Anshuman Pandey

pandey@umich.edu  
pandey.github.io/unicode

February 22, 2024

## 1 Introduction

This document supersedes the following proposals:

- L2/18-276: “Preliminary proposal to encode Book Pahlavi in Unicode” (Pandey)
- L2/21-090: “Advancing the encoding model for Book Pahlavi letters” (Pandey)

It incorporates information presented in the following documents:

- L2/02-449: “Revised proposal to encode the Avestan and Pahlavi script in the UCS” (Everson)
- L2/13-141: “Preliminary proposal to encode the Book Pahlavi script in [TUS]” (Pournader)
- L2/14-077R: “Proposal for Encoding Book Pahlavi in the Unicode Standard” (Meyers)
- L2/18-334: “A Critique of L2/18-276” (Meyers)
- L2/20-135: “Next Steps on Book Pahlavi” (Pournader & Hai)
- L2/20-141: “Response to ‘Next Steps on Book Pahlavi’ (L2/20-135)” (Zeini)
- L2/20-246: “Teeth and bellies: a proposed model for encoding Book Pahlavi” (Pournader)

The Unicode Technical Committee (UTC) maintains the ‘Topical Document List: Book Pahlavi’, which contains all documents related to this script: <https://unicode.org/L2/topical/bookpahlavi/>

This revised proposal presents a new encoding model and character repertoire for Book Pahlavi that enables the representation of text as it appears in the majority of manuscript and printed sources. A code chart follows p. 24 and character data is provided in § 8.

All of the Book Pahlavi text in this document has been composed using the proposed character repertoire, which supports the innate cursive joining nature of the script.

Specimens of Book Pahlavi text from the sources are forthcoming.

## 2 The Script

The attested repertoire for Book Pahlavi contains 25 letters, with Aramaic names as per scholarly convention:

𐭠	𐭡	𐭢	𐭣	𐭤	𐭥	𐭦	𐭧	𐭨	𐭩	𐭪	𐭫	𐭬
<i>aleph, heth</i>	<i>beth</i>	<i>gimel, daleth, yodh</i>	<i>curled gimel, daleth, yodh</i>	<i>old daleth</i>	<i>he</i>	<i>waw, ayin, nun, resh</i>	<i>zayin</i>	<i>kaph</i>	<i>old kaph</i>	<i>lamedh</i>	<i>stroked lamedh</i>	<i>looped lamedh</i>
𐭭	𐭮	𐭯	𐭰	𐭱	𐭲	𐭳	𐭴	𐭵	𐭶	𐭷	𐭸	𐭹
<i>hooked lamedh</i>	<i>old lamedh</i>	<i>mem- qoph</i>	<i>hooked mem- qoph</i>	<i>old nun</i>	<i>samekh</i>	<i>Indian samekh</i>	<i>pe</i>	<i>sadhe</i>	<i>shin</i>	<i>curled shin</i>		<i>taw</i>

Some of these letters have alternate forms:

𐭴	𐭵
<i>upright sadhe</i>	<i>final sadhe</i>

There are 3 atomic ligatures:

𐭶𐭵	𐭶𐭴	𐭶𐭵
<i>X<sub>1</sub></i>	<i>X<sub>2</sub></i>	<i>yodh- heth</i>

The following 7 diacritic signs are used for expressing specific phonetic values of letters (see § 7.2):

◌̇	◌̈	◌̈̈	◌̈̈̈	◌̈̈̈̈	◌̈̈̈̈̈	◌̂
<i>dot above</i>	<i>dot below</i>	<i>two dots above</i>	<i>two dots below</i>	<i>three dots above</i>	<i>three dots below</i>	<i>hat above</i>

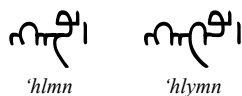
Various signs are used for punctuation:

𐭠	•	◦	⋄	⊕	⊖	⊗
<i>'otiose' sign</i>	<i>middle dot</i>	<i>middle circle</i>	<i>triangular three dots</i>	<i>triangular three circles</i>	<i>diamond four dots</i>	<i>diamond four circles</i>

Primary numbers are written using the following forms, but there is a full orthography for writing tens, compounds of tens and primary units, hundreds, and thousands (see § 7.3).

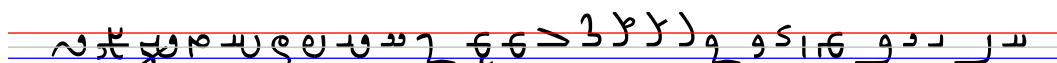
𐭠	𐭡	𐭢	𐭣	𐭤
1	2	3	4	5
𐭥	𐭦	𐭧	𐭨	𐭩
6	7	8	9	

Lastly, there is a convention for writing the name of Ahura Mazda's antagonist upside down:



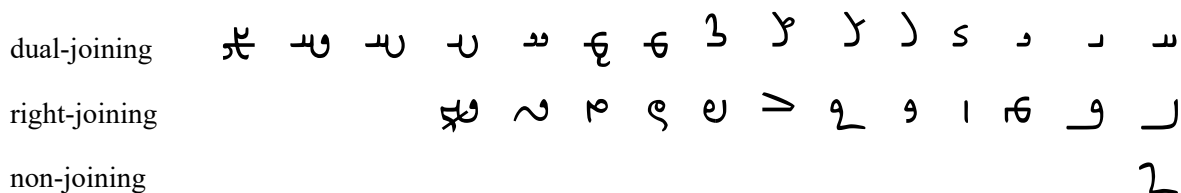
## 2.1 Directionality

Book Pahlavi is written from right to left, with lines that advance from top to bottom. Letters are written along a baseline, which is not readily apparent, but may be identified as the bottom of 𐭠, 𐭡, 𐭢, 𐭣, 𐭤, 𐭥; the resting spot for the heads of 𐭦, 𐭧, 𐭨, 𐭩, 𐭪; the cross-bar of 𐭫 and 𐭬. The alignment of all letters with the baseline (gray) is shown below. The head-height is best measured by the tops of 𐭠, 𐭡, 𐭢, etc., while the below-base is determined by the lowest points of 𐭦, 𐭧, 𐭨, etc. The swash terminals of 𐭭, 𐭮 run under the base, and in the case of 𐭯, the terminal extends beneath all letters that follow.

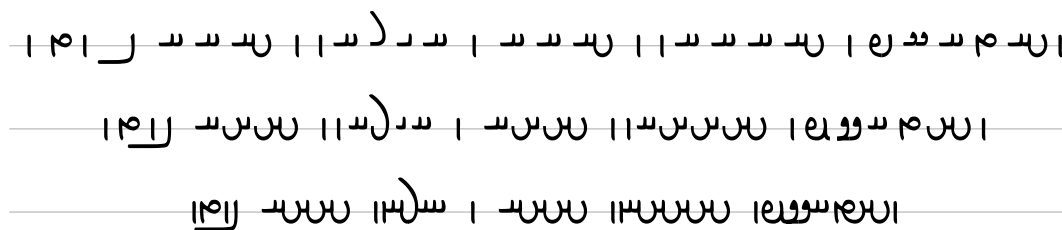


## 2.2 Joining behavior

Book Pahlavi is a cursive *abjad* script whose letters are dual-joining or right-joining:



As part of the normal cursive joining behavior of the script, a letter may be represented using a contextual form depending upon its position or by adjacent letters, or both. In the example below, the first line shows a string of independent letters, the second line shows the contextual forms of those letters when joining, and the third line shows the rendered forms of the contextual forms:



<wšt' sp' š'h'n' š'h w yln' š'h bwt'>

*wištāsp šāhān šāh ud ērān šāh būd*

Wištāsp was the king of kings and the king of the Iranians

The following two tables show the contextual forms of letters (table 1) and a classification of letters based on the complexity of their joining behavior (table 2). The lines highlighted red show the connection points for the positional forms of each letter.

		X <sub>n</sub>	X <sub>f</sub>	X <sub>m</sub>	X <sub>i</sub>
No change in shape	<i>beth</i>	𐭡	𐭡	—	—
	old <i>daleth</i>	𐭢	𐭢	—	—
	<i>kaph</i>	𐭣	𐭣, 𐭣	—	—
	old <i>kaph</i>	𐭤	𐭤	—	—
	hooked <i>lamedh</i>	𐭥	𐭥	𐭥	𐭥
	old <i>lamedh</i>	𐭦	𐭦	—	—
	old <i>nun</i>	𐭧	—	—	—
	Indian <i>samekh</i>	𐭨	𐭨	𐭨	𐭨
	<i>taw</i>	𐭩	𐭩	—	—
	<i>X</i> <sub>1</sub> ligature	𐭪	𐭪	—	—
	<i>X</i> <sub>2</sub> ligature	𐭫	𐭫	—	—
	<i>yodh-heth</i> ligature	𐭬	𐭬	—	—
Height adjustments for below-base or baseline connections	<i>he</i>	𐭭	𐭭, 𐭭	—	—
	<i>waw-nun-ayin-res</i>	𐭮	𐭮, 𐭮, 𐭮	—	—
	<i>zayin</i>	𐭯	𐭯, 𐭯, 𐭯	𐭯, 𐭯, 𐭯	𐭯, 𐭯, 𐭯
	<i>lamedh</i>	𐭰	𐭰, 𐭰	𐭰, 𐭰, 𐭰, 𐭰	𐭰, 𐭰, 𐭰
	stroked <i>lamedh</i>	𐭱	𐭱, 𐭱	𐭱, 𐭱, 𐭱, 𐭱	𐭱, 𐭱, 𐭱
	looped <i>lamedh</i>	𐭲	𐭲, 𐭲	𐭲, 𐭲, 𐭲, 𐭲	𐭲, 𐭲, 𐭲
	<i>mem-qoph</i>	𐭳	𐭳	𐭳, 𐭳, 𐭳	𐭳, 𐭳, 𐭳
	<i>samekh</i>	𐭴	𐭴, 𐭴	𐭴, 𐭴, 𐭴	𐭴, 𐭴
Stroke truncation or, no change	<i>pe</i>	𐭵	𐭵, 𐭵, 𐭵	—	—
	<i>sadhe</i>	𐭶, 𐭶	𐭶, 𐭶	—	—
Descent or curvature of terminal	<i>aleph-heth</i>	𐭷	𐭷	𐭷, 𐭷, 𐭷	𐭷, 𐭷, 𐭷
	<i>gimel-daleth-yodh</i>	𐭸	𐭸	𐭸, 𐭸	𐭸, 𐭸
	curled <i>gimel-daleth-yodh</i>	𐭹	𐭹	𐭹, 𐭹, 𐭹	𐭹, 𐭹, 𐭹
	<i>shin</i>	𐭺	𐭺	𐭺, 𐭺, 𐭺	𐭺, 𐭺, 𐭺
	curled <i>shin</i>	𐭻	𐭻	𐭻, 𐭻, 𐭻	𐭻, 𐭻, 𐭻

Table 1: Contextual forms of Book Pahlavi letters

		X <sub>n</sub>	X <sub>f</sub>	X <sub>m</sub>	X <sub>i</sub>
Simple	<i>beth</i>	𐭪	𐭪̌	—	—
	<i>old daleth</i>	𐭫	𐭫̌	—	—
	<i>kaph</i>	𐭬	𐭬̌, 𐭬̎	—	—
	<i>old kaph</i>	𐭭	𐭭̌	—	—
	<i>hooked lamedh</i>	𐭮	𐭮̌	𐭮̎	𐭮̎
	<i>old lamedh</i>	𐭯	𐭯̌	—	—
	<i>old nun</i>	𐭰	—	—	—
	<i>taw</i>	𐭱	𐭱̌	—	—
	<i>X<sub>1</sub></i>	𐭲	𐭲̌	—	—
	<i>X<sub>2</sub></i>	𐭳	𐭳̌	—	—
	<i>yodh-heth ligature</i>	𐭴	𐭴̌	—	—
Intermediate	<i>waw-nun-ayin-resh</i>	𐭵	𐭵̌, 𐭵̎, 𐭵̏	—	—
	<i>zayin</i>	𐭶	𐭶̌, 𐭶̎, 𐭶̏	𐭶̎, 𐭶̏, 𐭶̌, 𐭶̎	𐭶̎, 𐭶̏, 𐭶̌
	<i>lamedh</i>	𐭷	𐭷̌, 𐭷̎	𐭷̎, 𐭷̏, 𐭷̌, 𐭷̎	𐭷̎, 𐭷̏, 𐭷̌
	<i>stroked lamedh</i>	𐭸	𐭸̌, 𐭸̎	𐭸̎, 𐭸̏, 𐭸̌, 𐭸̎	𐭸̎, 𐭸̏, 𐭸̌
	<i>looped lamedh</i>	𐭹	𐭹̌, 𐭹̎	𐭹̎, 𐭹̏, 𐭹̌, 𐭹̎	𐭹̎, 𐭹̏, 𐭹̌
	<i>mem-qoph</i>	𐭺	𐭺̌	𐭺̎, 𐭺̏, 𐭺̌	𐭺̎, 𐭺̏, 𐭺̌
Complex	<i>aleph-heth</i>	𐭻	𐭻̌	𐭻̎, 𐭻̏, 𐭻̌	𐭻̎, 𐭻̏, 𐭻̌
	<i>gimel-daleth-yodh</i>	𐭼	𐭼̌	𐭼̎, 𐭼̏	𐭼̎, 𐭼̏
	<i>curled gimel-daleth-yodh</i>	𐭽	𐭽̌	𐭽̎, 𐭽̏, 𐭽̌	𐭽̎, 𐭽̏, 𐭽̌
	<i>he</i>	𐭾	𐭾̌, 𐭾̎	—	—
	<i>samekh</i>	𐭿	𐭿̌, 𐭿̎	𐭿̎, 𐭿̏, 𐭿̌	𐭿̎, 𐭿̏
	<i>Indian samekh</i>	𐮀	𐮀̌	𐮀̎	𐮀̎
	<i>pe</i>	𐮁	𐮁̌, 𐮁̎, 𐮁̏	—	—
	<i>sadhe</i>	𐮂, 𐮃	𐮂̌, 𐮂̎	—	—
	<i>shin</i>	𐮄	𐮄̌	𐮄̎, 𐮄̏, 𐮄̌	𐮄̎, 𐮄̏, 𐮄̌
	<i>curled shin</i>	𐮅	𐮅̌	𐮅̎, 𐮅̏, 𐮅̌	𐮅̎, 𐮅̏, 𐮅̌

Table 2: Categorization of Book Pahlavi letters by joining complexity

### 3 Complexities of the script

Book Pahlavi letters are fairly identifiable when written independently. However, it is a cursive joining script and certain letters take different shapes when they occur in initial, medial, or final positions in a word. The joining forms for some letters tend to obscure their identity, especially when they resemble other letters when written in sequence. As shown in table 2, Book Pahlavi letters can be classified as ‘simple’, ‘intermediate’, or ‘complex’ based on their shapes and joining behaviors:

- Simple: clearly recognizable and have no shape changes when joining
- Intermediate: clearly recognizable, but have height and other adjustments when joining
- Complex: identity may be ambiguous or obscured due to shape changes when joining

#### 3.1 Behavior of ‘intermediate’ letters

Some intermediate letters and their notable features include:

1. The *waw-nun-ayin-resh* generally occurs as its regular form *l*, but in certain contexts it takes a height-reduced *l* or curved *l* form:
  - Typically after *mem-qoph*, the *waw-nun-ayin-resh* is shortened as *l* to join with the left terminal of *mem-qoph*; notice the height difference of the two instances of *waw-nun-ayin-resh* in *murv* ‘bird’.
  - Numerous manuscripts exhibit a scribal preference for writing *waw-nun-ayin-resh* using the curved form *l* after *aleph-heth* and *daleth-gimel-yodh*, which is contrasted with the regular form, often within the same word, eg. *andarway* ‘interspace’; which also occurs as *andarway* without the stylized *waw-nun-ayin-resh*.
2. The *zayin* is written using *z* regular and *z* descending (shaped as *z* to connect to the baseline):
  - Before *daleth-gimel-yodh*: The regular form *z* is used in *pazdok*. The descending form *z* occurs in *ziyān*, where its terminal connected to *daleth-gimel-yodh* from below the baseline. The forms of *zayin* in these contexts is unpredictable.
  - Before *kaph*: A common Arameogram particle <ZK> is written as *z* <zk> *ān* using the descending form of *zayin*.
3. The *z* descending *zayin* is commonly used when writing the hundreds, eg. 100 *z*, 200 *z*, 300 *z*. The *z* regular form is generally not used.
4. The regular, stroked, looped forms of *lamedh* have regular *l*, *l*, *l*; descending *l*, *l*, *l*; and bellied forms *l*, *l*, *l* to enable connections with various letters.

5. The **𐭄** *mem-qoph* perhaps displays the widest variation when rendered in different positions. In order to enable connections with adjacent letters, its initial stroke may be vertically adjusted, or its body may be lowered or reduced in size, or its final stroke may be curved or raised or lowered.

- When initial, it connects to following letters with minimal change to its shape:
  - the baseline (𐭄𐭀 *mem-qoph* + *aleph-heth*)
  - with a descending stroke (𐭄𐭆 *mem-qoph* + *pe* / *sadhe*)
  - with a belly (𐭄𐭌 *mem-qoph* + *taw*) depending on shape of the latter
  - or variously, depending on shape of the latter, eg. *aleph-heth*: 𐭄𐭀, 𐭄𐭁
- When medial, its shape is modified to enable connections with preceding and following letters, and the shapes vary by scribal preferences:
  - For preceding letters:
    - \* the preceding letter is raised to meet the initial stroke: 𐭄𐭀𐭀 *aleph-heth* + *mem-qoph* + *aleph-heth*
    - \* the body of *mem-qoph* is lowered (𐭄 ← 𐭄), such that the origin of the head stroke connects more naturally at the baseline, eg. 𐭄𐭀𐭄 *aleph-heth* + *mem-qoph*; this lowering also facilitates connections with letters with ascenders, eg. 𐭄𐭌 *lamedh* + *mem-qoph*
    - \* in general, preceding letters do not connect to the body of *mem-qoph*, eg. \*𐭄𐭀
  - For following letters, the terminal stroke may be shaped variously:
    - \* descending to join to letters below the baseline, eg. with *pe* / *sadhe*: 𐭄𐭆 *aleph-heth* + *mem-qoph* + *pe* / *sadhe*
    - \* bellied to join to letters resting at the baseline, eg. with *taw*: 𐭄𐭌 *aleph-heth* + *mem-qoph* + *taw*; also with *aleph-heth*: 𐭄𐭌𐭀 *aleph-heth* + *mem-qoph* + *aleph-heth*
    - \* straight to join letters with vertical descenders, eg. with descending *gimel-daleth-yodh*: 𐭄𐭌𐭌 *aleph-heth* + *mem-qoph* + *aleph-heth*

- There are various stylized forms of lowered medial *mem-qoph*:

straight terminal stroke	𐭄𐭌	𐭄𐭌	𐭄𐭌	𐭄𐭌
descending terminal stroke	𐭄𐭌	𐭄𐭌	𐭄𐭌	𐭄𐭌
bellied terminal stroke	𐭄𐭌	𐭄𐭌	𐭄𐭌	𐭄𐭌
extended bellied terminal stroke	𐭄𐭌	𐭄𐭌	𐭄𐭌	𐭄𐭌

- When final, *mem-qoph* is written
  - lowered when preceded by a joining letter: 𐭄𐭀 *aleph-heth* + *mem-qoph*
  - at regular height, with the preceding letter raised: 𐭄𐭀 *aleph-heth* + *mem-qoph*
  - at regular height, when unconnected: 𐭄𐭀 *pe* + *mem-qoph*

- A sequence of *mem-qoph* + *mem-qoph* is generally represented as 𐭪𐭪, as in 𐭪𐭪 *mem-qoph* + *mem-qoph* + *waw-ayin-nun-resh*; but similar to the various stylizations of medial *mem-qoph*, scribes write the sequence in creative ways. A longer sequence of *mem-qoph*, could theoretically be written with a baseline that slopes southwest, eg. 𐭪𐭪𐭪.

### 3.2 Behavior of ‘complex’ letters

Complex letters exhibit several features that lead to potential ambiguous representations and interpretations, if the reader is unfamiliar with the orthography and morphology. Several cases are illustrated below:

1. Their independent forms may resemble sequences of other letters:

𐭪	<i>aleph-heth</i>	𐭪 + 𐭪	<i>daleth-gimel-yodh</i> + <i>daleth-gimel-yodh</i>
𐭪𐭪	<i>he</i>	𐭪 + 𐭪𐭪	<i>mem-qoph</i> + height-adjusted <i>waw-nun-ayin-resh</i>
𐭪𐭪	<i>samekh</i>	𐭪 + 𐭪	curled <i>daleth-gimel-yodh</i> + curled <i>daleth-gimel-yodh</i>
𐭪𐭪	Indian <i>samekh</i>	𐭪 + 𐭪𐭪	hooked curled <i>daleth-gimel-yodh</i> + <i>daleth-gimel-yodh</i>
𐭪𐭪	curled <i>shin</i>	𐭪 + 𐭪𐭪	hooked curled <i>daleth-gimel-yodh</i> + <i>aleph-heth</i>

2. Some letters have distinctive independent forms, but share the same final form. This is true for 𐭪 *pe* and 𐭪 *sadhe*, which may both be written as 𐭪 when final, in addition to their regular shapes.
3. They have contextual forms that obscure their identities, especially when adjacent to other complex letters. Some example words containing adjacent complex letters:

𐭪𐭪 <’y> *ay* ‘O!’; could be read as:  
 <𐭪 *aleph-heth*, 𐭪 *daleth-gimel-yodh*>  
 <𐭪 *daleth-gimel-yodh*, 𐭪 *aleph-heth*>

𐭪𐭪𐭪 <’sm’n> *āsmān* ‘sky’; the sequence 𐭪𐭪 could be read as  
 <𐭪 *aleph-heth*, 𐭪 alternate *samekh*>  
 <𐭪 *aleph-heth*, 𐭪 curled *daleth-gimel-yodh*, 𐭪 *daleth-gimel-yodh*>

𐭪𐭪𐭪𐭪 <’pykyh> *abēzagīh* ‘purity’; the sequence 𐭪𐭪 could be read as  
 <𐭪 *aleph-heth*, 𐭪 *pe*, 𐭪 curled *daleth-gimel-yodh*, 𐭪 *pe*>  
 <𐭪 *aleph-heth*, 𐭪 *sadhe*, 𐭪 curled *daleth-gimel-yodh*, 𐭪 *sadhe*>  
 <𐭪 *aleph-heth*, 𐭪 *pe*, 𐭪 curled *daleth-gimel-yodh*, 𐭪 *sadhe*>  
 <𐭪 *aleph-heth*, 𐭪 *sadhe*, 𐭪 curled *daleth-gimel-yodh*, 𐭪 *pe*>

𐭪𐭪𐭪𐭪 <drwyst> *drust* ‘healthy’; the sequence 𐭪𐭪 could be interpreted as  
 <𐭪 curled *daleth-gimel-yodh*, 𐭪 *samekh*>  
 <𐭪 *samekh*, 𐭪 curled *daleth-gimel-yodh*>  
 <𐭪 curled *daleth-gimel-yodh*, 𐭪 curled *daleth-gimel-yodh*, 𐭪 curled *daleth-gimel-yodh*>

𐭪𐭪𐭪𐭪 <g’h’n> *gāhān* ‘the Gathas’; the sequence 𐭪𐭪 could be read as  
 <𐭪 alternate *samekh*, 𐭪 *shin*, 𐭪 *aleph-heth*>



<⌞ curled *daleth-gimel-yodh*, ⌞ *daleth-gimel-yodh*, ⌞ shin, ⌞ *aleph-heth*>  
 <⌞ curled *daleth-gimel-yodh*, ⌞ *aleph-heth*, ⌞ *daleth-gimel-yodh*, ⌞ shin>

4. While there are typical joining patterns for complex letters, there are several irregular patterns:

- $\aleph$  *aleph-heth* +  $\gamma$  *daleth-gimel-yodh* typically takes the form  $\aleph\gamma$ , eg.  $\aleph\gamma$  <'dwk> *ēk* 'one'. But, in several words the letters connect without contextual forms, eg.  $\aleph\gamma$ , as in  $\aleph\gamma$  <'YK> *kū* 'where?, thus'. When this sequence is followed by *kaph*, the form  $\aleph\gamma$  may change to  $\aleph\gamma$ , where the *daleth-gimel-yodh* descends before connecting, as in  $\aleph\gamma\aleph$  <p'dk> *pādak* (see below for more information on the shaping of *daleth-gimel-yodh* before *kaph*).
- *gimel-daleth-yodh* + *aleph-heth*  $\aleph\gamma$  may be represented as a ligature  $\aleph\gamma$  at the end of words.
- $\gamma$  *daleth-gimel-yodh* +  $\gamma$  *daleth-gimel-yodh* is used for representing *zd* in preserved spellings. In such cases, the sequence is written as  $\gamma\gamma$ , which resembles *aleph-heth*. One important word is  $\gamma\gamma\gamma\gamma$  <'whrmzd> *ohurmazd* 'Ahura Mazda', where final *zd* is written  $\gamma\gamma$ , which is the independent or final form of *aleph-heth*. The sequence  $\gamma\gamma$  is represented as  $\gamma$  using the  $\gamma$  descending form of *daleth-gimel-yodh*, when connecting before letters at the baseline, eg.  $\gamma\gamma\gamma$  <zdy'n> *ziyān*.
- $\gamma$  *daleth-gimel-yodh* +  $\aleph$  *kaph* has unpredictable representations. The *daleth-gimel-yodh* may be written using both  $\gamma$  regular and  $\gamma$  tall forms. The tall form connects at the baseline to *kaph*, as in  $\aleph\gamma\gamma\gamma\gamma$  <p'dšnyk> *abāyišnīg* 'pleasing, attractive';  $\aleph\gamma\gamma$  <n'ylyk> *nāyrīg* 'adult woman';  $\gamma\gamma$  <kdk> *kayak*. But, also with its regular shape  $\gamma$ :  $\aleph\gamma\gamma\gamma\gamma$  <YKTYBWN-tn> *nipistan*;  $\aleph\gamma\gamma\gamma$  <wsyk'l> *vasīkār*;  $\aleph\gamma\gamma\gamma$  <kdkw't> *kai-kavar*. The usage of the  $\gamma$  regular and  $\gamma$  tall forms is not entirely contextual, that is, it is not based on letters that precede *daleth-gimel-yodh*, but a preservation of spelling conventions. Moreover, this variation attested in words, such as <nzdyk> *nazdīk* 'near', which has the contrastive forms  $\aleph\gamma\gamma$  and  $\aleph\gamma\gamma$  in the sources.
- The letter  $\aleph$  *pe* occurs in medial and final positions using both its  $\aleph$  regular and  $\aleph$  half forms (also used for medial and final  $\aleph$  *sadhe*). The medial forms of *pe* are not predictable:
  - After  $\aleph$  *aleph-heth*:
    - \* regular:  $\aleph\aleph\gamma\gamma$  <'pylwcglyh> *a-peroz-gareh*
    - \* half:  $\aleph\aleph$  <'p'c> *apac*
  - After  $\gamma$  *daleth-gimel-yodh*:
    - \* regular:  $\gamma\aleph$  <dpl> *davr*
    - \* half:  $\gamma\aleph$  <dpywr> *dipivar*
  - After  $\gamma$  *lamedh*:
    - \* regular:  $\gamma\aleph\gamma$  <hylpt> *herpat*
    - \* half:  $\gamma\aleph\gamma$  <gndlp> *Gandarv*
  - After  $\gamma$  *samekh*:
    - \* regular:  $\gamma\aleph\gamma\gamma$  <sp's> *spas*;
    - \* half:  $\gamma\aleph\gamma\gamma$  <spyn'y> *spenai*

## 4 Approach to the Encoding Model

While Book Pahlavi certainly has a set of rules that govern the joining behavior of its letters, it has several irregularities that present complications for defining a practical character-encoding model in Unicode.

Based on the few examples given above, it is clear that Book Pahlavi is a cursive joining script whose joining behavior presents several complications for interpretation. A writer can produce text using the rules that govern the joining behavior of each letter. But, due to the inherent ambiguities in the script, the resultant text may have multiple readings.

For this reason, a typical palaeographic or purely graphemic approach is not ideal. Instead, a hybrid model is needed. The goal of the proposed encoding model is to enable a user to represent Book Pahlavi in digital plain text as it appears in written and printed sources, using a repertoire that is graphically congruent with Book Pahlavi shapes. I propose a model based upon the following principles in order to achieve this:

### 4.1 Typology

Book Pahlavi is a right-to-left (RTL) cursive joining script. It requires an encoding model that supports these features, along with a character repertoire needed for proper representation of text.

### 4.2 Independent characters

A letter of the script that is clearly distinctive and cannot be confused with another letter or a sequence of other letters is to be encoded as an atomic character. This applies to:

𐭪 <i>beth</i>	𐭫 <i>lamedh</i>	𐭬 <i>hooked mem-qoph</i>
𐭧 <i>old daleth</i>	𐭫𐭮 <i>stroked lamedh</i>	𐭮 <i>old nun</i>
𐭥 <i>waw-nun-ayin-resh</i>	𐭫𐭮𐭮 <i>looped lamedh</i>	𐭮𐭮 <i>independent sadhe</i>
𐭥𐭮 <i>zayin</i>	𐭫𐭮𐭮𐭮 <i>hooked lamedh</i>	𐭮𐭮𐭮 <i>taw</i>
𐭥𐭮𐭮 <i>kaph</i>	𐭫𐭮𐭮𐭮𐭮 <i>old lamedh</i>	
𐭮𐭮𐭮 <i>old kaph</i>	𐭬𐭮 <i>mem-qoph</i>	

### 4.3 Letters not encoded independently

A letter whose shape resembles the rendered sequence of another letter or multiple letters is not encoded as an independent character. Instead it is to be represented using a sequence of letter-like elements, or ‘primitives’. The letters not included atomically in the repertoire are:

𐭪𐭮 <i>aleph-heth</i>	𐭬𐭮 <i>he</i>	𐭮𐭮𐭮 <i>pe</i>
𐭪𐭮𐭮 <i>gimel-daleth-yodh</i>	𐭬𐭮𐭮 <i>samekh</i>	𐭮𐭮𐭮𐭮 <i>shin</i>
𐭪𐭮𐭮𐭮 <i>curled gimel-daleth-yodh</i>	𐭮𐭮𐭮𐭮 <i>Indian samekh</i>	𐭮𐭮𐭮𐭮𐭮 <i>curled shin</i>

## 4.4 Character primitives

The above seven letters, as well as their contextual forms, are to be constructed using the following set of ‘primitives’:

𐭡	tooth	𐭢	curled tooth	𐭣	belly
𐭤	descending tooth	𐭥	descending curled tooth	𐭦	curled belly

As well as the half-form:

𐭧	final <i>pe-sadhe</i>
---	-----------------------

The descriptors ‘tooth’ and ‘belly’ are borrowed from Pournader’s L2/20-246. In L2/02-449, Everson used the term ‘archegrapheme’ for a similar set of characters. I use ‘tooth’ and ‘belly’ here as a matter of convenience; they may be replaced with more suitable terms. Similarly, I use ‘descending’ to describe the vertical lowering of a letter for enabling connections to letters below the baseline.

The ‘descending’ tooth characters 𐭤 and 𐭥 are differentiated from the belly characters 𐭣 and 𐭦 on account of their shape. For example, if a user needed to represent 𐭣, they would select <𐭣, 𐭧>. It would not be appropriate to treat 𐭤 as a contextual form of 𐭣, and to represent 𐭣 using <𐭣, 𐭧>. Similarly, for representing the form 𐭥, which might represent a descending *samekh* or a sequence of two *gimel-daleth-yodh* at word-final position, a user would expect to type <𐭥, 𐭧> and not <𐭦, 𐭧>. Burdening the belly with different forms in different contexts would prevent users from representing in plain text the written text. It would also require substitutions and additional rules for shaping in the font.

The descriptor ‘descending’ for 𐭤 and 𐭥 is assigned on the interpretation that they are forms of 𐭡 and 𐭢 with descending baselines. From an alternate interpretation, the ‘descending tooth’ and ‘descending curled tooth’ could be considered ‘half’ belly and ‘half curled’ belly, being truncated forms of 𐭣 belly and 𐭦 curled belly.

## 4.5 Alternate forms of letters

The following alternate forms are included as separate characters:

𐭨	independent upright <i>sadhe</i>	𐭩	final <i>sadhe</i>
---	----------------------------------	---	--------------------

## 4.6 Characters for contextual variants

The ‘descending’ and ‘belly’ forms of *zayin*, *lamedh*, stroked *lamedh*, looped *lamedh* that are used for forming different cursive connections are included as separate characters. Usage may vary by scribe and source text, but defining them as characters permits the representation of text as it appears in the sources:

𐭮	descending <i>zayin</i>	𐭮	descending stroked <i>lamedh</i>
𐭮	<i>zayin</i> belly	𐭮	stroked <i>lamedh</i> belly
𐭮	descending <i>lamedh</i>	𐭮	descending looped <i>lamedh</i>
𐭮	<i>lamedh</i> belly	𐭮	looped <i>lamedh</i> belly

Additional contextual forms of *waw-nun-ayin-res* are encoded as separate characters:

𐭮	short <i>waw-nun-ayin-res</i>	𐭮	curved <i>waw-nun-ayin-res</i>
---	-------------------------------	---	--------------------------------

Contextual forms are also provided for *mem-qoph*, which support the vertical adjustment that is conventional for the letter when medial and final:

𐭮	<i>mem-qoph</i>	𐭮	low <i>mem-qoph</i>
𐭮	descending <i>mem-qoph</i>	𐭮	descending low <i>mem-qoph</i>
𐭮	<i>mem-qoph</i> belly	𐭮	low <i>mem-qoph</i> belly

## 4.7 Atomic ligatures

The following atomic ligatures are proposed for encoding as independent characters:

𐭮	X <sub>1</sub>	𐭮	<i>yodh-heth</i>	𐭮	turned <i>ahriman</i>
𐭮	X <sub>2</sub>			𐭮	alternate turned <i>ahriman</i>

## 4.8 Punctuation

The following punctuation signs are proposed as distinctive characters for Book Pahlavi:

⋮	three dots	⋮	four dots
⊙	three circles	⊙	four circles

The 𐤀 ‘otiose sign’, or end-of-word sign is identical to 𐤀 *waw-ayin-nun-resh*, and is it be represented using the latter. At present, there is no justification for encoding the ‘otiose sign’ as a separate character. Other punctuation signs may be unified with existing Unicode common characters:

- U+00B7 MIDDLE DOT
- ○ U+25CB WHITE CIRCLE

## 4.9 ‘Kashida’

Extensions of the baseline are observed in several manuscripts, most commonly in the final stroke of letters in word-final position. The existing .U+0640 ARABIC TATWEEL is to be used for representing such features.

## 5 Application of the proposed encoding model

Applying the proposed model to the Book Pahlavi provides a repertoire that is not heavily reliant on a complex shaping engine. Instead, it provides characters that enable a user to represent text as it appears in Book Pahlavi materials, from manuscripts to printed books.

Accordingly, in order to represent the phrase shown at the beginning of this document:

אברהם שמואל טייטלבוים זצ"ל אב"ד ורעזבורג

*wištāsp šāhān šāh ud ērān šāh būd*

The following sequence of characters would be used, where SP is the common U+0020 SPACE:

[illegible]

As is evident in the above, the proposed encoding model enables a user to represent Book Pahlavi text simply by inspecting any given text and selecting the distinctive characters that correspond to the visual interpretation.

In L2/14-077R, Meyers showed how his model would be used for representing Pahlavi text, by using an excerpt from his figure 4.21 (shown below). He notes: “The passage [from Minug i Xrad] is handwritten in a fairly clear pedagogical style, making it easy for novices to read the text. At the same time the typeface of the handwriting employs some of the common stylistic ligatures.” (p. 57).

سورة ٢

ප්‍රකාශන අංක 1000

[illegible][illegible][illegible]

عائمتہ مسیحیوں کے لیے دعا  
 ۱۴۱۵ھ و ۱۴۱۶ھ کے لیے دعا  
 عائمتہ مسیحیوں کے لیے دعا

14

## 6 Proposed Repertoire

A practical Unicode repertoire for Book Pahlavi consists of the 57 characters shown in the table below:

- Proposed Unicode name for the character in the first column
- J: the joining behavior of the letter: dual-joining (D), right-joining (R), non-joining (N)
- X<sub>n</sub>: the independent or isolated form of the letter
- X<sub>f</sub>: the form of the letter in final position
- X<sub>m</sub>: the form of the letter in medial position
- X<sub>i</sub>: the form of the letter in initial position

<i>Primitives (6)</i>	J	X <sub>n</sub>	X <sub>f</sub>	X <sub>m</sub>	X <sub>i</sub>
BOOK PAHLAVI TOOTH	D	𐭠	𐭠-	𐭠-	𐭠
BOOK PAHLAVI DESCENDING TOOTH	D	𐭡	𐭡, 𐭡-	𐭡, 𐭡-	𐭡
BOOK PAHLAVI CURLED TOOTH	D	𐭢	𐭢-	𐭢-	𐭢
BOOK PAHLAVI DESCENDING CURLED TOOTH	D	𐭣	𐭣, 𐭣-	𐭣, 𐭣-	𐭣
BOOK PAHLAVI BELLY	D	𐭤	𐭤-	𐭤-	𐭤
BOOK PAHLAVI CURLED BELLY	D	𐭥	𐭥, 𐭥-	𐭥, 𐭥-	𐭥

<i>Letters (34)</i>	J	X <sub>n</sub>	X <sub>f</sub>	X <sub>m</sub>	X <sub>i</sub>
BOOK PAHLAVI LETTER BETH	R	𐭦	𐭦-	—	—
BOOK PAHLAVI LETTER OLD DALETH	R	𐭧	𐭧-	—	—
BOOK PAHLAVI LETTER WAW-NUN-AYIN-RESH	R	𐭨	𐭨-	—	—
BOOK PAHLAVI LETTER SHORT WAW-NUN-AYIN-RESH	R	𐭩	𐭩-	—	—
BOOK PAHLAVI LETTER CURVED WAW-NUN-AYIN-RESH	R	𐭪	𐭪-	—	—
BOOK PAHLAVI LETTER ZAYIN	D	𐭫	𐭫-	𐭫-	𐭫
BOOK PAHLAVI LETTER DESCENDING ZAYIN	D	𐭬	𐭬-	𐭬, 𐭬-	𐭬
BOOK PAHLAVI LETTER ZAYIN BELLY	D	𐭭	𐭭-	𐭭-	𐭭
BOOK PAHLAVI LETTER KAPH	R	𐭮	𐭮, 𐭮-	—	—
BOOK PAHLAVI LETTER OLD KAPH	R	𐭯	𐭯-	—	—
BOOK PAHLAVI LETTER LAMEDH	D	𐭰	𐭰-	𐭰-	𐭰
BOOK PAHLAVI LETTER DESCENDING LAMEDH	D	𐭱	𐭱-	𐭱, 𐭱-	𐭱
BOOK PAHLAVI LETTER LAMEDH BELLY	D	𐭲	𐭲-	𐭲-	𐭲

BOOK PAHLAVI LETTER STROKED LAMEDH	D	𐭣	𐭣	𐭣	𐭣
BOOK PAHLAVI LETTER DESCENDING STROKED LAMEDH	D	𐭣	𐭣	𐭣, 𐭣	𐭣
BOOK PAHLAVI LETTER STROKED LAMEDH BELLY	D	𐭣	𐭣	𐭣	𐭣
BOOK PAHLAVI LETTER LOOPED LAMEDH	D	𐭣	𐭣	𐭣	𐭣
BOOK PAHLAVI LETTER DESCENDING LOOPED LAMEDH	D	𐭣	𐭣	𐭣, 𐭣	𐭣
BOOK PAHLAVI LETTER LOOPED LAMEDH BELLY	D	𐭣	𐭣	𐭣	𐭣
BOOK PAHLAVI LETTER HOOKED LAMEDH	R	𐭣	𐭣	𐭣	𐭣
BOOK PAHLAVI LETTER OLD LAMEDH	R	𐭣	𐭣	—	—
BOOK PAHLAVI LETTER MEM-QOPH	D	𐭣	𐭣	𐭣	𐭣
BOOK PAHLAVI LETTER DESCENDING MEM-QOPH	L	𐭣			𐭣
BOOK PAHLAVI LETTER MEM-QOPH BELLY	D	𐭣			𐭣
BOOK PAHLAVI LETTER LOW MEM-QOPH	D	𐭣	𐭣	𐭣	
BOOK PAHLAVI LETTER DESCENDING LOW MEM-QOPH	D	𐭣		𐭣	
BOOK PAHLAVI LETTER LOW MEM-QOPH BELLY	D	𐭣		𐭣	
BOOK PAHLAVI LETTER HOOKED MEM-QOPH	D	𐭣	—	—	—
BOOK PAHLAVI LETTER OLD NUN	N	𐭣	—	—	—
BOOK PAHLAVI FINAL PE-SADHE	R	𐭣	𐭣	—	—
BOOK PAHLAVI LETTER INDEPENDENT SADHE	N	𐭣	—	—	—
BOOK PAHLAVI LETTER INDEPENDENT UPRIGHT SADHE	N	𐭣	—	—	—
BOOK PAHLAVI LETTER FINAL SADHE	R	𐭣	𐭣	—	—
BOOK PAHLAVI LETTER TAW	R	𐭣	𐭣	—	—

<i>Atomic Ligatures (5)</i>	J	X <sub>n</sub>	X <sub>f</sub>	X <sub>m</sub>	X <sub>i</sub>
BOOK PAHLAVI LIGATURE X1	R	𐭣𐭣	𐭣𐭣	—	—
BOOK PAHLAVI LIGATURE X2	R	𐭣𐭣	𐭣𐭣	—	—
BOOK PAHLAVI LIGATURE YODH-HETH	R	𐭣	𐭣	—	—
BOOK PAHLAVI LIGATURE TURNED AHRIMAN	N	𐭣𐭣𐭣	—	—	—
BOOK PAHLAVI LIGATURE ALTERNATE TURNED AHRIMAN	N	𐭣𐭣𐭣	—	—	—



*Combining Signs (7)*

BOOK PAHLAVI COMBINING DOT ABOVE	◌̇
BOOK PAHLAVI COMBINING DOT BELOW	◌̣
BOOK PAHLAVI COMBINING TWO DOTS ABOVE	◌̈
BOOK PAHLAVI COMBINING TWO DOTS BELOW	◌̣̣
BOOK PAHLAVI COMBINING THREE DOTS ABOVE	◌̣̣̣
BOOK PAHLAVI COMBINING THREE DOTS BELOW	◌̣̣̣̣
BOOK PAHLAVI COMBINING HAT ABOVE	◌̂

*Number (1)*

	J	X <sub>n</sub>	X <sub>f</sub>	X <sub>m</sub>	X <sub>i</sub>
BOOK PAHLAVI NUMBER ONE	R	𐭪	𐭪̣	—	—

*Punctuation (4)*

BOOK PAHLAVI PUNCTUATION THREE DOTS	⋮
BOOK PAHLAVI PUNCTUATION THREE CIRCLES	⊙⊙
BOOK PAHLAVI PUNCTUATION FOUR DOTS	⋮̣̣̣̣
BOOK PAHLAVI PUNCTUATION FOUR CIRCLES	⊙⊙⊙⊙

## 7 Encoded Representations

### 7.1 Palaeographical letters

Notably absent from the proposed Unicode repertoire are the following letters of Book Pahlavi. Given the fact that they resemble sequences of contextual forms of letters, they are not independently encoded. They are to be represented as sequences using the characters specified:

palaeographical letter	shape	character sequence
<i>aleph-heth</i>	𐭠𐭡	<𐭠 TOOTH, 𐭡 TOOTH>
<i>daleth-gimel-yodh</i>	𐭠	<𐭠 TOOTH>
curled <i>daleth-gimel-yodh</i>	𐭡	<𐭡 CURLED TOOTH>
<i>he</i>	𐭢	<𐭢 MEM-QOPH, 𐭣 SHORT WAW-NUN-AYIN-RESH>
<i>samekh</i>	𐭡𐭢	<𐭡 CURLED TOOTH, 𐭢 CURLED TOOTH>
Indian <i>samekh</i>	𐭢𐭡	<𐭢 CURLED BELLY, 𐭡 TOOTH>
<i>shin</i>	𐭢𐭠	<𐭢 BELLY, 𐭠 TOOTH, 𐭡 TOOTH>
curled <i>shin</i>	𐭢𐭡	<𐭢 CURLED BELLY, 𐭡 TOOTH, 𐭡 TOOTH>
<i>pe</i>	𐭣	<𐭣 DESCENDING TOOTH, 𐭤 FINAL PE-SADHE>

Contextual forms for complex letters are also to be composed using character primitives:

contextual form	shape	character sequence
descending <i>aleph-heth</i>	𐭠𐭡	<𐭠 TOOTH, 𐭡 DESCENDING TOOTH>
<i>aleph-heth</i> belly	𐭠𐭢	<𐭠 TOOTH, 𐭢 BELLY>
descending <i>gimel-daleth-yodh</i>	𐭠	<𐭠 DESCENDING TOOTH>
descending curled <i>gimel-daleth-yodh</i>	𐭡	<𐭡 DESCENDING CURLED TOOTH>
descending <i>samekh</i>	𐭡𐭢	<𐭡 DESCENDING CURLED TOOTH, 𐭢 DESCENDING CURLED TOOTH>
descending <i>shin</i>	𐭢𐭠	<𐭢 BELLY, 𐭠 TOOTH, 𐭡 DESCENDING TOOTH>
<i>shin</i> belly	𐭢𐭢	<𐭢 BELLY, 𐭠 TOOTH, 𐭢 BELLY>
descending curled <i>shin</i>	𐭢𐭡	<𐭢 CURLED BELLY, 𐭡 TOOTH, 𐭡 DESCENDING TOOTH>
curled <i>shin</i> belly	𐭢𐭢	<𐭢 CURLED BELLY, 𐭡 TOOTH, 𐭢 BELLY>

## 7.2 Combining signs

Combining signs are used as follows:

combining sign	usage	sequence	value
◌̇	DOT ABOVE	𐭪 <𐭪 DESCENDING TOOTH, 𐭪 FINAL PE-SADHE, ◌̇ DOT ABOVE>	<i>f</i>
◌̈	DOT BELOW	𐭪 <𐭪 CURLED TOOTH, ◌̈ DOT ABOVE>	<i>j</i>
◌̈̈	TWO DOTS ABOVE	𐭪 <𐭪 CURLED TOOTH, ◌̈̈ TWO DOTS ABOVE>	<i>g</i>
◌̈̈̈	TWO DOTS BELOW	𐭪 <𐭪 CURLED TOOTH, ◌̈̈̈ TWO DOTS BELOW>	<i>j, i, c</i>
◌̈̈̈̈	THREE DOTS ABOVE	𐭪 <𐭪 BELLY, 𐭪 TOOTH, 𐭪 TOOTH, ◌̈̈̈̈ THREE DOTS ABOVE>	<i>sh</i>
◌̈̈̈̈̈	THREE DOTS BELOW	𐭪 <𐭪 DESCENDING TOOTH, 𐭪 FINAL PE-SADHE, ◌̈̈̈̈̈ THREE DOTS BELOW>	<i>p, z</i>
◌̂	HAT ABOVE	𐭪 <𐭪 CURLED TOOTH, ◌̂ HAT ABOVE>	<i>d</i>

## 7.3 Numbers

Book Pahlavi numbers may be represented using characters in the proposed repertoire. If there is a requirement to preserve the numeric values of numbers, then these would need to be encoded as atomic characters.

The primary units are to be represented as follows:

1	𐭪	<𐭪 ONE>
2	𐭪𐭪	<𐭪 TOOTH, 𐭪 ONE>
3	𐭪𐭪𐭪	<𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE>
4	𐭪𐭪𐭪𐭪	<𐭪 TOOTH, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE>
5	𐭪𐭪𐭪𐭪𐭪	<𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE, 𐭪 TOOTH, 𐭪 ONE>
6	𐭪𐭪𐭪𐭪𐭪𐭪	<𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE>
7	𐭪𐭪𐭪𐭪𐭪𐭪𐭪	<𐭪 TOOTH, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE>
8	𐭪𐭪𐭪𐭪𐭪𐭪𐭪𐭪	<𐭪 TOOTH, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE>
9	𐭪𐭪𐭪𐭪𐭪𐭪𐭪𐭪𐭪	<𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE>

Ten through nineteen are represented as follows:

10	𐭌	<𐭌 OLD DALETH>
11	𐭎	<𐭎 CURLED BELLY, 𐭌 ONE>
12	𐭏	<𐭎 CURLED BELLY, 𐭌 TOOTH, 𐭌 ONE>
13	𐭐	<𐭎 CURLED BELLY, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 ONE>
14	𐭑	<𐭎 CURLED BELLY, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 ONE>
15	𐭒	<𐭎 CURLED BELLY, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 ONE, 𐭌 TOOTH, 𐭌 ONE>
16	𐭓	<𐭎 CURLED BELLY, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 ONE, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 ONE>
17	𐭔	<𐭎 CURLED BELLY, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 ONE, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 ONE>
18	𐭕	<𐭎 CURLED BELLY, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 ONE, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 ONE>
19	𐭖	<𐭎 CURLED BELLY, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 ONE, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 ONE, 𐭌 TOOTH, 𐭌 TOOTH, 𐭌 ONE>

Twenty through ninety are expressed as follows:

20	𐭗	<𐭗 LAMEDH>
30	𐭘	<𐭗 LAMEDH, 𐭘 DESCENDING CURLED TOOTH>
40	𐭙	<𐭙 CURLED BELLY, 𐭌 TOOTH>
50	𐭚	<𐭙 CURLED BELLY, 𐭌 TOOTH, 𐭘 DESCENDING CURLED TOOTH>
60	𐭛	<𐭘 DESCENDING CURLED TOOTH, 𐭙 CURLED BELLY, 𐭌 TOOTH>
70	𐭜	<𐭘 DESCENDING CURLED TOOTH, 𐭙 CURLED BELLY, 𐭌 TOOTH, 𐭘 DESCENDING CURLED TOOTH>
80	𐭝	<𐭘 DESCENDING CURLED TOOTH, 𐭘 DESCENDING CURLED TOOTH, 𐭙 CURLED BELLY, 𐭌 TOOTH>
90	𐭞	<𐭘 DESCENDING CURLED TOOTH, 𐭘 DESCENDING CURLED TOOTH, 𐭙 CURLED BELLY, 𐭌 TOOTH, 𐭘 DESCENDING CURLED TOOTH>

The hundreds are represented as follows:

100	𐭪𐭥	<𐭪 LAMEDH, 𐭥 DESCENDING ZAYIN>
100	𐭪𐭥	<𐭪 TOOTH, 𐭪 LAMEDH, 𐭥 DESCENDING ZAYIN>
200	𐭪𐭥𐭥	<𐭪 TOOTH, 𐭪 TOOTH, 𐭪 LAMEDH, 𐭥 DESCENDING ZAYIN>
300	𐭪𐭥𐭥𐭥	<𐭪 TOOTH, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 LAMEDH, 𐭥 DESCENDING ZAYIN>
800	𐭪𐭥𐭥𐭥𐭥	<𐭪 TOOTH, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 LAMEDH, 𐭥 DESCENDING ZAYIN>
900	𐭪𐭥𐭥𐭥𐭥	<𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 ONE, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 TOOTH, 𐭪 LAMEDH, 𐭥 DESCENDING ZAYIN>

Thousands are expressed as:

1000	𐭪𐭥𐭥	<𐭪 LAMEDH, 𐭥 FINAL OLD KAPH>
1000	𐭪𐭥𐭥	<𐭪 TOOTH, 𐭪 LAMEDH, 𐭥 FINAL OLD KAPH>
2000	𐭪𐭥𐭥𐭥	<𐭪 TOOTH, 𐭪 TOOTH, 𐭪 LAMEDH, 𐭥 FINAL OLD KAPH>

## 8 Character Data

### Character Properties: UnicodeData.txt

```

10BB0;BOOK PAHLAVI LETTER TOOTH;Lo;0;R;;;;;N;;;;;
10BB1;BOOK PAHLAVI LETTER DESCENDING TOOTH;Lo;0;R;;;;;N;;;;;
10BB2;BOOK PAHLAVI LETTER CURLED TOOTH;Lo;0;R;;;;;N;;;;;
10BB3;BOOK PAHLAVI LETTER DESCENDING CURLED TOOTH;Lo;0;R;;;;;N;;;;;
10BB4;BOOK PAHLAVI LETTER BELLY;Lo;0;R;;;;;N;;;;;
10BB5;BOOK PAHLAVI LETTER CURLED BELLY;Lo;0;R;;;;;N;;;;;
10BB6;BOOK PAHLAVI LETTER BETH;Lo;0;R;;;;;N;;;;;
10BB7;BOOK PAHLAVI LETTER OLD DALETH;Lo;0;R;;;;;N;;;;;
10BB8;BOOK PAHLAVI LETTER WAW-NUN-AYIN-RESH;Lo;0;R;;;;;N;;;;;
10BB9;BOOK PAHLAVI LETTER SHORT WAW-NUN-AYIN-RESH;Lo;0;R;;;;;N;;;;;
10BBA;BOOK PAHLAVI LETTER CURVED WAW-NUN-AYIN-RESH;Lo;0;R;;;;;N;;;;;
10BBB;BOOK PAHLAVI LETTER ZAYIN;Lo;0;R;;;;;N;;;;;
10BBC;BOOK PAHLAVI LETTER DESCENDING ZAYIN;Lo;0;R;;;;;N;;;;;
10BBD;BOOK PAHLAVI LETTER ZAYIN BELLY;Lo;0;R;;;;;N;;;;;
10BBE;BOOK PAHLAVI LETTER KAPH;Lo;0;R;;;;;N;;;;;
10BBF;BOOK PAHLAVI LETTER OLD KAPH;Lo;0;R;;;;;N;;;;;
10BC0;BOOK PAHLAVI LETTER LAMEDH;Lo;0;R;;;;;N;;;;;
10BC1;BOOK PAHLAVI LETTER DESCENDING LAMEDH;Lo;0;R;;;;;N;;;;;
10BC2;BOOK PAHLAVI LETTER LAMEDH BELLY;Lo;0;R;;;;;N;;;;;
10BC3;BOOK PAHLAVI LETTER STROKED LAMEDH;Lo;0;R;;;;;N;;;;;
10BC4;BOOK PAHLAVI LETTER DESCENDING STROKED LAMEDH;Lo;0;R;;;;;N;;;;;
10BC5;BOOK PAHLAVI LETTER STROKED LAMEDH BELLY;Lo;0;R;;;;;N;;;;;
10BC6;BOOK PAHLAVI LETTER LOOPED LAMEDH;Lo;0;R;;;;;N;;;;;
10BC7;BOOK PAHLAVI LETTER DESCENDING LOOPED LAMEDH;Lo;0;R;;;;;N;;;;;
10BC8;BOOK PAHLAVI LETTER LOOPED LAMEDH BELLY;Lo;0;R;;;;;N;;;;;
10BC9;BOOK PAHLAVI LETTER HOOKED LAMEDH;Lo;0;R;;;;;N;;;;;
10BCA;BOOK PAHLAVI LETTER OLD LAMEDH;Lo;0;R;;;;;N;;;;;
10BCB;BOOK PAHLAVI LETTER MEM-QOPH;Lo;0;R;;;;;N;;;;;
10BCC;BOOK PAHLAVI LETTER DESCENDING MEM-QOPH;Lo;0;R;;;;;N;;;;;
10BCD;BOOK PAHLAVI LETTER MEM-QOPH BELLY;Lo;0;R;;;;;N;;;;;
10BCE;BOOK PAHLAVI LETTER LOW MEM-QOPH;Lo;0;R;;;;;N;;;;;
10BCF;BOOK PAHLAVI LETTER LOW DESCENDING MEM-QOPH;Lo;0;R;;;;;N;;;;;
10BD0;BOOK PAHLAVI LETTER LOW MEM-QOPH BELLY;Lo;0;R;;;;;N;;;;;
10BD1;BOOK PAHLAVI LETTER HOOKED MEM-QOPH;Lo;0;R;;;;;N;;;;;
10BD2;BOOK PAHLAVI LETTER OLD NUN;Lo;0;R;;;;;N;;;;;
10BD3;BOOK PAHLAVI LETTER FINAL PE-SADHE;Lo;0;R;;;;;N;;;;;
10BD4;BOOK PAHLAVI LETTER INDEPENDENT SADHE;Lo;0;R;;;;;N;;;;;
10BD5;BOOK PAHLAVI LETTER INDEPENDENT UPRIGHT SADHE;Lo;0;R;;;;;N;;;;;
10BD6;BOOK PAHLAVI LETTER FINAL SADHE;Lo;0;R;;;;;N;;;;;
10BD7;BOOK PAHLAVI LETTER TAW;Lo;0;R;;;;;N;;;;;
10BD8;BOOK PAHLAVI LETTER LIGATURE X1;Lo;0;R;;;;;N;;;;;
10BD9;BOOK PAHLAVI LETTER LIGATURE X2;Lo;0;R;;;;;N;;;;;
10BDA;BOOK PAHLAVI LETTER LIGATURE YODH-HETH;Lo;0;R;;;;;N;;;;;
10BDB;BOOK PAHLAVI LETTER LIGATURE TURNED AHRIMAN;Lo;0;R;;;;;N;;;;;
10BDC;BOOK PAHLAVI LETTER LIGATURE ALTERNATE TURNED AHRIMAN;Lo;0;R;;;;;N;;;;;
10BDD;BOOK PAHLAVI COMBINING DOT ABOVE;Mn;230;NSM;;;;;N;;;;;
10BDE;BOOK PAHLAVI COMBINING DOT BELOW;Mn;220;NSM;;;;;N;;;;;
10BDF;BOOK PAHLAVI COMBINING TWO DOTS ABOVE;Mn;230;NSM;;;;;N;;;;;
10BE0;BOOK PAHLAVI COMBINING TWO DOTS BELOW;Mn;220;NSM;;;;;N;;;;;
10BE1;BOOK PAHLAVI COMBINING THREE DOTS ABOVE;Mn;230;NSM;;;;;N;;;;;
10BE2;BOOK PAHLAVI COMBINING THREE DOTS BELOW;Mn;220;NSM;;;;;N;;;;;

```

```

10BE3;BOOK PAHLAVI COMBINING HAT ABOVE;Mn;230;NSM;;;;;N;;;;;
10BE4;BOOK PAHLAVI NUMBER ONE;No;0;R;;;;;N;;;;;
10BE5;BOOK PAHLAVI PUNCTUATION THREE DOTS;Po;0;AL;;;;;N;;;;;
10BE6;BOOK PAHLAVI PUNCTUATION THREE CIRCLES;Po;0;AL;;;;;N;;;;;
10BE7;BOOK PAHLAVI PUNCTUATION FOUR DOTS;Po;0;AL;;;;;N;;;;;
10BE8;BOOK PAHLAVI PUNCTUATION FOUR CIRCLES;Po;0;AL;;;;;N;;;;;

```

### Linebreaking Properties: LineBreak.txt

```

10BB0..10BDC;AL # Lo [45] BOOK PAHLAVI TOOTH..BOOK PAHLAVI LIGATURE ALTERNATE TURNED AHRIMAN
10BDD..10BE3;AL # Cm [7] BOOK PAHLAVI COMBINING DOT ABOVE..BOOK PAHLAVI COMBINING HAT ABOVE
10BE4;AL # No BOOK PAHLAVI NUMBER ONE
10BE5..10BE8;AL # Po [4] BOOK PAHLAVI PUNCTUATION THREE DOTS..BOOK PAHLAVI PUNCTUATION THREE CIRCLES

```

### Arabic Shaping Properties: ArabicShaping.txt

```

10BB0; BOOK PAHLAVI TOOTH; D; No_Joining_Group
10BB1; BOOK PAHLAVI DESCENDING TOOTH; D; No_Joining_Group
10BB2; BOOK PAHLAVI CURLED TOOTH; D; No_Joining_Group
10BB3; BOOK PAHLAVI DESCENDING CURLED TOOTH; D; No_Joining_Group
10BB4; BOOK PAHLAVI BELLY; D; No_Joining_Group
10BB5; BOOK PAHLAVI CURLED BELLY; D; No_Joining_Group
10BB6; BOOK PAHLAVI BETH; R; No_Joining_Group
10BB7; BOOK PAHLAVI OLD DALETH; R; No_Joining_Group
10BB8; BOOK PAHLAVI WAW-NUN-AYIN-RESH; R; No_Joining_Group
10BB9; BOOK PAHLAVI SHORT WAW-NUN-AYIN-RESH; R; No_Joining_Group
10BBA; BOOK PAHLAVI CURVED WAW-NUN-AYIN-RESH; R; No_Joining_Group
10BBB; BOOK PAHLAVI ZAYIN; D; No_Joining_Group
10BBC; BOOK PAHLAVI DESCENDING ZAYIN; D; No_Joining_Group
10BBD; BOOK PAHLAVI ZAYIN BELLY; D; No_Joining_Group
10BBE; BOOK PAHLAVI KAPH; R; No_Joining_Group
10BBF; BOOK PAHLAVI OLD KAPH; R; No_Joining_Group
10BC0; BOOK PAHLAVI LAMEDH; D; No_Joining_Group
10BC1; BOOK PAHLAVI DESCENDING LAMEDH; D; No_Joining_Group
10BC2; BOOK PAHLAVI LAMEDH BELLY; D; No_Joining_Group
10BC3; BOOK PAHLAVI STROKED LAMEDH; D; No_Joining_Group
10BC4; BOOK PAHLAVI DESCENDING STROKED LAMEDH; D; No_Joining_Group
10BC5; BOOK PAHLAVI STROKED LAMEDH BELLY; D; No_Joining_Group
10BC6; BOOK PAHLAVI LOOPED LAMEDH; D; No_Joining_Group
10BC7; BOOK PAHLAVI DESCENDING LOOPED LAMEDH; D; No_Joining_Group
10BC8; BOOK PAHLAVI LOOPED LAMEDH BELLY; D; No_Joining_Group
10BC9; BOOK PAHLAVI HOOKED LAMEDH; R; No_Joining_Group
10BCA; BOOK PAHLAVI OLD LAMEDH; R; No_Joining_Group
10BCB; BOOK PAHLAVI MEM-QOPH; D; No_Joining_Group
10BCC; BOOK PAHLAVI DESCENDING MEM-QOPH; D; No_Joining_Group
10BCD; BOOK PAHLAVI MEM-QOPH BELLY; D; No_Joining_Group
10BCE; BOOK PAHLAVI LOW MEM-QOPH; D; No_Joining_Group
10BCF; BOOK PAHLAVI LOW DESCENDING MEM-QOPH; D; No_Joining_Group
10BD0; BOOK PAHLAVI LOW MEM-QOPH BELLY; D; No_Joining_Group
10BD1; BOOK PAHLAVI HOOKED MEM-QOPH; D; No_Joining_Group
10BD2; BOOK PAHLAVI OLD NUN; U; No_Joining_Group
10BD3; BOOK PAHLAVI FINAL PE-SADHE; R; No_Joining_Group
10BD4; BOOK PAHLAVI INDEPENDENT SADHE; U; No_Joining_Group
10BD5; BOOK PAHLAVI INDEPENDENT UPRIGHT SADHE; U; No_Joining_Group
10BD6; BOOK PAHLAVI FINAL SADHE; R; No_Joining_Group

```

10BD7; BOOK PAHLAVI TAW; R; No\_Joining\_Group  
10BD8; BOOK PAHLAVI X1; R; No\_Joining\_Group  
10BD9; BOOK PAHLAVI X2; R; No\_Joining\_Group  
10BDA; BOOK PAHLAVI YODH-HETH; R; No\_Joining\_Group  
10BDB; BOOK PAHLAVI TURNED AHRIMAN; U; No\_Joining\_Group  
10BDC; BOOK PAHLAVI ALTERNATE TURNED AHRIMAN; U; No\_Joining\_Group

## **9 Acknowledgments**

I express my gratitude to Roozbeh Pournader, Abe Meyers, Arash Zeini, Liang Hai, and Michael Everson for their previous and parallel contributions for encoding Book Pahlavi in Unicode. I am confident that, through continued communications, we will find the way forward for this script.

This project has been made possible in part by funding from the Adopt-A-Character program of the Unicode Consortium, and has been supervised by Dr. Deborah Anderson and Rick McGowan.

Earlier efforts were made possible in part by a grant (PR-268710-20) from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley).



	10BB	10BC	10BD	10BE
0	𐭪 10BB0	𐭫 10BC0	𐭬 10BD0	𐭮 10BE0
1	𐭩 10BB1	𐭬 10BC1	𐭭 10BD1	𐭯 10BE1
2	𐭪 10BB2	𐭫 10BC2	𐭮 10BD2	𐭰 10BE2
3	𐭫 10BB3	𐭬 10BC3	𐭭 10BD3	𐭯 10BE3
4	𐭬 10BB4	𐭭 10BC4	𐭮 10BD4	𐭩 10BE4
5	𐭭 10BB5	𐭮 10BC5	𐭯 10BD5	𐭰 10BE5
6	𐭮 10BB6	𐭯 10BC6	𐭰 10BD6	𐭱 10BE6
7	𐭯 10BB7	𐭰 10BC7	𐭱 10BD7	𐭲 10BE7
8	𐭰 10BB8	𐭱 10BC8	𐭲 10BD8	𐭳 10BE8
9	𐭰 10BB9	𐭱 10BC9	𐭲 10BD9	
A	𐭱 10BBA	𐭲 10BCA	𐭳 10BDA	
B	𐭲 10BBB	𐭳 10BCB	𐭴 10BDB	
C	𐭲 10BBC	𐭳 10BCC	𐭴 10BDC	
D	𐭲 10BBD	𐭳 10BCD	𐭴 10BDD	
E	𐭲 10BBE	𐭳 10BCE	𐭴 10BDE	
F	𐭲 10BBF	𐭳 10BCF	𐭴 10BDF	

**Primitives**

10BB0	𐭡	BOOK PAHLAVI TOOTH
10BB1	𐭢	BOOK PAHLAVI DESCENDING TOOTH
10BB2	𐭣	BOOK PAHLAVI CURLED TOOTH
10BB3	𐭤	BOOK PAHLAVI DESCENDING CURLED TOOTH
10BB4	𐭥	BOOK PAHLAVI BELLY
10BB5	𐭦	BOOK PAHLAVI CURLED BELLY

**Letters**

10BB6	𐭧	BOOK PAHLAVI BETH
10BB7	𐭨	BOOK PAHLAVI LETTER OLD DALETH
10BB8	𐭩	BOOK PAHLAVI LETTER WAW-NUN-AYIN-RESH = also used for the otiose sign or end-of-word marker
10BB9	𐭪	BOOK PAHLAVI LETTER SHORT WAW-NUN-AYIN-RESH
10BBA	𐭫	BOOK PAHLAVI LETTER CURVED WAW-NUN-AYIN-RESH
10BBB	𐭬	BOOK PAHLAVI LETTER ZAYIN
10BBC	𐭭	BOOK PAHLAVI LETTER DESCENDING ZAYIN
10BBD	𐭮	BOOK PAHLAVI LETTER ZAYIN BELLY
10BBE	𐭯	BOOK PAHLAVI LETTER KAPH
10BBF	𐭰	BOOK PAHLAVI LETTER OLD KAPH
10BC0	𐭱	BOOK PAHLAVI LETTER LAMEDH
10BC1	𐭲	BOOK PAHLAVI LETTER DESCENDING LAMEDH
10BC2	𐭳	BOOK PAHLAVI LETTER LAMEDH BELLY
10BC3	𐭴	BOOK PAHLAVI LETTER STROKED LAMEDH
10BC4	𐭵	BOOK PAHLAVI LETTER DESCENDING STROKED LAMEDH
10BC5	𐭶	BOOK PAHLAVI LETTER STROKED LAMEDH BELLY
10BC6	𐭷	BOOK PAHLAVI LETTER LOOPED LAMEDH
10BC7	𐭸	BOOK PAHLAVI LETTER DESCENDING LOOPED LAMEDH
10BC8	𐭹	BOOK PAHLAVI LETTER LOOPED LAMEDH BELLY
10BC9	𐭺	BOOK PAHLAVI LETTER HOOKED LAMEDH
10BCA	𐭻	BOOK PAHLAVI LETTER OLD LAMEDH
10BCB	𐭼	BOOK PAHLAVI LETTER MEM-QOPH
10BCC	𐭽	BOOK PAHLAVI LETTER DESCENDING MEM-QOPH
10BCD	𐭾	BOOK PAHLAVI LETTER MEM-QOPH BELLY
10BCE	𐭿	BOOK PAHLAVI LETTER LOW MEM-QOPH
10BCF	𐮀	BOOK PAHLAVI LETTER LOW DESCENDING MEM-QOPH
10BD0	𐮁	BOOK PAHLAVI LETTER LOW MEM-QOPH BELLY
10BD1	𐮂	BOOK PAHLAVI LETTER HOOKED MEM-QOPH
10BD2	𐮃	BOOK PAHLAVI LETTER OLD NUN
10BD3	𐮄	BOOK PAHLAVI LETTER FINAL PE-SADHE
10BD4	𐮅	BOOK PAHLAVI LETTER INDEPENDENT SADHE
10BD5	𐮆	BOOK PAHLAVI LETTER INDEPENDENT UPRIGHT SADHE
10BD6	𐮇	BOOK PAHLAVI LETTER FINAL SADHE
10BD7	𐮈	BOOK PAHLAVI LETTER TAW

**Ligatures**

10BD8	𐮉	BOOK PAHLAVI LIGATURE X1
10BD9	𐮊	BOOK PAHLAVI LIGATURE X2
10BDA	𐮋	BOOK PAHLAVI LIGATURE YODH-HETH
10BDB	𐮌	BOOK PAHLAVI LIGATURE TURNED AHRIMAN
10BDC	𐮍	BOOK PAHLAVI LIGATURE ALTERNATE TURNED AHRIMAN

**Combining signs**

10BDD	◌̇	BOOK PAHLAVI COMBINING DOT ABOVE
10BDE	◌̈	BOOK PAHLAVI COMBINING DOT BELOW

10BDF	◌̉	BOOK PAHLAVI COMBINING TWO DOTS ABOVE
10BE0	◌̊	BOOK PAHLAVI COMBINING TWO DOTS BELOW
10BE1	◌̋	BOOK PAHLAVI COMBINING THREE DOTS ABOVE
10BE2	◌̌	BOOK PAHLAVI COMBINING THREE DOTS BELOW
10BE3	◌̍	BOOK PAHLAVI COMBINING HAT ABOVE

**Number**

10BE4	𐮎	BOOK PAHLAVI NUMBER ONE
-------	---	-------------------------

**Punctuation**

10BE5	⋯	BOOK PAHLAVI PUNCTUATION THREE DOTS
10BE6	⊘	BOOK PAHLAVI PUNCTUATION THREE CIRCLES
10BE7	⋯	BOOK PAHLAVI PUNCTUATION FOUR DOTS
10BE8	⊘	BOOK PAHLAVI PUNCTUATION FOUR CIRCLES