

Advancing the encoding model for Book Pahlavi letters

Anshuman Pandey
pandey@umich.edu

April 23, 2021

1 Introduction

This document aims to provide a model for representing the letters of the Book Pahlavi script. Combining signs and numbers are not discussed here.

2 Description of the Letters

The basic repertoire for Book Pahlavi contains the following 14 letters:

𐭀	𐭁	𐭂	𐭃	𐭄	𐭅	𐭆	𐭇	𐭈	𐭉	𐭊	𐭋	𐭌	𐭍
<i>aleph</i>	<i>beth</i>	<i>gimel,</i>	<i>he</i>	<i>waw,</i>	<i>zayin</i>	<i>kaph</i>	<i>lamedh</i>	<i>mem,</i>	<i>samekh</i>	<i>pe</i>	<i>sadhe</i>	<i>shin</i>	<i>taw</i>
<i>heth</i>		<i>daleth,</i>		<i>ayin,</i>				<i>qoph</i>					
		<i>yodh</i>		<i>nun,</i>									
				<i>resh</i>									

It has 11 supplemental letters, which are alternate and historical forms:

𐭎	𐭏	𐭐	𐭑	𐭒	𐭓	𐭔	𐭕	𐭖	𐭗	𐭘	𐭙
'curled'	'old'	'old'	final	'stroked'	'looped'	'hooked'	'old'	'tall'	'Indian'	'curled'	
<i>gimel,</i>	<i>daleth</i>	<i>kaph</i>	'old'	<i>lamedh</i>	<i>lamedh</i>	<i>lamedh</i>	<i>lamedh</i>	<i>samekh</i>	<i>samekh</i>	<i>shin</i>	
<i>daleth,</i>			<i>kaph</i>								
<i>yodh</i>											

The repertoire also contains the following 3 atomic ligatures, which behave as letters:

𐭚	𐭛	𐭜
X_1	X_2	y'

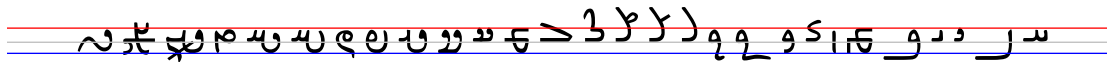
There names for letters are based on Aramaic analogues; however, there is no formal convention for describing alternate forms. The descriptors given above, such as 'old', 'curled', etc., have been assigned by the proposal author as a means for convenient reference.

2.1 Directionality

The script 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

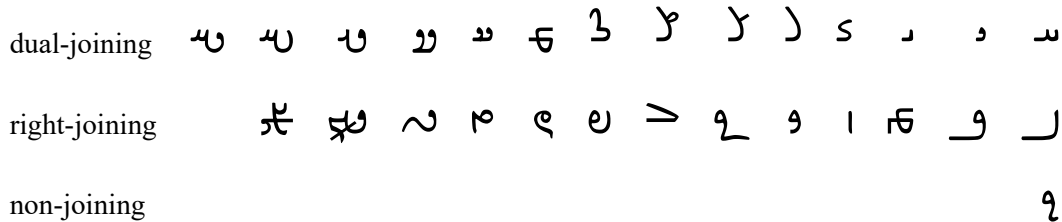
- the bottom of 𐭠, 𐭡, 𐭢, 𐭣, 𐭤
- the resting spot for the heads of 𐭥, 𐭦, 𐭧, 𐭨, 𐭩
- the cross-bar of 𐭪 and 𐭫.

The following shows the alignment of all letters with the baseline (gray). The head-height is measured by the tops of these letters, while the below-base is determined by 𐭥, 𐭦, 𐭩, etc. The swash terminals of 𐭬, 𐭭, 𐭮 run just under the below-base level, in order to accommodate letters that are written within the terminal.

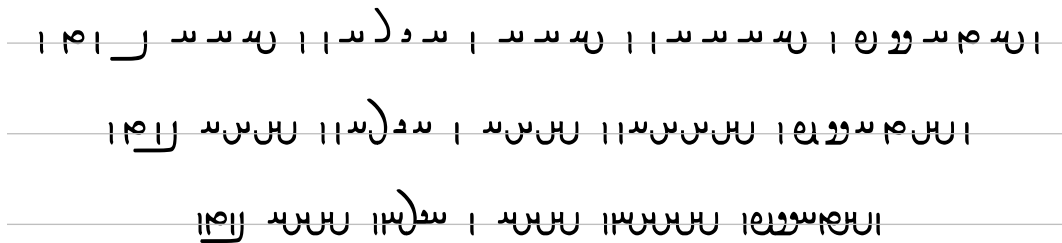


2.2 Joining behavior

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



During joining, a letter may be represented using a contextual or modified form, which is determined by its position within the cursive string, by adjacent letters, and in many cases, by both. In the example below, the first line shows a string of isolated letters, the second line shows the contextual forms of those letters, and the third line shows the shaped forms of the letters:



<wšt'sp' š'h'n' š'h w 'yl'n' š'h bwt'>
wištāsp šāhān šāh ud ērān šāh būd
 Wištāsp was the king of kings and the king of the Iranians.

The table below shows the contextual forms of Book Pahlavi letters, grouped according to joining patterns as analyzed by the proposal author. Two tables follow that, which show the joining behavior of dual-joining and right-joining letters.

		X_n	X_f	X_m	X_i
Minimal or no change in shape	<i>beth</i>	𐭪	𐭪	—	—
	‘old’ <i>daleth</i>	𐭫	𐭫	—	—
	<i>kaph</i>	𐭬	𐭬, 𐭬	—	—
	‘old’ <i>kaph</i>	𐭭	𐭭	—	—
	‘hooked’ <i>lamedh</i>	𐭮	𐭮	𐭮	𐭮
	‘old’ <i>lamedh</i>	𐭯	𐭯	—	—
	<i>samekh</i>	𐭰	𐭰	𐭰	𐭰
	‘tall’ <i>samekh</i>	𐭱	𐭱	𐭱	𐭱
	‘Indian’ <i>samekh</i>	𐭲	𐭲	𐭲	𐭲
	<i>taw</i>	𐭳	𐭳	—	—
	X_1	𐭴	𐭴	—	—
	X_2	𐭵	𐭵	—	—
	<i>y-h</i> ligature	𐭶	𐭶	—	—
Height adjustments for below-base or baseline connections	<i>waw-nun-ayin-res</i>	𐭷	𐭷, 𐭷	—	—
	<i>zayin</i>	𐭸	𐭸, 𐭸	𐭸, 𐭸, 𐭸	𐭸, 𐭸, 𐭸
	<i>lamedh</i>	𐭹	𐭹, 𐭹	𐭹, 𐭹, 𐭹	𐭹, 𐭹, 𐭹
	‘stroked’ <i>lamedh</i>	𐭺	𐭺, 𐭺	𐭺, 𐭺, 𐭺	𐭺, 𐭺, 𐭺
	‘looped’ <i>lamedh</i>	𐭻	𐭻, 𐭻	𐭻, 𐭻, 𐭻	𐭻, 𐭻, 𐭻
Vertical positioning of body	<i>he</i>	𐭼	𐭼, 𐭼	—	—
	<i>mem-qoph</i>	𐭽	𐭽, 𐭽	𐭽, 𐭽, 𐭽	𐭽, 𐭽
Truncation of strokes, or no change	<i>pe</i>	𐭾	𐭾, 𐭾, 𐭾	—	—
	<i>sadhe</i>	𐭿	𐭿, 𐭿	—	—
Descent of terminal	<i>aleph-heth</i>	𐮀	𐮀	𐮀, 𐮀, 𐮀	𐮀, 𐮀, 𐮀
	<i>gimel-daleth-yodh</i>	𐮁	𐮁	𐮁, 𐮁	𐮁, 𐮁
	‘curled’ <i>gimel-daleth-yodh</i>	𐮂	𐮂	𐮂, 𐮂, 𐮂	𐮂, 𐮂, 𐮂
	<i>shin</i>	𐮃	𐮃	𐮃, 𐮃, 𐮃, 𐮃	𐮃, 𐮃, 𐮃
	‘curled’ <i>shin</i>	𐮄	𐮄	𐮄, 𐮄, 𐮄	𐮄, 𐮄, 𐮄

Dual-joining letters

و	س	ن	ف	ط	ز	د	س	د	س	
س و	س س س س	ن و	ف و	ط و	ز و	د و	س و	د و د و د و د و	س و	س و
و و	س و	ن و ن و	ف و	—	ز و	د و	س و	د و د و و و و و	س و	د و
و و	س و س و	ن و	ف و	—	ز و	د و	س و	د و و و	س و	د و
و و	س و س و	ن و	ف و	—	ز و	د و	س و	د و و و	س و	د و
—	—	—	—	—	—	—	—	—	ط و	ط و
و و	س و	ن و ن و	ف و	—	ز و	د و	س و	د و د و د و د و	س و	د و
و و	س و	ن و	ف و	—	ز و ز و	د و د و	س و س و	د و و و و و	س و	د و
و و	س و	ن و	ف و	—	ز و	د و	س و	د و و و	س و	د و
و و	س و	ن و	ف و	—	ز و	د و	س و	د و و و	س و	د و

Right-joining letters

𐭀	𐭁	𐭂	𐭃	𐭄	𐭅	𐭆	𐭇	𐭈	𐭉	𐭊
𐭀	𐭁 𐭁	𐭂 𐭂 𐭂	𐭃	𐭄	𐭅	𐭆	𐭇	𐭈	𐭉	𐭊
𐭀	𐭁 𐭁	𐭂 𐭂 𐭂	𐭃	𐭄 𐭄 𐭄	—	𐭆 𐭆 𐭆	𐭇	𐭈	𐭉	𐭊
𐭀	𐭁 𐭁	𐭂 𐭂 𐭂	𐭃	𐭄 𐭄	—	𐭆	𐭇	𐭈	𐭉	𐭊
𐭀	𐭁 𐭁	𐭂 𐭂	𐭃	𐭄	—	𐭆	𐭇	𐭈	𐭉	𐭊
𐭀	𐭁 𐭁	𐭂 𐭂	𐭃	𐭄	—	𐭆	𐭇	𐭈	𐭉	𐭊
𐭀 𐭀	𐭁 𐭁	𐭂 𐭂	𐭃	𐭄 𐭄	—	𐭆 𐭆	𐭇	𐭈	𐭉	𐭊
𐭀 𐭀	𐭁 𐭁	𐭂 𐭂 𐭂	𐭃	𐭄 𐭄	—	𐭆	𐭇	𐭈	𐭉	𐭊
𐭀	𐭁	𐭂	𐭃	𐭄	—	𐭆	𐭇	𐭈	𐭉	𐭊
𐭀	𐭁 𐭁	𐭂 𐭂	𐭃	𐭄	—	𐭆	𐭇	𐭈	𐭉	𐭊

3 Summary of Graphemes

The Book Pahlavi letters may be categorized as those that have ‘simple’ or ‘complex’ graphemes. A summary of the distinctive graphemes for each letter, as shown in the above tables, is provided below:

Simple Letters		Complex Letters	
<i>beth</i>	𐭪	<i>aleph-heth</i>	𐭠, 𐭡, 𐭢
‘old’ <i>daleth</i>	𐭣	<i>a-y lig</i>	𐭤
<i>waw-nun-ayin-res</i>	𐭥, 𐭦	<i>gimel-daleth-yodh</i>	𐭧, 𐭨
<i>zayin</i>	𐭩, 𐭪, 𐭫	‘curled’ <i>gimel-daleth-yodh</i>	𐭬, 𐭭, 𐭮
<i>kaph</i>	𐭯	<i>he</i>	(𐭰 + 𐭱) 𐭲
‘old’ <i>kaph</i>	𐭳	<i>samekh</i>	(𐭴 + 𐭵) 𐭶
final ‘old’ <i>kaph</i>	𐭷	‘tall’ <i>samekh</i>	(𐭴 + 𐭵) 𐭸
<i>lamedh</i>	𐭹, 𐭺, 𐭻	‘Indian’ <i>samekh</i>	(𐭴 + 𐭵) 𐭹
‘stroked’ <i>lamedh</i>	𐭼, 𐭽, 𐭾	<i>pe</i>	𐭿, 𐮀
‘looped’ <i>lamedh</i>	𐭿, 𐮀, 𐮁	<i>sadhe</i>	𐮁, 𐮂
‘hooked’ <i>lamedh</i>	𐮃	<i>shin</i>	𐮄, 𐮅, 𐮆
‘old’ <i>lamedh</i>	𐮇	‘curled’ <i>shin</i>	𐮈
<i>mem-qoph</i>	𐮉		
<i>taw</i>	𐮋		
<i>X1</i>	𐮍		
<i>X2</i>	𐮏		
<i>y-h ligature</i>	𐮑		

‘Simple’ letters have graphical identities that are largely preserved while joining, despite elongation or shortening. ‘Complex’ letters may resemble sequences of other letters, or may be obscured when joined with certain adjacent letters, resulting in ambiguous representations.

The above graphemes could be used to reproduce Book Pahlavi text from transliterated sources. If a user understand basic rules of the script, these characters could also be used unambiguously to reproduce a Book Pahlavi document as encoded text. The accuracy of the latter, however, is dependently on the care and precision with which a source was written or printed.

4 Complexities of the script

The shaping behavior of Book Pahlavi letters presents several challenges for developing an encoding model. Some palaeographically distinct letters resemble shaped sequences of other letters:

𐭠	<i>aleph-heth</i>	𐭠 + 𐭠	<i>daleth-gimel-yodh</i> + <i>daleth-gimel-yodh</i>
𐭡	<i>he</i>	𐭠 + 𐭡	<i>mem-qoph</i> + height-adjusted <i>waw-nun-ayin-resh</i>
𐭢	<i>samekh</i>	𐭠 + 𐭢	‘curled’ <i>daleth-gimel-yodh</i> + ‘curled’ <i>daleth-gimel-yodh</i>
𐭣	‘tall’ <i>samekh</i>	𐭠 + 𐭣	descending ‘curled’ <i>daleth-gimel-yodh</i> + descending ‘curled’ <i>daleth-gimel-yodh</i>
𐭤	‘Indian’ <i>samekh</i>	𐭠 + 𐭤	descending ‘curled’ <i>daleth-gimel-yodh</i> + <i>daleth-gimel-yodh</i>
𐭥	‘curled’ <i>shin</i>	𐭠 + 𐭥	descending ‘curled’ <i>daleth-gimel-yodh</i> + <i>aleph-heth</i>

Some palaeographical letters have distinctive forms in some sources, while in other sources they may resemble shaped sequences of other letters:

‘curled’ <i>shin</i>	𐭥	𐭠 + 𐭥	descending ‘curled’ <i>daleth-gimel-yodh</i> + <i>aleph-heth</i>
	𐭦	𐭠 + 𐭦	short descending ‘curled’ <i>daleth-gimel-yodh</i> + back-sloping <i>aleph-heth</i>

Some letter combinations have alternate, multiple valid representations, which may occur concurrently. Some forms are used in preserved spellings. Usage of a particular form generally cannot be predicted.

<i>aleph</i> + <i>gimel-daleth-yodh</i>	𐭠𐭡, 𐭠𐭢, 𐭠𐭣, 𐭠𐭤, 𐭠𐭥 (medial)
<i>aleph</i> + <i>shin</i>	𐭠𐭥, 𐭠𐭦
<i>aleph</i> + <i>pe</i>	𐭠𐭧, 𐭠𐭨, 𐭠𐭩
<i>gimel-daleth-yodh</i> + <i>aleph-heth</i>	𐭡𐭠, 𐭢𐭠
<i>gimel-daleth-yodh</i> + <i>kaph</i>	𐭡𐭠, 𐭢𐭠, 𐭣𐭠
<i>gimel-daleth-yodh</i> + <i>pe</i>	𐭡𐭧, 𐭢𐭧, 𐭣𐭧
<i>zayin</i> + <i>aleph</i>	𐭠𐭫, 𐭠𐭬
<i>zayin</i> + <i>kaph</i>	𐭠𐭫, 𐭠𐭬
<i>lamedh</i> + <i>aleph</i>	𐭠𐭭, 𐭠𐭮

<i>lamedh + pe</i>	𐭪, 𐭫
<i>lamedh + sadhe</i>	𐭬, 𐭫
<i>mem-qoph + waw-nun-ayin-resh</i>	𐭮, 𐭯
<i>samekh + pe</i>	𐭰, 𐭱
<i>shin + pe</i>	𐭲, 𐭳
<i>shin + sadhe</i>	𐭴, 𐭳

Some letter sequences may be analyzed in multiple ways:

𐭮	𐭪 + 𐭮	<i>aleph-heth + daleth-gimel-yodh</i>
	𐭮 + 𐭪	<i>daleth-gimel-yodh + aleph</i>
𐭪	𐭪 + 𐭮	<i>aleph + pe</i>
	𐭬 + 𐭮	<i>aleph + sadhe</i>
𐭰	𐭰 + 𐭱	<i>'curled' daleth-gimel-yodh + samekth</i>
	𐭪 + 𐭰	<i>samekh + medial daleth-gimel-yodh</i>
𐭪	𐭪 + 𐭪	<i>daleth-gimel-yodh + pe</i>
	𐭬 + 𐭪	<i>daleth-gimel-yodh + sadhe</i>
𐭲	𐭪 + 𐭲	<i>shin + waw-nun-ayin-rest</i>
	𐭪 + 𐭮 + 𐭱	<i>daleth-gimel-yodh + aleph + waw-nun-ayin-rest</i>
𐭰	𐭪 + 𐭰	<i>samekh + pe</i>
	𐭪 + 𐭱 + 𐭱	<i>descending 'curled' daleth-gimel-yodh + descending 'curled' daleth-gimel-yodh + pe</i>
	𐭬 + 𐭰	<i>samekh + sadhe</i>
	𐭬 + 𐭱 + 𐭱	<i>daleth-gimel-yodh + daleth-gimel-yodh + sadhe</i>
𐭲	𐭪 + 𐭲	<i>shin + pe</i>
	𐭬 + 𐭲	<i>shin + sadhe</i>

5 Challenges for Defining the Encoding Model

In a typical ‘palaeographic’ encoding model for cursive joining scripts, all of the distinctive letters of a script would be encoded as characters, while the contextual forms would be handled at the font level. While many Book Pahlavi letters have straight-forward rules for joining, which are predictable, there are a handful of letters that have multiple representations when they are adjacent to certain other letter. For such pairs, the joining behavior cannot be predicted. The different representations of these letter pairs are found in historical spellings that are preserved in large numbers across the corpus. Accordingly, a suitable encoding model would be hybrid of the ‘palaeographical’ approach, and ‘graphemic’ and ‘graphetic’ systems. This model would provide character repertoire that includes some palaeographical letters, the contextual forms of some letters, and graphical primitives.

Of the various encoding models proposed for Book Pahlavi over the years, the latest document is “Teeth and bellies: a proposed model for encoding Book Pahlavi” (L2/20-246) by Roozbeh Pournader. It recommends using graphical elements to construct certain letters and letterforms instead of relying solely on shaping. Pournader’s model offers the most practical way forward. It eliminates ambiguities of character identity by encoding some contextual forms and by decomposing some letters into character primitives. The intent is to representing the overall graphical structure of Book Pahlavi text, rather than adhere to a one-to-one relationship between palaeographical letters and rendered text.

Among Pournader’s set of primitives are the characters:

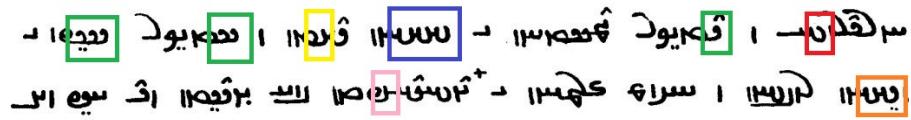
- 𐭪 ‘tooth’
- 𐭫 ‘curled tooth’
- 𐭬 ‘belly’
- 𐭭 ‘curled belly’

These characters would be used, for example, as follows:

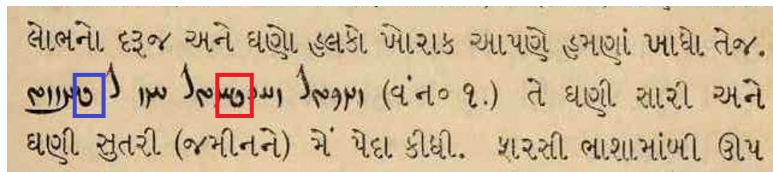
𐭪	𐭪 + 𐭪
𐭫	𐭪 + 𐭪 + 𐭪
𐭬𐭬𐭬𐭬	𐭬 + 𐭬 + 𐭬 + 𐭬 + 𐭬 + 𐭬 + 𐭬 + 𐭬
𐭬𐭭	𐭬 + 𐭬 + 𐭬 + 𐭬 + 𐭬 + 𐭭

This certainly provides for representation of the Book Pahlavi text in the plainest way possible. It follows the approach taken by Henrik Samuel Nyberg in his *A Manual of Pahlavi* (Otto Harassowitz: 1964). Nyberg provides a highly normalized method of writing Book Pahlavi that avoids calligraphic features found in manuscripts and printed materials. Letters with ‘teeth’ are written in a homogenous fashion:

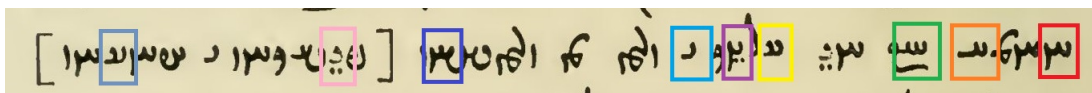
However, Nyberg takes care to differentiate the shapes of ‘bellies’, as shown in the example below. He draws distinctions between the belly of *shin* (outlined in red), the belly of *aleph-heth* (outlined in orange), and the belly of *pe* (outlined in pink). Also compare the shapes of the bellies of *shin* followed by *aleph-heth*, outlined in navy blue. These variations in the shapes of bellies are essential to the identity of letters.



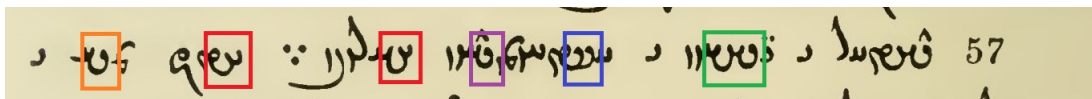
These distinctions, such as curvatures, are not whimsical; they are inherent aspects of the letterforms and carry semantic value. Note the similarity, but distinction between the size of the curled belly of the *shin* (red) and that of initial *gimel-daleth-yodh* (blue) in *A Grammar of the Pahlvi Language* (Bombay: 1871) by the *darab dastur* Peshotan Sanjana:



Likewise, not all ‘teeth’ are the same, as shown in the *Zand i javit sheda dad*, printed by Sanjana (Bombay: 1895), which contains *fargard* 1–9, 16 of the *Vendidad*. Note the distinctive representations of a sequence of two teeth. Compare initial *aleph-heth* (red) to its isolated / final form (orange), and compare these forms to the mediam *shin* in navy blue. Also, compare the form of initial *aleph-heth* in red with its shape in green. The form of initial *gimel-daleth-yodh* (blue gray, purple) also differs from initial *aleph-heth* (red).



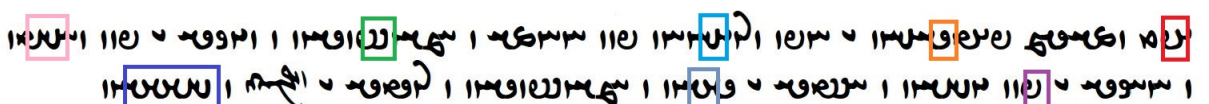
Similarly, not all ‘bellies’ are alike. Compare the belly of initial *aleph-heth* (red) with the shape of initial *shin* (green):



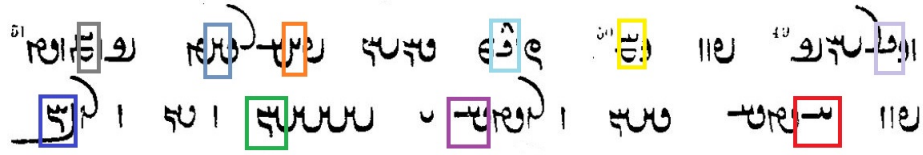
Similar distinctions between varieties of ‘teeth’ and ‘bellies’ are made by Prods Oktor Skjærvø in his *Introduction to Pahlavi* (Cambridge, MA: 2007): The ‘teeth’ of certain letters different from those of others. Compare the shape of the teeth of *shin* (green, sky blue) with *aleph-heth* (navy blue). Also note the differences in shape of intial *aleph-heth* (red) and the final form (light gray). Also compare the belly of *pe* (purple) with the others.



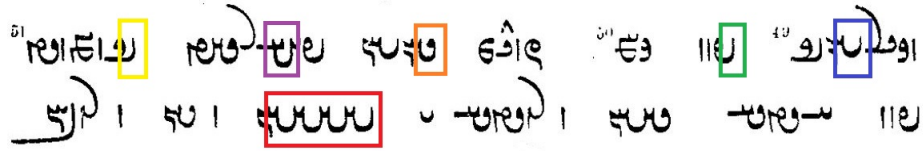
Skjærvø expresses the identity of letters by consistent usage of various ‘bellies’. Compare the shape of the belly of *shin* (pink) with *aleph-heth* (red). Also note the differences in shape of intial *aleph-heth* (red) and the final form (light gray).



The same distinctions are seen in the *Corpus of Pahlavi Texts*, edited by Jamaspji Dastur Minocherji Jamasp-Asana (Bombay: 1913), which was printed using an elementary metal type. Here, also, different representations of ‘teeth’ are illustrated:



The distinction between letters is maintained by consistent usage of different ‘bellies’:



Based on the examples shown above, it is evident that a model based on ‘teeth’ and ‘bellies’ is certainly viable. But, Pournader’s repertoire of ‘tooth’, ‘curled tooth’, ‘belly’, ‘curled belly’ needs to be extended to account for the semantic distinctions conveyed by variations of these shapes as attested across the corpus.

6 Proposed Encoding

The proposed model for Book Pahlavi is based encoding the following as distinctive characters:

Palaeographical letters		Teeth components	
<i>beth</i>	𐭪 r	<i>tooth</i>	𐭪 d
‘old’ <i>daleth</i>	𐭫 r	<i>curled tooth</i>	𐭫 d
<i>waw-nun-ayin-resh</i>	𐭬 r	<i>double tooth</i>	𐭬 d
<i>zayin</i>	𐭮 d	<i>triple tooth</i>	𐭮 d
<i>kaph</i>	𐭯 r		
‘old’ <i>kaph</i>	𐭰 r		
final ‘old’ <i>kaph</i>	𐭱 n		
<i>lamedh</i>	𐭲 d		
‘stroked’ <i>lamedh</i>	𐭳 d		
‘looped’ <i>lamedh</i>	𐭴 d		
‘hooked’ <i>lamedh</i>	𐭵 d		
‘old’ <i>lamedh</i>	𐭶 r		
<i>mem-qoph</i>	𐭷 d		
<i>pe</i>	𐭸 r		
<i>sadhe</i>	𐭹 r		
‘curled’ <i>shin</i>	𐭺 d		
<i>taw</i>	𐭻 r		
X_1	𐭼 r		
X_2	𐭽 r		
<i>g-d-y + a-h ligature</i>	𐭾 r		
		Belly components	
		<i>belly</i>	𐭿 d
		<i>half belly</i>	𐮀 d
		<i>straight belly</i>	𐮁 d
		<i>half straight belly</i>	𐮂 d
		<i>curled belly</i>	𐮃 d
		<i>half curled belly</i>	𐮄 d
		Contextual forms	
		short <i>waw-nun-ayin-resh</i>	𐮅 r
		<i>zayin</i> with half belly	𐮆 d
		<i>zayin</i> with belly	𐮇 d
		<i>lamedh</i> with half belly	𐮈 d
		<i>lamedh</i> with belly	𐮉 d
		‘stroked’ <i>lamedh</i> with half belly	𐮊 d
		‘stroked’ <i>lamedh</i> with belly	𐮋 d
		‘looped’ <i>lamedh</i> with half belly	𐮌 d
		‘looped’ <i>lamedh</i> with belly	𐮍 d
		final <i>pe-sadhe</i>	𐮎 r

The repertoire is based upon the following principles:

- palaeographic letters that are completely distinctive and cannot be confused with other letters or sequences of letters are encoded as independent characters
- any palaeographical letter that is identical to a rendered sequence of letters is decomposed into graphical primitives
- contextual variant forms of a given letter or primitive that are used for enabling connections with different letters are included as separate characters
- atomic ligatures are encoded as independent characters

In this model:

- The following letters are not independently encoded and are to be represented using a sequence of other characters:

<i>he</i>	𐭪	𐭪 + 𐭪
<i>samekh</i>	𐭫	𐭫 + 𐭫
‘tall’ <i>samekh</i>	𐭬	𐭫 + 𐭫
‘Indian’ <i>samekh</i>	𐭭	𐭫 + 𐭫
<i>shin</i>	𐭮	𐭮 + 𐭮

- The ‘tooth’ allographs 𐭫 and 𐭬, used for representing *gimel-daleth-yodh* in various contexts, are encoded as separate characters as they occur concurrently.
- In addition to 𐭫, the 𐭮 and 𐭯 are encoded as separate characters. Details on the latter two are provided below. The rule for usage of these letters is as follows: if a string contains two ‘teeth’, use 𐭮 instead of 𐭫 + 𐭫; if a string contains three ‘teeth’ use 𐭯 instead of 𐭫 + 𐭫 + 𐭫.
- The 𐭮 is encoded as a separate character in order to enable typographical support for different representations of *aleph-heth* in initial, medial, and final position.
- The ligature 𐭯 of *aleph-heth* + *gimel-daleth-yodh* is encoded as an atomic character in order to enable typographical support for different representations of it, as compared to 𐭮.
- Letters whose contextual forms consist of elongation of the body are also encoded separately, eg. 𐭪 and 𐭫 for writing 𐭪 when it occurs before certain letters: both 𐭮 and 𐭯 for 𐭪 *lamedh* + 𐭮 *aleph-heth*; and both 𐭮 and 𐭯 for 𐭮 *zayin* + 𐭮 *aleph-heth*

7 Acknowledgments

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. It was also 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).