Teeth and bellies: a proposed model for encoding Book Pahlavi

Roozbeh Pournader (WhatsApp)
September 7, 2020

Background

In Everson 2002, a proposal was made to encode a unified Avestan and Pahlavi script in the Unicode Standard. The proposal went through several iterations, eventually leading to a separate encoding of Avestan as proposed by Everson and Pournader 2007a, in which Pahlavi was considered non-unifiable with Avestan due to its cursive joining property. The non-cursive Inscriptional Pahlavi (Everson and Pournader 2007b) and the cursive Psalter Pahlavi (Everson and Pournader 2011) were later encoded too. But Book Pahlavi, despite several attempts (see the Book Pahlavi Topical Document list at https://unicode.org/L2/topical/bookpahlavi/), remains unencoded.

Everson 2002 is peculiar among earlier proposals by proposing six Pahlavi *archigraphemes*, including an *ear*, an *elbow*, and a *belly*. I remember from conversations with Michael Everson that he intended these to be used for cases when a scribe was just copying some text without understanding the underlying letters, considering the complexity of the script and the loss of some of its nuances to later scribes. They could also be used when modern scholars wanted to represent a manuscript as written, without needing to over-analyze potentially controversial readings.

Meyers 2014 takes such a graphical model to an extreme, trying to encode pieces of the writing system, most of which have some correspondence to letters, but with occasional partial letters (e.g. PARTIAL SHIN and FINAL SADHE-PARTIAL PE). Unfortunately, their proposal rejects joining properties for Book Pahlavi and insists that "[t]he joining behaviour of the final stems of the characters in Book Pahlavi is more similar to cursive variants of Latin than to Arabic". This is despite their discussion of "Joining side[s]" in their Table 2.1 (p. 11). As such, the proposal was not acceptable, as it was clear to Unicode's experts that the script would benefit from Arabic-like joining properties.
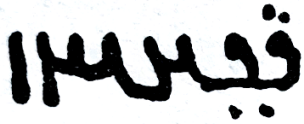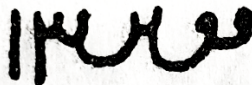
Reading the latest Book Pahlavi proposal, Pandey 2018, especially its sections 5.1 as well as the sections discussing proposed fixed-form characters, 6.2 and 6.4.2, and digging deeper into the material presented about the nuances of the script by Nyberg 1964, Amoozgar and Tafazzoli 1996, and Skjærvø 2008 led me to understanding that the same irregularities that Pandey has tried to fix by introducing fixed form letters are a key part of the writing system. The Book Pahlavi writing system is very graphical. Nyberg, for example, refers to "[t]he intricate process by which the Iranian scribes transformed Aramaic forms into *purely graphic signs* […]" (emphasis mine) when discussing heterograms (the quote is from Nyberg et al 1988.

1

Modern experts, when trying to reproduce Book Pahlavi manuscripts or discuss words, appear to care about two aspects a lot: if a belly is formed, and if there is a curl. The first distinction tends to make or break a word. Although there are some common patterns that can be discovered about when a belly is formed and when it's not (as described in Pandey 2018), there are irregularities and exceptions that tell me we should avoid over-analyzing the script for discovering them.

The formation of a belly appears to be a spelling convention that has some rules, as well as some exceptions. Instead of trying to hard-code those spelling conventions in a font or text shaping system, we should come up with different characters for each *element of writing*. Certain sequences of these elements may be common, and certain sequences may be rare or non-existent, but allowing them would let the modern scholar have the ability to choose the form they need, instead of their hand being forced by the font or restrictions of the typesetting software.

Teeth and bellies

Let's start with an example. The word *gēhān* <gyh'n'> according to my four major sources, as well as Pandey 2018, is written as follows:

Nyberg 1964 (with disambiguating dots, sorted under *gimel-daleth-yodh*)	
MacKenzie 1986 (sorted under *samekh*)	
Amoozgar and Tafazzoli 1996 (sorted under *samekh*)	
Skjærvø 2008	
Pandey 2018 (missing a final stroke)	

If someone sees the form in, say, MacKenzie 1986 or Skjærvø 2008 out of context, they will not know if the first two strokes are two *gimel-daleth-yodh*s, or a single *samekh*. MacKenzie 1986 and Amoozgar and Tafazzoli 1996 acknowledge this by even putting the word in their word list under *samekh*. Further inside the word, it's just teeth and bellies: somebody familiar with the word would know that there are two letters here: the first straight tooth and the first belly representing the first *aleph-heth*, and the final two straight teeth representing another *aleph-heth*. But they could theoretically be anything else: for example,

the belly and the two straight teeth after it may have represented *shin*. There's basically no way of knowing by looking at the spelling: one needs to check the transliterations provided by experts.

For this specific word, what is clear from the shapes is that the number of teeth and the existence of curls doesn't change. What is unclear is if the right-most *gimel* has a belly or not. It appears to form a belly in Nyberg and MacKenzie, although not in Amoozgar and Tafazzoli. In fact, Amoozgar and Tafazzoli may be listing a slightly different spelling.

The issue may be further glanced from the letter pair charts in Nyberg 1964 (p. 133), as well as Amoozgar and Tafazzoli 1996 (p. 56). Here's what they list for a sequence of two *gimel-daleth-yodh*s:

Nyberg   Amoozgar and Tafazzoli

I read this to mean the first letter may form a belly or not, and may lose its curly head, but probably not both (which would make the combination look too much like the beginning of *shin*). I assume the second letter's shape is not set in stone and may still get affected by the letter or letters that come after it. But all this does *not* appear to be completely arbitrary. We probably can't replace one of the three forms for another and have the same word.

Another way to look at this is that there is absolutely no way to distinguish an *aleph-het* or a *samekh* from a sequence of two *gimel-daleth-yodh*s without knowing the word. This may have been fine if any of these confusable cases were rare. But they are some of the most common letters and sequences in Book Pahlavi!

There are some patterns in the orthography, of course. For example, a careful analysis of word lists appears to indicate that two curl-less bellies cannot immediately follow each other. But looking deeper, it appears that manuscripts indeed contain them. Here are two examples from Skjærvø 2008, p. 9:

From these and several other examples I have concluded that the best model for encoding Book Pahlavi may be closer to Meyers 2014 than we had expected. It should be noted that neither Meyers's nor Pandey's model are able to represent the two Skjærvø examples above. But a simpler *teeth and bellies* model would.

The fundamental elements of the model I propose are four basic characters, which Everson 2002 called archigraphemes (these names are not formal character names, those can be decided later):

Tooth	Curled Tooth	Belly	Curled Belly
ل	د	ں	ں
- *g/d/y* - 1st or 2nd half of *a/h* - 2nd or 3rd thirds of *š* - 2nd half of bellied form of *s*	- *g/d/y* - 1st or 2nd half of non-bellied form of *s*	- bellied form of *g/d/y* - 2nd half of bellied form of *a/h* - 1st third of *š* in Iranian MSS - the tooth and belly before final *p*	- bellied form of *g/d/y* - 1st half of bellied form of *s* - 1st third of *š* in Indian MSS

All the characters would be dual-joining. Teeth appear in all positional forms, while bellies tend to appear only in initial and medial forms: they always tend to be followed by another character. The ezafe sign would be represented by the isolated form of <tooth>, whose stem would be elongated a little in final and isolated forms.

Here's how the letters made of teeth and bellies are represented in this model:
- *gimel-daleth-yeh*: <tooth>, <curled tooth>, <belly>, or <curled belly>
- *aleph-heth*: <tooth, tooth> or <tooth, belly>
- *samekh*: <curled tooth, curled tooth>, <curled tooth, curled belly>, <curled belly, tooth>, or <curled belly, curled belly>
- *shin*: <belly, tooth, tooth> in Iranian manuscripts, <curled belly, tooth, tooth> in Indian manuscripts

Here's a proposed encoding for the four slightly different spellings of *gēhān*. Note that this level of distinction may be unnecessary or undesired for this specific word, but it comes in handy for other words:

Nyberg 1964 — **curled belly**, two dots above, curled belly, two dots below, tooth, belly, tooth, tooth, *w/n/r*, *w/n/r*

MacKenzie 1986 — **curled belly**, curled belly, tooth, belly, tooth, tooth, *w/n/r*, *w/n/r*

Amoozgar and Tafazzoli 1996 — **curled tooth**, curled belly, tooth, belly, tooth, tooth, *w/n/r*, *w/n/r*

Skjærvø 2008 — **curled belly**, curled belly, tooth, belly, tooth, tooth, *w/n/r*, *w/n/r*

Similarly, the example given in Pandey 2018, pp. 8–9, ꮄ <š'h'n> *šāhān*, would be represented as <belly, tooth, belly, tooth, belly, tooth, belly, tooth, tooth, *waw-nun-ayin-resh*> with no need to analyze its letters.

Note that the experts' conceptualizations don't necessarily completely agree with each other or use all our characters. For example, in all of Skjærvø 2008's typeset Book Pahlavi examples there doesn't seem to be any curled tooth. This may be a limitation of his font, since in some of the examples included from manuscripts, such a distinction appears to exist (or it may be the case that Skjærvø doesn't believe the distinction is important):

In the above text, from Bundahišn (Skjærvø 2008, p. 152), the blue ovals indicate two normal teeth, while the red oval indicates a curled tooth.

On the contrary, Nyberg 1964 clearly has a distinction between curled and straight teeth in his conceptualization. For example, here's the word <LYLY'-1> from its page 1:

The character sequence representing the above word in our proposed model would be <*lamedh*, curled tooth, *lamedh,* curled belly, tooth, tooth, *beth*/1>. For the sake of comparison, here is the same word (minus the <-*beth*/1> suffix) in the other three sources:

MacKenzie 1986		*lamedh*, **tooth**, *lamedh,* belly/curled belly?, tooth, tooth
Amoozgar and Tafazzoli 1996		*lamedh*, **curled tooth**, *lamedh*, curled belly, tooth, tooth
Skjærvø 2008		*lamedh*, **tooth**, *lamedh*, curled belly, tooth, tooth

In MacKenzie 1986 and Amoozgar and Tafazzoli 1996 it's sometimes the line between curled tooth and curled belly that is not very clear. Considering that both these sources are handwritten, I would except that a choice would have been made every time if they were committing it to type.

As a side note, certain nuances of the writing system may also become representable if we choose this model. For example, a three-way distinction of the various shapes of the pair <*gimel-daleth-yeh*, *kaph*>, as seen in Nyberg 1964 (p. 132) and Amoozgar and Tafazzoli 1996 (p. 55) can be represented as follows:

  <curled tooth, kaph>

<tooth, kaph>

<belly, kaph>

It should be noted that the experts already make similar choices to our model in collating Book Pahlavi words, as they are forced to by the nature of the script. For example, here is what Nyberg 1964 says at the beginning of its word list (Amoozgar and Tafazzoli 1996 contains a structurally similar introduction to their word list):

Pahlavi Index.

Alphabetical order:

1. ‏ـں‎ 2. ‏ڔ‎ 3. ‏ڡ‎ 4. ‏ا‎ 5. ‏ک‎ 6. ‏ ‎ 7. ‏ٮ‎ 8. ‏ڡ‎ 9. ‏ں‎ ‏ٮ‎ 10. ‏ٯ‎ 11. ‏ٯ‎ 12. ‏ـں‎ 13. ‏م‎

To be observed:

1. ‏ـں‎ is ranged under no.1 even if it represents the ligature of ‏ڡ+ڡ‎ (v. p. ۱۳٤, C).

2. Abnormal forms of letters are ranged according to their external shape: ‏ک‎ under ‏ک‎ even if it is a miswritten ‏ا‎ or ‏ڡ‎; ‏ڡ‎ under ‏ڡ‎ even if it represents a dwarfed ‏ا‎, or ‏ک‎, or ‏ ‎; ‏ا‎ under ‏ا‎ even if it represents ‏ ‎= d (v. p. ۱۳۱).

3. ‏ڡ‎ is treated as one single letter irrespective of whether it represents ‏ ‎= ڡ̈, or d = ڡ̂, or y = ڡ̣, or j- = ڡ̊ , or any of the dwarfed letters mentioned above.

4. ‏ٯ‎°, representing both ‏ٯ‎ and ‏ٯ‎ in ligatures, is treated as one single letter, and ranged under ‏ٯ‎.

5. Final silent ‏ا‎ (v. p. ۱۳۱, no.7) is treated as the consonant ‏ا‎.

6. Final ‏ڡ‎°= -H in ideograms is treated as ‏ڡ‎+‏ا‎.

7. ‏ں‎ and ‏ٮ‎, both= s, are regarded as the same letter. As diacritical signs are used when they represent the ligatures of ‏ڡ+ڡ‎ (ڡ̂, ڡ̣, etc.) every unmarked ‏ں‎ or ‏ٮ‎ in Iranian words is to be looked for under no.9. If they occur in ideograms, which are mostly left without diacritical signs in this Manual (being themselves merely graphical signs, and no living elements of the language), cross references are given when necessary.

The teeth and bellies take care of *aleph-het*, *gimel-daleth-yodh*, *samekh*, and *shin*. Most other letters are straight-forward and can mostly follow the model of Pournader 2013 and Pandey 2018, except for *pe* and *sadhe* which share a final form.

Pe and *sadhe*

To represent *pe* and *sadhe*, we can:

1. Limit *sadhe* to a non-joining form, only to be used when not joining to the previous letter.
2. Make *pe* right joining with no extra teeth in its final form. These would be the final and isolated forms of pe:

<div align="center">

isolated   final

ﻊﻋ

</div>
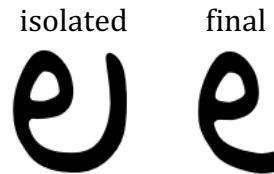
When there is an extra tooth or belly before a final *pe*, an explicit tooth or belly character would be used to represent the sequence. But in words where *pe* is used disconnected from its previous letter, no tooth or belly character would be used. The sequence <belly, *pe*> where the belly would be in initial positional form, should be avoided.

This is not the only model imaginable for *pe* and *sadhe*. I can imagine two other models, for a total of three:

A) The model presented above, with *pe* right-joining and *sadhe* non-joining. This has the problem of <initial belly, final *pe*> being indistinguishable from <isolated *pe*>.

B) Keeping *pe* right-joining and *sadhe* non-joining, but defining no isolated form for *pe*, requiring <belly, *pe*> to be used for an isolated *pe*. This is somehow similar to what Meyers 2014 proposes. This has the problem of *pe* becoming a final-only character, as well as a counter-intuitive requirement for entering isolated *pe*. But it avoids different ways to represent some words.

C) Making both *pe* and *sadhe* right-joining, with final *pe* having an extra tooth/belly compared to the final *sadhe*. This is close to the perception of users of the writing system. In this model, when there is no extra tooth/belly, *sadhe* would be used in the spelling, even if the letter *pe* is pronounced. This has the problem of <belly, final *sadhe*> being indistinguishable from <final *pe*>.

Here is an excerpt from the letter pair table in Amoozgar and Tafazzoli 1996 with proposed character sequences per models A and B above:

		sadhe	*pe*
ل	৭	sadhe	pe (model A) *belly*, pe (model B)
৬১	৬৬	tooth, belly, pe	tooth, tooth, belly, pe tooth, belly, pe
৬৬	৬৬	curled belly, pe	curled tooth, belly, pe curled belly, pe
৬৫	৬৫	zayin, pe	zayin, belly, pe zayin, pe
৬১	৬১	lamedh, pe	lamedh, belly, pe lamedh, pe
৬৬	৬৬	mem-qoph, pe	mem-qoph, belly, pe
৬৬	৬৬	curled tooth, *curled belly*, pe	curled tooth, curled tooth, *belly*, pe *curled belly, curled belly*, pe
৬৬	৬৬	belly, tooth, belly, pe	belly, tooth, tooth, belly, pe belly, tooth, belly, pe

Here's the same table from Nyberg 1964 per models A and B above, with slightly different spellings (italicized):

		sadhe	*pe*
?	ﻉ	sadhe	pe (model A) *tooth*, pe (model B)
ﻉﯩ	ﻉﯨﯿﺷ }	tooth, belly, pe	tooth, tooth, belly, pe tooth, belly, pe
ﻉﻮ	ﻉﻮﯿﻮ }	curled belly, pe	curled tooth, belly, pe curled belly, pe
ﻉﺲ	ﻉﯩﺲ }	zayin, pe	zayin, belly, pe zayin, pe
ﻩ	ﻉﻮﻮ }	lamedh, pe	lamedh, belly, pe lamedh, pe
ﻉﻟ	ﻉﻟ	mem-qoph, pe	mem-qoph, belly, pe
ﻉﻮﻮ	ﻉﻮﻮﻮﻮ }	curled tooth, *curled tooth*, pe	curled tooth, curled tooth, *tooth*, pe *curled tooth, curled tooth*, pe
ﻉﻮﯾ	ﻉﯩﯿﺷﺷ }	belly, tooth, belly, pe	belly, tooth, tooth, belly, pe belly, tooth, belly, pe

The examples from Skjærvø 2008 can now be represented too:

ﻉ + ﺍ > ﻉﯩﯾ , ﻮﯾ + ﺍ > ﺍﯩﯾﯾ

The left one is <tooth, belly, belly, pe> (model A or B), while the right one is <tooth, belly, belly, tooth, tooth>.

Other letters

1. All sources agree that the final stroke is indistinguishable from *waw-nun-ayin-resh*. They should be unified as one character.
2. All sources appear to agree that the letter *he* ⅁, only used in heterograms, is indistinguishable from the sequence *<mem-qoph, waw-nun-ayin-resh>*. That sequence should be used to represent *he*. Its "graphical side-form", as explained below (Nyberg 1964, p. 130), should be represented as *<mem-qoph, mem-qoph, waw-nun-ayin-resh>*:

Note. ₆₆° is a graphical side-form of ₆° = -H, cfr. ₆₆₁₆ = LPNH, inscr. ₆₆₁₀₆ .

3. The hooked *lamedh* ↳ and the old lamedh ⟍, as proposed by Pandey 2018, both appear to be used only in heterograms. They probably are glyph alternatives. But for the sake of more accurate representation of Book Pahlavi texts and helping the scholarly community, they should both be encoded. (Pandey 2018, p. 26, claims that "they occur concurrently". It would be great to see evidence of that, which would provide further justification for encoding both). Here are two examples where they appear on the same page in Skjærvø 2008 (p. 128) and Nyberg 1964 (p. 131):

[6] ⟍ is both *l* and *r*, but mostly *r*, replacing the old letter *r* (2 = L = 1) which had become too ambiguous. If it denotes *l* it may receive, as a diacritical sign, a stroke: ⟍ (in Iranian MSS), or a loop: ⟍ (in Indian MSS), or be written twice: ⟍, ⟍. The old form ↳ still occurs in ideograms: ↳ᵤ = ʾL = mā; ↳ᵤₙ = ʾHL (אחר) = pas; ↳ᵤG = MHL (מחר) = fratāk; ↳ₐS = ZKL (זכר) = nar. Also ⟍ : ⟍ᵤ .

4. As we are approaching a more graphetical encoding model, we should probably encode a separate looped lamedh ⟍ to represent [l] in Indian manuscripts too.
5. Nyberg 1964 p. 135, talks about a hooked *mem-qoph.* It should be encoded. We need to find out if the form only appears at end of words or can be used medially:

F. ₆° as the ending of the 1st pers. sing. of verbs is sometimes written with a hook beneath: ₆° . This hook can only be a small *w* of the type found in the Ps. (L): ₆° = -wm = -om.

6. It appears that the ligature x_2, typically used at the end of words, can connect to a letter after it. So, depending on if it can connect to its previous letter or not, the character should be encoded as either dual-joining or left-joining. Here's the example from Nyberg 1964, p. 136:

Note 1. If ℰ° means -ēnd the scribes often write ẋℰ°

It should also be noted that the character can take a combining hat above, as seen in Skjærvø 2008, p. 103:

The symbols <x_1> and <x_2>.

  Instead of the usual 3rd person singular and plural endings <-yt'> and <-ynd, -d>, we often find two symbols, ﻮ <x_1> and ﺮ ﺮ <x_2>. It is obvious that the two forms are from <-yt> and <-ynd>, respectively:

   ﻮ > ﻮ (*Psalms* ﻮ)   ﻮ > ﻮ and ﻮ > ﺮ (*Psalms* ﻮ)

7. Nyberg 1964, p. 134, says that *kaph* can occasionally join the following letter:

D. ɡ is sometimes connected with the following letter, especially ᴜ, e.g., in the plur. of words ending in ɡ°: ﺮ° =-īkān. In this case it may receive the dots of ة : ﺮ°. In fact, in Sasanian times it was pronounced -ɡ- (-ɣ-) in this position, thus -īkān as -īgān, later -īyān, written ﺮ° (v. above). It is sometimes also attached, in the form ة, to other letters: to ﻮ : ﻮ, ﻮ ; to ﻮ : ﻮ, ﻮ.

This needs to be investigated more and compared with manuscripts and other sources. We need to understand if the mid-word *kaph* in the suffix *-īkān* is just a curled belly or it is indeed a graphical *kaph*. If it's the latter, we would need to make *kaph* dual-joining instead of right-joining.

8. Nyberg 1964, p. 131, talks about an old form of *nun*:

Note 3. Instances of the old form of n: ᘯ, occur now and then in the MSS. The letter which in Madan's edition of the Dēnkart is falsely printed as an Arabic ẓ must be a ᘯ = 2 = ʟ = | = ᴡ or ʿain.

Skjærvø 2008, p. 104 seems to confirm this:

  Note also the 2nd singular ending ﻮ <-yd>, the diacritic on <d>, the archaic form of <n> in <YN-> similar to Psalter ʟ (inscriptions ﻮ), and the archaic form ﻮ <z> also found in old Pahlavi manuscripts. The form of <w> with the top curved left is also found in the oldest Pahlavi manuscripts.

A similar shape also appears in the Pahlavi alphabets listed at the end of two manuscripts of *Frahang i Pahlavīk*, appearing to correspond to Avestan *ń* (U+10B26 AVESTAN LETTER NYE). Here are the images from Nyberg et al 1988:

Alphabet of S₁ :

Alphabet of K :

This archaic *nun,* as well as other shapes mentioned in Skjærvø 2008, need to be investigated further, to see if additional characters need to be encoded.

9. On the other end of the spectrum are some forms found in the Pahlavi papyri. As there is very limited analysis of the Pahlavi papyri available to me, I can't make a conclusion either way if the script of the Pahlavi papyri should be unified with Book Pahlavi or not. If they are to be unified, we need to make decisions about certain form only found in the papyri, as mentioned by Nyberg 1964, p. 133:

Numbers

Book Pahlavi numbers have become unified with letters and are indeed sometimes indistinguishable from them. They have joining properties, and appear to be made of the same elements of writing:

- The number one, which looks identical to the letter *beth* and is also used at the end of words to represent the suffix *-ē*, should be represented by the letter *beth*.
- The numbers 2-9 should be represented by a sequence of teeth and <*beth*>s.
- The number 10 looks like the letter *kaph* as well as the old *daleth* (old *daleth* appears to be a better candidate). We need to investigate which one is the best candidate to unify it with. Note that the number 10 occasionally connects to the following character too, so the character used for it should be made left-joining or dual-joining. A combing hat above seems to be frequently used with the number.
- The number 20 should be represented by the letter *lamedh*.
- The number 40, 60, and 80 should be represented by a sequence of teeth and bellies.
- The numbers 30, 50, 70, and 90 should be represented by 20, 40, 60, or 80 followed by 10.
- The number 100 should be represented by <tooth/curly tooth, *lamedh*, *zayin*> or <*lamedh*, *zayin*> when the tooth is missing.
- The number 200, depending on the spelling, should be represented by <tooth, *beth*, *lamedh*, *zayin*> or <tooth, tooth, *lamedh*, *zayin*>

- The number 1000 should be represented by *<lamedh*, old *kaph>*.

Following is a list of numbers from Nyberg 1964, p. 173. Other sources contain a similar list of numbers, but are not as exhaustive as Nyberg:

	Parth. inscr.	Pers. inscr.	Psalter	Books
1		ܝ , ܝ		ܠ , ܠ
2	܁܁	܁܁	܁܁	܁܁
3	܁܁܁	܁܁܁	܁܁܁	܁܁܁
4	܁܁܁܁	܁܁܁܁	܁܁܁܁	܁܁܁܁
5			܁܁܁܁܁	܁܁܁܁܁
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
30				
40				
50				
60				
70				
80				
90				
100				
200				
500				
800				
1000				
millenium? 6000				

14

Rare combining marks

Five combining dot patterns in Book Pahlavi are well known and were proposed in Everson 2002 and Pournader 2013. Meyers 2014 brought our attention to three more combining marks we had previously missed. They are repeated in Pandey 2018 with some differences but no attestation. They appear to be used in an 1842 CE manuscript called MU 29 (see Jamasp Asa and Nawabi 1976 for a reproduction and Mazdapour 1999 and König 2008 for translation and analysis).

König 2008, which doesn't appear to be a complete translation, acknowledges in pp. 127–130 a combining dot above, used over *nun*, *pe*, *heth*, *daleth*, and *samekh*, as well as other combining marks used with other letters, but not a caron/hat below or three dots below. Zeini 2020 confirms the three-dots-below mark, which can be seen here used under *pe* or *sadhe* to represent or emphasize *p* or *č* sounds, similar to the Perso-Arabic script use of three dots below for پ and چ:

Mazdapour 1999

p. 124, footnote 11   p. 34, line -1

Jamasp Asa and Nawabi 1976

p. 5, line 2   p. 11, line 5

I believe this is enough evidence for encoding a combining dot above and a combining three dots below, and the MU 29 manuscript and its scholarly analysis justify encoding two characters to represent its orthography. Note that Joneidi 1981, in his Book Pahlavi-based writing system, which has found some amateur fan following, also uses a combining dot above. Some of his letters even coincide with the MU 29 orthography:

As for Meyers 2014's caron-below and Pandey 2018's hat-below, I agree with Zeini 2020's analysis that at least as far as MU 29 goes, it's just a quick way to write two dots below. Mazdapour 1999, pp. 34–35 discusses all the diacritics in MU 29. There, when discussing the shape, she talks about "a small crescent-shaped line" (خط هلالی کوچکی) that can go above or below letters. But almost every time it's mentioned, it is mentioned as an alternative to two dots. Here it is on page 34:

نامگُفته)؛ دو نقطه یا خط هلالی کوچکی در زیر حرف دنشانهٔ حرف «ی» (y) یا «ای» (ī/ē) یا
e / i) (مثلاً در ۱۵/۹: زکئیو ، nazdīk ، نزدیک؛ ۱۶/۸۹: ویلسپ، ērān، ایران، در «ایرانشهر»؛
۹/۹۰: بوم وینژس۱۱۱، Jud-kēšān ، جُـدکیشان، پیروان کیشهای دیگر؛ ۷/۹۱: بوفِسویس
Judāgīh ، جدایی، مفارقت؛ ۱۶/۹۱: ۱۱ایلپو ، u nērōg ، وَ نیرو)؛ هشت کوچکی بر بالای
حرف دنشانهٔ واج دال (d) ، و خط هلالی یا دو نقطه بر بالای آن نشانهٔ گاف (g) یا دال (d)
است (مـــانند ۵/۸: نوسرعازسرر، gōspandān ، گـــوسفندان؛ ۶/۸: ثریثا : da[hi]šn ، دهش،
آفرینش(نوسرورد ، gōspand ، گوسفند؛ ۱۴/۸: نسری، gāh ، گاه، هنگام و زمان).

And once again on page 35:

امّا نشانهای حروف از این چند تا درمیگذرد و مثلاً حتّی گاهی روی و (k) کمان کوچکی
یا دو نقطه میآید و نشان میدهد که باید آن را g خواند (۱۶/۹۰: ونرورهم šguft ، شگفت،
سخت؛ ۱۵/۵۳: رهدلروڳ۱ید ، pērōzgarīh ، پیروزگری؛ ۱۳/۶۴: ریوڳپ، bīmgēn ، بیمگین،
ترسان)، یا د با همان کمان میشود h (مثلاً در ۱۷/۹۰: ددگںری، wadīh ، بدی). همچنین، بنابر

Here are some of the Pahlavi words mentioned by Mazdapour 1999 from the manuscript itself (Jamasp Asa and Nawabi 1976) where the "crescent" form is used (note the difference in size and sharpness with the combining-hat-above, circled in blue, and that the crescent looks very much like the bottom two dots in the three-dots-above mark used over *shin*, circled in green, as well as the top two dots in the three-dots-below mark shown in the manuscript samples in previous page):

p. 8, line 5

p. 8, line 6

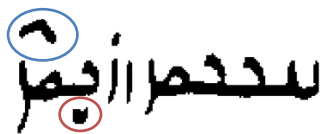p. 8, line 14

p. 53, line 15

p. 90, line 16

p. 90, line 17

And here are some examples of the "crescent" form appearing under some letters:

p. 2, line 12

p. 55, line 8

It is clear from the comparison with three-dots-above, three-dots-below, and hat-above, as well as analysis by Mazdapour 1999, König 2008, and Zeini 2020, that contrary to Meyers 2014 and Pandey 2018, this is not a hat-like character, but a simple quick way to write two dots, either above or below a letter. And from Mazdapour 1999's analysis, it is clear that they are indeed orthographically identical. Encoding the crescent-like character(s) would only be useful in representing the typeset text of Mazdapour's book (such niche applications may use U+0306 COMBINING BREVE and U+032E COMBINING BREVE BELOW). Digitization efforts of MU 29 would be better served by using the two-dots-above and two-dots-below characters.

Altogether, based on the evidence I've seen, I recommend seven combining marks for Book Pahlavi. A hat above, single-dot above and below, two-dots above and below, and three-dots above and below. These are the five common combining marks proposed in Pournader 2013 that are mentioned in every Book Pahlavi reference, plus dot-above and three-dots-below used by MU 29 (and Joneidi).

Bibliography

1. J. Amoozgar and A. Tafazzoli. 1996. Pahlavi Language: Literature, Grammatical Sketch, Texts and Glossary (زبان پهلوی: ادبیات و دستور آن). 2nd revised edition. Moin, Tehran.
2. Michael Everson. 2002. "Revised proposal to encode the Avestan and Pahlavi script in the UCS." UTC Document Register L2/02-449. The Unicode Consortium. https://www.unicode.org/L2/L2002/02449-n2556-avestan.pdf
3. Michael Everson and Roozbeh Pournader. 2007a. "Revised proposal to encode the Avestan script in the SMP of the UCS." UTC Document Register L2/07-006R. The Unicode Consortium. https://www.unicode.org/L2/L2007/07006r-n3197r-avestan.pdf
4. Michael Everson and Roozbeh Pournader. 2007b. "Proposal for encoding the Inscriptional Parthian, Inscriptional Pahlavi, and Psalter Pahlavi scripts in the SMP of the UCS." UTC Document Register L2/07-207R2. The Unicode Consortium. https://www.unicode.org/wg2/docs/n3286.pdf

5. Michael Everson and Roozbeh Pournader. 2011. "Proposal for encoding the Psalter Pahlavi script in the SMP of the UCS." UTC Document Register L2/11-147. The Unicode Consortium. https://www.unicode.org/L2/L2011/11147-n4040-psalter-pahlavi.pdf

6. Kh. M. Jamasp Asa and Mahyar Nawabi. 1976. MS. MU 29: Stories of Kersāsp, Tahmurasp & Jamshed, Gelshah & Other Texts (دستنویس ام او ۲۹: داستان گرشاسب، تهمورس و جمشید، گلشاه و متن‌های دیگر). Asia Institute of Pahlavi University. Shiraz. https://archive.org/details/fereydoun_MU29

7. Fereydoun Joneidi. 1981. *Nāme-ye Pahlevāni: Khodāmuz-e Khat va Zabān-e Pahlavi-e Aškāni, Sāsāni* (نامهٔ پهلوانی: خودآموز خط و زبان پهلوی اشکانی، ساسانی). Balkh. Tehran. https://commons.wikimedia.org/wiki/File:NPrg.pdf

8. Götz König. 2008. *Die Erzählung von Ṭahmuraṯ und Ğamšid: Edition des neupersischen Textes in Pahlavi-Schrift (MU 29) nebst zweier Parallelfassungen.* Harrassowitz Verlag. Wiesbaden. ISBN 978-3-447-05694-6.

9. D. N. MacKenzie. 1986. *A Concise Pahlavi Dictionary*. Oxford University Press. London. ISBN 0-19-713559-5. http://www.rabbinics.org/pahlavi/MacKenzie-PahlDict.pdf

10. Katayun Mazdapour. 1999 (=1378 AP). *Barresi-e Dastnevis-e M.U29: Dāstān-e Garšāsb, Tahmuras va Jamšid, Gelšāh va Matn-hāye Digar* (بررسی دستنویس م. او ۲۹: داستان گرشاسب، تهمورس و جمشید، گلشاد و متن‌های دیگر). Agah. Tehran. https://archive.org/details/MU29KatayounMazdapour

11. Abe Meyers. 2014. "Proposal for Encoding Book Pahlavi in the Unicode Standard. Version 1.2." UTC Document Register L2/14-077R. The Unicode Consortium. https://www.unicode.org/L2/L2014/14077r-book-pahlavi.pdf

12. Henrik Samuel Nyberg. 1964. A Manual of Pahlavi. Volume I: Texts, Alphabets, Index, Paradigms, Notes, and an Introduction. Otto Harrassowitz, Wiesbaden. Reprinted by Asatir, Tehran, 2003. ISBN 964-331-131-7.

13. Henrik Samuel Nyberg, Bo Utas, and Christopher Toll. 1988. *Frahang i Pahlavīk*. Otto Harrassowitz, Wiesbaden. ISBN 3-447-02671-5. https://bayanbox.ir/view/1235209569370403369/frahang-i-pahlavik.pdf

14. Anshuman Pandey. 2018. "Preliminary proposal to encode Book Pahlavi in Unicode." UTC Document Register L2/18-276. The Unicode Consortium. https://www.unicode.org/L2/L2018/18276-book-pahlavi.pdf

15. Roozbeh Pournader. 2013. "Preliminary proposal to encode the Book Pahlavi script in the Unicode Standard." UTC Document Register L2/13-141. The Unicode Consortium. https://unicode.org/L2/L2013/13141-book-pahlavi.pdf

16. Prods Oktor Skjærvø. 2008. "Introduction to Pahlavi". Cambridge, Mass. https://bayanbox.ir/view/8882150498859088732/Pahlavi-Primer-Prods-Oktor-Skjaerv.pdf

17. Arash Zeini. 2020. "Response to 'Next steps on Book Pahlavi' (L2/20-135)." UTC Document Register L2/20-141. The Unicode Consortium. https://www.unicode.org/L2/L2020/20141-book-pahlavi-resp.pdf