L2/20-141

Response to ‘Next steps on Book Pahlavi’ (L2/20-135)

Arash Zeini

8th May 2020

Introduction

Pournader & Hai (2020) might be right in saying that ‘Book Pahlavi may be the best-known complex script not yet encoded in Unicode’. However, I am inclined to add that Book Pahlavi is certainly one of the least researched and understood scripts of late antiquity. Much work is needed before we can make reliable statements about the palaeography of Book Pahlavi. The issues that have puzzled the Unicode experts are in part genuinely difficult to answer. But in my view, the wrong questions have also been asked at times.[1]

In this document I will attempt to answer the questions raised by Pournader & Hai (2020) which is a very welcome addition to the proposals as it consolidates the questions that have remained unanswered. Please note that my rendering of Middle Persian words and Pahlavi characters is limited by the possibilities offered by the available fonts, which are not designed to accurately represent the scribal idiosyncrasies. If any representations raise more questions, please contact me for clarifications.

Question 1

yh/1: The MP suffix *-īh*, transliterated <-yh>, is written 𐭩𐭥 in Book Pahlavi. Contrary to Meyers (2014:12), the correct analysis of 𐭩𐭥 is 𐭩 + 𐭣 or 𐭩 + 𐭫. Sometimes the scribes abbreviate or rather simplify this sequence of characters to 𐭩𐭦. This happens most commonly at the end of the line, and not in every manuscript. It is, for instance, very rare in manuscripts of the Pahlavi Yasna. Meyers

[1] See, for instance, Pournader & Hai (2020:6) on the issue of <b> as raised by Meyers (2014).

(2014:12) suggests that the simplification occurs 'when used at the end of the word'. As a suffix or ending, 𐭭 naturally occurs at the end of the word, but is not always simplified. The correct analysis of 𐭭 is yōḏ-ḥēṯ, while the analysis of its simplification, 𐭮, is more problematic. The fact that 𐭮 resembles the character ḥēṯ of Psalter Pahlavi is not helpful here. One could resort to the system used by MacKenzie (1990:xii) and transliterate the simplification of <-yh> to 𐭮 as <-y̱ẖ> or more elegantly as <-y͡h>.

I do not want to comment on the inclusion of this character in Unicode, but would like bring to the attention of the authors that this is not the only abbreviation or simplification known in Pahlavi.[2] The Pahlavi Documents, for instance, offer some more abbreviations. How do you propose to deal with the Pahlavi of the documents and the abbreviations and endings seen there?

It requires a dedicated investigation to ascertain the use of 𐭮 for the numerical value one and whether it would require or justify a code in Unicode. As to the question of the image included in Meyers (2014:50), I do not recommend the inclusion of West (1874) as evidence for the regular use of these numerals. This character is not considered in modern teaching material and contexts.

c/j: I am unsure I understand the question about non-joining and joining c/j/(p), as posited by Meyers (2014:11) and Pournader & Hai (2020:1). The character ṣāḏē, <c>, occurs in a non-joining manner in word initial position.[3] But it also occurs in a non-joining manner in a word final position (𐭩) after characters that do not join to the left. For instance, after <-k> in <ZK-c>, *ān-iz*, 𐭠𐭭𐭪𐭩 or 𐭠𐭭𐭩.
This is a feature of the script and the joining rules rather than a difference in character, scribal habit or hand. As expected we find various shapes of <c> in the manuscripts (see Figure 1).

(a) Pt4

(b) F2

(c) T6

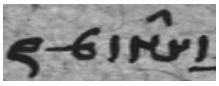
Figure 1: ZK-c

At the same time, the non-joining shape can replace the right-joining shape in word final position, as illustrated in Figure 2(a). In both preceding cases, the character remains the same, we only need to be able to disrupt the joining much like with the half-space for the Arabo-Persian script. Meyers (2014:11, 12) notes a variety in the shape of 𐭮 , which he views as a non-joining charachter. It is difficult to know why he thinks so, but most likely he is confusing shapes.

[2]For some examples, see Zeini (2015).

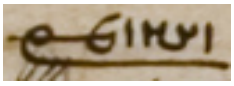
[3]For examples, see the entries in MacKenzie (1990) or any other glossary.

A right joining <c> in word final position that resembles the variety given by Meyers, , exists: . This variant attaches, for instance, to a word final <-m>. As the following examples show, its use is not consistent and subject to scribal preference:



(a) Pt4

(b) F2



(c) T6

Figure 2: ʿBYDWN-mc

I suspect that this variety does not require a dedicated code, as the difference between and is a difference that will be handled by the rendering engine.

If I understand the final question about ṣāḍē correctly, the answer is that the looped or open versions of this character are a matter of aesthetics and the hand of the scribes.

Question 2

The combination of and is attested with and without the protrusion. and can occur side by side on one line for <MN> or as part of other words.

Question 3

The Pahlavi diacritic marks or combining signs (Pandey 2018:5), much like the distribution of <y> with or without a head-loop, have not been properly investigated, and we do not have a systematic understanding of their frequency and consistency across manuscripts. The use of diacritic marks is generally considered a younger phenomenon. Individual scribes and manuscripts show habits of usage and distribution that cannot be generalised without a further analysis.

In general, we see five diacritic marks (un-circled in Pournader & Hai 2020) used across manuscripts with various degrees of consistency: . Meyers (2014:18) seems to confuse manuscript MU 29 with 'MU-16' when he refers to an edition by Mazdapur (1999), which is of MU 29.[4] This manuscript has also been edited by König (2008:127–130), who gives a range of diacritics for

[4]Due to Covid-19 lock down, I currently do not have access to a library to check the edition by Mazdapur (1999).

Text 3. He confirms the dot above, but not caron and three-dots below. My own prelimenary survey of the manuscript MU 29 seems to confirm that caron below is perhaps a misreading for two-dots below, while three-dots below is also attested in the manuscript. In the manuscript, New Persian چ is represented by a dot below <c>. Irrespective of these differences in reading, the question is whether the rendering of a New Persian text in Pahlavi script (MU 29) should be authoritative or relevant for the Unicode encoding of the Pahlavi script.

Question 4

See previous question.

Question 5

The use of punctuation varies from manuscript to manuscript and often even within a manuscript. Some manuscripts use one or several types of ornamentation to separate Avestan from Pahlavi and other ones to separate stanzas. Meyers (2014:47) shows one such case, where a '4-dot punctuation' separates Avestan from Pahlavi. Examples abound in the manuscripts. The use of such punctuation and ornamentation is not consistent. Nevertheless, if it was permissible from a Unicode point of view, one can envisage the inclusion of a number of punctuation and ornamentation marks to cover some of their functions.

Question 6

These are Gujarati numerals that are commonly found in the manuscripts. In this case it reads '117'.

Question 7

These are two different words. The first, 𐭠𐭥, is the heterogram <QDM> for *abar* 'up; on, over'. The second, 𐭧, is the heterogram <ME> for *čē*.

To highlight the richness of scribal habits, I present three instances of *abar* within two lines in the manuscript F2 alongside examples from Pt4 and T6 all from the same stanza:

(a) Pt4 (b) T6 (c) F2 (d) F2 (e) F2

Figure 3: *abar*

Question 8

I already discussed the diacritic marks in the above sections. The list given by Pournader & Hai (2020:5) is accurate as far as the scholarly discussions are concerned. If these are offered as combining signs, the advantage is that one can use these in cases where they are set wrongly in the manuscripts. The ability to render typographic mistakes with the help of combining diacritic signs would make the life of scholars much easier.

Question 9

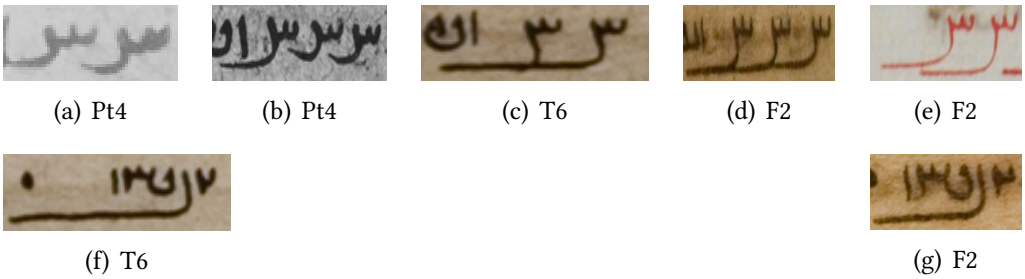
The same applies to 'mem-qoph', on which now see Zeini (2020:189). This would also be well placed as a combining diacritic sign.

I would add to this list perhaps the diacritic sign for <l>, be it as 𐭥 or 𐭦.

Further suggestions

I advise against encoding two different <b>, one with a smooth and one with a squarish curvature due to the assumption that these are distinct variants. They may be distinct, they may have been added to disambiguate, but this point has to be shown. Pournader & Hai (2020:6) rightly assume a number in the cyan box, namely the numeral 6. The curvature of <b> in this case is a matter of scribal hand and convenience rather than a distinction in the script. See following examples from the Pahlavi Yasna manuscripts with two instances of <b> in a word (Figure 4). The length and shape of <b> appears different in some numerals, but this is the artistic freedom of the scribes to adjust the length and curvature depending on context. In my view, two different codes are not justified.

The Pahlavi numerals are combinations of characters of the script. Whether they should be unified or disunified is perhaps a matter of technological rather than scholarly consideration. I remain open for discussions to see which would be the best solution.



(a) Pt4 (b) Pt4 (c) T6 (d) F2 (e) F2

(f) T6 (g) F2

Figure 4: Numbers

It should also be noted that the *ezafe* should not be encoded or otherwise treated as a distinct character. Some of the examples discussed by Meyers (2014) are based on prints rather than actual manuscript evidence. In general, the *ezafe*, *ī*, is a normal yōḏ, which can vary in looks in different manuscripts.

Not important in this context, but I did notice that the Pahlavi rendering and analysis of *ābādīh* in Pandey (2018:31) contains mistakes.

Bibliography

König, Götz. 2008. *Die Erzählung von Ṭahmuraṣ und Ǧamšid: Edition des neupersischen Textes in Pahlavi-Schrift (MU 29) nebst zweier Parallelfassungen*, vol. 14, Iranica. Wiesbaden: Harrassowitz Verlag.

MacKenzie, David Neil. 1990. *A concise Pahlavi dictionary*. Oxford: Oxford University Press.

Mazdapur, Katayun. 1999. *Dāstān-ī Garšāsb, Tahmuras va Ǯamšid, Gelšāh va matn-hāye digar: Barrasi-ī dastnevis-ī M.U. 29*. Tehran: Agah.

Meyers, Abe. 2014. *Proposal for Encoding Book Pahlavi in the Unicode Standard. Version 1.2.* The Unicode Consortium. UTC Document Register L2/14-077R.

Pandey, Anshuman. 2018. *Preliminary proposal to encode Book Pahlavi in Unicode*. The Unicode Consortium. UTC Document Register L2/18-276.

Pournader, Roozbeh & Liang Hai. 2020. *Next steps on Book Pahlavi*. The Unicode Consortium. UTC Document Register L2/20-135.

West, Edward William. 1874. *Glossary and inex of the Pahlavi texts of the Book of Arda Viraf, the Tale of Gosht-i Fryano, the Hadokht Nask, and to some extracts from the Din-Kard and Nirangistan*. London: Messers. Trübner and Co.

Zeini, Arash. 2015. Preliminary remarks on Middle Persian <nc> in the Pahlavi documents. In Anna Krasnowolska & Renata Rusek-Kowalska (eds.), *Studies on the Iranian world I. Before Islam*, 67–73. Krakow: Jagiellonian University Press.

Zeini, Arash. 2020. *Zoroastrian scholasticism in late antiquity: The Pahlavi version of the Yasna Haptaŋhāiti*, Edinburgh Studies in Ancient Persia. Edinburgh: Edinburgh University Press.