Preliminary proposal to encode the Book Pahlavi script in the Unicode Standard

Roozbeh Pournader, Google Inc.
July 24, 2013

1. Background

This is a proposal to encode the Book Pahlavi script in the Unicode Standard. Book Pahlavi was last proposed in the "Preliminary proposal to encode the Book Pahlavi script in the BMP of the UCS" (L2/07-234 = JTC1/SC2/WG2 N3294). Here is a short summary of the major differences in the character repertoire between that document and this proposal:

- **Characters only proposed here:** an alternate form of **d** and a second alternate form of **l** (see section 7), one atomic symbol and two atomic ligatures (section 8), combining dot below (sections 3 and 6);
- **Characters proposed in L2/07-234 but not proposed here:** ABBREVIATION TAA (representable as a character sequence, section 9), three archigraphemes (ambiguity, implementability, and multiple representation concerns, section 5), combining three dots below (no evidence of usage found), KASHIDA (unified with TATWEEL, sections 9 and 10), numbers (more information needed, section 10).

2. Introduction

Book Pahlavi was the most important script used in writing the Middle Persian language.[1] It probably started to be commonly used near the end of the Sassanian era (sixth century CE), evolving from the non-cursive Inscriptional Pahlavi. While the term "Book Pahlavi" refers to the script mostly surviving in books, Book Pahlavi has also been used in inscriptions, coins, pottery, and seals.

Together with the alphabetic and phonetic Avestan script, the Book Pahlavi abjad is of religious importance to the Zoroastrian community, as most of their surviving religious texts have been written in one or both of the scripts. As the main writing system for Middle Persian, the script is also critical in linguistic studies of Iranian languages and understanding the culture of Iran during the pre-Islamic and early Islamic eras. There has also been contemporary decorative usage (see Figure 1).

3. The cipher

Book Pahlavi is cipher-like to many modern readers, since the 22 originally Aramaic letters had merged into 14 letters in Book Pahlavi. For example, the Aramaic letters **w**, **n**, ʿ, and **r**, together with a final **y** that had lost its pronunciation and was used as a word ending "final stroke", had all merged into a vertical stroke: ‖.

---

[1] Middle Persian has also been written in Inscriptional Pahlavi, Psalter Pahlavi, Manichaean, and Avestan scripts.

Developments such as this led the scribes to invent ways to work around the limitations of the script, like reusing the lesser-used **l (ﻝ)** for [r], later needing to invent a diacritic (ﻝ) to distinguish [l] from the **l** that now usually meant [r]. Later, they also borrowed dots from Arabic to occasionally distinguish **g (ﺙ)**, **d (ﺙ)**, **y (ﻱ)**, and **j (ﻱ)**.

Then, there was the need to represent various phonemes that were not available in the alphabet. Some sounds were written using similarly soundly letters. Thus, the letter ʾ/**h (ﺱ)** was used to write the phonemes [a], [ā], [x], and [h]. This meant that the words such as [xār] (**hʾl**, "thorn") and [āhr] (ʾ**hl**, "fear") would both be written the same way, as ﺱﺱﻝ.

On top of that, several two-letter or three-letter combinations can be read in different ways. The letter **H (ﻉ)** can be read as the two-letter **mn**, **mw**, **mr**, or **mʿ** (ﺍ+ﻉ), and some ligatures can have up to nine different readings. Some letters or ligatures are also too similar to some other letters or ligatures, and could be misread or miswritten: for example, **g/d/y (ﺩ)** can sometimes be written very similarly to **b** (ﺟ); or in Indian manuscripts, **š (ﻭﺳ**, written ﻭﺳ in Iranian manuscripts) may not be distinguishable from the **gʾ/dʾ/yʾ/gh/dh/yh** ligature (ﻭﺳ).

A third level of confusion is added by historic spellings and Aramaic-based heterograms (Persian *hozvāreš*). For example the Middle Persian word [šudan], theoretically writable using four letters as ***ﭘﺳﻝ**, was written with eight letters as ˋ**ZLWNtn**ˋ (ﺍﻛﻩﺍﺻ), the Aramaic-based ˋ**ZLWN** ("to go") followed by the infinitive marker -**tn**ˋ.

Finally, there is the matter of psuedo-historical spellings, when a writer assumed a wrong spelling for a word based on similar words that were written using a traditional older spelling.

4. Encoding history

The "Pahlavi" script was originally proposed in 1993 by Becker and McGowan, to be unified with the Avestan script (a non-cursive script derived from Book Pahlavi and Psalter Pahlavi). Later proposals by Michael Everson, continuing to ask for unification of Avestan and "Pahlavi", appeared in 1998 and in 2002.

When I joined the efforts to encode Iranian scripts, I asked for disunification of Avestan from Pahlavi and to encode five different scripts instead of one, based on Iranianist classifications and differences in identity and behavior of the letters in the writing systems. Four proposals, by Everson and Pournader, appeared and were accepted by the UTC and the JTC1/SC2/WG2: Avestan (U+10B00..10B3F), Inscriptional Pahlavi (U+10B60..10B7F), Psalter Pahlavi (U+10B80..10BAF), and the closely related Inscriptional Parthian (U+10B40..10B5F). Manichaean (U+10AC0..10AFF) was also proposed and encoded. For a list of the proposals, see the Bibliography.

Book Pahlavi was deemed to be the most complex, but it is also perhaps the most important. A preliminary post-disunification proposal appeared in Everson, Pournader, and Durkin-Meistererenst 2007c. (That document should be consulted for additional background and samples.)

Due to the yet unencoded status of Book Pahlavi, some users have been misled to use the characters in the Inscriptional Pahlavi block instead, which have very different semantics and joining behavior. Because of that,

and the importance of the writing system to scholars, I am pushing for Unicode standardization of Book Pahlavi.[1]

5. Encoding model

This proposal does not attempt to solve the multi-layer and complex problems of properly and completely representing text written in Book Pahlavi. Instead, it tries to recognize and establish the units of Book Pahlavi writing, and help in its computerization efforts by providing an interchangeable and font-independent mechanism for representing Book Pahlavi texts. In that, it follows a "clean" mental model existing in the mind of its scholars and educated scribes, and does not try to represent every piece of potentially mangled text ever written in Book Pahlavi. Nyberg has been mostly followed as a clean model of Book Pahlavi, while recognizing some of the very distinct variant forms users may wish to represent, especially those based on the older forms of the letters.

I have chosen a middle ground between a one-to-one mapping from transliteration schemes of Book Pahlavi, and a graphical model of bellies and elbows (see West and Malandra). The first is not needed in a Unicode encoding (as the transliteration schemes suffice for it), and the second is best done in fonts or rendering engines. The model I've chosen matches the model of similar scripts already encoded in Unicode.

Even with the clean model, there may still exist several different ways to represent in Uniocode a piece of writing on paper, but that is unavoidable considering all the ambiguities of Book Pahlavi.

Variation in letters are handled in two different ways: Arabic-like dot diacritics used to distinguish letters that were unified in Book Pahlavi are proposed as combining marks, while alternative or older forms of letters that appear occasionally have been proposed as separate characters.

6. The repertoire

The proposed repertoire consists of:
- fourteen letters;
- four alternate forms of letters: one for **d**, two for **l**, and one for **k.**
- five combining diacritical signs, which are dot patterns typically used to distinguish letters that were unified in the writing system;
- three atomic ligatures/signs;
- a shaped word for Ahriman, the Zoroastrian "devil", which is commonly written turned, upside down:

    ꡁ instead of ʾ**hlmn**ʿ (ﺍﻫﻠﻤﻦ).

Character names follow the pattern of existing Aramaic-based Iranian scripts, although the letters are not called as such by the user community.

7. Alternate forms of letters

The most common alternate form of a letter is used for **l**: a stroke or ring has been added to the base shape to note that it is actually pronounced [l], since the letter had been requisitioned to be used as [r]. Iranian

---

[1]   After Book Pahlavi, perhaps the only major Iranian script not yet proposed for Unicode encoding is Sogdian.

manuscripts use the stroke form, while Indian manuscripts use the ring form, the same treatment as U+10B2E AVESTAN LETTER LE. (Some writers have instead used a doubled **l** to represent the [l] sound. For representing such cases, two Unicode characters would be used.)

Some other variants are based on older forms of letters. For example, ﻭ an old form of **d** (ﺩ), visually very similar to U+10B83 PSALTER PAHLAVI LETTER DALETH, is used in spelling a few words such as ﻮﻩﻠﭗ (**plpd**, "fraud").

Other than **d**, an old form exists for **l**, used in heterograms and visually similar to U+10B6B INSCRIPTIONAL PAHLAVI LETTER LAMEDH.[1] There is also an old form of **k**, visually similar to U+10B89 PSALTER PAHLAVI LETTER KAPH used in spelling heterograms, like ʾ**YK [**kū].

I am proposing encoding the alternate forms as separate characters. There are two other options: 1) encoding them as variation sequences, which would be hard on implementations, since the joining type of the alternate forms is sometimes different from the main letter. 2) Reusing codepoints from Psalter Pahlavi or Inscriptional Pahlavi, which would require joining across scripts, running to implementation difficulties, considering the architecture of some modern rendering engines.

For cases where an alternative form of a letter is identical to another Book Pahlavi letter or a sequence of letters, two choices could be made by the user or application: either markup could be used to specify the exact shape, or the other visually similar character(s) could be used for representing the text. For example, the word ﻣﺎﻩ, sometimes written instead of ﻘﺎﺩ (**cnd**), could be represented as <**g/d/y**, **w/n/ˈ/r**, **w/n/ˈ/r**, **g/d/y**, combining hat>. This could also be done when one wishes to represent "corruptions" (letters mistakenly written for another).

The distinction in Indian and Iranian manuscripts, in their display of **l** (ringed vs stroked) and **š** (ﻮﻩ vs ﺩﻩﺱ), is considered a stylistic variant, and is not proposed to be representable in plain text. Distinction in such presentation could be implemented using different fonts, or higher level markup. The same treatment is proposed for "the very cursive script of the papyri" (Nyberg, p. 133, Note 4), where simplified forms of some ligatures are found. This is similar to the model used for many scripts already encoded in Unicode, e.g. Arabic.

Except for the three atomic ligatures, where they are proposed for encoding as units, there is no mechanism suggested in the proposal for selecting alternative ligatures in Book Pahlavi (where more than one alternate form of a ligature exists and a user may wish to display or represent a specific form). Where such ligatures display the same as another letter or a sequence of letters, that other letter or sequence could be used to represent them. Other than that, such mechanisms are left to fonts and higher level protocols.

8. Atomic symbols and ligatures

Three atomic symbols/ligatures are proposed for encoding.

One is the common relative/possessive pronoun [i] (commonly called *kasre-ye "ezafe"* in modern Persian), written in Inscriptional and Psalter Pahlavi as the heterogram ZY, but changed in Book Pahlavi to a shorter I. Skjærvø 2007 considers this ligature to be the same as the letter **g/d/y** and transliterates it as **Y**, but I am

---

[1]There is also a sharper form of **l**, briefly mentioned in Nyberg (p. 131, line 13). I don't have enough information about it, but it may be considered a stylistic variant of the old form of **l**. Examples in Skjærvø 2007 appear to confirm that hypothesis.

proposing it for separate encoding, as a distinction is clearly seen by both Nyberg and the various typeset Book Pahlavi texts (where its glyph may appear as a semi-circle). It is proposed as a non-joining sign.

The ligatures x_1 and x_2 have been called by different names, including graphically analyzed -ẏṯ and -ṯy by MacKenzie. But most sources agree on naming them x_1 and x_2. Different opinions exist on what their parts are, but they are treated as units in most sources.

A fourth ligature, a hooked form of **m** visually similar to U+10B29 AVESTAN LETTER HME, is described in Nyberg (p. 135, under F). Since I did not encounter it in other sources, I am not yet proposing it for encoding.

Various other characters commonly used in Book Pahlavi have already been encoded. See the section "Shared characters", below.

9. Shared characters

The following characters, already encoded, are commonly used with the Book Pahlavi script:
- U+0640 ARABIC TATWEEL
- U+2E30 RING POINT
- U+2E31 WORD SEPARATOR MIDDLE DOT
- U+10B39 AVESTAN ABBREVIATION MARK
- U+10B3A TINY TWO DOTS OVER ONE DOT PUNCTUATION
- U+10B3B SMALL TWO DOTS OVER ONE DOT PUNCTUATION
- U+10B3C LARGE TWO DOTS OVER ONE DOT PUNCTUATION
- U+10B3D LARGE ONE DOT OVER TWO DOTS PUNCTUATION
- U+10B3E LARGE TWO RINGS OVER ONE RING PUNCTUATION
- U+10B3F LARGE ONE RING OVER TWO RINGS PUNCTUATION

The "BOOK PAHLAVI ABBREVIATION TAA", proposed in Everson, Pournader, and Durkin-Meisterernst 2007c is represented as the sequence <AVESTAN ABBREVIATION MARK, **w/n/ˁ/r**, **g/d/y**, combining hat, AVESTAN ABBREVIATION MARK>.

10. Numbers

Numbers are specially tricky to encode. The sources contain limited or contradictory information as to standard shapes of numbers (compare, for example, Nyberg, pp. 173–174, MacKenzie, p. 145, Amoozgar and Tafazzoli, pp. 58–59, Skjærvø 2007, pp. 97–98, and Everson, Pournader, and Durkin-Meisterernst 2007c, p. 12). Skjærvø 2007 says "[t]here is no complete description of the Pahlavi numerals".

Numbers also participate in joining, both with themselves and with letters. For example, a final joining number **1** is sometimes used to write the indefinite article -ē, as in **pyl1** ([pīlē], "an elephant"): ‎ريد.‎ The number **10** sometimes joins the following letters and numbers, and sometimes doesn't. For example, while it usually connects to 1 and 2, it only sometimes connects to 3 and 4. It doesn't connect to following letters either: compare **10wm** ([dahom], "tenth") ‎ô|ɕ‎ with the incorrect rendering ***** ‎ô|ɕ‎.

All or some numbers may be already be representable using characters proposed in this document. For example, the number **1** is visually similar or identical to the letter **b**,[1] the number **20** to the letter **l**, and the number **100** to the sequence **lz**.

Specially confusing is **10**, as it may be unifiable with the old form of **d** (the hat sometimes used on **10** is the hat typically used for **d**). The shape and joining behavior of the numbers **40-90** is also under-documented in the sources.

Numbers are not proposed for encoding in this proposal. I expect that further work would lead to a better understanding of Book Pahlavi numbers, and a later proposal would be forthcoming.

11. Joining and shaping

Book Pahlavi letters are traditionally divided into two sets, seven right-joining letters and seven dual-joining ones. The isolated forms of letters is usually very similar to their initial forms.

The letter **k** is traditionally considered right-joining, but it sometimes connects to a following letter (see Nyberg, p. 134, under 9D). Because of that, the letter is proposed as dual-joining. The common non-joining use of **k** would need the usage of ZWNJ, similar to the model used by U+10AC0 MANICHAEAN LETTER ALEPH (see Everson et al, 2011a, p. 3).

The joining behavior of the old **l** is not clear to me, since all the examples I ran into were using the letter in a word-final position. I am proposing it as right-joining. This can be changed if we find further information.

Book Pahlavi has tens of standard/obligatory ligatures, which is beyond this proposal to thoroughly document. Tables of basic ligatures is listed in Nyberg, p. 132–133 and various other sources. Some scans of such sources are available in Everson, Pournader, and Durkin-Meisterernst 2007c.

Nyberg (p. 136) differs with Skjærvø 2007 (p. 103) for the analysis of a form of x_2 that has a combining hat on top. Nyberg appears to consider it <x_2, old **d**, combining hat>, while Skjærvø 2007 seems to analyze it as <x_2, combining hat>. For the purpose of being able to represent both, x_2 is proposed as a left-joining character. When x_2 is followed by a joining character, a ZWNJ should be used after it to make sure it does not connect to the following character.

The U+0640 ARABIC TATWEEL may be used for elongation, if necessary.

12. Layout considerations

In typesetting Book Pahlavi, a baseline could be assumed, where the bottom of ʾ/**h** (ڛ) or **d/g/y** (د) lies. Letters like **l/r** (ﻝ) and **z** (ﺱ) can be considered to have ascenders, while several letters would have descenders. The baseline could align higher than or at the same vertical location as the baseline of modern scripts like Arabic and Latin, while it should align at the same location of the Avestan baseline, which Book Pahlavi commonly mixes with.

---

[1]In Nyberg and MacKenzie, but not in Skjærvø 2007.

The letter **m/Q** (ڡ) somehow moves the baseline vertically, as the attaching points on its left-side and right-side connections are not at the same height. For example, the sequence **mm** is written as ﻤﻤ. With **m/Q**, Book Pahlavi acts similar to the *nastaliq* style of writing Arabic, creating a slanted second baseline.

13. Font

The font used in this document is an amateurish font created by me to show the features of a "clean" Book Pahlavi model and distinguish or show similarities. Its design should *not* be used as a base to create computer fonts. For a more realistic font, see West and Malandra or several other sources in the bibliography.

14. Directionality

Book Pahlavi is written right-to-left and may mix with the Avestan script. Sometimes, with religious hymns of the Avesta, the older Avestan language written in the Avestan script is explained with the newer Middle Persian language written in Book Pahlavi (known as *zand*). Other times it happens the other way: the hard-to-decipher Middle Persian text in Book Pahlavi may be explained by transcribing it phonetically in the Avestan script (known as *pazand*). For examples, see Everson and Pournader 2007a, pp. 11 and 13.

As Avestan letters already have the bidirectional class "R" (Right-to-Left), I am proposing to use a similar value for Book Pahlavi.

The use of U+0640 ARABIC TATWEEL with Book Pahlavi may create a small occasional confusion for document authors, as it has the bidirectional class "AL" (Right-to-Left Arabic). But this confusion should not be a frequent problem, since elongation is expected to be used rarely and mostly between letters, and the difference in "AL" and "R" mostly affects only the behavior of modern numbers and punctuation following the strongly-typed character. If necessary, a U+200F RIGHT-TO-LEFT MARK could be used after the TATWEEL, to reset the directionality back to "R".

15. Sorting and collation

The Aramaic order is commonly used in scholarly lists of Book Pahlavi words. To help users find confusing words easily, some sources are more visually sorted: for example, **s** may be sorted as a sequence of <**g/d/y**, **g/d/y**>. In all sources consulted, **H** was sorted as if it was a sequence of <**m/Q**, **w/n/ʿ/r**>, and that's how it's also proposed here:

1. The primary order of letters is ﺱ < ﻝ < ﺩ < ﺍ < ﺱ < ﻭ < ﻥ < ﻤ < ﺒ < ﻥ < ﻕ < ﻣ.
2. The letter ﺡ should be collated as if it was the sequence <M, N>. A non-primary difference could be introduced to distinguish it from the sequence of two letters.
3. The alternate forms of letters should be primarily equivalent to the base letters, but have non-primary differences and get collated after them (in code point order, in case of **l**).
4. Due to the lack of consensus in their analyses, the two ligatures x_1 and x_2 should sort at the end of the letters, with primary weight differences: ﻣ < ﻤﻥ < ﻤﻝ.
5. The combining marks should follow the order of the letters they typically denote: ö < ô < ọ < ọ < ȯ (the mark for **j** is arbitrarily sorted after that of **y**, after Nyberg). They should have non-primary differences, but their difference should be at the same level at which ﻥ is distinguished from ﻝ.

6. The sign for **i** should be considered as an alternative form of **g/d/y** in collation (with a non-primary difference), but sorted before the letter, since it is usually listed in the word lists at the very beginning of the section for **g/d/y**.

7. The turned Ahriman should be treated in collation as primarily equal to the sequence of letters in its compatibility decomposition: <ﺳ, ﺳ, ﮞ, ﮒ, ﺍ, ﺍ>, with a non-primary difference.

16. Line breaking

All letters, ligatures, and signs should have the line breaking property AL (Alphabetic), as they all act letter-like. The combining marks should have the line breaking property CM (Combining Mark).

17. Confusability

Book Pahlavi could be extremely confusing. Experts commonly disagree about the reading of texts, starting the disagreement with the letters in a word. Making a model for strings that could be confusable in the Book Pahlavi script is quite an academic exercise. For some basic examples, see section 3. More extensive examples can be found in Nyberg and many of the references listed in the bibliography.

Several confusable character pairs and string pairs also exist with the closely related scripts Avestan, Psalter Pahlavi, and Inscriptional Pahlavi, and then the other cursive neighboring scripts, like Arabic, Syriac, Mandaic, and Manichaean.

18. Encoding model concerns

Sometimes, more than one ligature form is available for a combination of two or more characters. I am assuming that such choices are not explicit spelling choices and are thus acceptable alternatives. If that is proven to not be the case, new proposals could be made for new alternative forms of letters as characters (when there are joining class differences), for new atomic ligatures, or for new variation sequences for choosing specific forms (as in the Mongolian script).

Another concern is using the existing "Arabic" joining model in Unicode. While the model works very well for most of the script, the exceptions create a concern. As explained in section 11 (joining and shaping), the letter **k** and the ligature **x₂** do not typically join to their next letter on the left, but occasionally may do so, and such joining behavior appears to be a spelling decision. I am currently proposing these characters as always connecting to their next letter, unless explicitly forbidden to be broken by using a ZWNJ. But such frequent usage of ZWNJ is counter-intuitive. A better solution may be changing the "Arabic" joining model slightly by including a ZERO WIDTH CURSIVE CONNECTOR character that bridges over non-joining gaps and makes both sides connecting. Such a character would also be useful in better representation of several other joining scripts, including Manichaean and even Arabic.

19. Main character properties

UnicodeData.txt

```
10BB0;BOOK PAHLAVI LETTER ALEPH-HET;Lo;0;R;;;;;N;;;;;
10BB1;BOOK PAHLAVI LETTER BETH;Lo;0;R;;;;;N;;;;;
10BB2;BOOK PAHLAVI LETTER GIMEL-DALETH-YODH;Lo;0;R;;;;;N;;;;;
10BB3;BOOK PAHLAVI LETTER OLD DALETH;Lo;0;R;;;;;N;;;;;
```

```
10BB4;BOOK PAHLAVI LETTER HE;Lo;0;R;;;;;N;;;;;
10BB5;BOOK PAHLAVI LETTER WAW-NUN-AYIN-RESH;Lo;0;R;;;;;N;;;;;
10BB6;BOOK PAHLAVI LETTER ZAYIN;Lo;0;R;;;;;N;;;;;
10BB7;BOOK PAHLAVI LETTER KAPH;Lo;0;R;;;;;N;;;;;
10BB8;BOOK PAHLAVI LETTER OLD KAPH;Lo;0;R;;;;;N;;;;;
10BB9;BOOK PAHLAVI LETTER LAMEDH;Lo;0;R;;;;;N;;;;;
10BBA;BOOK PAHLAVI LETTER OLD LAMEDH;Lo;0;R;;;;;N;;;;;
10BBB;BOOK PAHLAVI LETTER L-LAMEDH;Lo;0;R;;;;;N;;;;;
10BBC;BOOK PAHLAVI LETTER MEM-QOPH;Lo;0;R;;;;;N;;;;;
10BBD;BOOK PAHLAVI LETTER SAMEKH;Lo;0;R;;;;;N;;;;;
10BBE;BOOK PAHLAVI LETTER PE;Lo;0;R;;;;;N;;;;;
10BBF;BOOK PAHLAVI LETTER SADHE;Lo;0;R;;;;;N;;;;;
10BC0;BOOK PAHLAVI LETTER SHIN;Lo;0;R;;;;;N;;;;;
10BC1;BOOK PAHLAVI LETTER TAW;Lo;0;R;;;;;N;;;;;
10BC2;BOOK PAHLAVI LIGATURE X1;Lo;0;R;;;;;N;;;;;
10BC3;BOOK PAHLAVI LIGATURE X2;Lo;0;R;;;;;N;;;;;
10BCB;BOOK PAHLAVI COMBINING TWO DOTS ABOVE;Mn;230;NSM;;;;;N;;;;;
10BCC;BOOK PAHLAVI COMBINING HAT ABOVE;Mn;230;NSM;;;;;N;;;;;
10BCD;BOOK PAHLAVI COMBINING TWO DOTS BELOW;Mn;220;NSM;;;;;N;;;;;
10BCE;BOOK PAHLAVI COMBINING DOT BELOW;Mn;220;NSM;;;;;N;;;;;
10BCF;BOOK PAHLAVI COMBINING THREE DOTS ABOVE;Mn;230;NSM;;;;;N;;;;;
10BD0;BOOK PAHLAVI SIGN I;So;0;;R;;;;;N;;;;;
10BD1;BOOK PAHLAVI TURNED AHRIMAN;Lo;0;R;<compat> 10BB0 10BB0 10BB9 10BBC 10BB5
          10BB5;;;;N;;;;;
```

ArabicShaping.txt

```
10BB0; BOOK PAHLAVI ALEPH-HET; D; No_Joining_Group
10BB1; BOOK PAHLAVI BETH; R; No_Joining_Group
10BB2; BOOK PAHLAVI GIMEL-DALETH-YODH; D; No_Joining_Group
10BB3; BOOK PAHLAVI OLD DALETH; R; No_Joining_Group
10BB4; BOOK PAHLAVI HE; R; No_Joining_Group
10BB5; BOOK PAHLAVI WAW-NUN-AYIN-RESH; R; No_Joining_Group
10BB6; BOOK PAHLAVI ZAYIN; D; No_Joining_Group
10BB7; BOOK PAHLAVI KAPH; D; No_Joining_Group
10BB8; BOOK PAHLAVI OLD KAPH; R; No_Joining_Group
10BB9; BOOK PAHLAVI LAMEDH; D; BOOK PAHLAVI LAMEDH
10BBA; BOOK PAHLAVI OLD LAMEDH; R; No_Joining_Group
10BBB; BOOK PAHLAVI LAMEDH WITH STROKE; D; BOOK PAHLAVI LAMEDH
10BBC; BOOK PAHLAVI MEM-QOPH; D; No_Joining_Group
10BBD; BOOK PAHLAVI SAMEKH; D; No_Joining_Group
10BBE; BOOK PAHLAVI PE; R; No_Joining_Group
10BBF; BOOK PAHLAVI SADHE; R; No_Joining_Group
10BC0; BOOK PAHLAVI SHIN; D; No_Joining_Group
10BC1; BOOK PAHLAVI TAW; R; No_Joining_Group
10BC2; BOOK PAHLAVI X1; U; No_Joining_Group
10BC3; BOOK PAHLAVI X2; L; No_Joining_Group
10BD0; BOOK PAHLAVI I; U; No_Joining_Group
10BD1; BOOK PAHLAVI TURNED AHRIMAN; U; No_Joining_Group
```

ScriptExtensions.txt

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

10B39..10B3F ; Avst Phlv # Po [7] AVESTAN ABBREVIATION MARK..LARGE ONE RING OVER
          TWO RINGS PUNCTUATION
```

Scripts.txt

```
10B39..10B3F ; AvestanCommon # Po [7] AVESTAN ABBREVIATION MARK..LARGE ONE RING
               OVER TWO RINGS PUNCTUATION
```

	10BB	10BC	10BD
0	млл	млл	ᴈ
1	ᴸ	ᵐ	ᩏᵚᴵ
2	ᴅ	ᴪ	
3	ᵠ	ᴪ	
4	ᴪ		
5	I		
6	ᵴ		
7	ᵠ		
8	ᵠ		
9	ᴽ		
A	ᴽ		
B	ᴽ	◌̈	
C	ᴪ	◌̂	
D	ᴪ	◌̤	
E	ᵱ	◌̣	
F	ᵴ	◌̈̇	

10BB0	ﺳ	BOOK PAHLAVI LETTER ALEPH-HET
10BB1	ﺟ	BOOK PAHLAVI LETTER BETH
10BB2	ﺩ	BOOK PAHLAVI LETTER GIMEL-DALETH-YODH
10BB3	ﻭ	BOOK PAHLAVI LETTER OLD DALETH

→ 10B83 psalter pahlavi letter daleth

10BB4	ﻫ	BOOK PAHLAVI LETTER HE
10BB5	١	BOOK PAHLAVI LETTER WAW-NUN-AYIN-RESH

• Also the final stroke

10BB6	ﺯ	BOOK PAHLAVI LETTER ZAYIN
10BB7	ﻭ	BOOK PAHLAVI LETTER KAPH
10BB8	ﻕ	BOOK PAHLAVI LETTER OLD KAPH

→ 10B89 psalter pahlavi letter kaph

10BB9	ﻝ	BOOK PAHLAVI LETTER LAMEDH
10BBA	ﻝ	BOOK PAHLAVI LETTER OLD LAMEDH

→ 10B8A psalter pahlavi letter lamedh

10BBB	ﻝ	BOOK PAHLAVI LETTER L-LAMEDH

• Used to represent explicit [l]

10BBC	ﻉ	BOOK PAHLAVI LETTER MEM-QOPH
10BBD	ﺳ	BOOK PAHLAVI LETTER SAMEKH
10BBE	ﻉ	BOOK PAHLAVI LETTER PE
10BBF	ﻕ	BOOK PAHLAVI LETTER SADHE
10BC0	ﺳ	BOOK PAHLAVI LETTER SHIN
10BC1	ﻉ	BOOK PAHLAVI LETTER TAW
10BC2	ﺳ	BOOK PAHLAVI LIGATURE X1
10BC3	ﻉ	BOOK PAHLAVI LIGATURE X2
10BCB	◌̈	BOOK PAHLAVI COMBINING TWO DOTS ABOVE
10BCC	◌̂	BOOK PAHLAVI COMBINING HAT ABOVE
10BCD	◌̤	BOOK PAHLAVI COMBINING TWO DOTS BELOW
10BCE	◌̣	BOOK PAHLAVI COMBINING DOT BELOW
10BCF	◌̇	BOOK PAHLAVI COMBINING THREE DOTS ABOVE
10BD0	ﺟ	BOOK PAHLAVI SIGN I
10BD1	ﻟﻫ	BOOK PAHLAVI TURNED AHRIMAN

• The actual glyph is larger than the one shown in the chart
≈ <compat> 10BB0 ﺳ 10BB0 ﺳ 10BB9 ﻝ 10BBC ﻉ 10BB5 ١
10BB5 ١

Acknowledgments

Bibliography

- Becker, Joe, and Rick McGowan. 1992–1993. "Unicode Technical Report #3: Exploratory Proposals", revision 2. The Unicode Consortium. http://www.unicode.org/reports/tr3-2/.
- Chrisomalis, Stephen. 2010. *Numerical Notation: A Comparative History*. New York: Cambridge University Press.
- J. Duchesne-Guillemin. 2011. "Ahriman". In *Encyclopædia Iranica*, edited by Ehsan Yarshater. http://www.iranicaonline.org/articles/ahriman.
- Everson, Michael. 1998. "Proposal to encode Avestan in the BMP of ISO/IEC 10646." *ISO/IEC JTC1/SC2/WG2* N1684 and *UTC Document Register* L2/98-031. http://www.unicode.org/L2/L1998/98031.pdf.
- Everson, Michael. 2002. "Proposal to encode Avestan in the BMP of ISO/IEC 10646." *ISO/IEC JTC1/SC2/WG2* N2556 and *UTC Document Register* L2/03-072. http://www.unicode.org/L2/L2003/03072-n2556-avestan.pdf.
- Everson, Michael, and Roozbeh Pournader. 2006. "Proposal to encode the Avestan script in the BMP of the UCS." *ISO/IEC JTC1/SC2/WG2* N3178 and *UTC Document Register* L2/06-335. http://www.unicode.org/L2/L2006/06335-n3178-avestan.pdf.
- Everson, Michael, and Roozbeh Pournader. 2007a. "Revised proposal to encode the Avestan script in the SMP of the UCS." *ISO/IEC JTC1/SC2/WG2* N3197R and *UTC Document Register* L2/07-006R. http://www.unicode.org/L2/L2007/07006r-n3197r-avestan.pdf.
- Everson, Michael, and Roozbeh Pournader. 2007b. "Proposal for encoding the Inscriptional Parthian, Inscriptional Pahlavi, and Psalter Pahlavi scripts in the SMP of the UCS." *ISO/IEC JTC1/SC2/WG2* N3286 and *UTC Document Register* L2/07-207. http://www.unicode.org/L2/L2007/07207-n3286-parthian-pahlavi.pdf.
- Everson, Michael, Roozbeh Pournader, and Desmond Durkin-Meisterernst. 2007c. "Preliminary proposal to encode the Book Pahlavi script in the BMP of the UCS." *ISO/IEC JTC1/SC2/WG2* N3294 and *UTC Document Register* L2/07-234. http://www.unicode.org/L2/L2007/07234-n3294-pahlavi.pdf.
- Everson, Michael, Desmond Durkin-Meisterernst, Roozbeh Pournader, and Shervin Afshar. 2011a. "Second revised proposal for encoding the Manichaean script in the SMP of the UCS." *ISO/IEC JTC1/SC2/WG2* N4029 and *UTC Document Register* L2/11-123. http://www.unicode.org/L2/L2011/11123-n4029-manichaean.pdf.
- Everson, Michael, and Roozbeh Pournader. 2011b. "Proposal for encoding the Psalter Pahlavi script in the SMP of the UCS." *ISO/IEC JTC1/SC2/WG2* N4040 and *UTC Document Register* L2/11-147. http://www.unicode.org/L2/L2011/11147-n4040-psalter-pahlavi.pdf.
- Haug, Martin, and E. W. West. 1874. Glossary and Index of the Pahlavi Texts of the Book of Arda Viraf. Bombay: Government Central Book Depot, and London: Messrs. Trübner and Co.
- Jamaspji Asa, Destur Hoshangji, and Martin Haug. 1870. *An Old Pahlavi-Pazand Glossary*. Bombay: Government Central Book Depot, and London: Messrs. Trübner and Co.
- MacKenzie, David Neil. 1986. *A Concise Pahlavi Dictionary*. London: Oxford University Press.

- Nyberg, Henrik Samuel. 2003. *A Manual of Pahlavi*. Tehran: Asatir. Originally Wiesbaden: Otto Harrassowitz, 1964.
- Skjærvø, Prods Oktor. 1996. "Aramaic Scripts for Iranian Languages". In *The World's Writing Systems*, edited by Peter T. Daniels and William Bright, pp. 515–535. New York: Oxford University Press.
- Skjærvø, Prods Oktor. 2007. "Introduction to Pahlavi". Cambridge, Mass. Accessed July 8, 2013. http://shr.receptidocs.ru/docs/5/4314/conv_1/file1.pdf
- West, Emily, and William Malandra. "Ancient Iranian Font Project: Pahlavi Fonts". St. Catherine University. Accessed July 1, 2013. http://minerva.stkate.edu/offices/academic/classics.nsf/pages/pahlavifonts.

- آموزگار، ژاله، و تفضلی، احمد. ۱۳۷۵. *زبان پهلوی: ادبیات و دستور آن*. ویرایش ۲، تهران: معین.
- ابوالقاسمی، محسن. ۱۳۸۰. *تاریخ زبان فارسی*. ویرایش ۲. تهران: سازمان مطالعه و تدوین کتاب‌های علوم انسانی دانشگاهها (سمت).
- بیضایی، بهرام. ۱۳۷۸. مرگ یزدگرد: *نمایشنامه* . تهران: انتشارات روشنگران و مطالعات زنان.
- عریان، سعید. ۱۳۸۲. *متن‌های پهلوی*. گردآورنده جاماسب جی دستور منوچهر جی جاماسب آسانا؛ پژوهش سعید عریان. تهران: سازمان میراث فرهنگی کشور.
- مکنزی، دیوید نیل. ۱۳۷۸. فرهنگ کوچک *زبان پهلوی*. ترجمۀ مهشید میرفخرایی. تهران: پژوهشگاه علوم انسانی و مطالعات فرهنگی.
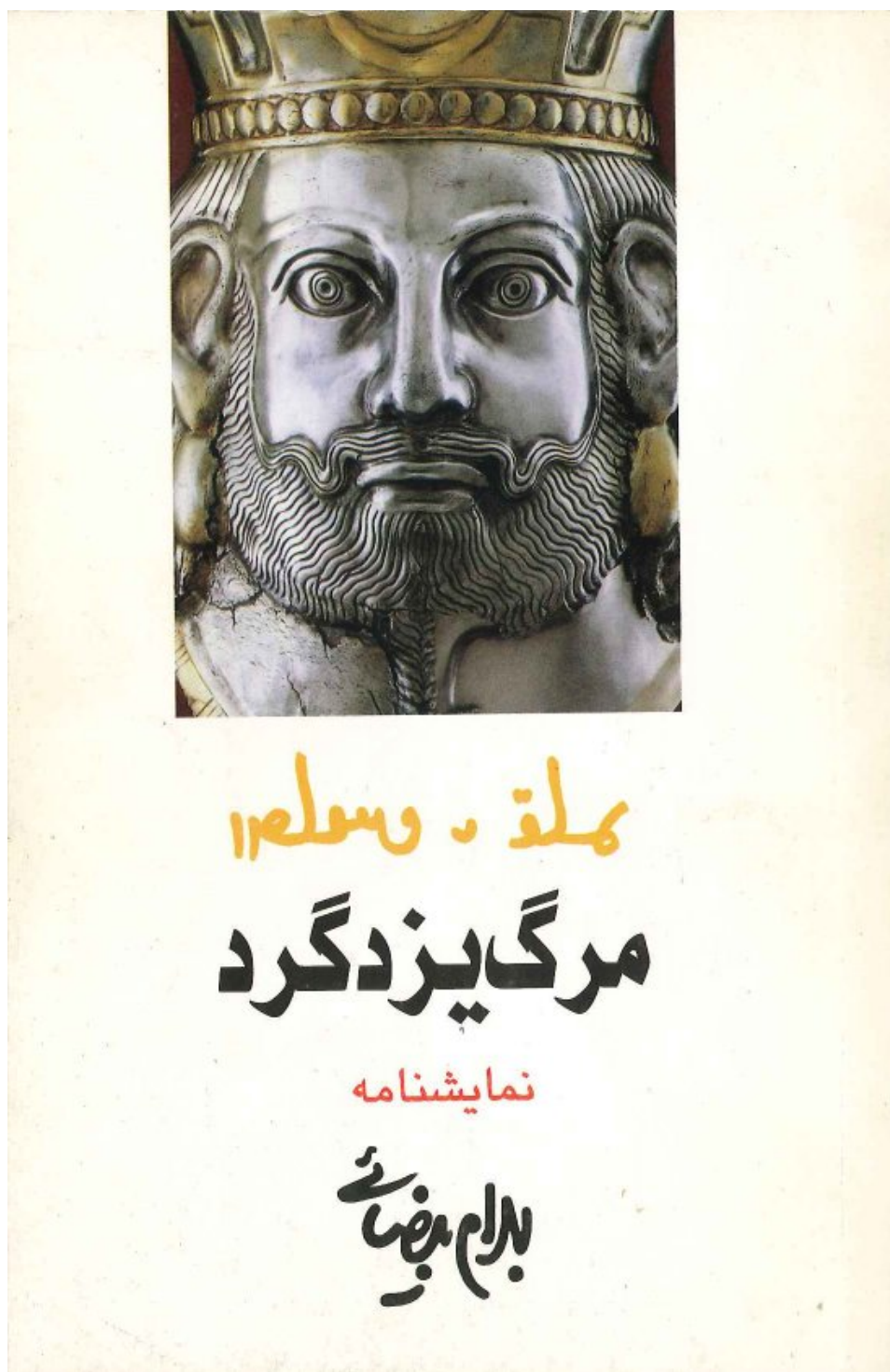
Figure 1. The front cover of the 1999 edition of the book مرگ یزدگرد (*Death of Yazdgerd*), a famous play by Bahram Beyzaie, written and performed in the 1970s and made into a movie by the *auteur*. The Book Pahlavi text could be transliterated in the Nyberg system as **mlg ZY yzdglt**ʾ. It would be represented as <10BBC, 10BB9, 10BB2, 10BCB, 0020, **10BD0**, 0020, 10BB2, **10BB0**, 10BB2, 10BB9, 10BC1, 10BB5>. Note the usage of the sign for [i] and the "simplification" of **zd** to ʾ/**h**: see Nyberg, p. 134, 9(B)(a).

IV   P r e f a c e.

[Book Pahlavi script characters]

And in accordance with this arrangement, the letters with all their usual modifications and compounds, will be found in the following order, observing that the forms enclosed in parentheses are classed along with that one which precedes the parenthesis:

[Book Pahlavi script characters and compounds listed across multiple lines]

A few of these compounds occur in two places in the list, because they can be variously read.

The references to AV. (Ardâ Vîrâf), GF. (Gôsht-i Fryânô), and Hn. (Hâdôkht Nask) are to the *chapter* and *sentence* as numbered in the texts; when several chapters are mentioned together (as 2, 6, 18,

Figure 2. A page from Haug and West 1874, showing some of the Book Pahlavi combinations supported by the font used to typeset the book. Note that this is 19th century technology, and a modern font is expected to do a much better job. The page discusses the sorting order used in the word list, which is based on the sorting order of the Persian alphabet (in the Arabic script). This is different from the modern Aramaic-based sorting order recommended in this proposal.

(١)   I

ᚠ

[Book Pahlavi script text]

S١.   [Book Pahlavi script text]

[Book Pahlavi script text]

Ṭ.

A.١   [Book Pahlavi script text]

[Book Pahlavi script text]

[Book Pahlavi script text]

S٢.A٢   [Book Pahlavi script text]

[Book Pahlavi script text]   5

[Book Pahlavi script text]

[Book Pahlavi script text]

[Book Pahlavi script text]

[Book Pahlavi script text]

[Book Pahlavi script text]   10

[Book Pahlavi script text]

[Book Pahlavi script text]

[Book Pahlavi script text]

[Book Pahlavi script text]

S٣.   ١٥ [Book Pahlavi script text]

A.٣   [Book Pahlavi script text]

[Book Pahlavi script text]

[Book Pahlavi script text]

Figure 3a. The opening of Kar-Namag i Ardashir i Pabagan, hand-written by Nyberg. Note the usage of U+2E30 RING POINT as sentence-ending punctuation. The three different forms of **l** (stroked, normal, old) can be seen at the beginning of line 9 of the text.

MN

hac

k'ln'mk' z̲y̲. 'lthšdl z̲y̲ p'pk'n'.
kār-nāmak i Artaxšēr i Pāpakān.

PWN šM z̲y̲ d't'l 'whrmz̲d z̲y̲ l'd'wmnd z̲y̲ GDH-'wmnd.
pat nām i dātār Ohurmazd i rāyōmand i x^uarrahōmand.

I.

PWN k'l šM-k' z̲y̲ 'lhtšdl z̲y̲ p'pk'n' 'ytwn' npšt' YK‘YMWN-'t' 'YK 'ḤL MN mlg
pat kār-nāmak i Artaxšēr i Pāpakān ētōn nipišt estāt kū pas hac marg

z̲y̲ 'lksndl z̲y̲ hlwm'yk 'yl'n' štr' 200 w 40 ktk' hwt'y yḥwwn-t'. sph'n' w
i Alaksandar i hrōmāyik ērān šahr 200 u 40 katak-x^uatāi būt. Spāhān u

p'ls w kwstkyh' z̲y̲ 'wbš nzdyktl PWN YDH z̲y̲ 'ldw'n' srd'l bwt'. p'pk'
Pārs u kustakihā i aviš nazdiktar pat dast i Ardavān sardār būt. Pāpak

mlcp'n' w štr'd'l z̲y̲ p'ls bwt' w MN gwm'ltk' z̲y̲ 'ldw'n' yḥwwn-t'. 'ldw'n
marzpān u šahridār i Pārs būt u hac gumārtak i Ardavān būt. Ardavān

PWN st'hl YTYBWN-st'. w p'pk' l'd hyc prdnd z̲y̲ šM bwlt'l L' bwt'. w s's'n'
pat Staxr nišast. u Pāpak rād hēc fradand i nām-burtār nē būt. u Sāsān

šp'n' z̲y̲ p'pk' yḥwwn-t' w hmw'l LWTH KYN'-'n' yḥwwn-t' w MN twhmk' z̲y̲
šupān i Pāpak būt u hamvār apāk gōspandān būt u hac tōhmak i

d'l'y z̲y̲ d'l'd'n' yḥwwn-t'. w BYN dwšhwt'dyh z̲y̲ 'lksndl ‘L wlyk' w nyh'n'
Dārāi i Dārāyān būt. u andar dušx^uatāyēh i Alaksandar ō virēk u nihān-

lwbšnyh YK‘YMWN-'t' w LWTH kwltyk̲'n' šp'n'n' SGYTN-t'.
ravišnēh estāt u apāk kurtikān šupānān raft.

p'pk' L' YD‘YTN-st' 'YK s's'n' MN twhmk' z̲y̲ d'l'y z̲y̲ d'l'd'n' YLYDWN-t'
Pāpak nē dānist kū Sāsān hac tōhmak i Dārāi i Dārāyān zāt

YK‘YMWN-yt'. p'pk' LYLY'-1 PWN hwmn ḤZYTN-t' cygwn 'MT hwlšyt' MN L‘YŠH z̲y̲
estēt. Pāpak šap-ē pat x^uamn dīt cigōn ka x^uaršēt hac sar i

s's'n' BR' t'pt' w hm'k gyh'n' lwšnyh ‘BYDWN-x₁. 'ḤRN LYLY' 'ytwn' ḤZYTN-t'
Sāsān bē tāft u hamāk gēhān rōšnēh kart. an šap ētōn dīt

cygwn' 'MT s's'n' PWN pyl-1 z̲y̲ 'l'stk' z̲y̲ spyt' YTYBWN-st' YK‘YMWN-'t' w KR'
cigōn ka Sāsān pat pīl-ē i ārāstak i spēt nišast estāt u har

MNW BYN kyšwl pyl'mwn' z̲y̲ s's'n' YK‘YMWN-d ‘SGDH 'wbš YDLWN-d w st'dšn'
kē andar kišvar pērāmōn i Sāsān estēnd, namāc aviš barēnd u stāyišn

w 'pryn' hm'y ‘BYDWN-d. 'ḤRN stykl LYLY' hmgwnk' 'ytwn' ḤZYTN-t' cygwn
u āfrīn hamē kunēnd. an sitīkar šap hamgōnak ētōn dīt cigōn

Figure 3b. The opening of Kar-Namag i Ardashir i Pabagan, as transliterated and transcribed by Nyberg. Note that there's a typo in line 1 of the text, mis-tranliterating ʾ**lthšdl** as ʾ**lhtšdl**.

10BBC (**m/Q**), 10BB5 (**w/n/ʿ/r**)

10BB7 (**k**), 10BB0 (ʾ**/h**), 10BB9 (**l**), 10BB5 (**w/n/ʿ/r**), 10BB0 (ʾ**/h**), 10BBC (**m/Q**), 10BB7 (**k**), 10BB5 (**w/n/ʿ/r**), 0020 (space), 10BD0 (**i**), 0020 (space), 10BB0 (ʾ**/h**), 10BB9 (**l**), 10BC1 (**t**), 10BB0 (ʾ**/h**), 10BC0 (**š**), 10BB2 (**g/d/y**), 10BCC (hat above), 10BB9 (**l**), 0020 (space), 10BD0 (**i**), 0020 (space), 10BBE (**p**), 10BB0 (ʾ**/h**), 10BBE (**p**), 10BB7 (**k**), 10BB0 (ʾ**/h**), 10BB5 (**w/n/ʿ/r**), 10BB5 (**w/n/ʿ/r**), 2E30 (ring point)

10BBE (**p**), 10BB5 (**w/n/ʿ/r**), 10BB5 (**w/n/ʿ/r**), 0020 (space), 10BC0 (**š**), 10BBC (**m/Q**), 0020 (space), 10BD0 (**i**), 0020 (space), 10BB2 (**g/d/y**), 10BCC (hat above), 10BB0 (ʾ**/h**), 10BC1 (**t**), 10BB0 (ʾ**/h**), 10BB9 (**l**), 0020 (space), 10BB0 (ʾ**/h**), 10BB5 (**w/n/ʿ/r**), 10BB0 (ʾ**/h**), 10BB5 (**w/n/ʿ/r**), 10BBC (**m/Q**), **10BB0** (ʾ**/h**), 0020 (space), 10BD0 (**i**), 0020 (space), 10BB9 (**l**), 10BB0 (ʾ**/h**), 10BB2 (**g/d/y**), 10BCC (hat above), 10BB0 (ʾ**/h**), 10BB5 (**w/n/ʿ/r**), 10BBC (**m/Q**), 10BB5 (**w/n/ʿ/r**), 10BB2 (**g/d/y**), 10BCC (hat above)

10BD0 (**i**), 0020 (space), 10BB2 (**g/d/y**), 10BCB (two dots above), 10BB2 (**g/d/y**), 10BCC (hat above), 10BB4 (**H**), 10BB0 (ʾ**/h**), 10BB5 (**w/n/ʿ/r**), 10BBC (**m/Q**), 10BB5 (**w/n/ʿ/r**), 10BB2 (**g/d/y**), 10BCC (hat above), 2E30 (ring point)

10BBE (**p**), 10BB5 (**w/n/ʿ/r**), 10BB5 (**w/n/ʿ/r**), 0020 (space), 10BB7 (**k**), 10BB0 (ʾ**/h**), 10BB9 (**l**), 0020 (space), 10BC0 (**š**), 10BBC (**m/Q**), 10BB7 (**k**), 10BB5 (**w/n/ʿ/r**), 0020 (space), 10BD0 (**i**), 0020 (space), 10BB0 (ʾ**/h**), 10BB9 (**l**), 10BC1 (**t**), 10BB0 (ʾ**/h**), 10BC0 (**š**), 10BB2 (**g/d/y**), 10BCC (hat above), 10BB9 (**l**), 0020 (space), 10BD0 (**i**), 0020 (space), 10BBE (**p**), 10BB0 (ʾ**/h**), 10BBE (**p**), 10BB7 (**k**), 10BB0 (ʾ**/h**), 10BB5 (**w/n/ʿ/r**), 10BB5 (**w/n/ʿ/r**), 0020 (space), 10BB0 (ʾ**/h**), 10BB2 (**g/d/y**), 10BCD (two dots below), 10BC1 (**t**), 10BB5 (**w/n/ʿ/r**), 10BB5 (**w/n/ʿ/r**), 10BB5 (**w/n/ʿ/r**), 0020 (space), 10BB5 (**w/n/ʿ/r**), 10BBE (**p**), 10BC0 (**š**), 10BC1 (**t**), 10BB5 (**w/n/ʿ/r**)

10BB2 (**g/d/y**), 10BB7 (**k**), 10BB5 (**w/n/ʿ/r**), 10BB2 (**g/d/y**), 10BBC (**m/Q**), 10BB5 (**w/n/ʿ/r**), 10BB5 (**w/n/ʿ/r**), 10BB0 (ʾ**/h**), 10BC1 (**t**), 10BB5 (**w/n/ʿ/r**), 0020 (space), 10BB0 (ʾ**/h**), 10BB2 (**g/d/y**), 10BB8 (old **k**), 0020 (space), 10BB0 (ʾ**/h**), 10BB0 (ʾ**/h**), 10BBA (old **l**), 0020 (space), 10BBC (**m/Q**), 10BB5 (**w/n/ʿ/r**), 0020 (space), 10BBC (**m/Q**), 10BB9 (**l**), 10BB2 (**g/d/y**), 10BCB (two dots above), 0020 (space), 10BD0 (**i**), 0020 (space), 10BB0 (ʾ**/h**), 10BBB (l-**l**), 10BB7 (**k**), 10BBD (**s**), 10BB5 (**w/n/ʿ/r**), 10BB2 (**g/d/y**), 10BCC (hat above), 10BB9 (**l**), 0020 (space), 10BD0 (**i**), 0020 (space), 10BB0 (ʾ**/h**), 10BB9 (**l**), 10BB5 (**w/n/ʿ/r**), 10BBC (**m/Q**), 10BB0 (ʾ**/h**), 10BB2 (**g/d/y**), 10BCD (two dots below), 10BB7 (**k**), 0020 (space), 10BB0 (ʾ**/h**), 10BB2 (**g/d/y**), 10BCD (two dots below), 10BB9 (**l**), 10BB0 (ʾ**/h**), 10BB5 (**w/n/ʿ/r**), 10BB5 (**w/n/ʿ/r**), 0020 (space), 10BC0 (**š**), 10BC1 (**t**), 10BB5 (**w/n/ʿ/r**), 10BB5 (**w/n/ʿ/r**)

Figure 3c. The title and the first two lines of text from Figure 3a, in proposed Unicode characters.

<table>
<tr><td align="center">

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646

</td></tr>
</table>

A. Administrative

1. Title**:** *Proposal to Encode the Book Pahlavi Script*
2. Requester's name: **Roozbeh Pournader**
3. Requester Type (Member body/Liaison/Individual contribution): **Individual Contribution**
4. Submission date: **July 24, 2013**
5. Requester's reference, if applicable: **N/A**
6. Choose one of the following:
    This is a complete proposal: **Yes**
    (or) More information will be provided later: **No**

B. Technical – General

1. Choose one of the following:
    a. This proposal is for a new script (set of characters): **Yes**
        Proposed name of script: **Book Pahlavi**
    b. The proposal is for addition of character(s) to an existing block: **No**
        Name of existing block: **N/A**
2. Number of characters in proposal: **Twenty seven (27)**
3. Proposed category: **C. Major extinct**
4. Is a repertoire including character names provided? **Yes**
    a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? **Yes**
    b. Are the character shapes attached in a legible form suitable for review? **Yes**
5. Fonts related:
    a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? **Roozbeh Pournader**
    b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): **Roozbeh Pournader, Google Inc.**
6. References:
    a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? **Yes**
    b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? **Yes. See also N3294.**
7. Special encoding issues:
    Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please endorse information)? **Yes**

8. Additional information:
Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.orgUTH for such information on other scripts. Also see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? **Yes**
    If YES explain: **They were preliminary proposals.**
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?
    If YES, with whom?
    If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? **Yes**
    Reference: **N/A**
4. The context of use for the proposed characters (type of use; common or rare): **common**
    Reference: **N/A**
5. Are the proposed characters in current use by the user community? **Yes**
    If YES, where? Reference: **Scholarly community, education, religious usage by Zoroastrian community. Important concentration of religious user community in Iran and India.**
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? **No**
    If YES, is a rationale provided? **N/A**
    If YES, reference: **N/A**
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? **Yes**
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? **No**
    If YES, is a rationale for its inclusion provided? **N/A**
    If YES, reference: **N/A**
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? **No**
    If YES, is a rationale for its inclusion provided? **N/A**
    If YES, reference: **N/A**
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? **Yes**
    If YES, is a rationale for its inclusion provided? **Yes. The proposed characters have different identities and joining behavior**
    If YES, reference: **See above**
11. Does the proposal include use of combining characters and/or use of composite sequences? **Yes**
    If YES, is a rationale for such use provided? **Yes**
        If YES, reference: **See the proposal**
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? **N/A**
        If YES, reference: **N/A**
12. Does the proposal contain characters with any special properties such as control function or similar semantics? **No**
    If YES, describe in detail (include attachment if necessary): **N/A**
13. Does the proposal contain any Ideographic compatibility characters? **No**
    If YES, are the equivalent corresponding unified ideographic characters identified? **N/A**
    If YES, reference: **N/A**