

A Critique of L2/18-276

Abe Meyers*

November 30, 2018

Contents

| | | |
|----------|--|----------|
| 1 | Introduction | 1 |
| 2 | Multiple incompatible representations | 2 |
| 2.1 | <gimel-daleth-yodh> + <shin> vs <aleph-heth> + <aleph-heth> | 3 |
| 2.2 | <Fixed-aleph> + <gimel-daleth-yodh> vs <fixed-gimel-daleth-yodth> + <aleph> | 3 |
| 2.3 | <gimel-daleth-yodth> + <gimel-daleth> vs <samekh> | 3 |
| 2.4 | <pe> vs <sadhe> | 4 |
| 3 | Miscellaneous issues | 4 |
| 3.1 | Joining of <aleph-heth> | 5 |
| 3.2 | Missing alternate form of <gimel-daleth-yodh> | 5 |
| 3.3 | Inclusion of <HE> | 5 |
| 3.4 | Joining of <zayin> | 5 |
| 3.5 | Old lamedth | 5 |
| 4 | The dogma of shape-shifting and the problem of good-enough | 5 |
| 5 | Bibliography | 6 |

1 Introduction

It has been a source of delight that after a dormant period of four years, since the submission of my proposal to encode Book Pahlavi in the Unicode

*abraham.meyers AT orientology DOT ca

standard, there has been some renewed activity in the community. The recent preliminary proposal by Dr. Anshuman Pandey (L2/18-276) might therefore signal a resurgence of activities towards the noble goal of encoding of Book Pahlavi in the Unicode standard.

I started reading the work of Dr. Pandey with enthusiasm and in anticipation of further improvement and suggestions and perhaps “discovery” of new characters. It was indeed pleasant to see a relatively thorough classification of the visual joining of the stem of the characters of Book Pahlavi, while taking the base-line into consideration. Such studies will be very beneficial for the future type designers of Book Pahlavi—although I have doubts about the applicability of this study to the level of abstraction pertaining to the Unicode standard.

Furthermore, the recognition of the need to include the character `<alternate sin>`, as first suggested by L2/14-077, was a source of satisfaction for the present author.

The case for the inclusion of “old lamedth” may also be considered a step forward, provided that the author of L2/18-276 supplies sources for the usage.

Unfortunately, this is as far as one can go in granting somewhat of an unqualified support to this proposal. Upon reading the proposal, one quickly suspects that Dr. Pandey has not consulted, as thoroughly as one might have hoped, the previous proposals on this subject. As a trivial example, one sees omissions of characters that have been identified in manuscripts by L2/14-077 (e.g. characters 10BD7, 10BB5).

More importantly one sees many retrogressions that could have been avoided had more attention been paid to the existing works. In this critique, I shall respectfully identify some of the shortcomings that I perceive with Dr. Pandey’s proposal. I shall avoid mentioning issues that have already been pointed out by others at the Unicode Technical Committee sessions, for the *most* part.

2 Multiple incompatible representations

The gravest shortcoming of L2/18-276, is that the proposed encoding schema by Dr. Pandey, lends itself to multiple incompatible representations for a single piece of writing (say a word) in Book Pahlavi. Such possibility is going to be nothing short of disastrous for any digitization efforts of the extant Book Pahlavi material. One need not even look outside L2/18-276 to find full words that lend themselves to multiple incompatible representations as

we shall see below.

Before giving a more detailed account of a number of such cases—as there could be more classes of such cases—I emphasize once again that this class of problems is **not** inevitable. Had the author based his work on L2/14-077 such problems could have been easily avoided.

2.1 <gimel-daleth-yodh> + <shin> vs <aleph-heth> + <aleph-heth>

In the proposed encoding by the author of L2/18-276, the combination <gimel-daleth-yodh> + <shin> results in an identical shape to that of <aleph-heth> + <aleph-heth>. Interestingly enough, an example of such identity for a full word can be seen in the proposal document itself. In section 6.4 the word *Myš* (sheep) can also be read as the word *Mah* (moon), shown in section 6.12 (p. 28), leading to two different encodings.

2.2 <Fixed-aleph> + <gimel-daleth-yodh> vs <fixed-gimel-daleth-yodth> + <aleph>

A similar case as to the above holds for this combination. <fixed-aleph> + <gimel-daleth-yodh>, cannot be distinguished from <fixed-gimel-daleth-yodth> + <aleph-heth>. This in turn leads to the possibility of multiple incompatible representations of a single piece of writing.

Furthermore, the fact that the author had to resort to the use of <fixed-aleph> and <fixed-gimel> characters should already be considered a warning sign that is indicative of deeper problems in the design.

2.3 <gimel-daleth-yodth> + <gimel-daleth> vs <samekh>

This one is the result of the classic mistake of including a digraph as a separate character. It is contrary to the Unicode Technical Committee recommendations. The extra curving that the parts have, in the font used by Dr. Pandey, have no semantic meaning in Pahlavi manuscripts and cannot be used as a means of disambiguation. A simple look at the manuscripts or glossaries should settle the case. For example one sees that in Mackenzie (p. 165) “deg” and “sag” are the very same entry. In Dr. Pandey’s encoding, the relevant part of “deg” is represented as <gimel-daleth-yodth> + <gimel-daleth-yodth>, and that of “sag” as <samekh>. Again, this is a major source of trouble for any digitization effort. One cannot even digitize that entry of Mackenzie using this proposed encoding, without making some arbitrary choices.

The following note by Nyberg (p. 137) should dispel any doubts that might have remained on the fact that <samekh> and <gimel-daleth-yodth> + <gimel-daleth> should be treated as identical:

As diacritical signs are used when [<samekh> and <alternate-samekh>] *represent the ligature* [(emphasis is mine)] of [<gimel-daleth-yodth> + <gimel-daleth-yodth>] every unmarked [<gimel-daleth-yodth> + <gimel-daleth-yodth>] should be referred to section 9 [even if it does not represent the phoneme s].

Essentially, one does **not** rely on extra curving and minor variations to distinguish <gimel-daleth-yodth> + <gimel-daleth-yodth> from <samekh> but rather from diacritics. Without the diacritics <gimel-daleth-yodth> + <gimel-daleth-yodh> should be treated as identical to <samekh>. See Nyberg (p. 133, Note 7) for more examples.

2.4 <pe> vs <sadhe>

the transliterated letters **p** and **c**, frequently have identical representations in Pahlvi manuscripts when they are preceded by certain characters. Examples of this identity can readily be inferred even by looking at some of the samples that are provided in the L2/18-276 by Dr. Pandey. For example, one can see that following the joining behaviour put forward in that document, <gimel-daleth-yodth> + <pe> will be identical to <gimel-daleth-yodth> + <sadhe>.

I mention in passing that the joining behaviour of <pe> (if we consider it a separate character as Dr. Pandey has) is not even consistent in Pahlvi manuscripts. For example, what is usually transliterated as “lp” can look both like what L2/18-276 has put forward in section 6.14, or it can look like what is transliterated as “lc” as it is shown in section 6.15 (See Nyberg, p. 132).

3 Miscellaneous issues

The fundamental issue discussed in the previous section should cast serious doubt on the suitability of the proposal in its current form. In this section, I shall list some of the less significant issues with the proposal.

3.1 Joining of <aleph-heth>

The curving of the final stem of <aleph-heth> when preceding <taw>, <pe> or <sadhe> has no semantic consequence, and carries no extra information. This is purely an aesthetic issue and not a matter of an obligate ligature.

3.2 Missing alternate form of <gimel-daleth-yodh>

The proposal is missing the alternate form of <gimel-daleth-yodh>, in which, the bulge at the start of the character is missing. For an example refer to the table provided by Nyberg (p. 132), where the joining of <gime-daleth-yodh> to itself is discussed. the second <gime-daleth-yodh> can have either of the two variants.

3.3 Inclusion of <HE>

The Aramaic HE is clearly a digraph and is even treated as such in all major Pahlavi glossaries, (as an example see MacKenzie and Nyberg). The inclusion of the digraph as a separate character because of the perceived “semantic value” is untenable. It is in clear violation of the unicode recommendation. What the proposal is advocating for, is akin to inclusion of “sh” for the phoneme š , because of its semantic value.

3.4 Joining of <zayin>

The curving of the final stem of <zayin> has no semantic consequence and carries no extra information.. This is purely as aesthetic issue and not a matter of an obligate ligature.

3.5 Old lamedth

Citations or justifications are needed for this new proposed character. One should also consider if this issue can be dealt with through the alternate glyph mechanism at the font level, as we already have a variant of lamedth that is purely used for heterograms.

4 The dogma of shape-shifting and the problem of good-enough

The conflation of the realm of phonology and that of writing systems is a major barrier to the suitability of encoding schemas. With the exception

of L2/14-077, all proposals to encode Book Pahlavi have suffered from this flaw. It is again this conflation that has led the respective authors down the rabbit hole of introducing complicated shape-shifting conditions, and introducing novelties such as fixed-form characters.

Interestingly enough, despite all the complexity introduced in this proposal, it is still conceded that the encoding is suitable for representing [only] the “typical” renditions of Book Pahlavi [and not presumably pieces of texts with idiosyncrasies] (See L2/18-300) . It is my hope to have shown that unfortunately in its current form, this less ambitious goal cannot be met either.

On a more general note, such concessions would not have been necessary had Dr. Pandey based his work on L2/14-077, which can reliably and uniquely encode the vast majority of the extant Pahlavi material if not all (the present author is not aware of such examples).

Finally I should emphasize that perhaps with the exception of <partial-shin> character in L2/14-077, all the other alphabetical characters in that proposal L2/14-077 correspond to a well-defined phonemes in Pahlavi. Hence, it should be clear that I am by no means suggesting that phonology should be thrown out the window entirely, just because one is interested in the writing system.

5 Bibliography

- Mackenzie, DN. Concise Pahlavi Dictionary, Oxford University Press. 1986
- Nyberg, HS. A Manual of Pahlavi I, Tehran: Asatir.