Iranica Antiqua, vol. LVII, 2022
doi: 10.2143/IA.57.0.3291506

ON NEWLY PROPOSED PROTO-ELAMITE SIGN VALUES

BY

Kathryn KELLEY[1], Logan BORN[2], M. Willis MONROE[3], Anoop SARKAR[2]
([1]University of Bologna, Bologna, Italy; [2]Simon Fraser University, Burnaby, Canada; [3]University of British Columbia, Vancouver, Canada)

Abstract: Proto-Elamite (PE) is a mostly undeciphered writing system of late 4th / early 3rd millennium Iran. Although known to be a bookkeeping system that employs many "object" signs and numerical notations, it has long been hypothesized that the script could contain some signs which functioned as a "syllabary", which would make PE a very early example of glottographic writing.

A recent article on the decipherment of the much-later Linear Elamite (LE) script (Desset et al. 2022) proposed a set of correspondences between the LE signs and graphically similar PE signs, suggesting that future work should explore whether LE sign values may be successfully applied to PE signs to yield a greater decipherment. The current article applies the values of Desset et al. to the existing PE corpus to begin to test this theory. Contextual information in PE facilitates examining whether some LE signs and values may be of phylogenetic relationship to PE ones. The article explores the PE signs and sign sequences in question, offering observations on patterns of use with implications for future work towards investigating hypothetical sound values for PE signs.

Keywords: Proto-Elamite, Linear Elamite

Introduction

In a recent article in *Zeitschrift für Assyriologie*, Desset et al. (2022)[1] propose that the "proto-Elamite and Linear Elamite scripts were probably not two different writing systems, but the same system at two different chronological stages of evolution", and that one might:

> proceed in a regressive way, starting from the vocalic, consonantal, and syllabic values established for the LE signs (see Figs. 6 and 10), and trying to apply these "readings" to their graphic counterparts in

[1] Henceforth Desset et al. refers to this paper.

> the earlier PE writing (see Fig. 7a-c). The same signs may have been used with similar or identical phonemic values to record the names of the persons involved in the transactions and administrative work documented in the late 4th millennium BCE PE tablets
>
> *— Desset et al. 2022: 53*

The authors provide a chart with correspondences between Linear Elamite (LE) and some proto-Elamite (PE) signs, along with proposed sound values. No description is given for the methodology by which signs in the two scripts were equated. The authors do not specify whether the PE signs are hypothesized to maintain the same sound values as the LE correlates in their chart for each occurrence, and they make no claim to present a complete list of an original PE (alpha-)syllabic sign set. The figure rather stands to present a renewed proposal for possible sign equivalencies and (to an unclear extent) sign values.

The full LE corpus itself is not yet available for study in transliteration using the readings of Desset et al. (2022). However, the much larger PE corpus is readily available in a working version of transliterations in nearly its entirety. For this study, the newly proposed LE-PE sign equivalencies along with the LE sound values are directly applied to the PE corpus to begin to investigate their plausibility within PE. Our initial mapping is strategically simplistic (see Commentary below) – the proposed values are applied to all the sign instances in PE regardless of contextual cues or plausibility; the separation of candidate personal name (PN) sequences, and the distinction between "syllabic signs" and other sign types has been a continual project within PE studies for decades. Our current approach serves to highlight the broader use patterns of the signs in question and represents a first step in assessing the equivalencies and values of Desset et al.

An assumption of this analysis is that Desset et al.'s sound values are substantially correct in the context of LE writing; however, the mapping to PE remains to be demonstrated; it would undoubtedly be nuanced in ways that could not be fully captured by Table 7 of Desset et al. 2022. The following work provides an analysis of this *particular* assignment of sounds to PE signs.

Although Desset et al. refer to a "genetic link between Early (PE) and Late (LE) Proto-Iranian writing and the consequent continuous tradition of writing in Iran", current evidence indicates a large gap in time between the PE and LE writing systems, which poses a challenge to the hypothesis that

LE scribes would have intimate knowledge of the PE accounting system, including sign meanings and possible sound values. Additionally, it is known that proto-cuneiform writing of the late 4th and early 3rd millennium was largely, if probably not entirely, non-glottographic (most recently, Krebernik 2021). The contemporary PE system shows evidence of some similar bookkeeping mechanisms, and the commonly-held hypothesis that a sub-set of PE signs functioned as a syllabary remains as yet unproven.

With these cautions in mind, this work presents the full proto-Elamite corpus with the proposed sign values of Desset et al. inserted, in order to support scholars in examining the mapping. Also provided are selections of the data in tables with some summary comments and initial analysis. The data presented has been derived from the proto-Elamite corpus prepared and hosted by the *Cuneiform Digital Library Initiative*, based on a working sign list created by J. Dahl. The fully remapped data can be explored on github at https://github.com/sfu-natlang/pe-sign-value-data.

Selections from the corpus

Several selections from the dataset may facilitate assessment of the proposed values. Table 1 shows the overlap between the proto-Elamite signs in Tables 7a-b of Desset et al. and the partially-identified proto-Elamite syllabary in Dahl 2019. Dahl conceives of around 100 signs in his (still hypothetical) syllabary, but identifies only 61 of them by name (those occurring at least 10 times, Dahl 2019: 47 Fig. 13). Dahl 2019 identified these signs as being part of a potential syllabary but did not hypothesize values for these signs. 22 signs occur both in the list by Desset et al. and in the list by Dahl, and 39 signs are identified as possible syllables by Dahl but not by Desset et al. 70 signs are identified by Desset et al. and are not included in Dahl's published sample of syllabic signs. Of these, some rare (<10 attestation) signs may be considered candidates for syllabary by Dahl, while others are definitely not (Dahl, personal communication). Seventeen further PE signs (Fig. 7b and 7c) are given a mapping to LE but no reading by Desset et al. and thus may not necessarily be conceived of as indicating sound values by those authors.

We next extract every contiguous span of 2 or more values drawn from the table in Desset et al. (2022: Fig. 7), in the interest of linguistic analysis. We present these as a "word list" in Table 2 together with the number of times each span occurs. Though this is called a word list we remain agnostic

as to whether these are truly complete *words*, or if the involved signs function to indicate sound values. This table only includes the most common spans (see github for the complete list including links to texts where each span occurs) and it removes them from any surrounding signs which have not been included in Desset et al.'s mapping. For example, the sequence *pu-sha-a* (M347 $M377_e$ M218)[2], occurring four times, originates from the following full proto-Elamite text entries:

MDP 6, 376 / P008157: $M111_e$ M388 M347 $M377_e$ M218 $M36+1_{N30D}$, 2_{N1}
MDP 17, 233 / P008431: M347 $M377_e$ M218 M288, 2_{N30C}
MDP 17, 445 / P008643: M347 $M377_e$ M218, ⌜6_{N1}⌝
MDP 26, 231 / P008919: ⌜$M111_a$⌝? M388 M347 $M377_e$ ⌜M218⌝ M36, 2_{N1}

The context of this particular string indicates it might be a "personal name", which is sometimes followed by an object sign (in these cases providing information on the capacity measures that follow), and once preceded by the candidate person/class-marker M388. Specialists have hypothesized that if PE does contain syllabic sequences, they are likely to be personal names. Promising spans in table 2 should always be checked against tablet context, as demonstrated here, to establish whether a personal name is plausible in each instance.

For comparison, Table 3 presents contiguous sequences of the common syllabary signs presented in Dahl 2019 (similarly de-contextualised), on the assumption that the graphotactical analysis undertaken by that author (as partly outlined in Dahl 2019: 82-84) has identified a strong set of candidate syllabic sequences to analyze. Dahl does not propose readings for these signs, so this table maintains the sound values from Desset et al., and preserves the original PE sign name if Desset et al. do not assign it a value.

Tables 2 and 3 exclude sequences in which proposed syllables are interspersed with other signs, appear next to unreadable broken signs, or appear as the only sign in an entry (proto-Elamite entries are explicitly delimited by numerical notations). These situations happen with considerable frequency,

[2] The notational system for proto-Elamite used here differs from the CDLI conventions, the latter which are retained in the github data. Thus the traditional publication form $M377_e$ refers to M377~e in the corpus data; half-brackets for broken signs appear there as a following #. Sign numbers lower than 100 are preceded with 0 or 00 (e.g. M002 not M2); and numerical notations are marked in the format 1(N01) rather than $1N_1$.

the latter resulting in single-syllable or single-consonant "words" which are difficult to interpret without tablet-wide context. To highlight this, all cases where a sign from Desset et al.'s table occurs as a single-sign entry are separated, and presented in Table 4.

Tables 2 and 3 present only a "brute" measure of syllabic sign use plausibility, since PNs in PE might have been written using a mixed logographic-syllabic system such as that familiar from cuneiform. In Table 5, the sequences from Table 2 are extended to include any adjacent sign for which no mapping is given; the unmapped sign can occur at the beginning or end of the sequence, or in between two distinct sequences from Table 2. This gives some flexibility in assessing proposed values that might appear in the context of as-yet to be identified syllabic signs, logograms or other sign types. Table 2 thus reveals cases such as *zu*-M318$_{a1}$-u_2 where a sign with no reading may be an unrecognized part of a longer syllabic sequence. Rather than assessing this table in its entirety, it may be more fruitful to analyze focused subsets; Tables 5a, 5b, and 5c suggest excerpts of possible interest. For comparison, Table 6 applies the same treatment to the common syllables from Dahl 2019: it thus records sequences of Dahl's proposed syllables, plus one adjacent sign which does not occur in Dahl's list.

Finally, Table 7 has a list of all complete PE entries (excluding those that contain only one sign) which are readable in their entirety using the sound values in Desset et al. 2022. These entries contain no breaks and no signs without a proposed reading. Tablet-wide context may still be important for interpreting these entries. Since PE entries are separated by numerals, a word boundary at the beginning and end of each of these strings can be assumed (though additional word boundaries may occur elsewhere within the entry). As complete and unbroken strings, these entries should provide the cleanest and most unambiguous test of the extent to which the sound values from Desset et al. 2022 can be applied to proto-Elamite to yield readable words or phrases – so long as one checks tablet context for clues such as the presence of probable object signs in the final position.

This brief study presents abridgments of each table; the complete tables are available on github in the tables/ directory.

Commentary

In Table 1, some of the differences between the two proposed syllabaries reflect the inclusion or exclusion of signs annotated by Dahl with alphabetic subscript. For example, Dahl includes M223 where Desset et al.

include instead the visually-similar $M223_b$. Dahl's sign list is agnostic about the relationship between signs with the same number but differentiated by letters (whereas additional numbers indicates a suspected graphical variant, e.g. $M269_3$ or $M269_{a1}$). Our own research has suggested that a certain number of these alphabetic variants could be merged while others should be kept distinct (Born et al. 2019; Born et al. 2021).

Desset et al. identify certain pairs of PE variants as mapping to distinct LE signs: for example, M220 is equated with LE su_2, and $M220_n$ with LE *ra*. In other cases, two wholly distinct signs from Dahl's list have been equated, such as M2 and M387, both associated with LE *na*. M2 and its variants often occur at the end of entries, and it may function to introduce capacity measures in a derived numerical system (Dahl 2019: 77). M387 and letter variants can also occur in final position, but are additionally attested in a wide variety of other locations; it has been described as both an "owner sign" and a possible syllabic sign (Dahl 2019: 77, 84). M2 and M387 can be used in the same texts, and possibly even the same entry (MDP 17, 271 / P008469). Neither the sign clustering methods applied in Born et al. 2019 nor the PE sign embeddings obtained by a multimodal language model in Born et al. 2021 show evidence of a close relation between these signs. There are thus compelling contextual cues to resist a merger of M2 and M387 in PE, and it is not clear what encourages Desset et al. to correlate both of these signs to the same later LE sign.

A feature of note in the syllabary of Desset et al. is the presence of eighteen PE *hapax legomena*. Many of these have been suggested to group with other signs (with greater or lesser impact on the frequency: $M41_c$ is merged with M41 attested 17 times and $M214_c$ attested twice; $M134_b$ is merged with M134 itself attested twice) while some remain unique (e.g. M253, LE correlate assigned the value *ki*, M243 *l* and $M302_e$ li_3). It would be quite remarkable if a sign with a single attestation across 26 thousand tokens in the currently known corpus (approximately 12,000 numerical signs; 14,000 non-numerical signs) were to prove resilient enough to be transferred with a similar value into a much later writing system. Seven further signs ($M257_a$, $M218_a$+M101, M483, M344, $M316_e$, $M218_a$+M218, $M57_{a3}$) do not occur in the CDLI corpus, some of them deriving from earlier versions of the CDLI transliterations (M483 has been re-interpreted as $M346_d$, or two M346 "sheep" signs superimposed, appearing as an object sign in the livestock account P009524; Dahl, personal communication).

Turning to Table 2, some 2- or 3-sign sequences recur with moderate frequency (up to 23 times for M4 M218). However, there is a rapid falloff, and in the full table 230 of the 340 total sequences occur only once. A common problem in decipherment is that even a random assignment of sounds to signs will produce some short, readable "words". It is much more challenging to find an assignment which produces long words or interpretable phrases. Unlike the contiguous syllabic representation of language in LE, syllabic sequences in PE are typically hypothesized to exist primarily for isolated proper nouns, themselves potentially represented in a mixed logographic-syllabic way. It may also be worth considering the extent to which signs functioned as phonetic complements to logograms, as has been suggested for certain combinations in proto-cuneiform. We should be wary to over-rely on short and rarely repeated strings for the purpose of decipherment, though to some extent this will be necessary given that PE contains very few long, repeated strings to act as guides (Born et al. 2019). This fact is emphasized by comparison to Table 3, where even the PE specialist's syllabary fails to produce many long or frequent "words". Many distinct singularly-attested candidate PNs (*i-ki*$_2$*-pu-ri*$_2$, incorporating a possible Elamite *igi* "brother"?) should appear across the corpus if they are to provide significant evidence for the proposed mappings. Although our initial survey of the results does not identify a critical mass of such candidates, a further analysis is needed.

In Table 2, there are only two 4-sign sequences which occur more than once: *la-ha-a-ha* is found in P009032 and P008987, realized by the PE signs M4 M263 M218 M263$_{(a)}$, and *sha-pu-u*$_2$*-pu*$_2$ occurs in P008184, P008260, and P008688 realized by M377$_e$ M347 M371 M288. However, these strings highlight how existing knowledge about proto-Elamite "object signs" reduces the number of sequences that could reasonably be hypothesized to be read as full sequences of sound values (as explained below). Object signs have previously been identified by studying tablet format, summary lines, and other clues.

Both P009032 and P008987 include multiple other entries ending in M263$_{(a)}$, and P009032 includes an entry counting M263 by itself, indicating that the final M263 in the 4-grams is a counted object sign. These texts also contain unambiguous capacity measures. Together with the fact that M263 resembles a vessel, in context this sign appears to be a counted object containing some product quantified in the capacity system, and therefore not a final syllable *ha*. The same is true of the *ha* in *me-lu-ha*

(P008239 in Table 7) – not, it seems, an instance of the later geographical name. However, M263 with a syllabic reading may be likely in the middle of the above sequence (M4 M263 M218, or *la-ha-a*) and in other texts (Dahl 2019: 84 cites 11 texts).

The frequency of strings with pu_2 (including *sha-pu-u*$_2$*-pu*$_2$ noted above) reflects the prevalence of the object sign M288, very often occurring at the end of PE entries. There are indicators within the PE texts themselves that nearly all uses of this sign are about distributions of grain (Dahl 2018), and the sign itself may depict a kind of grain container. If M288 has a syllabic use, it likely comprises only a fraction of its many occurrences. Thus pu_2 in Tables 2 and 7 should perhaps be ignored in most cases, especially where it occurs at the end of a string (the vast majority of instances, including multiple instances of *na-pu*$_2$, certainly not reflecting Elamite /nap/ "god"). Therefore, taken at face value, the mapping by Desset et al. produces no strings of candidate 4-sign "words" with more than a single occurrence.

Desset (2012: 27) is aware of the conventional understanding of M288, but has suggested that M288 may sometimes be used "in a phonetic way (rebus principle)", citing text MDP 7, 246 / P008444 (Desset 2016: 84). In that fragmentary text, M288 appears near the beginning of a sequence of signs which constitutes the final entry in the text; the poor state of the text make analysis difficult (recent imaging by the CDLI reveal no traces of the first few text entries on the reverse as drawn by V. Scheil). Another cited possible use in an anthroponymic sequence (Desset 2012: 62 table 6) in P008134 probably employs a modified form of M288 (Dahl's M288~c, which may be described as M288-*gunû* in Mesopotamian terms), and appears in a text with a long single entry and subscript. Although M288$_c$ here does appear after M388 (a hypothesized person-marker), the entry also ends in M288, thus the sequence may be an example of a poorly understood "framing" behaviour that occurs particularly in some long first entries. Thus evidence for syllabic use of M288 remains scant.

In terms of the LE-PE correspondence, M288 is arguably not a natural graphical comparison to the relatively rare Linear Elamite sign: early forms of cuneiform GUM would be closer to most of the attested LE forms. Unfortunately, the Linear Elamite sign is missing on the "school text" syllabary presented in Desset et al. 2022: 38. Given the identification of M288 as a grain container, the sound value /pu/ may evoke a connection to *pa-u-mi-ráš* "ein Kornspeicher" or *pu-pu-ma* "zum Einfüllen(?)" (Hinz & Koch 1987: 174, 235), except that the former is of apparent Median origin and the latter known only from Achaemenid texts.

Browsing the complete table, one does observe many apparent "plene" spellings, i.e. sequences with matching vowels (*ne-e*, *ri$_2$-i*, *la-a-ha*). However sequences of adjacent non-matching vowels also appear. Specifically, there are 20 unique matching (C)V-V sequences attested 111 times in total, and 49 mis-matching sequences attested 64 times in total. On the one hand the matching vowel sequences may seem promising, but the implications of such proposed spellings must be thoroughly examined with reference to the possible types of writing systems that would employ them – this is beyond the scope of the present note. Previous work by Desset (2016: 82-87) expressed interest in PE "anthroponymical sequences", and it is possible that the choices made for PE-LE equivalencies in Desset et al. (which could be selected from a large number of similar-looking PE signs) were guided in part by an attempt to produce realistic looking sound sequences in the PE corpus. It may be notable that the most frequently occurring (C)V-V sequences in the corpus over-represent "matching" vowels – the highest being *la-a* appearing 23 times independently (as well as in longer strings) – whereas the non-matching sequences tend to be of less frequency each, such as *la-i* appearing only twice independently, *la-u* once, but also *lu-u* once (refer to the complete data of table 2). Finally, without as yet drawing conclusions from this fact, the proposed "vowel" signs in LE are correlated with PE signs that are among the most frequently attested in the corpus[3] (frequencies obtained at https://mrlogarithm.github.io/pe-pc-datasets-interface/pe.html):

Reading	Sign	Attestations	Frequency Rank
a	M218	525	3
e	M96	212	13
i	M66	243	10
i	M131	22	123
u	M57	95	28
u	M57$_{a3}$	(not attested in our corpus)	
u$_2$/w	M371	309	4

Other sound combinations seem less promising. In one text (P008702), the proposed sound values yield the string *m-t* from PE M246 M374$_c$. This is preceded by M157, a "header" sign which is not likely to have a syllabic

[3] Note some of these signs (excepting unattested M57$_{a3}$) may function as object signs in PE (see table 4) in addition to any other hypothetical sound-value uses.

reading, and followed immediately by a numeral. The assigned values thus seem to spell a phonetically unlikely word /mt/. Similarly, M218+M288 M4 M218, constituting an entire entry in P009179, receives the reading *m-la-a*. In order to make each of these sequences plausible, modifications to some of the values would be required. Similar examples in other texts show that the proposed LE readings cannot be directly ported to PE without producing some unlikely consonant clusters. Notably, most (though not all) of these clusters involve LE *m*, which Desset et al. recognize as the correlate to *8* distinct PE signs (more than twice as many as any other sign), and which exhibits a significant degree of graphical variation even in Desset et al.'s LE signary.

Table 4 highlights signs from Desset et al.'s list which occur as the only sign in an entry. In such entries, the lone sign is typically read as a counted object, and thus is presumably not used for its sound value. In total, the table records well over 700 entries comprising lone signs from Desset et al.'s list, including 322 instances of pu_2, reflecting M288's established use in labeling capacity measures. 61 PE signs from Desset et al. (2022: Fig. 7a-b) are used this way, out of a total of 86 PE signs with LE correlates given sound values in their list. This raises the question of the existence and frequency of polyvalence in PE, with signs functioning with sound values in some contexts and as object signs in other contexts. Such a phenomenon would be in line with long-standing theories about the mechanisms for expanding sign uses in early Near Eastern writing (Englund 2001: 101; Englund 1988: 131 fn. 9). However, it also remains possible that the signs in question are *always* "object" signs or similarly non-glottographic.

Conclusion

As further research revises the proto-Elamite signlist and clarifies PE sign meanings, the conventional transliterations of the corpus will need to be amended. A process of selectively applying sign values to promising sequences may continue in the amended corpus, guided by increasingly better understandings of the texts and knowledge of contextual information. This study makes the full corpus available for study with the newly-proposed LE-PE sign equivalencies, with the aim to allow specialists to analyze the full data, to collaborate on understanding the PE corpus, and to form their own conclusions about the PE-LE correspondences observed by Desset et al. 2022.

While overall conclusions on the proposed LE-PE sign value correspondences are premature at this point, the following initial observations are most notable: There are significant differences between the proposed syllabaries of Dahl (2019) and the signs of Desset et al. (2022: Fig. 7a-c). Some of these differences highlight the ongoing need to revise the PE signlist with better understanding of the relationship between alphabetic variants. A sizeable minority of the signs identified in Desset et al. 2022 are hapaxes or rare signs (even occasionally non-existent in the current CDLI corpus), and their transfer into LE with similar sign values can only be supported by substantial linguistic evidence from the PE sign sequences. Targeted studies of possible sign values should focus on sequences selected by contextual clues within PE. At the same time, the limited set of repeated n-grams (which becomes greatly narrowed as object signs are discluded from potential syllabic sequences) continues to pose a challenge to the linguistically-based decipherment project. Finally, the majority of newly proposed sign correspondences (61 of the 86 total) refer to PE signs which can stand independently in entries, indicating that if sound values do apply to these signs, they are likely to do so only in some instances.

The presence of a syllabary in the PE texts remains an intriguing hypothesis, and more work is needed in order to delineate the most likely "syllabic" sequences, as well as to understand the relationship between the hypothetical syllabary signs and the overall nature of the PE writing system.

Tables

Table 1

Overlap between the proposed proto-Elamite syllabary in Dahl 2019 and the signs with LE correlates identified in Desset et al. 2022. A tick indicates that the sign is included in the respective list; a question mark indicates that the sign is rare (<10 attestations) and thus not included in Dahl's list, which only shows the most common of the hypothesized syllables. Sign images courtesy of J. Dahl via the CDLI.

Sign	Sign Image	Dahl 2019	Desset et al. 2022
M1		✓	
M2			✓
M4		✓	✓

Sign	Sign Image	Dahl 2019	Desset et al. 2022
$M29_b$		✓	
M32		✓	✓
$M32_a$			✓

Sign	Sign Image	Dahl 2019	Desset et al. 2022
M6			✓
M9		✓	✓
M24		✓	✓
$M24_a$			✓
M29			✓
$M48_k$		✓	
M57			✓
$M57_a$		✓	
$M57_{a3}$			✓
$M57_b$		✓	
M58			✓
M63			✓
M66		✓	✓
M96		✓	✓
$M97_h$		✓	
M99		✓	✓
M101		✓	
M134			✓
$M134_b$			✓
$M143_d$			✓

Sign	Sign Image	Dahl 2019	Desset et al. 2022
M33		✓	
M35		✓	
M41			✓
$M41_c$			✓
$M48_c$		✓	
$M53_a$			✓
$M102_d$		✓	
$M105_{ab}$			✓
M106		✓	
M106+M288		✓	
$M107_a$			✓
M109		✓	✓
M110		✓	
M122			✓
M129			✓
M131			✓
M131+M388		✓	
M218+M288		✓	
$M218+M288_f$			✓
M218+M320		✓	✓

Sign	Sign Image	Dahl 2019	Desset et al. 2022
M155			✓
M157$_{a}$			✓
M175			✓
M204$_{g}$			✓
M205$_{a}$			✓
M206$_{b}$			✓
M214$_{c}$			✓
M218		✓	✓
M223		✓	
M223$_{b}$			✓
M224$_{a}$			✓
M226$_{c}$			✓
M226$_{ca}$			✓
M228$_{b}$		✓	
M228$_{ga}$		✓	
M230		✓	
M230$_{a}$		✓	
M240		✓	
M240$_{c}$		✓	
M257$_{a}$			✓
M259		✓	✓

Sign	Sign Image	Dahl 2019	Desset et al. 2022
M218$_{a}$+M101			✓
M218$_{a}$+M218			✓
M218$_{b}$		✓	
M219		✓	✓
M220		✓	✓
M220$_{n}$			✓
M221		✓	
M222		✓	
M242$_{b}$		✓	
M243$_{e}$			✓
M243$_{g}$			✓
M243$_{i}$			✓
M246$_{b}$			✓
M246$_{h}$			✓
M248			✓
M248$_{a}$			✓
M250$_{ba}$		✓	
M253			✓
M254$_{a}$		✓	
M295$_{ka}$			✓
M296		✓	

Sign	Sign Image	Dahl 2019	Desset et al. 2022
M262		✓	✓
M263		✓	✓
$M263_{a}$			✓
$M278_{c}$			✓
$M281_{c}$			✓
$M286_{a}$			✓
$M286_{b}$			✓
M288			✓
$M288_{f}$			✓
M318			✓
$M318_{a}$		✓	
$M319_{c}$			✓
M344			✓
M347			✓
M348+M4			✓
M352n		✓	✓
M352o		✓	
M356			✓
M371		✓	✓

Sign	Sign Image	Dahl 2019	Desset et al. 2022
M296+M296		✓	✓
M297			✓
M298			✓
M301		✓	✓
$M302_{e}$			✓
$M308_{e}$			✓
M315		✓	
$M316_{a}$			✓
$M316_{e}$			✓
M372		✓	✓
$M374_{c}$			✓
$M377_{e}$		✓	✓
M380		✓	
$M386_{a}$		✓	
M387		✓	✓
$M387_{a}$		✓	
$M387_{ef}$		✓	
M390		✓	
M391			✓

Sign	Sign Image	Dahl 2019	Desset et al. 2022
$M371_c$			✓
M455			✓
M483			✓

Sign	Sign Image	Dahl 2019	Desset et al. 2022
M419		✓	
M471			✓

Table 2

20 most common sound sequences resulting from the application of values in Desset et al. 2022. Some of these might be parts of longer "words" for which some signs do not have proposed syllabic or logographic readings. Complete table available at https://github.com/sfu-natlang/pe-sign-value-data/tree/public/tables/wordlist_desset.md

Sequence	Original Signs	Number of Attestations
la-a	M4 M218	23
e-pu$_2$	M96 M288	23
zu-u$_2$	M9 M371	16
ha-a	M259 M218 M263 M218 $M263_a$ M218	16
i-pu$_2$	M66 M288	15
u$_2$*- pu*$_2$	M371 M288	14
sha-pu-u$_2$	M377 M347 M371 $M377_e$ M347 M371	13
na-a	M387 M218	12
a-pu$_2$	M218 M288	11
a-ha	M218 M259 M218 M263 M218 $M263_a$	10

Sequence	Original Signs	Number of Attestations
m-pu$_2$	M218+M288 M288 M224 M288 $M246_b$ M288	8
pu-u$_2$	M347 M371	7
i-ki$_2$	M66 $M352_n$	7
i-i	M66 M66	7
i-ha	M66 M259 M66 M263 M131 M263	7
a-ta	M218 M219	7
ta-a	M219 M218	6
pu-ta	M347 M219	6
na-pu$_2$	M387 M288	6
i-a	M66 M218 M131 M218	6

Table 3

20 most common sequences of syllabic signs from Dahl 2019, applying any sound values suggested by Desset et al. 2022. Note that *la-a* appears 16 times in this table (row 1, column 1) versus 23 times in table 2. This is because Dahl considers $M97_h$ a potential syllabic sign whereas Desset et al. do not, so that seven further uses of *la-a* are accounted for within the string $M97_h$ M4 M218 in this table (row 1, column 2). Complete table available at https://github.com/sfu-natlang/pe-sign-value-data/tree/public/tables/wordlist_dahl.md

Sequence	Original Signs	Number of Attestations
la-a	M4 M218	16
ha-a	M259 M218, M263 M218	16
zu-u_2	M9 M371	14
na-a	M387 M218	9
ta-a	M219 M218	8
i-ki_2	M66 $M352_n$	8
i-$M352_o$	M66 $M352_o$	8
sha-a	$M377_e$ M218	7
i-i	M66 M66	7
$M240_e$	M240 M96	7

Sequence	Original Signs	Number of Attestations
$M97_h$-*la-a*	$M97_h$ M4 M218	7
hu$_2$-u_2	M99 M371	6
a-ta	M218 M219	6
a-ha	M218 M259 M218 M263	6
$M242_b$-*e*	$M242_b$ M96	6
u_2-*zu*-u_2	M371 M9 M371	5
M390-a	M390 M218	5
$M387_{ef}$-*ha*	$M387_{ef}$ M263	5
$M386_a$-M240-*e*	$M386_a$ M240 M96	5
M380-u_2	M380 M371	5

Table 4

PE signs with LE correlates in Desset et al. 2022 which occur as the only sign in an entry. Complete table available at https://github.com/sfu-natlang/pe-sign-value-data/tree/public/tables/singletons_desset.md

Sound Value	Original Signs	Single-Sign Attestations
pu$_2$	M288 M288$_f$	322
ri$_2$	M297 M298	104
zu	M9	81
na	M2 M387	61
ha	M263 M263$_a$	42
p	M6	31
she	M32	26
a	M218	25
m	M218+M288 M218+M320 M224 M243 M243$_e$ M243$_i$ M246 M246$_b$	20
nu	M391	19
h$_2$	M122	13
u$_2$	M371	9
pi	M53$_a$ M105$_{ab}$	9
p$_2$	M286 M286a M286$_b$	9
u	M57	8
ni	M41	2
z	M129	1
su	M205	1
ri	M63	1
pu	M347	1
me	M206$_b$	1

Sound Value	Original Signs	Single-Sign Attestations
t	M374$_c$	8
ra	M220$_n$ M223	8
e	M096	8
ta	M219	7
li	M319 M319$_c$	6
zu$_2$	M143 M143$_d$ M155	5
su$_2$	M220	4
pi$_2$	M278	4
mi	M134 M134$_b$	3
la	M4	3
hu	M262	3
ti	M218+M218	2
te	M316$_a$	2
shi	M32$_a$	2
sha	M377	2
lu	M301	1
li$_3$	M302	1
li$_2$	M107$_a$	1
i	M131	1
hi	M356	1

Table 5

Contiguous sequences of PE signs with LE correlates in Desset et al. 2022, including one adjacent sign for which no reading is given. Includes the readings that are implied by the sound values in Desset et al. 2022. This table reports the 20 most common instances; complete table available at https://github.com/sfu-natlang/pe-sign-value-data/tree/public/tables/skipgrams_desset.md

Sequence	Original Signs	Number of Attestations
M388-*a*	M388 M218	24
$M56_f$-pu_2	$M56_f$ M288	21
M388-*na*	M388 M387	15
M240-*e*	M240 M96	15
M124-*a*	M124 M218	13
$M242_b$-*e*	$M242_b$ M96	12
na-M69a	M387 $M69_a$	11
M390-*a*	M390 M218	11
M388-*i*	M388 M66 M388 M131	10
M388-*ha*	M388 M259 M388 M263 M388 $M263_a$	10

Sequence	Original Signs	Number of Attestations
zu-$M3_b$	M9 $M3_b$	9
M388-*pu*	M388 M347	9
M124-pu_2	M124 M288	9
$M56_e$- pu_2	$M56_e$ M288	9
na-M346	M387 M346	8
na-$M263_{b1}$	M387 $M263_{b1}$	8
i-$M352_o$	M66 $M352_o$	8
$M380_b$-u_2	$M380_b$ M371	8
M128-*e*	M128 M96	8
M124-*zu*	M124 M9	8

Table 5a

Alternative sample from the full Table 5, sorted by sequence length rather than number of attestations.

Sequence	Original Signs
$M48_k$-*ta-a-*u_2*-sha-pu-*u_2	$M48_k$ M219 M218 M371 M377e M347 M371
*pu-ra-*M320-*ne-e-*pu_2	M347 $M223_b$ M320 $M281_c$ M96 M288
*na-na-sha-*M295-*e-i*	M387 M387 M377 M295 M96 M66
*na-i-*M309-*she-e-ha*	M387 M131 M309 M32 M96 M263
M104-*i-ha-lu-i-*pu_2	M104 M66 M259 M301 M66 M288
i-$M352_o$-*a-ha-a-*pu_2	M66 $M352_o$ M218 M259 M218 M288
i-la-ha-a-$M57_b$-*ha*	M66 M4 M263 M218 $M57_b$ $M263_a$
*hu-*M380-u_2*-la-zu-u*	M262 M380 M371 M4 M9 M57
ha-$M281_f$*-sha-pu-*u_2*-*pu_2	M263 $M281_f$ $M377_e$ M347 M371 M288
*u-*M240-*sha-pu-*u_2	M57 M240 $M377_e$ M347 M371
u_2*-a-*su_2-M131+M388-*ha*	M371 M218 M220 M131+M388 M263

Sequence	Original Signs
tu-la-M263$_{l}$-*a-pu$_2$*	M226$_{c}$ M4 M263$_{l}$ M218 M288
ta-M48$_{k}$-*e-pu$_2$-ri$_2$*	M219 M48$_{k}$ M96 M288 M297
sha-pu-u$_2$-M388-*sha*	M377$_{e}$ M347 M371 M388 M377$_{e}$
sha-pu-M146-*m-a*	M377$_{e}$M347 M146 M218+M288 M218
ra-M320-*ne-e-pu$_2$*	M223$_{b}$ M320 M281$_{c}$ M96 M288
pu-M97$_{h}$-*la-a-ha*	M347 M97$_{h}$ M4 M218 M263
p-M388-*sha-a-sha*	M6 M388 M377$_{e}$ M218 M377$_{e}$
na-sha-M295-*e-i*	M387 M377 M295 M96 M66
na-hu$_2$-a-M338$_{m}$-*i*	M387 M99 M218 M338$_{m}$ M66

Table 5b

Alternative sample from the full Table 5, showing signs with no proposed reading between signs with a proposed reading.

Sequence	Original Signs	Number of Attestations
zu-M318$_{a1}$-*u$_2$*	M9 M318$_{a1}$ M371	3
u$_2$-M332$_{d}$-*a*	M371 M332$_{d}$ M218	3
la-M263$_{1}$-*a*	M4 M263$_{1}$ M218	3
na-M240-*e*	M387 M240 M96	2
lu-M296+M296-*u*	M301 M296+M296 M57	2
hu$_2$-M390-*a*	M99 M390 M218	2
ha-M387$_{ef}$-*i-pu$_2$*	M263 M387$_{ef}$ M66 M288	2
ha-M48$_{c}$-*e*	M263 M48$_{c}$ M96	2
a-M386$_{a}$-*ra*	M218 M386$_{a}$ M223	2
a-M386$_{a}$-*ha*	M218 M386$_{a}$ M259 M218 M386$_{a}$ M263$_{a}$	2
zu$_2$-M44$_{a}$-*p*	M143 M44$_{a}$ M6	1
zu-M320$_{n}$-*u$_2$*	M9 M320$_{n}$ M371	1
zu-M318$_{a2}$-*u$_2$*	M9 M318$_{a2}$ M371	1
zu-M314-*sha-a*	M9 M314 M377$_{e}$ M218	1
zu-M281$_{f}$-*e-ha*	M9 M281$_{f}$-M96 M263	1
zu-M259$_{1}$-*a-pu$_2$*	M9 M259$_{1}$ M218 M288	1
zu-M102$_{d}$-*la*	M9 M102$_{d}$ M4	1
zu-M24-*u$_2$*	M9 M24 M371	1
u$_2$-a-su$_2$-M131+M388-*ha*	M371 M218 M220 M131+M388 M263	1
u$_2$-M388-*sha*	M371 M388 M377$_{e}$	1

Table 5c

Alternative sample from the full Table 5, showing sequences containing M387 or any of its variants.

Sequence	Original Signs	Number of Attestations
ha-M387$_{ef}$-*i*-*pu*$_2$	M263 M387$_{ef}$ M66 M288	2
ta-M387$_c$-*i*	M219 M387$_c$ M66	1
ri$_2$-M387$_a$	M297 M387$_a$	1
ki$_2$-M387$_a$-*h*$_2$-*u*$_2$	M352$_n$ M387$_a$ M122 M371	1
hu$_2$-M387$_a$-*ri*$_2$	M99 M387$_a$ M297	1
hu-M387$_{ef}$-*pu*$_2$	M262 M387$_{ef}$ M288	1
hu-M387$_{ca}$ -*ri*$_2$	M262 M387$_{ca}$ M297	1
ha-M387$_{ef}$-*pu*$_2$	M263$_a$ M387$_{ef}$ M288	1
ha-M387$_c$	M263 M387$_c$	1
h$_2$-M387$_c$	M122 M387$_c$	1
e-M387$_{eh}$-*la*	M96 M387$_{eh}$ M4	1
e-M387$_a$-*a*-*ha*	M96 M387$_a$ M218 M263	1
e-M387$_a$	M96 M387$_a$	1
a-*e*-M387$_a$-*a*-*ha*	M218 M96 M387$_a$ M218 M263	1
a-M387$_c$	M218 M387$_c$	1

Table 6

Contiguous sequences of common syllables from Dahl 2019 with sound values in Desset et al. 2022, and including one adjacent sign which is not identified as a common syllable. This table reports the 20 most common instances; complete table available at https://github.com/sfu-natlang/pe-sign-value-data/tree/public/tables/skipgrams_dahl.md

Sequence	Original Signs	Number of Attestations
a-*pu*$_2$	M218 M288	36
pu-*u*$_2$	M347 M371	29
e-*pu*$_2$	M96 M288	26
u$_2$-*pu*$_2$	M371 M288	23
M388-*a*	M388 M218	21
i-*pu*$_2$	M66 M288	19
sha-*pu*-*u*$_2$	M377$_e$ M347 M371	17
na-M69$_a$	M387 M69$_a$	11
M388-*na*	M388 M387	11

Sequence	Original Signs	Number of Attestations
M1-M388	M1 M388	11
zu-$M3_b$	M9 $M3_b$	9
na-$M263_{b1}$	M387 $M263_{b1}$	9
M388-*she*	M388 M32	9
M124-a	M124 M218	9
na-M346	M387 M346	8
m-pu_2	M218+M288 M288	8
a-ri_2	M218 M297	8
M388-*ha*	M388 M259 M388 M263	8
$M380_b$-u_2	$M380_b$ M371	8
M128-e	M128 M96	8

Table 7

Alphabetized list of all PE entries greater than one sign in length for which every sign has a proposed reading in Desset et al. 2022.

Tablet	Reading	Original Sequence
P008271	*a-ha-*u_2*-ra-a*	M218 M259 M371 $M223_b$ M218
P009014	*a-ha*	M218 M263
P009014	*a-*hu_2	M218 M99
P009257	*a-la-a*	M218 M4 M218
P008144	*a-la-a*	M218 M4 M218
P008229	*a-m-*pu_2	M218 M218+M288 M288
P008049	*a-*pi_2*-a*	M218 $M295_{ka}$ M218
P008423	*a-ra-a*	M218 M223 M218
P008101	*a-she*	M218 M32
P009257	*a-*su_2*-e*	M218 M220 M96
P009537	*a-ta*	M218 M219
P009423	*a-ta*	M218 M219
P009089	*e-*pu_2	M96 M288
P008914	*e-*pu_2	M96 M288
P008740	*e-pu-*u_2	M96 M347 M371
P008242	*e-ru*	M96 M372
P008442	*ha-a*	M259 M218
P009018	*ha-la-a*	M259 M4 M218

Tablet	Reading	Original Sequence
P009330	*ha-ta-a*	M259 M219 M218
P008065	*ha-zu-u*$_2$	M259 M9 M371
P008914	*ha-a-pu*$_2$	M263 M218 M288
P009376	*ha-r-u*$_2$	M263 M318 M371
P008626	*hi-na-a*	M356 M387 M218
P008444	*hu*$_2$*-u*$_2$	M99 M371
P008067	*hu-a*	M262 M218
P008688	*hu-ha-a-pu*$_2$	M262 M259 M218 M288
P008521	*hu-hu-e*	M262 M262 M96
P008444	*i-ki*$_2$	M66 M352$_n$
P009375	*i-ki*$_2$*-pu-ri*$_2$	M66 M352$_n$ M347 M297
P008355	*i-m-i-pu*$_2$	M66 M246$_b$ M66 M288
P008942	*i-na-a-pu*$_2$	M66 M387 M218 M288
P008083	*i-ri*$_2$	M66 M297
P008186	*i-t*	M66 M374$_c$
P008060	*i-hu*$_2$*-u*$_2$	M131 M99 M371
P009185	*i-ri*$_2$	M131 M297
P008136	*ki*$_2$*-i*	M352$_n$ M66
P009089	*ki*$_2$*-zu-u*$_2$*-pu*$_2$	M352$_n$ M9 M371 M288
P008081	*la-a-la*	M4 M218 M4
P009375	*la-su-u*	M4 M58 M57
P008974	*la-zu-u*$_2$	M4 M9 M371
P009233	*li-a-ri*$_2$	M319 M218 M297
P008001	*li-she*	M319 M32
P008345	*lu-ha*	M301 M263$_a$
P008975	*m-a*	M218+M288 M218
P009179	*m-la-a*	M218+M288 M4 M218
P008010	*m-pu*$_2$	M218+M288 M288
P008258	*m-pu*$_2$	M218+M288 M288
P008990	*m-pu*$_2$	M218+M288 M288
P008702	*m-t*	M246 M374$_c$
P009228	*m-ha*	M246$_b$ M263
P008834	*m-pu*$_2$	M246$_b$ M288
P008239	*me-lu-ha*	M206$_b$ M301 M263
P008914	*na-ha-a-pu*$_2$	M387 M259 M218 M288

Tablet	Reading	Original Sequence
P008982	*na-ha*	M387 M263
P008501	*na-ha*	M387 M263
P009000	*na-ha*	M387 M263
P008085	*na-hu-e*	M387 M262 M96
P008117	*na-m*	M387 $M224_a$
P009524	*na-mi*	M387 M134
P008713	*na-na*	M387 M387
P234794	*na-na-hu-e*	M387 M387 M262 M96
P008904	*na-p*	M387 M6
P008712	*na-pi*$_2$	M387 $M278_c$
P008713	*na-pu*$_2$	M387 M288
P009127	*na-pu*$_2$	M387 M288
P009127	*na-pu*$_2$	M387 M288
P008995	*na-pu*$_2$	M387 M288
P009127	*na-pu*$_2$	M387 M288
P008566	*na-zu-u*$_2$	M387 M9 M371
P008852	*ni-pu*$_2$	M41 M288
P009086	*ni-a*	$M214_c$ M218
P008922	*nu-na*	M391 M387
P008620	*nu-pu*$_2$	M391 M288
P008220	*nu-she*	M391 M32
P008250	*nu-she*	M391 M32
P009223	*pi*$_2$*-na*	M248~A M2
P008275	*pu*$_2$*-sha*	M288 $M377_e$
P009545	*pu*$_2$*-ta*	M288 M219
P008643	*pu-sha-a*	M347 $M377_e$ M218
P008431	*pu-sha-a-pu*$_2$	M347 $M377_e$ M218 M288
P008122	*ra-i-a-pu*$_2$	M223 M66 M218 M288
P008998	*ri*$_2$*-pu*$_2$	M297 M288
P009088	*sha-i-lu-u*	M377 M66 M301 M57
P009296	*sha-pu-u*$_2$	$M377_e$ M347 M371
P008164	*sha-z-ha-a*	$M377_e$ M129 M263 M218
P008085	*she-e*	M32 M96
P008331	*she-i-na*	M32 M66 M2

Tablet	Reading	Original Sequence
P008688	*she-i-pu$_2$*	M32 M66 M288
P008438	*she-na-a-pu$_2$*	M32 M387 M218 M288
P008428	*she-u$_2$*	M32 M371
P008634	*she-u$_2$*	M32 M371
P008979	*shu-ha-a-na-ha*	M226$_{ca}$ M259 M218 M387 M263
P008964	*su$_2$-she-she*	M220 M32 M32
P008444	*su-ri$_2$-i*	M58 M297 M66
P008136	*ta-na-a*	M219 M387 M218
P009156	*u$_2$-ha*	M371 M263
P009262	*u$_2$-zu-u$_2$*	M371 M9 M371
P008319	*u$_2$-zu-u$_2$*	M371 M9 M371
P008082	*u-ki$_2$-la-a*	M57 M352$_n$ M4 M218
P009127	*na-pu$_2$*	M387 M288
P009156	*u-su-u*	M57 M58 M57
P008444	*u-su-u*	M57 M58 M57
P008688	*u-wa-u-pu$_2$*	M57 M455 M57 M288
P009014	*zu-ha-a-ha*	M9 M259 M218 M263
P009156	*zu-ha*	M9 M263
P009156	*zu-ha*	M9 M263$_a$
P009257	*zu-la-ha-e*	M9 M4 M263 M96
P009539	*zu-pu$_2$*	M9 M288
P009018	*zu-u$_2$-a*	M9 M371 M218
P008528	*zu-zu*	M9 M9

Bibliography

BORN, L., KELLEY, K., KAMBHATLA, N., CHEN, C., & SARKAR, A., 2019. Sign clustering and topic extraction in proto-Elamite, in: *Proceedings of the Third Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* (LaTeCH-ClfL-2019), Minneapolis: 122-132 (https://aclanthology.org/volumes/W19-25/).

BORN, L., KELLEY, K., MONROE, M.W. & SARKAR, A., 2021. Compositionality of Complex Graphemes in the Undeciphered Proto-Elamite Script using Image and Text Embedding Models, *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021*: 4136-4146 (https://aclanthology.org/volumes/2021.findings-acl/).

DAHL, J., 2018. The proto-Elamite writing system, in: Álvarez-Mon J., Basello G.P. & Wicks Y. (eds.), *The Elamite World*, London – New York: 383-396.

DAHL, J., 2019. *Tablettes et fragments proto-élamites* (TCL 32), Paris.
DESSET, F., 2012. *Premières écritures iraniennes: les systèmes proto-élamite et élamite linéaire*, Naples.
DESSET, F., 2016. Proto-Elamite writing in Iran, *Archéo-Nil* 26: 67-104.
DESSET, F., 2018. Linear Elamite Writing, in: Álvarez-Mon J., Basello G.P. & Wicks Y. (eds.), *The Elamite World*, London – New York: 397-415.
DESSET, F., TABIBZADEH, K., KERVRAN, M., BASELLO, G.P. & MARCHESI, G., 2022. The Decipherment of Linear Elamite Writing, *Zeitschrift für Assyriologie und vorderasiatische Archäologie* 112: 11-60.
ENGLUND, R.K., 1988. Administrative Timekeeping in Ancient Mesopotamia, *Journal of the Economic and Social History of the Orient* 31: 121-185.
ENGLUND, R.K., 2001. The state of decipherment of proto-Elamite, in: Houston S. (ed.), *The First Writing*, Cambridge: 100-149.
HINZ, W. & KOCH, H., 1987. *Elamisches Wörterbuch* (Archäologische Mitteilungen aus Iran. Ergänzungsband 17), Berlin.
KREBERNIK, M., 2021. Early cuneiform writing and language, in Gabriel G., Overmann K. & Payne A. (eds.), *Signs – Sounds – Semantics. Nature and Transformation of Writing Systems in the Ancient Near East* (Wiener Offene Orientalistik 13), Münster: 9-26.