Title: Proposal to encode the Northern Palaeohispanic script

From: Joan Ferrer i Jané, Grup LITTERA, Universitat de Barcelona; Noemí Moncunill, Grup LITTERA, Universitat de Barcelona; Javier Velaza, Grup LITTERA, Universitat de Barcelona and Deborah Anderson, SEI, UC Berkeley

Date: 1st January 2020

This is a revised proposal for Northern Palaeohispanic. Changes from the earlier proposal L2/19-332 are identified by yellow highlighting in this proposal. Changes include one new character, two new figures, a new bibliographic reference, improved captions, and some corrections.

Earlier documents include:

L2/19-332 Proposal to encode the Northern Palaeohispanic script
L2/19-045 New charts for Northern and Southern Palaeohispanic

The following documents were based on a unified script of Northern and Southern Palaeohispanic.

L2/18-283 Proposal to encode the Palaeohispanic script
L2/18-030 Proposal to encode the Palaeohispanic script
L2/17-129 Proposal to encode the Palaeohispanic script

The two preliminary proposals below were only for what is later called “Northern Palaeohispanic”:

L2/15-120 Preliminary proposal to encode the north-eastern Iberian script
L2/15-012 Preliminary proposal to encode the north-eastern Iberian script
From: Joan Ferrer i Jané, Grup LITTERA, Universitat de Barcelona; Noemí Moncunill, Grup LITTERA, Universitat de Barcelona; Javier Velaza, Grup LITTERA, Universitat de Barcelona and Deborah Anderson, SEI, UC Berkeley

Date: 1st January 2020

Title: Proposal to encode the Northern Palaeohispanic script

1. Summary

Palaeohispanic scripts are attested in the Iberian Peninsula by ca. 2,700 inscriptions dating from the 7th century BC to the 1st century AD. They were used to write at least four different local languages: Celtiberian, Iberian, the south-western or Tartessian language and probably as well Turdetanian. Although the longest inscription contains ca. 500 words, most long texts rarely reach over 50 words; others, on the other hand, are very short and contain just a personal name or abbreviations.

The Palaeohispanic script family consists of several scripts that can be divided into 2 types: the northern (with ca. 2,500 inscriptions) and the southern group (with ca. 170 inscriptions). That’s why we codify them according to two different standards:

1. Northern Palaeohispanic, which includes the northeastern Iberian and the Celtiberian script.
2. Southern Palaeohispanic, which includes the south-eastern Iberian script, the south-western or Tartessian script, the Turdetan script and the Espanca abecedary.

All Palaeohispanic writing systems are characterised by a similar corpus of signs and by the coexistence of alphabetic and syllabic characters. Moreover, all of them share a common ancestor, which might ultimately arise from the Phoenician alphabet (see fig. 1). However, the differences between the two groups are too deep to be appropriately processed into a unique Unicode character set. The main obstacle is the different degree of decipherment between the two groups, which is almost complete for the northeastern Iberian group and still incomplete for at least a third of the signs attested in the southern scripts. The second obstacle is that a large number of signs shared by these two groups actually have different values: it is the case for most of the vocalic signs, as well as for other frequent signs for which there is clear consensus on their value.

The decipherment of the northern Palaeohispanic scripts was accomplished at the beginning of the 20th century by Manuel Gómez-Moreno (1922, 1949). Nevertheless, some aspects were not entirely deciphered until very recent dates. Such is the case of a variant of this script called the dual system, which consists of the use of signs with two variants, each of them with its own distinctive value, differing from each other in presenting an additional stroke (e. g. $\overline{X} = da$ and $\overline{X} = ta$). In fact, some concrete features of this system are even still undergoing research. Unlike the northern scripts, the southern scripts have not been fully deciphered, since there are many signs for which there is no agreed value among specialists. The non-deciphered characters will be named in this proposal after an arbitrary code, as it is currently in use in the specialized bibliography.

The Palaeohispanic inscriptions are being edited and digitalized in the Hesperia open access database (http://hesperia.ucm.es/), in the framework of a research project carried out by a team of scholars from different Spanish universities.

2. Background

This proposal, together with the one devoted to “Southern Palaeohispanic”, is an updated version of three earlier proposals: "L2/15-120-Preliminary proposal to encode the northeastern Iberian script";
"L2/15-119 Preliminary proposal to encode the southern Palaeohispanic scripts"; and L2/18-283 "Proposal to encode the Palaeohispanic script".

The last of them was an attempt to codify all the Palaeohispanic scripts together, under a single encoding. However, after submitting this possibility for analysis with several Unicode specialists, we have come to the conclusion that a separate codification for the southern and northern scripts makes it possible to better reproduce the graphematic system in the different scripts. The revised charts in the current proposal now reflect repertoires that reflect the graphemes in the writing systems. Therefore, two documents are produced, this one for the northern Palaeohispanic script, and another one for the southern Palaeohispanic script.

With the proposed repertoires, all characters in Northeastern Iberian and Celtiberian scripts can be represented. As with Old Italic, fonts will be used to represent the different alphabets of Iberia.


3. Structure

All Palaeohispanic scripts are semisyllabic: vowels, nasals, laterals, sibilants and trills are alphabetical, whereas characters for plosives are syllabic.

There is another feature shared by most Palaeohispanic scripts: the possibility they offer to differentiate some similar signs with close phonetic value by an additional stroke; rather than a mere diacritic (although originally it might have been so), this stroke tends to be an integrating component of the sign itself. This subset of scripts with a larger number of variants has been labelled as “dual”. The recent discovery of dual abecedaries confirms that these dualities were integrated in the standard scripts, where the pair of signs appears always in the same order: the complex variant, with its additional stroke, in the first place and, thereafter, the simple one.

These dualities can affect different sets of sounds: plosives, continuous consonants and even vowels. However, and despite their autonomous apparition in the abecedaries, only for the first ones the phonetic opposition (in this case, voiced – voiceless) between the simple and the complex has been confirmed. Therefore, according to the extension of the use of dualities, it is possible to identify different subsets within every script:

- The northeastern Iberian script can be divided into: i. extended dual, when dualities affect plosives, continuous consonants and vowels; ii. standard dual, when it only affects plosives, and iii. non-dual, when the script lacks dualities.

- The Celtiberian script, either in its eastern or western variant, can be divided into: i. standard dual, or ii. non-dual.

The standard script for Unicode has been built taking into account an inventory of signs as large as possible, including all dual variants confirmed in the extant abecedaries; in general terms, the glyphs for the proposed characters to be encoded match the glyphs of the northeastern Iberian dual extended script.

In addition, the Unicode repertoire also considers as meaningful a three-elements variability for the signs ke and ka in the standard or extended dual script: / /  and / / . Indeed, some long standard dual inscriptions show the simultaneous use of three variants of the ke sign. That’s the case of the lead sheet from Castelló (F.6.1) where two-stroke ke ( ) coexist with one-stroke ( ) and no-marked variants.
In a similar way, on the lead sheet from Ensérune (B.1.373*; Hesperia HER.02.773) a two-dot variant of the ke sign (Collider) coexists with a one-dot (Collider) and a no-marked variant (Collider). Furthermore, it needs to be pointed out that the most common word in the lead sheet from Castelló with the two-stroke ke, balke, was written in this same way in painted inscriptions of the extended dual type from Llíria. This behavior can also be observed in the lead sheet from Los Villares (F.17.2) where three simultaneous variants of the ka sign are used (Collider/Collider/Collider). This suggests that the three-element variation for these two characters is also a characteristic of the extended dual script, although it is not explicitly documented in any of the known abecedaries.

The number of significant variants is a critical issue in the definition of each writing; from our point of view, the proposed codification cannot leave aside epigraphic evidence, based on the point of view of the original user, and add characters that did not exist as autonomous forms. Although a superficial look at the corpus may suggest some of these characters might have existed, this is caused by the fact that we are unifying under a single codification different writings and, within them, different epigraphic schools. The main variants used by the different scripts and epigraphic schools will be expressed by means of different fonts which will allow the variability of writing to be better represented.

That’s the case of the ti (Collider/Collider/Collider) and to (Collider/Collider/Collider) characters, which show an apparently three-elements variation; however, the extant abecedaries, which represent the real set of signs, confirm that there are only two meaningful signs, that is only a two-elements distinction1. The most common dual opposition found in Palaeohispanic inscriptions, used almost in all the Iberian territory (blue dots in the map of fig. 0), is expressed as Collider/Collider and Collider/Collider, as confirmed in the Tos Pelat abecedary (Fig. 5). There is, however, a residual use where the opposition is expressed as Collider/Collider and Collider/Collider, as shown in the Castellet de Bernabé’s abecedary (Fig. 4). This exceptional use is mainly found in a small area around Llíria (València, red dots in the map). In this tradition, the glyphs Collider/Collider express the value represented by Collider/Collider in the most common tradition and the glyphs Collider/Collider express the value represented by Collider/Collider in the most widespread tradition. That’s why our proposal only encodes two values corresponding to the main glyphs used in the most common tradition (Collider/Collider and Collider/Collider).

Regarding the above mentioned residual tradition in Llíria and surroundings, the glyphs Collider/Collider will be always codified as Collider and Collider, as they always represent the simple value. On the other hand, the glyphs Collider/Collider will be codified always as they look like, except when they coexist in the same inscription with Collider/Collider (note however that this is a very rare situation, which occurs up till now only once in the corpus). In this particular case, in order not to lose the graphematic opposition between the marked and unmarked characters (expressed here as Collider/Collider and Collider/Collider, instead of the standard Collider/Collider and Collider/Collider), the glyphs Collider/Collider will be codified as Collider/Collider, as shown in fig. 3.

A similar thing occurs with other characters which apparently seem to display a more than two elements variation, such as the vowel o Collider/Collider/Collider/Collider/Collider. In this case, the signs with more than one additional stroke have not been encoded separately, since they can be always considered an allograph of the glyph with only one additional bar. The opposition between the marked and unmarked variant of this character is scarcely documented: three times in abecedaries and only twice in other types of inscriptions. The Tos Pelat’s abecedary contains at least two abecedaries (Fig. 5), partially in the form of a palimpsest, all of them displaying only two meaningful signs of the vowel o, which is consistent with what happens with the rest of vowels (see Fig. 5). However, one of the Tos Pelat’s abecedaries

1 Ferrer i Jané 2019.
use the pair \( \overline{\text{H}} / \overline{\text{N}} \) while in the same inscription \( \overline{\text{H}} / \overline{\text{N}} \) is also used; in the Castellet de Bernabé abecedary (Fig. 4) the same kind of opposition is expressed as \( \overline{\text{H}} / \overline{\text{N}} \). On the other hand, in inscriptions other than abecedaries, this opposition appears once under the pair \( \overline{\text{H}} / \overline{\text{N}} \) and once under the pair \( \overline{\text{H}} / \overline{\text{N}} \). For all these reasons, our codification only considers two different values which will be represented by the two more common glyphs \( \overline{\text{H}} / \overline{\text{N}} \). The simple variant, \( \overline{\text{H}} \), will be used in all inscriptions representing the unmarked variant, and the complex variant, \( \overline{\text{N}} \), will be used only in inscriptions with an explicit two elements opposition, representing the marked variant, as shown in Fig. 3.

![Geographic distribution of the less common variants of the ti sign. Blue dots (four strokes variants). Red dots (two stroke variants).](image)

**Fig. 0.-** Geographic distribution of the less common variants of the ti sign. Blue dots (four strokes variants). Red dots (two stroke variants).

### 4. Direction of script

The proposed default direction of the script is left to right, which is the predominant direction in inscriptions of the northern group. In order to render texts right to left, users should use RLO and PDF overrides, or other mechanisms as described in UAX #9 ([http://www.unicode.org/reports/tr9/proposed.html](http://www.unicode.org/reports/tr9/proposed.html)). If the default direction of the script is overridden, the glyphs in the font should be mirrored from those presented here.

---

\(^2\) Ferrer i Jané 2015, fig. 3 and 6.
5. Character names

The character names are based by default on the northeastern script, which is the script with more inscriptions and the one whose decipherment is completed. However, since the values are not usually the same in each Palaeohispanic script, notes are used to specify the value of the character in every script. For instance:

The lack of notes implies that the sign has the same value in all scripts. For instance:

1022B  PALAEOHISPANIC NORTHERN LETTER N1

When a sign is exclusive of one single script it is specified that way:

10234  PALAEOHISPANIC NORTHERN NUMERAL A

• Northern Palaeohispanic

The marked signs in the "dual" scripts are named with the terminology "with additional stroke" or " with two additional strokes ". For instance:

10202  PALAEOHISPANIC NORTHERN LETTER E WITH ADDITIONAL STROKE

• Northeastern Iberian (Dual Extended)

1020F  PALAEOHISPANIC LETTER KA WITH TWO ADDITIONAL STROKES

• Northeastern Iberian (Dual)

When the phonetic value of a sign is unknown, we use the conventional code used in the specialized bibliography. For instance:

10232  PALAEOHISPANIC NORTHERN LETTER S87

Please note that the names of the letters don't always reflect the transcription system used in the discursive parts of the proposal. These are the correspondences:

S1 = s
S2 = ś
S1 WITH ADDITIONAL STROKE = ŝ
N1 = n
N1 WITH ADDITIONAL STROKE = ň
N2 = m
N3 = m̌
N4 = ṁ
R1 = r
R2 WITH ADDITIONAL STROKE = ř
R2 = ř
A = a
A WITH ADDITIONAL STROKE = á
A2 = â
E = e
E WITH ADDITIONAL STROKE = é
I = i
I WITH ADDITIONAL STROKE = í
O = o
O WITH ADDITIONAL STROKE = ó
U = u
U WITH ADDITIONAL STROKE = ú
KA = ga
KA WITH ADDITIONAL STROKE = ka
KA WITH TWO ADDITIONAL STROKES = ķa
KE = ge
KE WITH ADDITIONAL STROKE = ke
KE WITH TWO ADDITIONAL STROKES = ĵe
KI = gi
KI WITH ADDITIONAL STROKE = ki
KO = go
KO WITH ADDITIONAL STROKE = ko
KU = gu
KU WITH ADDITIONAL STROKE = ku

6. Numbers

Iberian metrological expressions are basically formed by groups of vertical bars (equivalent to the sign ba) to generate the numerical component of the expression: I = 1, II = 2, III = 3, IIII = 4, IIIII = 5. The accumulation of bars can reach up to 20 elements (F.17.1). Occasionally these bars can appear together with a sign similar to Greek Π, which appears to be acting as an auxiliary base, perhaps with the value of 5 or 6.

Some metrological expressions use a specific L-shaped sign, which does not match any other character of the Iberian script; the numerical value for that sign is still uncertain. This sign also appears in metrological expressions on painted amphora inscriptions from Vieille-Toulouse (for instance L III) and in lead-sheet inscriptions from Iàtova (for instance L Π IIIII [F.20.2]).

Finally, some coin inscriptions present value marks, which, in some cases, have an equivalent symbol formed by the initial of the unit followed by the numerical component. In the case of undikesken coins, quarters show the – sign and halves the = sign, which is actually a reduplication of the former (¼ + ¼ = ½).

7. Punctuation

The most common word separator consists of two vertical dots. Nevertheless, the oldest epigraphic tradition tends to use rather 3 or more vertical dots; in the most recent inscriptions on stone, on the other hand, the use of an isolated dot is frequent, imitating the Roman style. Finally, the vertical bar can also be used and, in some rare cases, just a blank.
Although the different word separators used in the northeastern Palaeohispanic scripts are already-encoded characters, as 205A for the two-dots punctuation, we consider it is better to codify the word separator concept with a unique code specific to this codification. This will allow to customize the word separator in each font representing the different northeastern traditions.

8. Order

For the code chart: vowels will appear in the alphabetical order a, e, i, o, u; plosives in the usual alphabetical order b, k/g, t/d; and continuous consonants in the alphabetical order l, m, n, r, s. The marked-sign pairs will be grouped together, the marked character preceding the unmarked, as appears in the northeastern Iberian abecedaries. The conflictive T-shaped sign is grouped together with nasals, as it actually appears in the attested abecedaries. The conflictive sign in the shape of an Iberian l (â) with an additional stroke is placed together with this sign, as it appears in the Castellet de Bernabé’s abedecary. Numerals are grouped at the end after letters.

The proposed order for sorting is as follows: a, á, â, ba, be, bi, bo, bu, da, de, te, di, ti, do, to, du, tu, e, é, ga, ka, ka, ge, ke, ki, go, ko, ku, i, í, l, m, n, o, ó, r, ř, š, s, š, u, ú, ŕ, m, S87. Specific exceptions to the alphabetical order are as follows:

- Consecutive order for simple sibilant (s) and sibilant with an additional stroke (š);
- Consecutive order for voiceless and voiced plosives in order to keep together the dual and non-dual transcriptions of the same elements (for instance, the word ektar / egiar).
- Consecutive order for m and n, since they are signs that can alternate (for instance iunstir / iumstir).
- Consecutive order for the supposed nasal ñ and ñ, after the two signs for u, since the characteristic vocalic component of ñ can be usually identified as u (for instance mbaf / VMAR).

9. Unicode Character Properties

All the Palaeohispanic letters from 10200 to 10238 are as below.

10200;NORTHERN PALAEOHISPANIC LETTER A WITH ADDITIONAL STROKE;Lo;0;L;;;;;N;;;;;
10201;NORTHERN PALAEOHISPANIC LETTER A;Lo;0;L;;;;;N;;;;;
10202;NORTHERN PALAEOHISPANIC LETTER E WITH ADDITIONAL STROKE;Lo;0;L;;;;;N;;;;;
10203;NORTHERN PALAEOHISPANIC LETTER E;Lo;0;L;;;;;N;;;;;
10204;NORTHERN PALAEOHISPANIC LETTER I WITH ADDITIONAL STROKE;Lo;0;L;;;;;N;;;;;
10205;NORTHERN PALAEOHISPANIC LETTER I;Lo;0;L;;;;;N;;;;;
10206;NORTHERN PALAEOHISPANIC LETTER O WITH ADDITIONAL STROKE;Lo;0;L;;;;;N;;;;;
10207;NORTHERN PALAEOHISPANIC LETTER O;Lo;0;L;;;;;N;;;;;
10208;NORTHERN PALAEOHISPANIC LETTER U WITH ADDITIONAL STROKE;Lo;0;L;;;;;N;;;;;
10209;NORTHERN PALAEOHISPANIC LETTER U;Lo;0;L;;;;;N;;;;;
1020A;NORTHERN PALAEOHISPANIC LETTER BA;Lo;0;L;;;;;N;;;;;
All the NORTHERN PALAEOHISPANIC numerals and fractions from 10235 to 10239 are as below.

10233;NORTHERN PALAEOHISPANIC NUMERAL ONE;No;0;L;;;;1;N;;;;;
10234;NORTHERN PALAEOHISPANIC NUMERAL A;So;0;L;;;;;N;;;;;
10235;NORTHERN PALAEOHISPANIC NUMERAL B;So;0;L;;;;;N;;;;;
10236;NORTHERN PALAEOHISPANIC FRACTION ONE QUARTER;No;0;L;;;;;1/4;N;;;;;
10237;NORTHERN PALAEOHISPANIC FRACTION ONE HALF;No;0;L;;;;;1/2;N;;;;;

The word separator will be encoded as below.

10238;NORTHERN PALAEOHISPANIC SEPARATOR ;No;0;L;;;;1/2;N;;;;;
<table>
<thead>
<tr>
<th>c</th>
<th>1020</th>
<th>1021</th>
<th>1022</th>
<th>1023</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>A WITH ADDITIONAL STROKE</td>
<td>A</td>
<td>KA WITH ADDITIONAL STROKE</td>
<td>( \Psi )</td>
<td>( \Psi )</td>
</tr>
<tr>
<td>1</td>
<td>D</td>
<td>A</td>
<td>KA</td>
<td>TO WITH ONE ADDITIONAL STROKE</td>
<td>S2</td>
</tr>
<tr>
<td>2</td>
<td>E WITH ADDITIONAL STROKE</td>
<td>( \epsilon )</td>
<td>KE WITH TWO ADDITIONAL STROKES</td>
<td>TU</td>
<td>S87</td>
</tr>
<tr>
<td>3</td>
<td>E</td>
<td>( \epsilon )</td>
<td>KE WITH ADDITIONAL STROKE</td>
<td>TU WITH ADDITIONAL STROKE</td>
<td>ONE</td>
</tr>
<tr>
<td>4</td>
<td>I WITH ADDITIONAL STROKE</td>
<td>C</td>
<td>KE</td>
<td>NUMERAL A</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>I</td>
<td>KI WITH ADDITIONAL STROKE</td>
<td>( \iota )</td>
<td>NUMERAL B</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>O WITH ADDITIONAL STROKE</td>
<td>( \omega )</td>
<td>KI</td>
<td>A2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>O</td>
<td>KO WITH ADDITIONAL STROKE</td>
<td>N2</td>
<td>FRACTION ONE QUARTER</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>U WITH ADDITIONAL STROKE</td>
<td>( \upsilon )</td>
<td>KO</td>
<td>N1</td>
<td>FRACTION ONE HALF</td>
</tr>
<tr>
<td>9</td>
<td>U</td>
<td>KU WITH DOT</td>
<td>N1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>BA</td>
<td>KU</td>
<td>N3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>BE</td>
<td>TA WITH ADDITIONAL STROKE</td>
<td>N4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>BI</td>
<td>TA</td>
<td>R1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>BO</td>
<td>TE WITH ADDITIONAL STROKE</td>
<td>R2 WITH ADDITIONAL STROKE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>BU</td>
<td>TE</td>
<td>R2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>KA WITH TWO ADDITIONAL STROKES</td>
<td>TI WITH ONE ADDITIONAL STROKE</td>
<td>S1 WITH ADDITIONAL STROKE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10. Proposed Characters
10200  NORTHERN PALAEOHISPANIC LETTER A WITH ADDITIONAL STROKE
  • Northeastern Iberian (Dual extended)
10201  NORTHERN PALAEOHISPANIC LETTER A
10202  NORTHERN PALAEOHISPANIC LETTER E WITH ADDITIONAL STROKE
  • Northeastern Iberian (Dual extended)
10203  NORTHERN PALAEOHISPANIC LETTER E
10204  NORTHERN PALAEOHISPANIC LETTER I WITH ADDITIONAL STROKE
  • Northeastern Iberian (Dual extended)
10205  NORTHERN PALAEOHISPANIC LETTER I
10206  NORTHERN PALAEOHISPANIC LETTER O WITH ADDITIONAL STROKE
  • Northeastern Iberian (Dual extended)
10207  NORTHERN PALAEOHISPANIC LETTER O
10208  NORTHERN PALAEOHISPANIC LETTER U WITH ADDITIONAL STROKE
  • Northeastern Iberian (Dual extended)
10209  NORTHERN PALAEOHISPANIC LETTER U
1020A  NORTHERN PALAEOHISPANIC LETTER BA
1020B  NORTHERN PALAEOHISPANIC LETTER BE
1020C  NORTHERN PALAEOHISPANIC LETTER BI
1020D  NORTHERN PALAEOHISPANIC LETTER BO
1020E  NORTHERN PALAEOHISPANIC LETTER BU
1020F  NORTHERN PALAEOHISPANIC LETTER KA WITH TWO ADDITIONAL STROKES
10210  NORTHERN PALAEOHISPANIC LETTER KA WITH ADDITIONAL STROKE
10211  NORTHERN PALAEOHISPANIC LETTER KA
10212  NORTHERN PALAEOHISPANIC LETTER KE WITH TWO ADDITIONAL STROKES
10213  NORTHERN PALAEOHISPANIC LETTER KE WITH ADDITIONAL STROKE
10214  NORTHERN PALAEOHISPANIC LETTER KE
10215  NORTHERN PALAEOHISPANIC LETTER KI WITH ADDITIONAL STROKE
10216  NORTHERN PALAEOHISPANIC LETTER KI
10217  NORTHERN PALAEOHISPANIC LETTER KO WITH ADDITIONAL STROKE
10218  NORTHERN PALAEOHISPANIC LETTER KO
10219  NORTHERN PALAEOHISPANIC LETTER KU WITH DOT
1021A  NORTHERN PALAEOHISPANIC LETTER KU
1021B  NORTHERN PALAEOHISPANIC LETTER TA WITH ADDITIONAL STROKE
1021C  NORTHERN PALAEOHISPANIC LETTER TA
1021D  NORTHERN PALAEOHISPANIC LETTER TE WITH ADDITIONAL STROKE
1021E  NORTHERN PALAEOHISPANIC LETTER TE
1021F  NORTHERN PALAEOHISPANIC LETTER TI WITH ADDITIONAL STROKE
10220  NORTHERN PALAEOHISPANIC LETTER TI
10221  NORTHERN PALAEOHISPANIC LETTER TO WITH ADDITIONAL STROKE
10222  NORTHERN PALAEOHISPANIC LETTER TO
10223  NORTHERN PALAEOHISPANIC LETTER TU WITH ADDITIONAL STROKE
10224  NORTHERN PALAEOHISPANIC LETTER TU
10225  NORTHERN PALAEOHISPANIC LETTER N1
10226  NORTHERN PALAEOHISPANIC LETTER A2
10227  NORTHERN PALAEOHISPANIC LETTER N2
10228  NORTHERN PALAEOHISPANIC LETTER N1 WITH ADDITIONAL STROKE
10229  NORTHERN PALAEOHISPANIC LETTER N1
1022A  NORTHERN PALAEOHISPANIC LETTER N3
1022B  NORTHERN PALAEOHISPANIC LETTER N4
1022C  NORTHERN PALAEOHISPANIC LETTER R1
1022D  NORTHERN PALAEOHISPANIC LETTER R2 WITH ADDITIONAL STROKE
1022E  NORTHERN PALAEOHISPANIC LETTER R2
1022F  NORTHERN PALAEOHISPANIC LETTER S1 WITH ADDITIONAL STROKE
10230  NORTHERN PALAEOHISPANIC LETTER S1
10231  NORTHERN PALAEOHISPANIC LETTER S2
10232  NORTHERN PALAEOHISPANIC LETTER S87
10233  NORTHERN PALAEOHISPANIC NUMERAL ONE
10234  NORTHERN PALAEOHISPANIC NUMERAL A
10235  NORTHERN PALAEOHISPANIC NUMERAL B
10236  NORTHERN PALAEOHISPANIC FRACTION ONE QUARTER
10237  NORTHERN PALAEOHISPANIC FRACTION ONE HALF
10238  NORTHERN PALAEOHISPANIC SEPARATOR
### 11. Glyph Variation Chart

<table>
<thead>
<tr>
<th>Encoded glyph</th>
<th>Transcription (see pp. 4-5)</th>
<th>Northern</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Iberian</td>
<td>Celtiberian</td>
</tr>
<tr>
<td></td>
<td>NI+</td>
<td>NI++</td>
</tr>
<tr>
<td>á</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>a</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>é</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>e</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>í</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>i</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>ó</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>o</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>ú</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>u</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>ba</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>be</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>bi</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>bo</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Code</td>
<td>Character</td>
<td>Print</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>1020E</td>
<td>bu</td>
<td></td>
</tr>
<tr>
<td>1020F</td>
<td>ka</td>
<td></td>
</tr>
<tr>
<td>10210</td>
<td>ka</td>
<td></td>
</tr>
<tr>
<td>10211</td>
<td>ga</td>
<td></td>
</tr>
<tr>
<td>10212</td>
<td>ke</td>
<td></td>
</tr>
<tr>
<td>10213</td>
<td>ke</td>
<td></td>
</tr>
<tr>
<td>10214</td>
<td>ge</td>
<td></td>
</tr>
<tr>
<td>10215</td>
<td>ki</td>
<td></td>
</tr>
<tr>
<td>10216</td>
<td>gi</td>
<td></td>
</tr>
<tr>
<td>10217</td>
<td>ko</td>
<td></td>
</tr>
<tr>
<td>10218</td>
<td>go</td>
<td></td>
</tr>
<tr>
<td>10219</td>
<td>ku</td>
<td></td>
</tr>
<tr>
<td>10220</td>
<td>gu</td>
<td></td>
</tr>
<tr>
<td>10221</td>
<td>ta</td>
<td></td>
</tr>
<tr>
<td>10222</td>
<td>da</td>
<td></td>
</tr>
<tr>
<td>10223</td>
<td>te</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1021E</td>
<td>de</td>
<td>1021F</td>
</tr>
<tr>
<td>-------</td>
<td>----</td>
<td>-------</td>
</tr>
<tr>
<td>10223</td>
<td>tu</td>
<td>10224</td>
</tr>
<tr>
<td>10228</td>
<td>ñ</td>
<td>10229</td>
</tr>
<tr>
<td>1022D</td>
<td>r</td>
<td></td>
</tr>
<tr>
<td>1022E</td>
<td>ř</td>
<td>ř</td>
</tr>
<tr>
<td>-------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>1022F</td>
<td>š</td>
<td>š</td>
</tr>
<tr>
<td>10230</td>
<td>s</td>
<td>s</td>
</tr>
<tr>
<td>10231</td>
<td>ММ</td>
<td>ММ</td>
</tr>
<tr>
<td>10232</td>
<td>S87</td>
<td></td>
</tr>
<tr>
<td>10233</td>
<td>numeral 1</td>
<td></td>
</tr>
<tr>
<td>10234</td>
<td>numeralA</td>
<td></td>
</tr>
<tr>
<td>10235</td>
<td>numeralB</td>
<td></td>
</tr>
<tr>
<td>10236</td>
<td>Fraction 1/4</td>
<td></td>
</tr>
<tr>
<td>10237</td>
<td>Fraction 1/2</td>
<td></td>
</tr>
<tr>
<td>10238</td>
<td>Separator</td>
<td></td>
</tr>
</tbody>
</table>

NI+ = Northeastern Iberian dual standard
NI++ = Northeastern Iberian dual extended
NI- = Northeastern Iberian non dual
WC+ = Western Celtiberian dual
WC- = Western Celtiberian non dual
EC+ = Eastern Celtiberian dual
EC- = Eastern Celtiberian non dual
12.- Images.

Fig. 1.- Comparison between Phoenician, Southern Palaeohispanic and Northern Palaeohispanic.

Fig. 2.- Palaeohispanic scripts. Most recent proposal of diffusion and Genealogic model (Ferrer i Jané 2018).

Fig. 3.- Example showing an inscription with code points and expected letter shapes in online/print publications (Castellet de Bernabé: fragment of a dual extended abecedary).

---

3 The two signs of the duality of the vowel a are apparently equal, but, it does not make any sense, as it is a dual abecedary. We suppose that the second a is drew in a clumsy way but is in fact the sign D.
Fig 4. Castellet de Bernabé’s abecedary (Extended dual script). See fig. 3 for its codification.

Fig 5. Tos Pelat’s abecedary (F.13.77*) (Extended dual script). Above, detail of dualities for vowels and trill. Below: General view.

\[4\] The two signs for the vowel a are apparently the same, which does not make sense, as it is a dual abecedary, where a pair of marked and unmarked characters would be expected. We suppose that the second one is drawn
Fig 6.-Bolvir’s abecedary (Standard dual script). Above: general view. Below: Detailed pictures.

here clumsily but should actually be the D sign.
Fig 7.- Ger’s abecedary (Standard dual script).
Fig. 8.- La Tor de Querol’s abecedary (Standard dual script).

Fig. 9.- Simplified abecedaries from Can Rodon (Non-dual script).
Fig. 10.- Esquirol's abecedary (Non-dual script).

Fig. 11.- Lead sheet from Ullastret (C.2.4) (dual script).

Fig. 12.- Lead sheet from La Balaguera (F.17.*) (dual script).
Fig 13.- Lead sheet from Castellet de Bernabé (F.13.75*). Extended dual abecedary with explicit duality for ḳ.

Fig 14.- Stone plaque from Empúries (non-dual script).
Fig 15.- Stele from Vispesa bearing a non-dual northeastern Iberian inscription.

Fig 16.- Bronze coin from undikesken (non-dual script), with mark of value: e= (1/2)
Fig 17.- Bronze coin from undikesken (non-dual script), with mark of value: e- (1/4)

Fig 18.- Ceramic vase from Terrassa bearing the name talskubilos (non-dual script).
Fig 19.- Spindle-whorl from Gebut (non-dual script, right to left).

Fig 20.- Painted inscription in a ceramic vase from Llíria (F.13.5) (Extended dual script).
Fig 21.- Stamps on dolium from Pech Maho (B.7.32) (Standard dual script).

Fig 22.- Vase from Joncosa (D.18.1*) (non-dual script).
Fig 23.- Lead plaque from Yatóva (F.20.1) (non-dual script) with numerals (red circle).
Fig 24.- Northeastern Iberian script (de Hoz 2011).
13. Bibliography


MALUQUER, J. (1968): Epigrafia prelatina de la península ibérica, Barcelona.


14. Acknowledgement

Project supported by a 2017 Leonardo Grant for Researchers and Cultural Creators, BBVA Foundation (The Foundation accepts no responsibility for the opinions, statements and contents included in the project and/or the results thereof, which are entirely the responsibility of the authors) and “Hesperia: lenguas, epigrafía y onomástica paleohispánica” (FFI2015-63981-C3-1-P), MINECO/FEDER. It has been also partly supported by a grant from the United States National Endowment for the Humanities (PR-253360-17), which funds the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment for the Humanities.
### A. Administrative

1. **Title:** Proposal to encode the **NORTHERN PALAEOHISPANIC** script

2. **Requester’s name:** Joan Ferrer, Noemi Moncunill, Javier Velaza, and Deborah Anderson

3. **Requester type (Member body/Liaison/Individual contribution):** Liaison

4. **Submission date:** FILL IN

5. **Requester’s reference (if applicable):**

6. **Choose one of the following:**
   - This is a complete proposal: 
     - x
   - (or) More information will be provided later:

### B. Technical – General

1. **Choose one of the following:**
   - a. This proposal is for a new script (set of characters):
     - Proposed name of script: NORTHERN PALAEOHISPANIC
     - x
   - b. The proposal is for addition of character(s) to an existing block:
     - Name of the existing block:

2. **Number of characters in proposal:** 58

3. **Proposed category (select one from below - see section 2.2 of P&P document):**
   - A-Contemporary
   - B.1-Specialized (small collection)
   - B.2-Specialized (large collection)
   - C-Major extinct
   - D-Attested extinct
   - x E-Minor extinct
   - F-Archaic Hieroglyphic or Ideographic
   - G-Obscure or questionable usage symbols

4. **Is a repertoire including character names provided?**
   - 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

---

Footnote:

5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
   
   The Atelier National de Recherche Typographique (Thomas Huot-Marchand and Arthur Francietta)

   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   
   Fill in

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

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 enclose 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.org 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

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Has this proposal for addition of character(s) been submitted before?</td>
<td>yes</td>
</tr>
<tr>
<td>If YES explain</td>
<td>L2/15-119, L2/15-120, and L2/15-012</td>
</tr>
<tr>
<td>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.)?</td>
<td>yes</td>
</tr>
<tr>
<td>If YES, with whom?</td>
<td>Scholars in Spain, Portugal, France, Germany, and UK</td>
</tr>
<tr>
<td>If YES, available relevant documents:</td>
<td></td>
</tr>
<tr>
<td>3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?</td>
<td>yes</td>
</tr>
<tr>
<td>Reference:</td>
<td>See proposal</td>
</tr>
<tr>
<td>4. The context of use for the proposed characters (type of use; common or rare)</td>
<td>rare</td>
</tr>
<tr>
<td>Reference:</td>
<td>See proposal</td>
</tr>
<tr>
<td>5. Are the proposed characters in current use by the user community?</td>
<td>yes</td>
</tr>
<tr>
<td>If YES, where? Reference:</td>
<td>Books, articles, database, etc.</td>
</tr>
<tr>
<td>6. After giving due considerations to the principles in the P&amp;P document must the proposed characters be entirely in the BMP?</td>
<td>no</td>
</tr>
<tr>
<td>If YES, is a rationale provided?</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?</td>
<td>yes</td>
</tr>
<tr>
<td>8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?</td>
<td></td>
</tr>
<tr>
<td>If YES, is a rationale for its inclusion provided?</td>
<td>yes</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td>See proposal</td>
</tr>
<tr>
<td>9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?</td>
<td>no</td>
</tr>
<tr>
<td>If YES, is a rationale for its inclusion provided?</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>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?</td>
<td>yes</td>
</tr>
<tr>
<td>If YES, is a rationale for its inclusion provided?</td>
<td>yes</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td>See proposal</td>
</tr>
<tr>
<td>11. Does the proposal include use of combining characters and/or use of composite sequences?</td>
<td>no</td>
</tr>
<tr>
<td>Question</td>
<td>Answer</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>If YES, is a rationale for such use provided?</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?</td>
<td>no</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>12. Does the proposal contain characters with any special properties such as control function or similar semantics?</td>
<td>no</td>
</tr>
<tr>
<td>If YES, describe in detail (include attachment if necessary)</td>
<td></td>
</tr>
<tr>
<td>13. Does the proposal contain any Ideographic compatibility characters?</td>
<td>no</td>
</tr>
<tr>
<td>If YES, are the equivalent corresponding unified ideographic characters identified?</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
</tbody>
</table>