Preliminary proposal to encode the north-eastern Iberian script for the Unicode standard

Joan Ferrer – Noemí Moncunill – Javier Velaza

LITTERA GROUP

University of Barcelona

1. Introduction

The north-eastern Iberian script is attested in the north-eastern quarter of the Iberian Peninsula from the second half of the 5th century B.C. to the 1st century A.D. in more than 2.000 inscriptions in Iberian language (Untermann 1975; 1980; 1990; Campmajo, Untermann 1991; Panosa 1993; Hesperia[1]). The corpus consists of long texts on lead sheets (Fig. 2, 11, 12 and 13), which mainly correspond to commercial letters and administrative documents; small inscriptions on ceramic vases, usually with the name of the proprietary (Fig. 18); coin inscriptions with the indication of the mint, personal names and marks of value (Fig. 16 and 17); steles and stone plaques (Fig. 13 and 14) with funerary formulae; stamps on ceramic with the name of the producer (Fig. 21); painted inscriptions on ceramic of indeterminate linguistic content (Fig. 1 and 22); rock inscriptions with likely votive formulae (Fig. 3, 4, 5 and 7); inscriptions on personal objects such as spindle-whorls (Fig. 6 and 19), weights, or even arms, etc.

The decipherment of this script was accomplished at the beginning of the 20th century by Manuel Gómez-Moreno (1922, 1949). This achievement was essentially possible thanks to the existence of coin inscriptions, which were, in some cases, bilingual, and, in some other cases, clearly linkable to place names known trough Greek and Latin ancient sources. The emerge of some Iberian inscriptions written in an adaptation of the Greek alphabet –the most remarkable is the lead of La Serreta d'Alcoi (G.1.1), which revealed the phonetics of the Iberian language– was also helpful at this regard, as well as the existence of a Latin inscription with an extensive relation of Iberian personal names, known as the *Turma Salluitana* (*CIL* I^2, 709). Nevertheless, some aspects of the Iberian script had not been 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 a proper value, differing one from each other in presenting an additional stroke (Maluquer 1968; de Hoz 1985; Correa 1992; Quintanilla 1992; Ferrer i Jané 2005). In fact, some concrete aspects of this system are even still in process of research, as we will explain in the following sections (Ferrer i Jané 2013b; in press 2014a).

The north-eastern Iberian script belongs to the palaeohispanic scripts family, together with the Celtiberian script (Untermann 1997), the south-eastern Iberian script (Untermann 1990) and the south-western or Tartessian script (Untermann 1997). In addition, it is necessary to consider the existence of the Espanca abecedary (Untermann 1997, J.25.1; de Hoz 2010, 488), found in the same territory where the south-western or Tartessian script is attested. Nevertheless, this abecedary does not exactly match neither the Tartessian script nor the south-eastern Iberian script. This family of writing systems is characterized by both a similar corpus of signs and the coexistence of alphabetic and syllabic signs.

The Celtiberian script is clearly an adaptation of the north-eastern Iberian script to the particularities of the Celtiberian language. Just as its model, it also presents the so-called dual variant or dual system (Ferrer i Jané 2005; Jordán 2005; 2007). This script can be considered as a subset of the north-eastern Iberian, which would allow using the Iberian north-eastern Unicode to represent Celtiberian inscriptions as well. This script is documented between the end of the 3rd century B.C. and the early 1st century A.D. in nearly two hundred inscriptions attested in the interior of the Iberian Peninsula.

[1] hesperia.ucm.: Database with critical editions of all the palaeohispanic inscriptions (Luján, Orduña forthcoming; Orduña, Luján, forthcoming; Velaza, forthcoming). It is a collaborative project carried out by a team of researchers from the Universities of Zaragoza, Complutense of Madrid, Basque Country and Barcelona.

The south-eastern Iberian script is used in *ca.* 70 inscriptions attested in the south-eastern quarter of the Iberian Peninsula, likely to represent the Iberian language, from the 4th century B.C. to the 1st century B.C. However, it is also possible that the most western inscriptions were written in a different language (de Hoz 2011, 707). Although this script is fundamentally used to represent the Iberian language, it is noticeably different from the north-eastern Iberian scripts: both writing systems present an amount of exclusive signs and, in some cases, the shared signs have different values.

On the other hand, the south-eastern script is very close to the south-western or Tartessian script, employed in a hundred inscriptions in a language of unknown filiation attested in the south-western corner of the Iberian Peninsula, perhaps since a period running from the 7th up to the 4th century B.C. Some scholars (de Hoz 2010, 521; Correa 2009, 276) use the denomination of *Tartessian* in a restrictive way to identify only the script in which the Tartessian-core-zone inscriptions are written, saving the denomination *south-western* for the inscriptions of the western peripheral zone, which comprehend the main body of the group.

Hence, considering the indicated differences between the southern scripts and the north-eastern Iberian script, and also their different degree of decipherment, which is much more deficient for the two southern scripts, it is advisable that the south-eastern Iberian script and the south-western or Tartessian script configure a Unicode codification apart.

It is usually accepted that at the last analysis the palaeohispanic scripts derive from the Phoenician alphabet (Untermann 1990, 135; Correa 2009, 276; de Hoz 2010, 485; 2011, 200; Rodríguez Ramos 2004, 39), although some authors claim the influence of the Greek alphabet as well (Untermann 1990, 135). The traditional thesis holds that the north-eastern Iberian script derives, in one way or another, from the south-eastern Iberian script (de Hoz 2010, 423), which, in turn, would derive from the south-western or Tartessian script. Nevertheless, if the hypothesis of the existence of a dual system also in the south-eastern Iberian script was confirmed, then the most economical hypothesis would be to put both Iberian scripts at the same level and to postulate the existence of a common ancestor which would also be dual. This characteristic would not match the scripts that have been traditionally identified as possible ancestors, since neither the south-western or Tartessian script nor the Espanca abecedary are dual scripts (Ferrer i Jané 2010).

2. Characteristic of the script

The north-eastern Iberian script consists of vocalic signs, **a**, **e**, **i**, **o, u**; syllabic signs for the dental plosives, **ta, te, ti, to, tu**, velar plosives, **ka, ke, ki, ko, ku,** and labial plosives, **ba**, **be**, **bi**, **bo**, **bu;** and consonantal signs for the nasals, **m**, **n**, lateral, **l**, sibilants, **s, ś,** and trills, **r, ŕ**.

There are still some other signs of controversial value, which seem to combine both a vocalic and a consonantal component. Such is the case of a relatively frequent sign, transcribed as **ḿ,** which has been commonly integrated into the Iberian signs repertoires. It is also the case of two other extremely rare signs, initially considered as allographs of others, but that have later been confirmed as autonomous in appearing in some of the abecedaries discovered in the last years. One of them is the sign with the shape of a **T** or **I**, which is transcribed as **m̌**, due to its proximity to the nasal signs, in a general way, and with **ḿ,** in particular, due to the fact that it appears next to this sign in Ger's abecedary (Ferrer i Jané 2013a; 2014b). The other controversial sign is  , which would fit as a marked variant of **l**. This sign appears in Castellet de Bernabé's abecedary together with a normal **l** sign, and it is transcribed as **â** (Rodríguez Ramos 2001, 286; Ferrer i Jané 2009, 474; 2014b), taking into account its apparent vocalic nature and the fact that it always appears after **l**.

The oldest inscriptions (5th to 3rd century B.C.) are written in a script modality called *dual*, which uses, as mentioned above, two variants for every single sign. These variants differ one from each other in the addition or lack of a stroke.

The most common dual system (which would come up to 39 signs [Fig. 9]) duplicates only dental syllabic signs, **ta/da, te/de, ti/di, to/do, tu/du**, and velar syllabic signs, **ka/ga, ke/ge, ki/gi, ko/go, ku/gu.** On the other hand, dualities for the labial signs, **ba**, **be**, **bi**, **bo**, **bu**, are not detected, which is due to the low productivity of /p/ in the Iberian language, as it can be particularly seen in the longest Greco-Iberian inscriptions, as the lead sheet from La Serreta d'Alcoi (G.1.1). This dual variant is attested in Ger's (Fig. 4), La Tor de Querol's (Fig. 5) and probably Bolvir's (Fig. 3) abecedaries and it probably represents roughly a 35% of total. Nevertheless, in short inscriptions it is not always clear whether the system being used is the dual or non-dual abecedary.

Another variant of the dual system, which would reach 46 signs (Fig. 10), expands the repertoire of dualities to vowels, **a/á**, **e/é**, **i/í**, **o/ó,u/ú** and to some continuous consonants, **s/ŝ** and **ŕ/ř**. This variant is attested in the Castellet of Bernabé's (Fig. 1) and Tos Pelat's (Fig. 2) abecedaries. The inscriptions showing dualities also for continuous consonants and vocals are scarce: they represent only 5% of the total and are confined to the surroundings of Llíria (Valencia) (Ferrer i Jané 2013b; in press 2014a).

On the other hand, the most recent inscriptions (2nd and 1st centuries B.C.) no longer show dualities; the abecedary remains reduced to 29 signs instead (Fig. 8), mostly using only the unmarked variant of each pair. Near 60% of inscriptions in the corpus use a non-dual script.

The inscriptions are usually written from left to right, but occasionally from right to left instead. Only around 30 texts (out of 2.000 inscriptions in north-eastern Iberian script) are written from right to left, although most of them are ceramic stamps in which the model was left-to-right. One of the most remarkable exceptions is Bolvir's abecedary on rock (Fig. 3).

3. Ordering

3.1. Order in the code chart

The north-eastern Iberian abecedaries do not always use the same sign order. Among the five north-eastern Iberian dual abecedaries attested (Castellet de Bernabé's [Fig. 1; Ferrer i Jané 2009; Velaza 2006], Tos Pelat's [Fig. 2; Burriel et al. 2011], Bolvir's [Fig. 3; Ferrer i Jané 2013a; 2013b], Ger's [Fig. 4; Ferrer i Jané 2013a; 2013b] and La Tor de Querol's [Fig. 5; Ferrer i Jané 2014c]), the initial sequence **kututidibabita** appears in the last two of them. They represent, together with that from Bolvir, the only three abecedaries preserving, with certainty, the initial sequence. In addition, a probabilistic method has been recently proposed to identify non-dual abecedaries in other inscriptions (Ferrer i Jané 2014b); two of the abecedaries identified following these criteria, Esquirol's and Can Rodon's abecedaries, begin with the sequence **kutukiŕbitatiko**.

The inexistence of a unique sequence in the Iberian abecedaries forces to adopt an ad-hoc order, grouping signs according to their value. Therefore, vowels will appear in the alphabetic order **a**, **e**, **i**, **o**, **u**; plosives in the usual alphabetic order **b**, **k**, **t**; and continuous consonants in the alphabetic order **l**, **m**, **n**, **r**, **s**. The marked-sign pairs will be grouped together, the marked character preceding the unmarked, as it appears in the north-eastern Iberian abecedaries. The conflictive **T**-shaped sign is grouped together with nasals, as it actually appears in the attested abecedaries. Numerals are grouped at the end, and the last sign is saved as word separator. The conflictive sign in the shape of a complex Iberian **l** is placed together with this sign, as it appears in Castellet de Bernabé's abecedary.

3.2. Order for sorting

Published Iberian lexicons (Tovar 1951; Siles 1985; Velaza 1991, Silgo 1994; Moncunill 2006, 24) use the Latin alphabetical order for the alphabetisation of Iberian transcribed texts, though with some small changes depending on the author and regarding the treatment of voiceless and voiced occlusive sounds, as well as that of sibilants and trills.

Hence, in general terms the order proposed follows the alphabetical order of Iberian texts transcribed into the Latin alphabet. The exceptions to this principle are due to the aim of maintaining together groups of signs with similar values. For this reason, the order proposed would be as follows: **a**, **á**,**â, ba**, **be**, **bi**, **bo**, **bu**, **da, ta, de, te, di, ti, do, to, du, tu**, **e**, **é,ga, ka, ge, ke, gi, ki, go, ko, gu, ku**, **i**, **í**, **l**, **m, n**, **o**, **ó**, **r**, **ŕ**, **ř, ś, s, ŝ**, **u, ú, ḿ**, **m̌**. Specific exceptions to the alphabetic order are the following ones:

- Consecutive order for simple sibilant (**s**) and complex sibilant **(ŝ)**;
- 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 **ekiar** / ***egiar***).
- Consecutive order for **m** and **n**, since they are signs that can alternate (for instance **iunstir** / **iumstir**).
- Consecutive order for the supposed nasal **ḿ** and **m̌**, after the two signs for **u**, since the characteristic vocalic component of **ḿ** can be usually identified as **u** (for instance **ḿbaŕ** / VMAR).

4. Numbers

Iberian metrological expressions (Untermann 1990, 146, de Hoz 1981; 2011, 191) 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).

There is a subset of metrological expressions, which usually does not exceed six vertical bars, that appears together with a sign similar to Greek **Π.** This element appears to be acting as an auxiliary base, perhaps with the value of 5 (Lejeune 1983, 33; de Hoz 2011, 195), in a decimal context and in accordance with its value in the Greek acrophonic number system (**Π** = *penta*), or maybe with the value of 6 (Ferrer i Jané 2014, 65) optimizing its value in a duodecimal context.

Usually the numerical expressions formed by a **Π** followed by several bars appear together with characters of the basic corpus, **a**, **o**, **ki**, **e**, **be**, **l, ti**, **ḿ** and **ka**, which these numerals seem to quantify. These characters could express measurement units in different metrological systems and in most of the cases they probably stand for the initial of the unit name, for instance, **e** for **etar**, and **ki** for **kitar** (Rodríguez Ramos 2005, 45 i 63; Orduña 2005, 499; Ferrer i Jané 2011b, 101; 2014), so it does not seem necessary to encode them as different shapes. Nor do we consider necessary to encode the sign **s** as an independent shape**,** as attested on painted amphora inscriptions from Vieille-Toulouse, forming groups of up to 4 elements: **ssss** (Lejeune 1983; de Hoz 2011, 195).

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 (Untermann 1990, 147). 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]).

Some coin inscriptions present value marks (Ferrer i Jané 2007), 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 sign - and halves sign **=**, which is actually a reduplication of the former (¼ + ¼ = ½). These signs can present several other variants depending on the mint, as it is in **śaitabi,** where < is used for quarters and **<<** for halves. Sign – also appears in some metrological expressions in lead sheets from Iàtova (for instance **ḿ** · **III** · **ḿ**– [F.20.2]).

Numerical symbols not represented as characters							
5? / 6? / ?		?		1/4	–	1/2	=

5. Word separators

Most part of the two thousand of the north-eastern Iberian inscriptions are very short and do not need word separators, but long texts (*ca.* 200 items) do use them. 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. Sporadically the vertical bar can also be used and, in some rare cases, just a blank (Simon 2011).

6. Characters

The standard script has been built taking into account an inventory of signs as large as possible, including all dual variants confirmed in the attested abecedaries, and thus choosing duality as an answer to one of the main doubts posed about how the codification of north-eastern Iberian script should be done (Comas, Moncunill 2009).

Following the criteria and main objectives of the Unicode standards, multiple variants of a single sign have not been included (Untermann 1980, 49; 1990, 246; Rodríguez Ramos 2004, 143; de Hoz 2011, 743), but just the signs with different values; such principle has not been followed in an officious proposal of Unicode encoding recently realised (Huertas 2009). The choice of the most representative variant for each sign has been done according to their concurrency frequency in presumably dual inscriptions.

The most common convention to transcribe dualities for dental and velar plosives signs is to use the voiced signs for simple variants and to save voiceless ones for complex variants (for instance, **ge** / **ke** and **de** / **te**). Such convention is based on the Iberian phonetics attested in Greco-Iberian inscriptions and in Greek and Latin inscriptions, where simple variants appear represented by voiced signs and complex ones by voiceless signs. An alternative convention to represent this opposition is to arbitrarily maintain the representation of the voiceless and to stress the vowel of the marked variant: for instance **ke** / **ké** and **te** / **té**. In any case, these conventions are of little significance as for the Unicode encoding, since they only come up in the name given to each character.

Velar Syllabic signs									
ka		ke		ki		ko		ku	
ga		ge		gi		go		gu	
Dental Syllabic signs									
ta		te		ti		to		tu	
da		de		di		do		du	
Labial Syllabic signs									
ba		be		bi		bo		bu	

The standard script takes also into account dualities for all vowels, as attested in Tos Pelat's abecedary for all five vowels, and in Castellet de Bernabé's abecedary, for **a** and **o**. Duality for sign **e** explicitly appears in some painted inscriptions from Llíria. However, the detected graphic opposition between marked and unmarked vowels has not found any a plausible phonetic explanation so far (Ferrer i Jané 2013b, 451; in press 2014a).

Vowels									
á		**é**		**í**		**ó**		**ú**	
a		**e**		**i**		**o**		**u**	

The standard Unicode script also includes dualities for continuous consonants explicitly attested in some of the preserved abecedaries. Such is the case of the duality of trill **r** attested in Tos Pelat's abecedary and in several long inscriptions on lead. On the analogy of the use of this exact same duality in south-eastern Iberian script, where **ř** appears most of the time in an intervocalic context, it has been proposed (Ferrer i Jané 2010, 101; 2013b, 448; in press 2014a) that the marked variant in north-eastern Iberian script, **ř,** was the multiple trill, whereas the unmarked variant, **ŕ**, should be the simple one. This is also the case of sibilant **s**, attested in Castellet de Bernabé's abecedary and sporadically in some other inscriptions. On the analogy of the value of the mark of occlusive consonants, it has been proposed (Ferrer i Jané 2013b, 445; in press 2014a) that the marked sibilant, **ŝ**, should be the voiced one (*fortis*), whereas the unmarked variant, **s**, should be the voiceless one (*lenis*).

Consonants with dual variants			
ř		**ŝ**	
ŕ		**s**	

The rest of continuous consonants, **r**, **ś**, **m**, **n** and **l,** do not appear dual variants in currently known abecedaries.

Consonants without dual variants									
r		**ś**		**m**		**n**		**l**	

Although it exists at least one Iberian inscription (Ferrer i Jané in press 2014a) and two Celtiberian inscriptions with an isolated marked variant of **n**, similar to what is attested in south-eastern Iberian script, the more plausible hypothesis is that sign **m** in north-eastern Iberian script was actually a marked variant of **n** and, therefore, the explicitly marked variants of **n** were indeed allographs of **m**.

There is a scarcely attested sign,  , which according to its shape could work as a marked variant of **l** (X5; Rodríguez Ramos 2001, 287; S79; de Hoz 2011, 190 and 744); in fact, in the Castellet de Bernabé's abecedary it is found pairing with **l**, although inverted with regard to the usual complex-simple order. It has commonly been (Untermann 1990, 246) considered as a variant of **e** (**e7**) or **ka** (**ka7**). In inscriptions, it appears most of the times in the same order, pairing with preceding **l**; for this reason the hypothesis of this sign being a complex variant of **l** must be rejected, although it could have been so originally. Therefore it is necessary to reject

also its transcription as **l'**. In some texts where attested, it seems plausible to think that this sign had a vocalic component related to lateral **l**. Finally, it is arbitrarily represented in this proposal as **â,** instead of **á** as displayed in other works (Rodríguez Ramos 2001, 286; Ferrer i Jané 2009, 474) in order to avoid the confusion with the complex variant of vowel **a**.

Sign  (S75; de Hoz 2011, 190 and 744), which appears as  or under a double-arrow shape in early inscriptions (S77; de Hoz 2011, 189 and 744), has commonly been regarded as a variant of one of the nasals signs, either **m** or **ḿ** (Untermann 1990, 247, **ḿ5),** or as a non-deciphered sign, but its appearance in Ger's abecedary and probably also in Tor de Querol's abecedary (Ferrer i Jané 2013ta; 2014b), together with **m** and **ḿ**, proves its independence and its relationship with nasals, as it appears next to **ḿ**.

The transcription sign  as **ḿ** does not reflect its real value either, since it is accepted that it might have both a nasal and a vocalic component. Yet, we keep the usual transcription **ḿ** out of tradition.

Signs with problematic values with both a vocalic and a consonantal component					
m̌		**ḿ**		**â**	

Additionally, there are several extremely rare signs for which it cannot be said whether they are independent signs or local variants for any already known signs. However, as they do not appear in any of the attested abecedaries and considering their low concurrency frequency, it seems reasonable not to encode them in the standard Unicode, except for sign S87.

hapax			
S87		**B'oide**	

Sign S87, which is similar to a herringbone with two,  , or three strokes,  , (X4; Rodríguez Ramos 2001, 284; S87; de Hoz 2011, 744), is the most common one of this group. This sign is sometimes transcribed as **e**, as in coin inscription **sesars** (A.44); however, it could also be either an inverted variant of the complex shape of sign **u** (Ferrer i Jané 2011a, 218, note 8), a hypothetic variant of the sign for **bo** (Ferrer i Jané, Garcés 2005, 987) or even a sign with another different value.

The sign  also belongs to this group (X1/X2; Rodríguez Ramos 2001, 282; S76; de Hoz 2011, 744). For instance, it appears in the *ostrakon* of Pontós (C.3.1), where it is interpreted as a complex variant of **ke**.

Another conflictive character (Velaza 2009, 617) is the second sign attested in the coin inscription **arsaos** (A.37). In spite of its irregular shape,  /  , similar to a **a6** (Untermann 1990, 246), it is usually considered an allograph of **r**, and thus it does not need to be encoded separately.

It has neither been taken into account in the Unicode repertoire the possibility that some dualities actually show three elements of variability, as in the lead sheet from Castelló (F.6.1) and perhaps in the one from Ensérune (B.1.373*), where variants for sign **ke** composed of two strokes, in the first case, and of two dots, in the second one, are found (Orduña 2013, 518, note 9). Such variants have been usually considered to be allographs of the complex variant, but it

must be noted that they appear together in the same text, alternating both with purely unmarked variants and with variants marked with a single stroke or dot. This could be an indication that the three-element variability could also be internally significant. Although this phenomenon is not explicitly attested in any other inscription, such explanation could also be applicable to other signs, which, according to their known shapes, could turn out to have three working variants, instead of the two defined by the dual system (Ferrer i Jané 2010, 107, note 122). However, since evidences are currently scarce at this regard, these variants have not been included in this proposal.

In a similar way, in some other inscriptions, for instance in a *kálathos* from Castelillo de Alloza (E.4.2) and in the lead sheet from Olriols, Tamarit de Llitera (Ferrer i Jané, Garcés 2005, 988, note 10), two reflected variants of the sign **ki**, in particular **ki1**,  , and **ki6**,  , coexist. Again, it is not clear if this must be considered as a deliberate attempt to differentiate the sounds or if it is just the result of the natural variability in manual writing.

Hence, the group of characters proposed to encode the north-eastern Iberian script would be as expressed in the following table. A place for a hypothetical complex **ś** has been saved, not attested in north-eastern Iberian script, but existing in south-eastern Iberian script. The rest of saved codes are displayed in the fourth column.

	000105C		000105D		000105E		000105F	
0		**á**		**ga**		**do**		S87
1		**a**		**ke**		**tu**		
2		**é**		**ge**		**du**		
3		**e**		**ki**		**l**		
4		**í**		**gi**		**ã**		
5		**i**		**ko**		**m**		
6		**ó**		**go**		**n**		
7		**o**		**ku**		**ḿ**		
8		**ú**		**gu**		**m̌**		
9		**u**		**ta**		**r**		
A		**ba**		**da**		**ř**		
B		**be**		**te**		**ŕ**		
V		**bi**		**de**		**ŝ**		numeral
D		**bo**		**ti**		**s**		numeral
E		**bu**		**di**				1/4
F		**ka**		**to**		**ś**		1/2

7. Characters Names

Over the last years, the designation *north-eastern Iberian* script has been the most common name used among scholars, prevailing over *Levantine Iberian* script. In order to establish the terminology for each character, the name of the script is displayed in the first place, followed by its transcription or proposed value. In order to avoid problems with the special

characters in the text file, transcriptions **a, á, e, é, i, í, o, ó, u, ú, r, ŕ, ř, s, ś, ŝ, m, ḿ, m̌** are represented respectively as **a1**, **a2**, **e1**, **e2**, **i1**, **i2**, **o1**, **o2**, **u1**, **u2**, **r1**, **r2**, **r3**, **s1, s2, s3, m1, m2**, **m3**.

Glyph	NAME
	NE IBERIAN LETTER E2
	NE IBERIAN LETTER E1
	NE IBERIAN LETTER KA
	NE IBERIAN LETTER GA
	NE IBERIAN LETTER M2
	NE IBERIAN LETTER N
	NE IBERIAN VALUE MARK 1-4

8. Unicode character properties (UnicodeData.txt)

The General_Category property is established at LO (Lowercase, Other) for the characters, NO (Numbers, Oher) for the numerals and ZS (Separator, Space) for the word separator.

The Canonical_Combining_Class is established at 0 (Not_Reordered) for all elements, since the provided characters cannot be combined.

The property Bidi_class is established at L (Left to Rigth) for all elements, since this is the natural direction of writing for the provided glyphs. The property Bidi_mirrored is established at N (No) for all elements, since this situation never occurs in Iberian.

For the two numerals that represent fractions ½ and ¼ their respective values are indicated in the third field of the Numeric_Value property, since they are fractions. For the two numerical elements whose value is still unknown no value is indicated.

The Simple_Uppercase_Mapping and Simple_Lowercase_Mapping properties are left blank, since the distinction between uppercase and lowercase letters does not exist in Iberian.

000105C0**;**NE IBERIAN LETTER A2;LO;0;L;**;;;;**N;**;;;;**

000105C1;NE IBERIAN LETTER A1;LO;0;L;**;;;;**N;**;;;;**

000105C2;NE IBERIAN LETTER E2;LO;0;L;**;;;;**N;**;;;;**

000105C3;NE IBERIAN LETTER E1;LO;0;L;**;;;;**N;**;;;;**

000105C4;NE IBERIAN LETTER I2;LO;0;L;**;;;;**N;**;;;;**

000105C5;NE IBERIAN LETTER I1;LO;0;L;**;;;;**N;**;;;;**

000105C6;NE IBERIAN LETTER O2;LO;0;L;**;;;;**N;**;;;;**

000105C7;NE IBERIAN LETTER O1;LO;0;L;**;;;;**N;**;;;;**

000105C8;NE IBERIAN LETTER U2;LO;0;L;**;;;;**N;**;;;;**

000105C9;NE IBERIAN LETTER U1;LO;0;L;**;;;;**N;**;;;;**

000105CA;NE IBERIAN LETTER BA;LO;0;L;**;;;;**N;**;;;;**

000105CB;NE IBERIAN LETTER BE;LO;0;L;**;;;;**N;**;;;;**

000105CC;NE IBERIAN LETTER BI;LO;0;L;**;;;;**N;**;;;;**

000105CD;NE IBERIAN LETTER BO;LO;0;L;;;;;N;;;;;
000105CE;NE IBERIAN LETTER BU;LO;0;L;;;;;N;;;;;
000105CF;NE IBERIAN LETTER KA;LO;0;L;;;;;N;;;;;
000105D0;NE IBERIAN LETTER GA;LO;0;L;;;;;N;;;;;
000105D1;NE IBERIAN LETTER KE;LO;0;L;;;;;N;;;;;
000105D2;NE IBERIAN LETTER GE;LO;0;L;;;;;N;;;;;
000105D3;NE IBERIAN LETTER KI;LO;0;L;;;;;N;;;;;
000105D4;NE IBERIAN LETTER GI;LO;0;L;;;;;N;;;;;
000105D5;NE IBERIAN LETTER KO;LO;0;L;;;;;N;;;;;
000105D6;NE IBERIAN LETTER GO;LO;0;L;;;;;N;;;;;
000105D7;NE IBERIAN LETTER KU;LO;0;L;;;;;N;;;;;
000105D8;NE IBERIAN LETTER GU;LO;0;L;;;;;N;;;;;
000105D9;NE IBERIAN LETTER TA;LO;0;L;;;;;N;;;;;
000105DA;NE IBERIAN LETTER DA;LO;0;L;;;;;N;;;;;
000105DB;NE IBERIAN LETTER TE;LO;0;L;;;;;N;;;;;
000105DC;NE IBERIAN LETTER DE;LO;0;L;;;;;N;;;;;
000105DD;NE IBERIAN LETTER TI;LO;0;L;;;;;N;;;;;
000105DE;NE IBERIAN LETTER DI;LO;0;L;;;;;N;;;;;
000105DF;NE IBERIAN LETTER TO;LO;0;L;;;;;N;;;;;
000105E0;NE IBERIAN LETTER DO;LO;0;L;;;;;N;;;;;
000105E1;NE IBERIAN LETTER TU;LO;0;L;;;;;N;;;;;
000105E2;NE IBERIAN LETTER DU;LO;0;L;;;;;N;;;;;
000105E3;NE IBERIAN LETTER L;LO;0;L;;;;;N;;;;;
000105E4;NE IBERIAN LETTER A3;LO;0;L;;;;;N;;;;;
000105E5;NE IBERIAN LETTER M1;LO;0;L;;;;;N;;;;;
000105E6;NE IBERIAN LETTER N;LO;0;L;;;;;N;;;;;
000105E7;NE IBERIAN LETTER M2;LO;0;L;;;;;N;;;;;
000105E8;NEIBERIAN LETTER M3;LO;0;L;;;;;N;;;;;
000105E9;NE IBERIAN LETTER R;LO;0;L;;;;;N;;;;;
000105EA;NE IBERIAN LETTER R3;LO;0;L;;;;;N;;;;;
000105EB;NE IBERIAN LETTER R2;LO;0;L;;;;;N;;;;;
000105EC;NE IBERIAN LETTER S3;LO;0;L;;;;;N;;;;;
000105ED;NE IBERIAN LETTER S1;LO;0;L;;;;;N;;;;;
000105EF;NE IBERIAN LETTER S2;LO;0;L;;;;;N;;;;;
000105F0; NE IBERIAN LETTER S87; LO;0;L;;;;;N;;;;;
000105FB;NE IBERIAN NUMERAL L;NO;0;L;;;;;N;;;;;
000105FF;NE IBERIAN NUMERAL P;NO;0;L;;;;;N;;;;;

000105FD;NE IBERIAN VALUE MARK 1/4;NO;0;L;;;;**1/4;**N;;;;;
000105FE;NE IBERIAN VALUE MARK 1/2;NO;0;L;;;;**1/2;**N;;;;;
000105FF;NE IBERIAN WORD SEPARATOR;ZS;0;L;;;;;N;;;;;

9. Bibliography

BURRIEL, J. Mª, MATA, C., FERRER I JANÉ, J, RUIZ, A.L. VELAZA, J., PEIRÓ, Mª A., ROLDÁN, C., MURCIA, S., DOMÉNECH, A. (2011): "El plomo escrito del Tos Pelat (Moncada, Valencia)", *Palaeohispanica* 11, 191-224.

COMES, R., MONCUNILL, N. (2009): "Propuesta de integración del signario ibérico en el proyecto Unicode", *Palaeohispanica* 9, 63-64.

CAMPMAJO, P. UNTERMANN, J. (1991): "Corpus des gravures ibériques de Cerdagne", *Ceretania* 1, 39-59.

CORREA, J.A. (1992): "Representación gráfica de la oposición de sonoridad en las oclusivas ibéricas (semisilabario levantino)", *AION* 14, 253-292.

CORREA, J.A. (2009): "Identidad, cultura y territorio en la Andalucía prerromana a través de la lengua y la epigrafía", en: F. Wulff Alonso y M. Álvarez Martí-Aguilar (coords.), *Identidades, culturas y territorios en la Andalucía prerromana*, Málaga, 273-295.

FERRER I JANÉ, J. (2005): "Novetats sobre el sistema dual de diferenciació gràfica de les oclusives", *Palaeohispanica 5*, 957-982.

FERRER I JANÉ, J. (2007): "Sistemes de marques de valor lèxiques sobre monedes ibèriques", *Acta Numismàtica* 37, 53-73.

FERRER I JANÉ, J. (2010): "El sistema dual de l'escriptura ibèrica sud-oriental", *Veleia* 27, 69-113.

FERRER I JANÉ, J. (2011a): "Ibèric **baikar**: un nou testimoni en un escif àtic de Sant Julià de Ramis", *Excavacions arqueològiques a la muntanya de Sant Julià de Ramis* 4, 203-217.

FERRER I JANÉ, J. (2011b): "Sistemas metrológicos en textos ibéricos (1): del cuenco de La Granjuela al plomo de La Bastida". *E.L.E.A.* 11, 99-130.

FERRER I JANÉ, J. (2013a): "Deux alphabets duals rupestres de Cerdagne", *Sources – Les cahiers de l'Âne Rouge* 1, 9-18.

FERRER I JANÉ, J. (2013b): "Els sistemes duals de les escriptures ibèriques", *Palaeohispanica* 13, 445-459.

FERRER I JANÉ, J. (2014): "A propòsit d'un sisè de bronze de baitolo amb la llegenda be II", *Acta Numismàtica* 44, 57-69.

FERRER I JANÉ, J. (e.p. 2014a): "Las dualidades secundarias de la escritura ibérica nororiental", *ELEA* 14.

FERRER I JANÉ, J. (2014b): "Ibèric kutu i els abecedaris ibèrics", *Veleia* 31, 227-259.

FERRER I JANÉ, J. (2014c): "Deux noveaux alphabets ibères rupestres de Cerdagne", *Sources – Les cahiers de l'Âne Rouge* 2, 11-20.

FERRER I JANÉ, J., GARCÉS, I. (2005): *El plom ibèric d'Olriols (Sant Esteve de Llitera, Osca)*. Acta Palaeohispanica IX, (Barcelona, 2004). *Palaeohispanica* 5, 983-994.

GÓMEZ-MORENO, M. (1922): "De epigrafía ibérica: El plomo de Alcoy", Revista de Filología Española, Tomo IX, 342-366.

GÓMEZ-MORENO, M. (1949): *Misceláneas, Historia, Arte, Arqueología*, Madrid 1949.

HOZ J. de (1985): "El nuevo plomo inscrito de Castell y el problema de las oposiciones de sonoridad en ibérico", *Symbolae Ludouico Mitxelena septuagenario oblatae*, 443-453.

HOZ, J. de, (1981): "Algunas precisiones sobre textos metrológicos ibéricos", *APL* 16, 475-486.

HOZ, J. DE (2010): *Historia lingüística de la Península Ibérica en la antigüedad. I. Preliminares y mundo meridional prerromano*. Madrid: Consejo Superior de Investigaciones Científicas.

HOZ, J. DE (2011): *Historia lingüística de la Península Ibérica en la antigüedad. II. El mundo ibérico preromano y la indoeuropeización*. Madrid: Consejo Superior de Investigaciones Científicas.

HUERTAS, C. J. (2009): Codificació informàtica del signari ibèric nord-oriental, Barcelona.

JORDÁN, C. (2005). ¿Sistema dual de escritura en celtibérico?, *Actas del IX Coloquio Internacional de Lenguas y Culturas Paleohispánicas (Barcelona 2004), Acta Palaeohispanica IX*, *Palaeohispanica*, 5: 1013-1030.

JORDÁN, C. (2007): Estudios sobre el sistema dual de escritura en epigrafía no monetal celtibérica, *Palaeohispanica*, 7: 101-142.

LEJEUNE, M. (1983): "Vieille-Toulouse et la métrologie ibèriques", *Revue Archeologique de Narbonnaise*, 16, 29-37.

LUJÁN, E. R., ORDUÑA, E. (e.p.), "Implementing a database for the analysis of ancient inscriptions: the Hesperia electronic corpus of Palaeohispanic inscriptions", en Tara Andrews – Caroline Macé (eds.), Methods and means for digital analysis of ancient and medieval texts and manuscripts (volumen de la colección Lectio. Studies in the transmission of texts and ideas), Lovaina, Peeters, en prensa.

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

MONCUNILL, N. (2007): *Lèxic d'inscripcions ibèriques (1991-2006)*, Tesi Doctoral, Barcelona, Universitat de Barcelona.

ORDUÑA, E. (2005): "Sobre algunos posibles numerales en textos ibéricos", *Palaeohispanica* 5, 491-506.

ORDUÑA, E., (2006): *Segmentación de textos ibéricos y distribución de los segmentos*, Madrid, Departamento de Filología Clásica. Facultad de Filología de la UNED.

ORDUÑA, E. (2013): "Los numerales ibéricos y el vascoiberismo", *Palaeohispanica* 13, 517-529.

ORDUÑA, E., LUJÁN E.R. (e.p.). Philology and technology in the Hesperia databank, *Journal of History, Literature, Science and Technology* (JHLiST).

PANOSA, M.I. (1993): "Nuevas inscripciones ibéricas de Cataluña", *Complutum* 4, 175-222.

QUINTANILLA, A. (1992): "Sobre la notación en la escritura ibérica del modo de articulación de las consonantes oclusivas", *Studia Palaeohispanica et Indogermánica J. Untermann ab Amicis Hispanicis Oblata, Aura Saecula* 10, Universitat de Barcelona, 239-250.

RODRÍGUEZ RAMOS, J. (2001): "Signos de lectura problemática en la escritura ibérica", *Archivo Español de Arqueología* 74, Madrid, 281-290.

RODRÍGUEZ RAMOS, J. (2004): *Análisis de epigrafía íbera*, Anejos de Veleia, Series Minor 12, Vitoria-Gasteiz, Universidad del País Vasco.

RODRÍGUEZ RAMOS, J. (2005): "Introducció a l'estudi de les inscripcions ibèriques", Revista de la Fundació Privada Catalana per l'Arqueologia ibèrica 1, 13-144.

SILES, J. (1985): *Léxico de inscripciones ibéricas*, Madrid, Epigrafia Hispánica, Ministerio de Cultura.

SILGO, L. (1994): *Léxico ibérico*, Valencia.

SIMON, I. (2011): "Interpunctiones palaeohispanicae", *Epigraphica* LXXIII, 1-2, 87-108.

TOVAR, A. (1951): *Léxico de las inscripciones ibéricas*, Madrid.

UNTERMANN, J. (1975): *Monumenta Linguarum Hispanicarum. I Die Münzlegenden*, Wiesbaden.

UNTERMANN, J. (1980): *Monumenta Linguarum Hispanicarum.* II *Die iberischen Inschriften aus Sudfrankreich*c

UNTERMANN, J. (1990): *Monumenta Linguarum Hispanicarum Band III. Die iberischen Inschriften aus Spanien*, Wiesbaden.

UNTERMANN, J. (1997): *Monumenta Linguarum Hispanicarum* IV *Die tartessischen, keltiberischen und lusitanischen Inschriften*, Wiesbaden.

VELAZA, J. (1991): *Léxico de las inscripciones ibéricas: 1976-1989*, Universitat de Barcelona, Barcelona.

VELAZA, J. (2006): "Chronica epigraphica iberica VII (2004-2005)", *Palaeohispanica* 6, 2006, 303-327.

VELAZA, J. (2009): "Epigrafía y literacy paleohispánicas en territorio vascón: Notas para un balance provisional", *Palaeohispanica* 9, 61-622.

VELAZA, J. (e.p.): "Hesperia: un nou concepte de corpus epigràfic online", *Auriga*.

10.- Images.

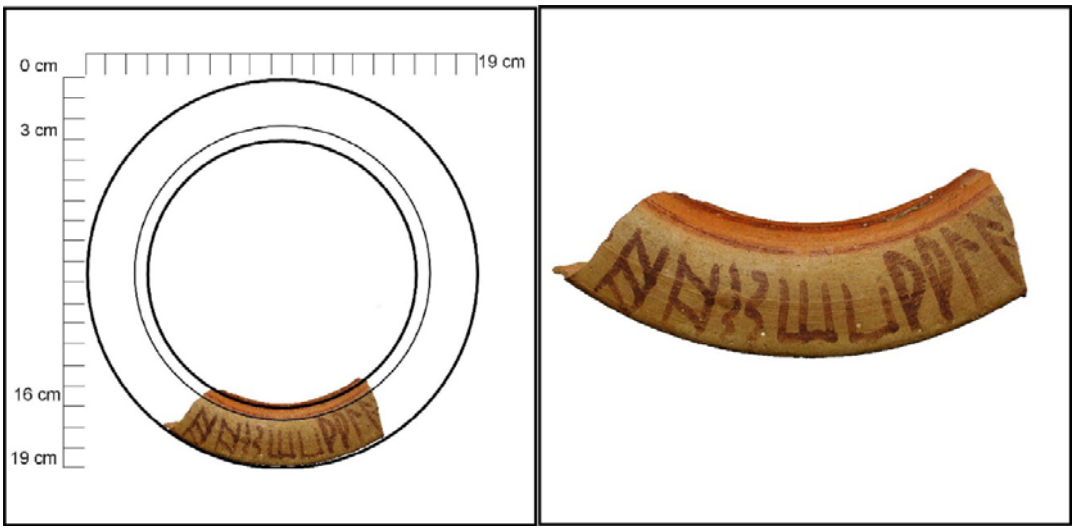


Fig 1.-Castellet de Bernabé's abecedary (Extended dual script).

Fig 2.- Tos Pelat's abecedary (F.13.77*) (Extended dual script). Above, detail of dualities for vowels and trill.

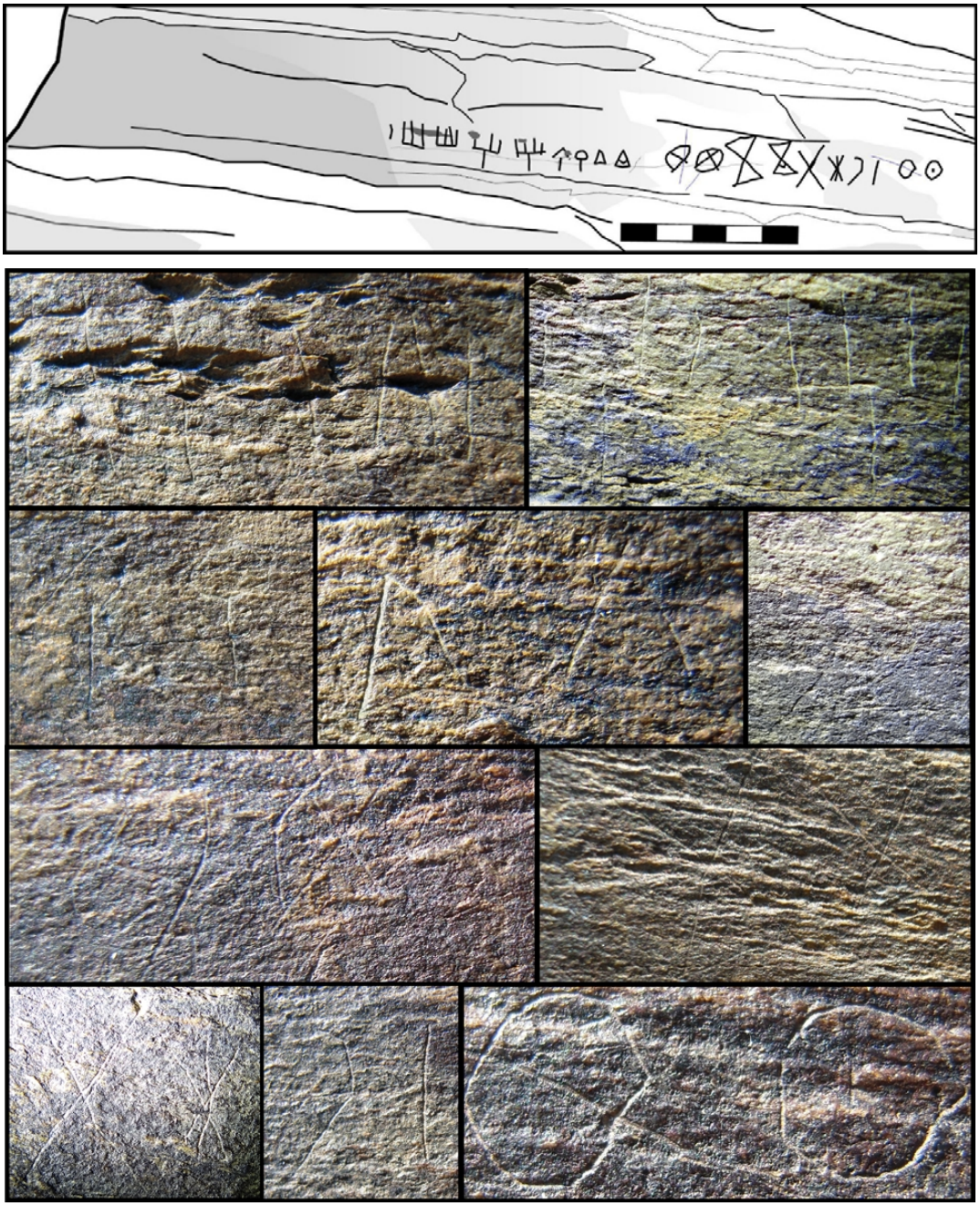


Fig 3.-Bolvir's abecedary (Standard dual script).

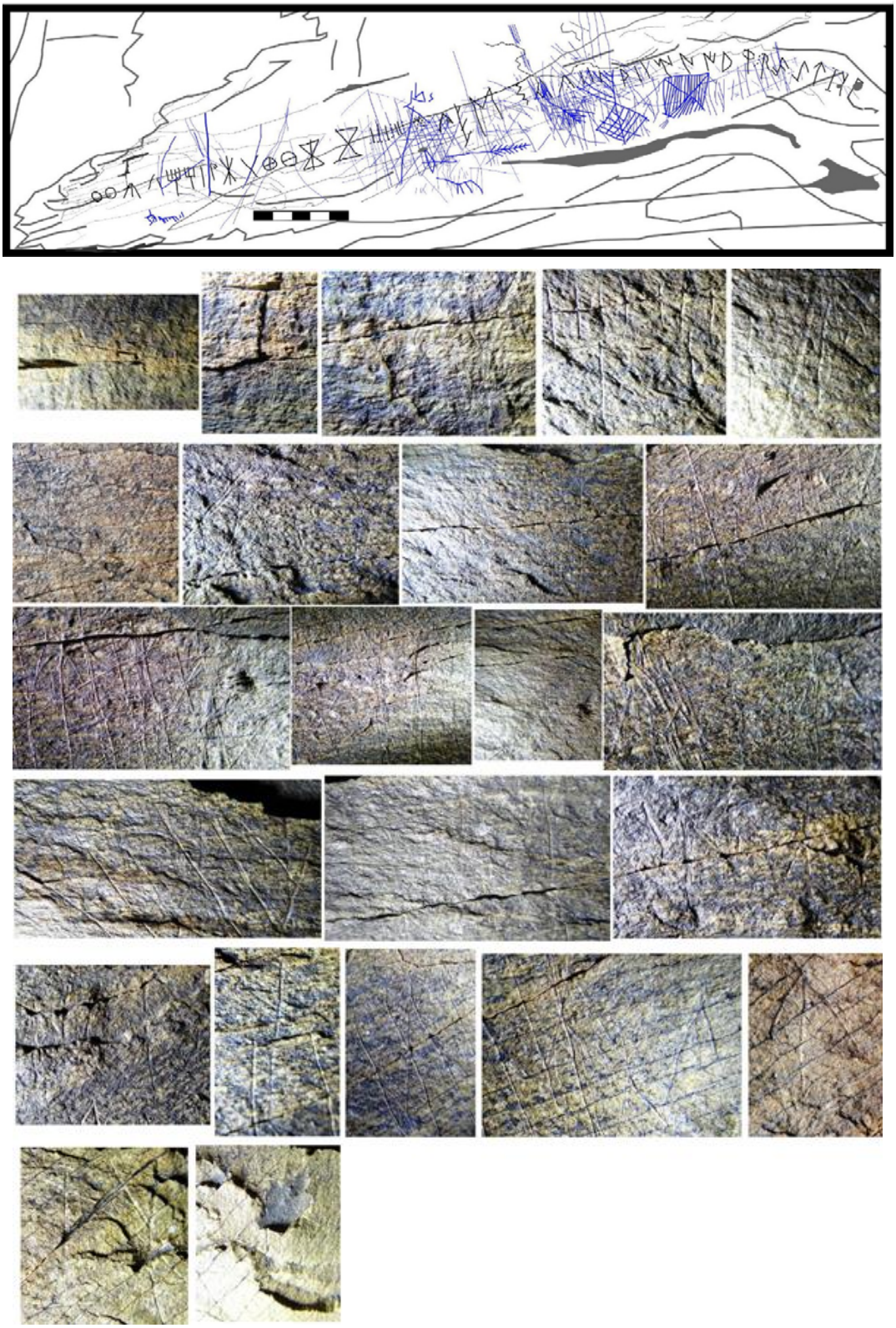


Fig 4.-Ger's abecedary (Standard dual script).

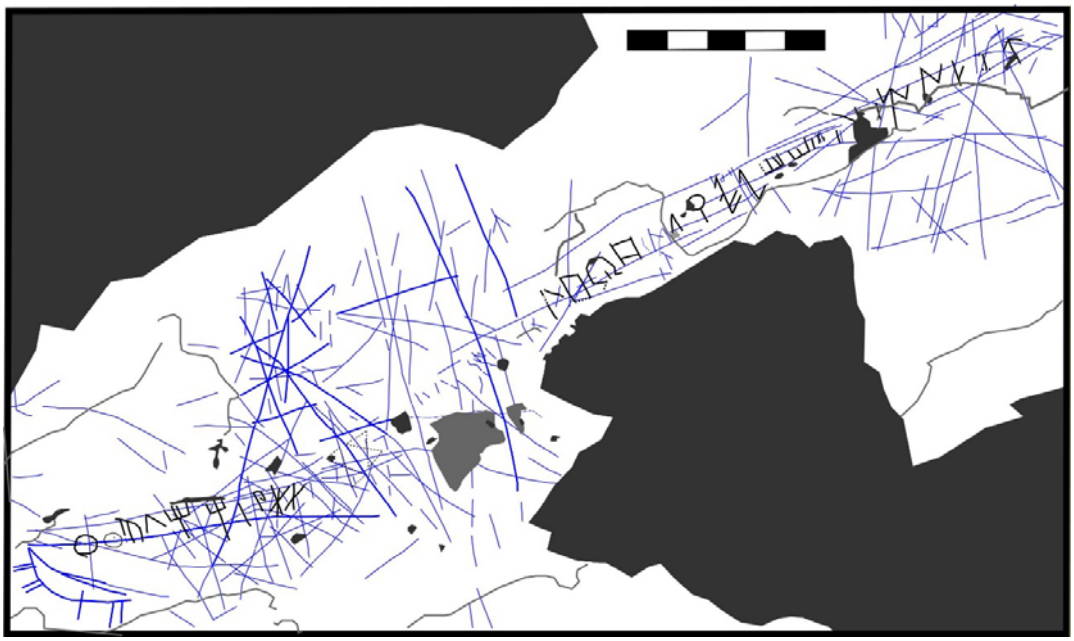
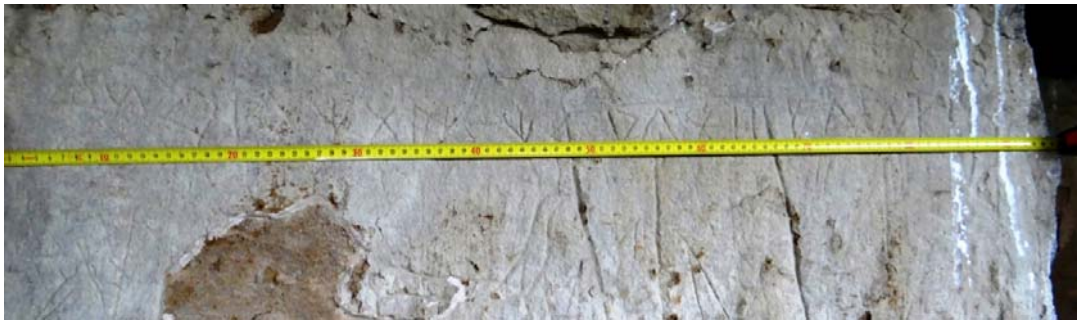


Fig 5.-La Tor de Querol's abecedary (Standard dual script).

Fig. 6.-Simplified abecedaries from Can Rodon (Non-dual script).



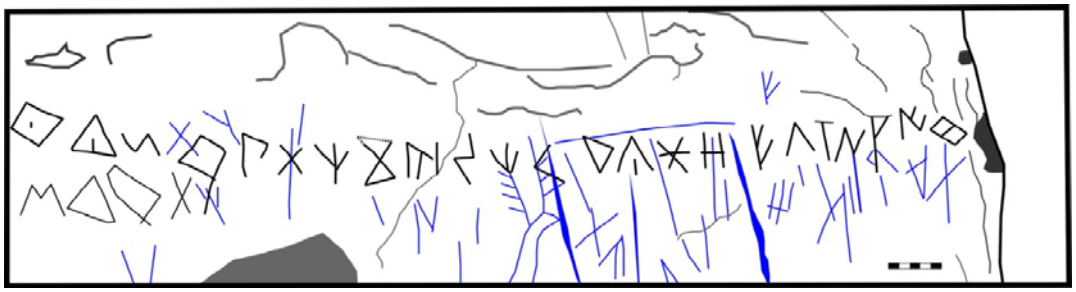


Fig. 7.-Esquirol's abecedary (Non-dual script).

		G/K	B	D/T				
A					S		Ś	
E					Ŕ		R	
I					M		N	
O					Ḿ		?	
U					L			

Fig. 8.- A possible non-dual abecedary.

		K	G	B	T	D				
A							S		Ś	
E							Ŕ		R	
I							M		N	
O							Ḿ		?	
U							L			

Fig. 9.-A possible dual standard abecedary.

				K	G	B	T	D						
Á		A							Ŝ		S		Ś	
É		E							Ř		Ŕ		R	
Í		I									M		N	
Ó		O									Ḿ			
Ú		U							Â		L			

Fig. 10.- A possible dual extended abecedary.

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 **ŕ**.

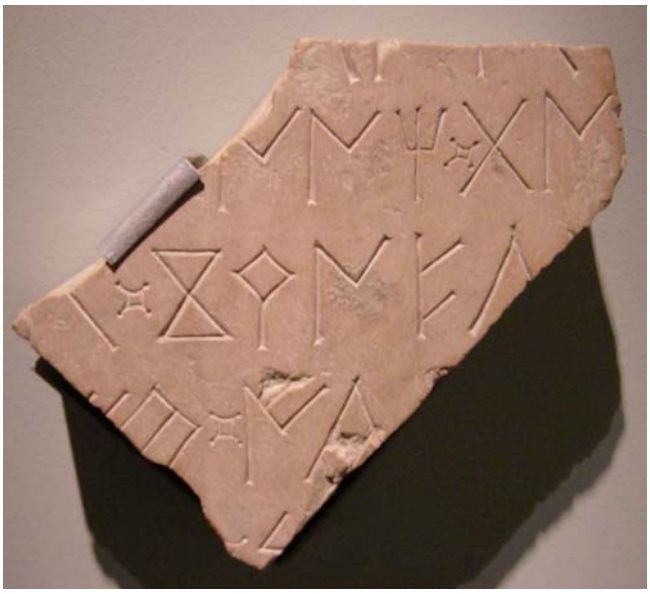


Fig 14.-Stone plaque from Empúries (non-dual script).

Fig 15.- Stele from Vispesa.

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 with 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).

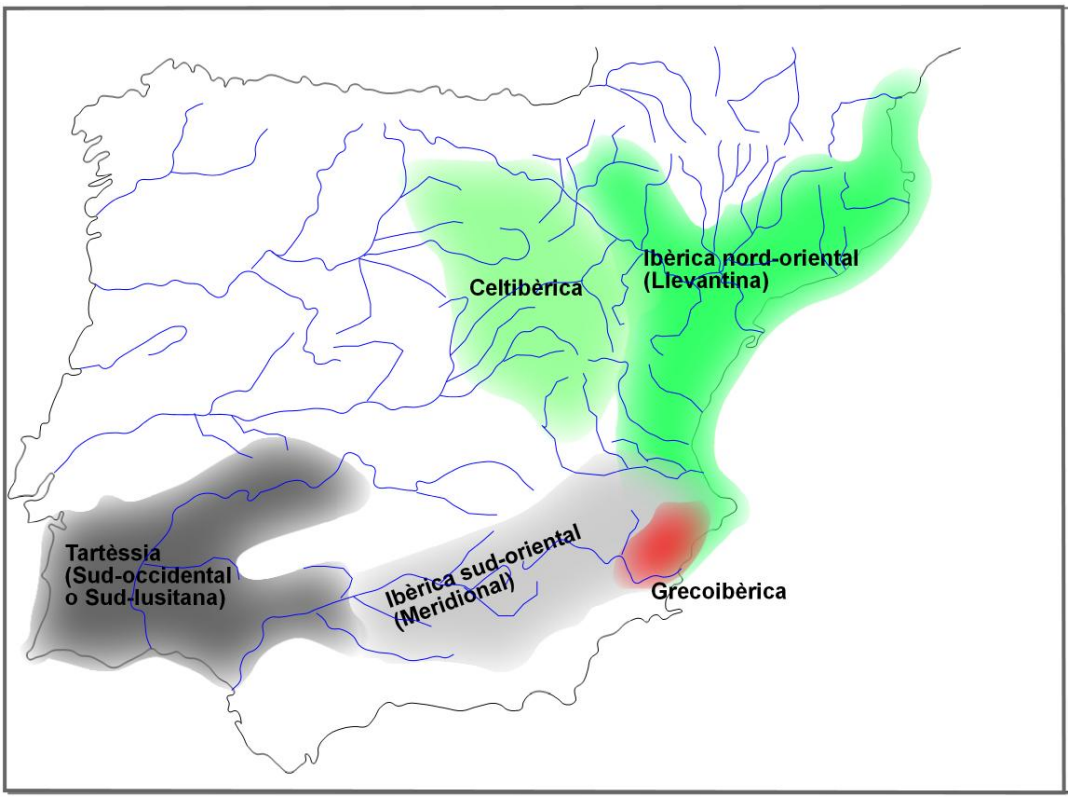


Fig. 23.- Palaeohispanic scripts.

A. Administrative

1. Title

Proposal to encode the North-eastern Iberian script in the SMP of the UCS.

2. Requester's name

Joan Ferrer i Jané, Noemí Moncunill, Javier Velaza

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

Researchers, University of Barcelona (Grup Littera)

4. Submission date

2015-03-10

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

1b. Proposed name of script

Northeastern Iberian.

1c. The proposal is for addition of character(s) to an existing block

No.

1d. Name of the existing block

2. Number of characters in proposal

52.

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attestedextinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category C.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

¿?

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

¿?

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. 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. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist incorrect understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casinginformation, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combiningbehaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibilityequivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such informationon other scripts. Also see Unicode Character Database http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html andassociated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in theUnicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, otherexperts, etc.)?

¿?

2b. If YES, with whom?

¿?

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, orpublishing use) is included?

Palaeohispanists and other scholars.

13

4a. The context of use for the proposed characters (type of use; common or rare)

Rare.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

Scholarly publications.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

No.

6b. If YES, is a rationale provided?

6c. If YES, reference

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposedcharacters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

Yes.

10b. If YES, is a rationale for its inclusion provided?

Yes.

10c. If YES, reference

North-western Iberian is related to other scripts derived from Phoenician script and some of their glyphs are alike.

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC

10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?