Eugenio R. Luján
(Universidad Complutense de Madrid)

**HESPERIA: THE ELECTRONIC CORPUS OF PALAEO-HISPANIC INSCRIPTIONS AND LINGUISTIC RECORDS**

Abstract: The Hesperia project is being currently developed at the Universidad Complutense de Madrid. It is a digitization project aiming at producing an electronic corpus of all the inscriptions in Greek and pre-Roman languages from ancient Hispania (Spain and Portugal). It also includes all the onomastic records in the pre-Roman languages of that area. This paper provides a general overview of the project with some examples of the various types of files used in it. It also mentions future developments of the electronic corpus and directions of research.

Key words: epigraphy, inscriptions, pre-Roman languages, onomastics, databases

1. **Introduction**

Since 1997 a team of the Department of Greek Philology and Indo-European Linguistics of the Universidad Complutense de Madrid, lead by Professor Javier de Hoz, has undertaken the task of producing a comprehensive electronic corpus of all the linguistic records related to the Iberian peninsula (Spain and Portugal) in Antiquity excepting Latin and Phoenician epigraphy.

The corpus thus includes inscriptions written in the following languages:

1. Iberian (a language with no known cognates which died out in Antiquity),
2. Celtiberian (a Celtic language),
3. Lusitanian (an Indo-European, non-Celtic language),
4. the unidentified language of some of the so-called “southern” inscriptions,
5. the unidentified language of the south-western (or “Tartessian”) inscriptions;
6. Greek.

It must also be taken into account that, besides that linguistic complexity, the inscriptions that we have to deal with have an additional level of variation, given that

---

1 This paper is part of the research project BFF2003-09872-C02-01, which has the financial support of the Spanish Ministry of Science and Technology. Some of the arguments and ideas which appear in it were originally formulated by others members of the team, especially by its director, Professor Javier de Hoz.
2 In collaboration with other researchers from the University of the Basque Country and the University of Barcelona.
3 Latin inscriptions are by far the largest set of inscriptions preserved from ancient Hispania. They were gathered in the second volume of the Corpus Inscriptionum Latinarum (CIL II), of which a new revised edition is currently being published. New findings, corrections of readings, new interpretations, etc. of inscriptions written in Latin and Greek alphabets and coming from Spain and Portugal are systematically surveyed in the journal Hispania Epigraphica, directed by I. Velázquez and currently published by the Universidad Complutense de Madrid. As for Phoenician epigraphy from Spain, it has also been covered by other projects. See [4].
they are not all written in the same writing system. The following writing systems were used for those languages:

1. Latin;
2. Greek;
3. “classical” (or Levantine) Iberian script;
4. “southern” script;
5. “south-western” script.

The last three belong clearly to the same family of scripts and must be divergent evolutions from the same original system; however, they are not identical with each other and must thus be considered different scripts.

The extant possible combinations of language and writing can be best summarized in a table:

<table>
<thead>
<tr>
<th>Language</th>
<th>Latin script</th>
<th>Greek script</th>
<th>“Classical” Iberian script</th>
<th>“Southern” script</th>
<th>“South-western” script</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iberian</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Celtiberian</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Lusitanian</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Southern” language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>“South-western” language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Greek</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Table 1. Languages and scripts of the Hesperia electronic corpus

Currently the total amount of inscriptions known in all these varieties of languages and scripts is beyond 2,000.

The corpus, however, must also include other non-epigraphical linguistic records related to these or any other language(s) spoken in Spain before the arrival of the Romans, like Basque or Basque-like languages or the “Old-European” substrate, among other possibilities. We have the following types of linguistic records:

1. Glosses of classical writers who mention some Hispanic words,
2. Names: personal names, god names, place names, ethnic names. God names appear exclusively on inscriptions; personal names mainly do so, too, but we also find a few personal names in Latin and Greek sources. As for place names and ethnonyms, many of them appear on inscriptions, basically on Latin

---

4 Detailed information about all these languages and scripts of ancient Spain and Portugal can be found in the volumes of J. Untermann’s *Monumenta Linguarum Hispanicarum (MLH)* [13]. Recent advances in this field have been summarized by de Hoz [9].

5 See figure 8 at the end of the paper for the signs of the classical Levantine script and the southern one.

6 The arrival of the Romans, however, did not mean that these languages and scripts were given up immediately. Instead Roman writing practices must have influenced indigenous ones and must have induced an increase of the number of certain types of inscriptions. We have Iberian inscriptions until the 2nd century CE.

7 See Gorrochategui’s study of Aquitanian [6], a language cognate with Basque, and Villar’s books [14, 15] for the possibility of other Indo-European languages different from Celtiberian and Lusitanian. We must also bear in mind the possibility that other Celtic languages different from Celtiberian were spoken in *Hispania*. 
inscriptions, but there are also some important Latin and Greek sources for them, like the works by Pliny and Strabo, and, specially, Ptolemy’s *Geography*. This is the kind of material that we have to work with – a very heterogeneous material from all points of view.

### 2. Designing the electronic corpus: objectives and scope

The design of the electronic corpus had to be made in accordance with the various objectives that it was intended to be used for. I will briefly review those objectives.

1. **First**, one of the main goals was to produce an edition of the Palaeo-Hispanic inscriptions which did not become outdated immediately after – or even before – its publication. In 1997 the fourth and last volume of the impressive *Monumenta Linguarum Hispanicarum (MLH)* by Jürgen Untermann was published. In the span of time going from 1975 – the date of publication of the first volume of this work, which included the inscriptions on coins – and 1997 hundreds of new inscriptions have been found. So even if the *MLH* are still the standard edition of the Palaeo-Hispanic inscriptions, they do not include all the inscriptions known to date. Clearly what was needed was a type of edition that could be easily updated, so that new findings could be added to the corpus in a short time after they appeared and they did not have to wait until the publication of additional volumes of *addenda*. The traditional printed edition had thus to be given up in favor of the electronic edition. The aim of the team is thus to provide the scientific community with a constantly updated corpus of inscriptions.

2. An electronic edition also complies with another important objective of such a corpus – preservation. This aspect of preservation is very important to all digitization projects. Digitizing documents, inscriptions in this case, is a guarantee that even if the original document itself came to be lost, at least most of the relevant information that it contains will be kept for the future. The transmission and copying of electronic information becomes easier every day and the possibility that copies of an electronic corpus stored at far away places are destroyed at the same time becomes more remote.

3. A third important objective – peculiar to this project when compared to other digitization projects currently in progress – is that it must also contribute to a better understanding of the languages of the inscriptions included in it. Obviously this is not the case with the few ancient Greek inscriptions from Spain. Lusitanian and Celtiberian inscriptions can at least be understood in their more general traits – both are Indo-European languages and can thus benefit from the long tradition of comparative linguistics when trying to understand how their grammars work and what their words mean. Iberian, however, has no known cognate and it is still a language that we do not properly understand. Indeed, some progresses toward the understanding of its grammar have been made along the 20th century, so that the value of a few suffixes is approximately known, as shown in the following table.

---

8 See the paper by de Hoz – Luján [10] for the list of new Iberian inscriptions. To those the new Celtiberian inscriptions should be added [11], as well as some inscriptions on coins absent from the first volume of *MLH* [13] and now published in the comprehensive corpus by García-Bellido et alii [5].

9 See volume III.1 of *MLH* [13] as well as the paper by Correa [1].
<table>
<thead>
<tr>
<th>SUFFIX</th>
<th>VALUE/MEANING</th>
</tr>
</thead>
<tbody>
<tr>
<td>-en</td>
<td>possession</td>
</tr>
<tr>
<td>-sk en</td>
<td>Genitive plural</td>
</tr>
<tr>
<td>-‘yi</td>
<td>‘me’?</td>
</tr>
<tr>
<td>-ta</td>
<td>locative?</td>
</tr>
<tr>
<td>-ka</td>
<td>agent?</td>
</tr>
</tbody>
</table>

Table 2. Identified Iberian suffixes

We also have a good understanding of the naming formula, which allows for the identification of personal names in the inscriptions, but Iberian, for the most part, remains a language that we can read but we cannot translate. The design of the database had accordingly to be made so as to become a tool for the linguistic analysis, too.

4. Furthermore, the electronic *corpus* had also to contribute to a better understanding not only of the languages of the inscriptions, but also of the writing systems, for some of them are still not totally deciphered. We can obviously read inscriptions in Greek and Latin scripts and we can also read inscriptions written in the classical Levantine Iberian script\(^ {10} \), but things get more complicated when we come to the southern and south-western scripts. There are some signs of the southern script whose value is still disputed, but at least we can control linguistically the assignation of values given that some of the inscriptions written in this script are in Iberian language, so some sequences that we know in the classical Levantine script re-appear here. As for the south-western script, although there is a certain agreement regarding the value of some signs, there is not yet a general consensus for many of them\(^ {11} \). The electronic *corpus* must also contribute to an easier handling of the inscriptions which may in the end result in a better understanding of the script and the definitive assignment of phonetic values to its signs\(^ {12} \).

We have already seen (§ 1) that we have to deal with two basic types of documents in this *corpus* – inscriptions and onomastic records. Given that the kind of information that we need to provide about them is different, we finally decided that it would be more convenient to have two different types of files, even if linked to each other. Both types of files are currently run in the programme FileMaker 5.0, although we are exploring now the possibility of migrating to another programme.

\(^{10}\) In the classical Levantine script there is just one sign, conventionally transcribed as Y (imitating the shape of the sign) or alternative as m, whose phonetic value is not totally clear. From the contexts in which it appears it has been inferred that it must be a vocalic element combining the phonetic traits of nasality and velarity.

\(^{11}\) For the main proposals of interpretation for these inscriptions can be found in the works by Untermann, *MLH IV* [13], Correa [2], de Hoz [7], Correia [3], and Rodríguez Ramos [12].

\(^{12}\) See one of the planned developments of the database in § 5.
3. The epigraphical database

This is indeed the most complex file, especially because, as opposed to previous editions of Palaeo-Hispanic inscriptions, we decided to include, whenever it is possible, very detailed information about the archaeological context in which every inscription was found, besides linguistic, paleographic, epigraphic, and philological information about them. The file is organized into various sections, as can be seen in figure 1. The files are filled in by specialists of different disciplines, mainly archaeologists and philologists, which are responsible for the various sections.

The upper part (see figures 1-5) is kept throughout all the sections, so that the inscription that we are dealing with can be quickly identified. In this first section, the general information about the inscription is provided: type of material, technique used for making the inscriptions, measures both of the object and the inscriptions...

In the second section we provide the text of the inscriptions with the variant readings if any (figure 2). Variants must not be understood here in the same way as in classical scholarship when dealing with various manuscripts transmitting the same text. In this case it refers to divergent readings proposed by different scholars for particular signs of the inscriptions.

As it was impossible at a first stage to have our own direct readings of every inscription we decided to take Professor Untermann’s *Monumenta Linguarum Hispanicarum* as a starting point, so that we could thus have at least the text of the entire corpus of Palaeo-Hispanic inscriptions in the database and begin to use it as a tool for a better and more systematic analysis of the inscriptions. Nevertheless, we did not
just reproduce Untermann’s readings – we controlled and changed them when we felt that it was needed on the basis of photographs and/or proposals by other scholars. Using a terminology borrowed from Textual Criticism, at this stage we did not have a critical edition, but at least a revised one. On the other hand, our first task was entering in the database the inscriptions which had been published after the completion of Untermann’s *MLH*, having thus a complete corpus of inscriptions which was not available elsewhere. This is a task that we have kept on doing regularly, because, as I said above, fortunately there are findings of new Iberian, Celtiberian, or “Tartessian” inscriptions every year.

At the same time, we have begun to study in museums and private collections the inscriptions, having thus our own photographs and readings, which we progressively enter into the database. Indeed, in the fields of the database we reflect at which stage every file is, allowing thus the user to know who is responsible for the reading. I would like to stress that readings by other authors are not just discarded, but transferred to the critical apparatus of the file. We always try to keep in mind that even if we can read Iberian (at least the most widespread variant of its script, see § 1) we do not understand it for the most part, so we cannot simply reject alternative readings and interpretations that might prove to be the right ones in the end.

In the next section (see figure 3) a philologist provides the data concerning the language, the paleography and any other epigraphic or linguistic relevant information.
Our aim is also to provide graphic material of every inscription, at least a photograph and a sketch, as shown in figure 4.
As I said, we are very concerned with the archaeological context of the inscriptions, which can help a lot in their study and provide important clues. So there is a specific section of the file devoted to that (see figure 5).

Figure 5. Hesperia: inscription file – archaeological data.

The final section includes the bibliography on that particular inscription. The programme in which we run the database supports complex searches. So it is very easy now to know how many inscriptions can be dated in a given century, which of them have been found in the same type of context (burial site, town, etc.), which have been written on the same material, and so on.

As I have stressed several times, when electronically editing Iberian inscriptions we need to bear in mind that one of the most important aims of such an edition must be to contribute to a deeper understanding of the language, providing the means for the electronic parsing of texts and the devising of automatic analyses. That involves that the variants of readings registered in the critical apparatus should also be taken into account when carrying out that kind of analysis, for, as opposed to what happens when making a critical edition of a Latin or a Greek text, in most cases we do not have a clue as to what reading is the right one. In fact, we do not usually know even where a word or a morpheme begins and ends. This has a direct bearing on the critical edition, since we need to introduce in the main text markers that can direct the parsing programme to the critical apparatus when a variant reading exist, so that that variant can also be considered a possibility. Taking into consideration those variants, however, introduces great complexity in the programming, for there may exist variants at various places of the same inscriptions and this results in a multiplication of the possibilities of
combination. It should also be taken into account that Iberian is written for the most part in a semi-syllabary that has no different signs for voiced and voiceless stops – although we know that they existed in the language, as shown by the inscriptions in Latin and Greek scripts – and may also leave aside r and l following a stop. That means that the sign ka can stand for ka, ga, kra, gra and, in final position, even just for k. We also have to take into account Iberian inscriptions written in the so-called “southern” system, which is different from the classical, eastern one, in Latin script, and in a variant of a Greek, Ionian alphabet.

For the moment, one of the members of the team, E. Orduña, has created a parsing programme that allows for some complex searches and represents a first step towards a systematic analysis of the Iberian language, which – we hope – will contribute to improve our understanding of the language. This programme has been written in Perl and allows for searches using different linguistic criteria that can be combined together. In a very simple and convenient way morphological structures that are similar to those that we know in Iberian can be found easily and very quickly.

4. The onomastic database

In the onomastic database there is indeed no place for photographs nor archaeology, but, instead, it is very important to be able to access to such information as what other names are attested on the same inscription, which variants of the same name can be found in different sources, etc. The kind of file for all the names is basically the same, but some adaptations had been necessary for each type.

In figure 6 below we have the file that we use for god names.
As can be seen with this example, this type of file may also be very complex. This happens mainly when the same name is attested several times in different sources, which is the case, for instance, with the goddess *Ataecina*, an indigenous deity of ancient *Hispania* that a large number of Latin inscriptions were consecrated to. The file has various parts. In the first one we provide the general information – a regularized form of the name which will serve for the indexes, the classification of the name (god name, personal name, place name...), and other related names (e.g., epithets of a god). We then provide all the variants of that name in the various sources and also the equivalent among the Roman gods if it is attested in the sources (in this example, *Ataecina* is identified with the Roman goddess *Proserpina* in some inscriptions). Finally, we provide a detailed historical and linguistic commentary with the appropriate bibliography.

Up to now mainly the god names had been introduced in the database. We are currently working with place names and we have also begun to work with personal names.

5. **Future developments**

Besides the task of completing the databases and carefully revising all the information included in them we have already planned and begun to work in two directions, which can serve as a hint of what future developments and application this electronic *corpus* may have.
a) First, under the direction of Professor Javier de Hoz a map has been produced in which all the places of finding of the inscriptions included in the database have been plotted, as shown in figure 7.

This map consists of various layers, each of them containing the places of finding of specific kinds of inscriptions. What we would like to do in the near future is linking the map to the database of inscriptions, so that maps can be produced automatically when a query using some given criteria is made in the database (e.g. inscriptions of the 2nd century BC, Iberian inscriptions having the sequence *eban* in their texts, etc.).

Figure 8. Signs of the southern (left) script and the classical Iberian script (right) according to de Hoz [8]
b) As stated above, we also intend to use the databases as a mean to gain some insight into languages and/or scripts not fully understood yet. In this sense we are working on the development of a “decipherment” device for the south-western inscriptions, which we have conceived as an aid for researchers working on that corpus of inscriptions who may want to check if their proposed phonetic values for the various signs are coherent or not. We will provide a grid with the different signs and it will be possible to fill in that grid with the values assigned by a particular researcher to the signs. Those values will then be automatically applied to all the inscriptions of the corpus so that the researchers may check in a few seconds if the readings resulting from their proposals are likely or not.

References


erlujan@filol.ucm.es