



Universiteit  
Leiden  
The Netherlands

## The pre-Roman elements of the Sardinian lexicon

Swanenvleugel, C.

### Citation

Swanenvleugel, C. (2025, February 12). *The pre-Roman elements of the Sardinian lexicon*. LOT dissertation series. LOT, Amsterdam. Retrieved from <https://hdl.handle.net/1887/4180290>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4180290>

**Note:** To cite this publication please use the final published version (if applicable).

---

## 1 Sardinia and its linguistic background

---

Sardinia is the second-largest island of the Mediterranean Sea. Its history has been defined by its relative isolation as an island, as well as by contacts with the three landmasses surrounding it: the Italian Peninsula, the Iberian Peninsula, and North Africa. Currently an autonomous region of Italy, Sardinia has belonged to different spheres of influence at different times. Most of these have left linguistic traces in one way or another. In chronological order, Sardinia was part of, or influenced by, the following foreign political powers: Carthage, Rome, the Vandal Kingdom, the Byzantine Empire, the Pisan and Genovese merchant republics, the Crown of Aragon, the Spanish Empire, the House of Savoy, and finally the Italian state. By far the strongest influence was exerted by the city of Rome, as evidenced by the linguistic romanization of Sardinia. With the possible exception of the Vandal Kingdom, all other aforementioned political powers have left lasting imprints, albeit in varying degrees, on the Sardinian language (see Putzu 2019 for an overview).<sup>1</sup>

Before the Carthaginian and Roman dominance, Sardinia had been inhabited for multiple millennia. The prehistoric Sardinians left a rich archeological record, but we have no direct evidence of the language(s) they spoke. Although Sardinia was eventually completely romanized linguistically, traces of the native pre-Roman language(s) survived in the Sardinian lexicon and toponymy. These linguistic substrate traces have been the subject of many scholarly investigations since the early 20<sup>th</sup> century (see § 1.3). The present study aims to contribute to this field of research, by synthesizing and evaluating the lexical evidence for the Sardinian substrate, its phonological and morphological features, and its linguistic affiliation.

### 1.1 The languages of Sardinia

The native language of most of Sardinia is Sardinian, which is the direct continuation of Latin spoken in Sardinia. As such, it is considered a separate language or language branch in the Romance family tree (Bossong 2016: 65; Mensching & Remberger 2016: 270; Lupinu 2023: 27). Modern Sardinian is characterized by considerable internal variation. Its varieties are usually classified into a number of dialect groups. The main distinction is between Logudorese in the north and Campidanese in the south. Nuorese in the center-east is either considered a

---

<sup>1</sup> For an overview of the history of Sardinia, see Guidetti (1987).

sub-variety of the former, or a variety of its own. In the west, the Arborese dialects are transitional between Logudorese and Campidanese (Virdis 1988; Mensching & Remberger 2016: 270; Molinu & Floricic 2017). Overviews of the Sardinian language, its linguistics, and its philology, can be found in Wagner (1997 [1950]), Atzori (1982), Contini & Tuttle (1982), Blasco Ferrer (1984a), Jones (1997), Bosson (2008: 225–245), Mensching & Remberger (2016), Blasco Ferrer, Koch & Marzo (2017), and Lupinu (2023). The appendix to this volume contains an overview of some of the major historical phonological developments of Sardinian as a whole, and of its individual dialect groups.

Besides Sardinian, other Romance languages are spoken in smaller areas: the Sardo-Corsican dialects of Sassarese and Gallurese in the north (Maxia 2017), Catalan in and around the city of Alghero (Dessi Schmid 2017), and the Tabarchino variety of Ligurian in the Sulcis archipelago in the southwest (Toso 2017)(see Figure 1.1.1). These language varieties have gained a presence in Sardinia in post-classical times and are a testimony to its diverse linguistic contacts throughout history.

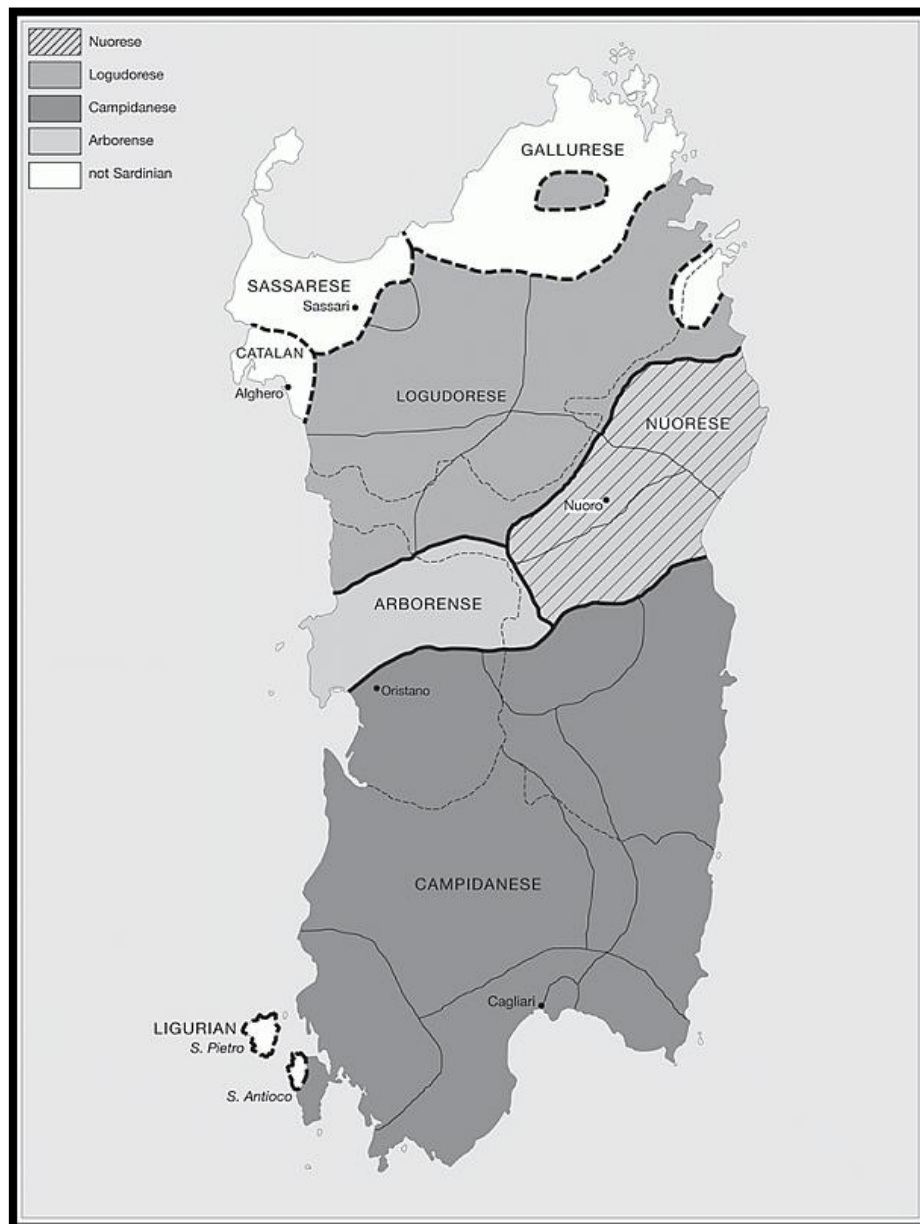


Figure 1.1.1: The languages of Sardinia (Mensching & Remberger 2016: Map 17.1; modified version of Viridis 1988: 905)

## 1.2 The romanization of Sardinia

In 238 BCE, Sardinia was annexed by the Roman republic, after which it entered a process of administrative, cultural, and linguistic romanization. This was a gradual process. It is clear that the Romans did not abolish the institutions and infrastructure left behind by the Carthaginians, but rather integrated them into the new political order (Meloni 1987). Roman authority was established relatively quickly in the more densely populated fertile plains in the south and west of the island. The more central mountainous regions have been argued to have resisted significantly longer (Dyson 1985; Rowland 1985).<sup>2</sup>

Various hypotheses have been posited regarding the linguistic romanization of Sardinia. Wagner (1928; 1997: 105–120) argues for an early romanization of Sardinia on the basis of certain lexical archaisms. He further notes lexical variation between northern and southern Sardinia (e.g. *éliye* < Lat. *ēlex* next to *iliži* < Lat. *īlex* ‘holm-oak’ etc.), which he interprets as resulting from two separate waves of romanization. The first of these would have reached all of Sardinia, while the second was more limited to the south (Wagner 1928: 8). However, this does not explain why some of the more archaic formations are found in the south, and the innovative ones in the north: e.g. Camp., Barb. *fórru* < Lat. *fornus* vs. Log., Nuor. *fúrru* < Lat. *furnus* ‘oven’ (Wagner 1928: 8–14). Blasco Ferrer (1989: 26) disagrees with Wagner’s (1928) two-step romanization, but does agree that an initial wave of romanization reached all the way into the mountainous center of the island (Blasco Ferrer 1989: 16–23). He dates the Latin of this first wave to the time of Plautus and Terence (i.e. 3<sup>rd</sup>–2<sup>nd</sup> century BCE), and assumes it to be closely related to the Latin of Africa (Blasco Ferrer 1989: 34–40). Later waves of Latin would have continued to influence Sardinia’s more accessible regions, with some particular local sources of innovation at various points in time (Blasco Ferrer 1989: 74–76). Thus, Wagner and Blasco-Ferrer agree that the presumed conservative nature of the dialects of Sardinia’s mountainous center is the result of the first wave of romanization, after which it developed in relative isolation. This stance is also taken by Lupinu (2005), who essentially follows Wagner.<sup>3</sup>

A different model is proposed by Viridis (2018: 12–25; 2020), who argues that the Campidanese dialects do in fact preserve some more archaic features than the

<sup>2</sup> But recently more nuanced views have been put forward (Trudu 2012; Farre 2016; Canu 2016).

<sup>3</sup> On the question of the archaicity of Sardinian as a whole, see Mensching (2004).

northern Sardinian dialects.<sup>4</sup> Viridis argues that the Campidanese dialects therefore continue a more archaic stage of Latin, indicative of an earlier time of romanization, even if the modern Logudorese-Nuorese dialects are more conservative.<sup>5</sup> An important realization in Viridis' hypothesis is that greater conservatism cannot be equated with greater archaicity of language varieties, or vice versa. This crucial distinction is absent from Wagner's (1928) and Blasco Ferrer's (1989) accounts. It is in any case clear that the linguistic romanization of Sardinia was not a single process, as it was initiated and completed at different times in different places.<sup>6</sup>

### 1.3 The Sardinian substrate

The romanization of Sardinia was at the same time its Indo-Europeanization. While the starting date of 238 BCE is early in terms of the former, it is comparatively late with regard to the latter. It has long been noted that the Sardinian toponymy and lexicon contain numerous elements that cannot satisfactorily be etymologized within Latin (cf. Wagner 1932; 1997: 254–280 for an overview). It is probable that at least a part of these stem from the language or languages spoken on Sardinia at the time of the Roman conquest. These non-inherited elements have been extensively studied by various authors, often in the context of substrates in the wider geographical region. For a detailed history of the study of linguistic substrates in Sardinia and elsewhere in the Mediterranean, see Devoto (1955), Serra (1956), Paulis (2008), and especially Craddock's (1969: 18–47) critical review of almost two centuries of literature on the topic. In what follows I will present an overview of the most important literature regarding the Sardinian substrate specifically. Previous work on the Sardinian substrate can roughly be divided into three categories, all of which have contributed to the ongoing debate regarding Sardinia's linguistic prehistory in different ways:

1. Studies that focus on the specificities of the non-inherited linguistic material within Sardinia, without necessarily comparing it to phenomena found outside Sardinia. These are discussed in § 1.3.1.
2. Studies that observe and compile similarities in non-inherited linguistic material between Sardinia and various other areas. This category com-

---

<sup>4</sup> The examples include some of the lexical variants discussed by Wagner (1928), as well as phonological and morphological features of Campidanese.

<sup>5</sup> Cf. Mastino (1993), who shows that literacy, and probably fluency in Latin, was lower in the center than in the coastal areas of Roman Sardinia.

<sup>6</sup> The attestations of Punic inscriptions well into the imperial period is one indication that the definitive romanization must have been a lengthy process (Adams 2003: 209).

prises the bulk of the studies in the earliest phases of substrate research. An overview of these studies is given in § 1.3.2, and the relevant lexical evidence is discussed in § 11.

3. Studies that argue for the identification of the pre-Roman language of Sardinia as a relative of some other, attested, language. These present a clearly defined linguistic scenario that can be tested by examination of the data. The various proposals in this category are discussed in § 1.3.2.1 and in § 10.

Various terms have been used to refer to the pre-Roman languages of Sardinia. Most common is the term “Paleo-Sardinian”. The term “Nuragic”, based on the eponymous archeological culture, is used specifically by Pittau (1984; 1995). Both of these terms may give the impression that we are dealing with a single, homogeneous language. While this is possible (if we exclude Punic), it is by no means certain (cf. § 2.3.1). As such, I prefer to use the more neutral, be it less compact, term “pre-Roman Sardinian” to refer to the non-inherited features of the modern Sardinian varieties of putative substrate origin.

### 1.3.1 Features of the Sardinian substrate

The language-internal analysis of possibly non-inherited features does not immediately shed light on the linguistic affiliations of any pre-Roman language in Sardinia, but it does allow us to differentiate between what should and what should not be compared, i.e. which features may indeed be of pre-Roman origin, and which are due to more recent developments. Moreover, a detailed understanding of the distribution of substrate features is essential in recovering the pre-Roman linguistic dynamics within Sardinia.

#### 1.3.1.1 The Sardinian pre-Roman lexicon

With the exception of Punic loans (§ 1.3.2.1), only a single word from a pre-Roman language is directly attested and identifiable; i.e. the form *NURAC* (§ 5.2.2), inscribed on the nuraghe *Aidu Entos* in Bortigali (Gasperini 1992: 303–306). As a result, our main source of evidence for language contact between Sardinian Latin and other, local languages, is the lexicon of the Romance varieties of Sardinia. Collections of Sardinian lexical material of proposed pre-Roman origin can be found in Wagner (1932; 1997: 254–280) and Hubschmid (1953). Important etymological sources regarding the Sardinian lexicon are Wagner’s *Dizionario etimologico sardo* (DES), and Paulis’ (1992) work on Sardinian plant names. The phonological and morphological features of Sardinian lexical items of pre-Roman origin, have usually been discussed per individual etymon, across

many studies (cf. § 8 and § 9). No complete overview and evaluation of all proposed pre-Roman phenomena in the lexicon has as of yet been given. In chapter § 8, I discuss phonological features found in words of presumed pre-Roman origin, and evaluate which of these can be projected back to a pre-Roman language. In chapter § 9, the evidence for proposed pre-Roman morphological features in the Sardinian lexicon is reviewed.

### 1.3.1.2 Pre-Roman toponyms in Sardinia

Another important line of evidence are Sardinian place names, many of which have been argued to be of pre-Roman origin. Overviews can be found in Serra (1956: 181–198; 1960) Paulis (1987; 2008), Wolf (1998a), and Pittau (2013). From Meyer-Lübke's (1925) discussion of pre-Roman toponyms of the Iberian Peninsula onwards, various scholars have relied on toponymic evidence to posit phonological or morphological properties of pre-Roman Sardinian languages as well (e.g. Terracini 1927; Bertoldi 1928; Serra 1960; Hubschmid 1963a; Wolf 1998a; Paulis 2008), or to argue for certain linguistic affiliations of the substrate (Bertoldi 1928; Serra 1954a; Hubschmid 1963a; Pittau 1995; Blasco Ferrer 2017). For methodological reasons outlined in § 2.1.2.2.1, this study will focus primarily on the lexical evidence, rather than toponymic evidence.

### 1.3.1.3 Substrate-induced features of Sardinian

Besides lexical and toponymic elements of potential pre-Roman origin, it has been proposed that some of the phonological characteristics of Sardinian too might have arisen by influence from one or more local substrate languages. Most commonly discussed in this regard are the following phenomena:

- Loss of inherited *\*f* in the Barbaricine dialects (e.g. Bertoldi 1937a: 169; Wagner 1997: 285; Iribarren Argaiiz 1997: 338; Paulis 2008: 34; Blasco Ferrer 2010a: 85).
- Development of inherited *\*k* as *ʔ* in the Barbaricine dialects (e.g. Wagner 1997: 285).
- Vowel prothesis before inherited initial *\*r-* in Campidanese (e.g. Bertoldi 1937a: 169; Wagner 1997: 285; Iribarren Argaiiz 1997: 338; Paulis 2008: 34; Viridis 2020: 35).

Both the loss of *\*f* and the rise of a prothetic vowel before initial *\*r-* have been seen as evidence for the presence of a Basque or Iberian substrate in Sardinian (Bertoldi 1937a: 169; Blasco Ferrer 2010a: 85; Viridis 2020: 35). Wagner (1997: 285) compares the presence of *ʔ* (as the outcome of *\*k* in various Sardinian dialects) to the phonological systems of Afro-Asiatic languages.



However, both the loss of *\*f* and the development *\*k > ʔ* have been shown to be of more recent nature in the Barbaricine dialects, and are thus not informative on the phonological characteristics of any substrate language (cf. Paulis 1984a: lxxx–lxxxv, xc–xcvii; Wolf 2011: 613; Loporcaro 2013: 57). The idea of vowel prothesis before initial *\*r-* due to some Iberian substrate influence is more resilient (e.g. recently Viridis 2020: 35). Basque does not allow word-initial *r-*, adding a prothetic vowel in Latin and Romance loans (e.g. Bq. *errege* ‘king’ << Lat. *regem*), a feature that is probably shared with ancient Aquitanian and Iberian (Michelenena 1990: 332–333). Proving that the Sardinian phenomenon must be caused by substrate influence from a pre-Roman language of the Iberian Peninsula is difficult. Paulis (1984a: xxx fn. 45) notes that similar phenomena are found elsewhere in Romance too, and restrictions against initial */r/* are cross-linguistically common (Labruno 2021). It is therefore possible that the development of a prothetic vowel before *\*r-* in Campidanese is simply the result of Sardinian-internal phonological developments.<sup>7</sup> A pre-Roman Iberian origin of this phenomenon remains speculative.

### 1.3.2 The Mediterranean substrate

The study of the pre-Roman linguistic situation in Sardinia cannot be seen separately from the study of such linguistic substrates in the wider geographical region. From the early 20<sup>th</sup> century onwards, numerous studies have been conducted on the pre-Roman elements in Romance languages. Especially in the early stages of substrate research, many proposals revolved around the concept of a Mediterranean unity, in cultural as well as linguistic terms, before the Indo-Europeanization of Europe (cf. Ribezzo 1919: 104; 1920: 83; Bertoldi 1928: 222; Terracini 1936: 59). Initially, the hypothesized prehistorical linguistic connections between various regions were to a large extent based on the similarities between certain toponyms, as well as on traditional views on the origins of peoples recorded by Greek and Roman authors (e.g. Philipp 1920: 2486; Ribezzo 1920: 87–97; Terracini 1927; Bertoldi 1928).<sup>8</sup> In most studies on this “Mediterranean” substrate, Sardinia has had a prominent role (cf. Devoto 1955; Craddock 1969: 18–47; Wigman 2023: 18–20). As the details of the Mediterranean hypothesis have always remained somewhat vague with regard to the concrete historical and linguistic scenarios that are to be inferred, it is no surprise that diverging proposals have been put forward to account for the Sardinian attested data. The

<sup>7</sup> Cf. Paulis (1984a: xxxi–xxxii) for a possible scenario.

<sup>8</sup> On the use of toponyms as evidence for substrate connections, cf. § 2.1.2.2.

proposals usually involve “layers” or “currents” of linguistic influence, connecting different regions in the Mediterranean in varying chronological orders.

One of the first hypotheses that were firmly grounded in linguistic analysis is Terracini's (1927: 145–146). On the basis of an analysis of the endings found in Sardinian toponyms, he argues that Sardinia was “intersected by two currents in particular: one that unites it with Africa and Iberia, and another that seems to have its center in the northern Tyrrhenian region”,<sup>9</sup> the first of which would have been stronger and more durable.<sup>10</sup> Terracini (1927: 146) believes that these “Afro-Iberian” toponyms in Sardinia, which would possibly be related to Punic presence in Sardinia, should be able to be analyzed with the help of Berber and Libyan. Bertoldi (1928) discusses pre-Roman suffixes and roots which he identifies in various Sardinian toponyms. Largely in agreement with Terracini, he posits a “Libyco-Sardo-Iberian” current (Bertoldi 1928: 241). Later, he connects this to the Libyan and Iberian migrants and mercenaries in Sardinia described by Pausanias (X: 17; Bertoldi 1937a: 148, 151).<sup>11</sup> He also interprets Terracini's “northern Tyrrhenian current” as an east-to-west “Anatolian-Aegean-Iberian” current (Bertoldi 1928: 250; 1939: 100). The idea behind this “Anatolian-Aegean-Iberian current” is based on a combination of “Ibero-Sardinian”, “Pyrenaeal-Alpine”, and “Pyrenaeal-Caucasian” comparisons, all of which are compared to Sardinian toponyms as well (Bertoldi 1928: 249–250). With regard to the pre-Roman substrate of Italy, Bertoldi (1936: 296; 1939: 100) makes a distinction between an “Aegean-Anatolian” element introduced with the (putative) Etruscan colonization, and a native “Italo-Rasennic” element that was somehow assimilated to Etruscan due to their mutual linguistic affinity.<sup>12</sup> This is contradictory to his idea that the original pre-Roman language of Sicily (i.e. Sicanian) was related to languages in Iberia and Sardinia, suggesting a connection to his “Anatolian-Aegean-Iberian current”, but that it was unrelated to Etruscan (Bertoldi 1937a: 169). Unfortunately, it is not made clear how the idea of all these independent linguistic “currents” relates to the idea of a single Mediterranean-wide substrate, to which Bertoldi (1928: 222) subscribes. In later articles, Bertoldi takes the idea of a Mediterranean

<sup>9</sup> “[...] intersecata particolarmente da due correnti: una che l'unisce all'Africa e all'Iberia, l'altra che pare aver il suo centro nel Tirreno settentrionale.”

<sup>10</sup> Cf. Bottiglioni (1929: 101), who identifies toponyms of Iberian, Ligurian, and Etruscan origin in Corsica, agreeing well with Terracini's (1927: 145–146) “north Tyrrhenian” idea.

<sup>11</sup> Bertoldi (1937a: 151–152) relates this to Wagner's (1932: 223–224) proposed correspondences between the Sardinian substrate and Berber languages.

<sup>12</sup> This is especially curious because Bertoldi's “Rasennic” is derived from Etr. *Rasenna*, the Etruscan endonym denoting Etruria (Bonfante & Bonfante 1983: 218).

substrate as a given (e.g. Bertoldi 1929; 1930), without specifying the extent of this phenomenon, or the historical reality that would have been behind it. In yet other articles (Bertoldi 1931; 1932; 1933), numerous lexical comparisons are made, without a clear conclusion as to the “current” in which the posited wide-spread word families would have originated (cf. Craddock 1969: 37–39 for a discussion of Bertoldi’s various strata). In contrast to Bertoldi’s intricate scenarios, Alessio returns to a more clearly defined view of the Mediterranean substrate as a genealogical linguistic unit (e.g. Alessio 1935: 141–143; 1944: 105), although he allows for internal dialectal variation (Alessio 1944: 107–108). However, this does not mean that Alessio’s hypotheses are more concrete or easier to verify, and he does in fact operate with an additional, more recent substrate layer in Sardinia that would have its provenance in North-Africa (Alessio 1957: 263). In general, Alessio frequently revisits Bertoldi’s comparisons and adds less convincing comparanda to the equation. I agree with Craddock (1969: 42) that “Alessio leaves even the impartial reader with a far less favorable impression than in the case of Bertoldi”.

Most specific, but also most complex, in terms of stratification of the substrate lexicon, are the studies by Hubschmid, who also dedicated considerable attention to the Sardinian substrate specifically (Hubschmid 1953). His comparisons of putative pre-Roman forms are, if not convincing, supported by an impressive amount of lexical material from and by detailed formal, semantic, and etymological analysis of said material. More than Bertoldi and Alessio, Hubschmid attempts to explain the observed linguistic facts with historical scenarios — although often of speculative nature (e.g. Hubschmid 1950a: 225–226; 1950b: 133). Hubschmid does not depart from the *a priori* assumption of a unified Mediterranean substratum, but posits various pre-Indo-European language families that would at various times have expanded and receded in the Mediterranean region. Among his hypotheses are that of a “Eurafrican” substrate, and a more recent and wide-spread “Hispano-Caucasian” substrate, both of which would have had some relation to Basque (Hubschmid 1953: 91; 1960a: 24, 40; 1965: 156). The Euraf-rican substrate is based mainly on comparisons between Sardinian, Berber and Basque, as well as on comparanda from Gallo-Romance and Insular Celtic. The Hispano-Caucasian substrate includes comparisons between words from Sardinian, Basque, Occitan, Alpine Romance varieties, Greek, Caucasian languages, even reaching Burushaski (Hubschmid 1953: 109). These two language families (as Hubschmid sees them) would have caused intricate contact situations with each other and with yet other languages at various times in the past. This becomes clearest in Hubschmid (1963a: 147–148):

“In Sardinia we can assume at least six Paleo-Sardinian linguistic strata:

- 1) an oldest Eurafrican substrate, which in the south encompassed Northwest-Africa, and that reached until Lusitania in the west, to Gaul in the north, to Italy in the east;
- 2) an Iberian substrate, related to Eurafrican, but stemming rather from Hispania;
- 3) a Hispano-Caucasian substrate stemming from the east, which reached until Lusitania in the west, Gaul in the north, the Caucasus in the east, and which in the south probably encompassed parts of North-Africa too;
- 4) a Tyrrhenian or Etruscan substrate, stemming from the east as well;
- 5) a Libyan “Hamitic” substrate related to Berber, or an early Semitic substrate;
- 6) perhaps another, unknown, very old Paleo-Sardinian substrate.”<sup>13</sup>

On the whole, no amount of linguistic data adduced by the substratists discussed has been able to lend wide-spread acceptance to any configuration of the Mediterranean hypothesis. The main problem is that the scenarios proposed are so all-encompassing and abstract, and at the same time so vaguely defined, that they are almost impossible to prove or disprove — even if some of their proposals may someday prove to be close to the truth. This is exacerbated by the lack of a clearly delineated methodology as to how formally and semantically different two comparanda are allowed to be until the comparison is rejected. Another weakness is the heavy reliance on toponyms as evidence.<sup>14</sup>

Even if the resulting hypotheses fail to convince, the large amounts of lexical data gathered from a wealth of disparate sources especially by Bertoldi, Alessio and Hubschmid, are impressive in their own right, and I have made grateful use of these scholars’ efforts in this study. The many cross-Mediterranean comparisons with relevance to the Sardinian substrate that have been proposed are discussed in § 11. On the basis of the geographic distribution of the accepted comparisons, I discuss conclusions that can potentially be drawn on any prehistoric linguistic connections in § 11.4 – 11.5.

---

<sup>13</sup> (Translation mine).

<sup>14</sup> See § 2.1.2 for a discussion of these methodological issues.

## 1.3.2.1 A note on “Ligurian”

The notion of an ancient “Ligurian” language often recurs in the context of Mediterranean substrate elements. Classical sources mention a people called *Liguri* in Latin, Λίγυες in Greek. Even in antiquity, ideas on the ethnological, political, and linguistic identity of the *Liguri*, varied and sometimes contradicted each other. Some sources place *Liguri* in the Iberian Peninsula, others between the Iberian Peninsula and the Rhône, and yet others east of the Rhône into northern Italy (see Bats 2019 on the use of Gr. Λίγυες). In addition, there are mentions of a Ligurian presence in Corsica and in Sicily (Cuno 1878: 90–150; Müllenhoff 1892: 171–193; Kretschmer 1905: 108–109; Mullen 2013: 27). In the 19<sup>th</sup> century, the *Liguri* were widely held to be a pre-Indo-European people (Kretschmer 1905: 109–110).<sup>15</sup>

There are no inscriptions that can be attributed with certainty to a language by the name of Ligurian. The sole evidence of a language spoken by the *Liguri* (provided that the term *Liguri* indeed refers to the speakers of a single language) is limited to personal names and toponyms (Mullen 2013: 26). This very limited record has paradoxically not stopped authors from indulging in various types of far-reaching speculations. Classifications of the Ligurian language have ranged from pre-Indo-European to a separate branch of Indo-European to a variety of Celtic. Cuno (1878: 91) is the first to identify the *Liguri* as a Celtic people. Based on inscriptional evidence, Kretschmer (1905: 99, 127) favored classification as an independent Indo-European branch, close to Celtic.<sup>16</sup> The idea of Ligurian being an unattested Indo-European language is continued in more recent scholarship (e.g. Hubschmid 1953: 116–117; Paulis 1992: 305; Dworkin 2012: 20; Mullen 2013: 29). Most frequently however, the term “Ligurian” features as a more general label referring to substrate phenomena roughly in the area of the southwestern Alps, i.e. modern-day Liguria and surroundings (e.g. Battisti 1928; 1960; Bertoldi 1928; 1937b; Alessio 1944; 1949; Ambrosini 1983). It is not always clear whether Ligurian is used purely as a geographic indicator (as in Alessio 1957: 238), or as a specific hypothetical pre-Indo-European language (as in Bertoldi 1937b: 140). It is sometimes proposed that there was a connection between Ligurian and Iberian (Bertoldi 1928: 229; Alessio 1935: 141–142; Hubschmid 1953: 116), or that Ligurian was also found in Sicily (e.g. Alessio 1944: 96 fn. 15). Notably, Ligurian has also

<sup>15</sup> This idea was partly rooted in contemporary racial theories (cf. Kretschmer 1905: 109).

<sup>16</sup> However, the inscription on the basis of which Kretschmer (1905: 99, 127) classifies Ligurian as an Indo-European language is now considered to be Lepontic (cf. Stifter et al. 2022: VB-3.1).

been mentioned as one of the linguistic substrates of Sardinia (e.g. Alessio 1944: 96 fn. 15; Devoto 1955).

In short, both in ancient and in modern times there has been considerable disagreement on the geographic localization of the *Liguri* or Λίγυρες, and on the linguistic classification of these people's unattested language. Mullen (2013: 26–27) rightly calls for caution in attributing any specific linguistic or historical reality to the term “Ligurian”. As such, the term “Ligurian” has lost much of its usefulness as a label for substrate phenomena, regardless of whether these substrates are of Indo-European or pre-Indo-European nature.

#### 1.3.2.2 A note on “Iberian”

Like in the case of “Ligurian”, the term “Iberian” is frequently used as a label for certain substrate phenomena by proponents of the Mediterranean substrate hypothesis. “Iberian-Sardinian” substrate features have been posited time and time again (Forsyth Major 1893: 154; Wagner 1907a: 408; 1932: 238; 1997: 259; Bertoldi 1928: 229, 238–239; Terracini 1936: 67; Bertoldi 1937b: 144, 151, 160, 169; Alessio 1944: 263; Devoto 1955: 219; Battisti 1960: 377–378; Hubschmid 1960a: 7; Paulis 1992: 21). These have usually been explained as testifying to a migration from the Iberian Peninsula to Sardinia in prehistoric times (e.g. already Wagner 1907a: 408). In this context, frequent reference is made to Pausanias' (X: 17: 5) remark that there was at some point an Iberian migration to Sardinia led by Norax (Frazer 2012a: 523 [1898]). It is difficult to verify if this semi-mythical migration could explain the linguistic facts.

Unlike Ligurian, however, Iberian refers to an actually attested language. Iberian was a non-Indo-European language, attested in ca. 2300 inscriptions, found from the southeastern coast of Spain to the Languedoc region in France (Moncunill & Velaza 2020: 593). De Hoz (2009; 2011: 56–57) argues that Iberian was spoken natively in the southeastern part of the Iberian Peninsula, and that it expanded to the northeast as a commercial lingua franca. Velaza (2019: 162) on the other hand argues that Iberian was in fact spoken as a native language across the area where it is attested, even if it would have expanded just prior to the earliest inscriptional evidence. Whatever the case, there is onomastic evidence documenting the existence of yet another, non-Indo-European language spoken in the northeastern part of the area where Iberian is attested (Untermann 1969). Untermann notes structural similarities of non-Iberian personal names in Iberian inscriptions in southwestern France to personal names recorded more eastward along the modern French Mediterranean coast. Despite being aware of the issues surrounding that term (cf. § 1.3.2.1), he labels this onomastic phenomenon “Ligu-

rian” (Untermann 1969: 8–9). On the basis of this non-Iberian onomastic material, Gorrochategui and Vallejo (2019: 360) posit the existence of an “autochthonous language in Roussillon and Catalonia, different from Iberian”. Beyond the onomastic evidence, lexical substrate connections between Iberian and Ligurian (and Sardinian) have been suggested (Kretschmer 1905: 109–110; Battisti 1928: 672; Alessio 1935: 141–142; Terracini 1936: 67).

Iberian was long thought to be related to Basque (Orduña 2019: 219–221). However, with the decipherment of the Iberian script came the realization that Basque is irrelevant to understanding Iberian texts (Orduña 2019: 223). Basque and Iberian do not appear to be closely related, or in fact related at all. Any structural similarities may be explained as resulting from language contact (Gorrochategui 1995; Trask 1995: 79).<sup>17</sup> The conflation of Iberian and Basque as constituting a single “substrate layer” is present in many publications by proponents of the Mediterranean substrate hypothesis (e.g. Bertoldi 1937b; Alessio 1944: 109 fn. 114; Hubschmid 1953: 117–118; Battisti 1960: 352). It persists even in more recent scholarship (e.g. Blasco Ferrer 2017: 81). However, the fact that Iberian and Basque were not (closely) related implies that comparisons of pre-Roman linguistic features of Sardinian to Basque (§ 10.3) cannot be used as direct evidence in favor of an Iberian migration to Sardinia. Conversely, pre-Roman correspondences between Sardinian and languages in the Iberian Peninsula other than Basque (§ 11.2.1.1) are no direct evidence in favor of the presence of a Vasconic language in pre-Roman Sardinia.

Paulis (2008) notes some similarities between features of the Sardinian substrate and of the Iberian language. The phonological constraints that appear to operate in the pre-Roman toponyms of the Barbagia di Ollolai, as described by Wolf (1998a: 28–35), are similar to those observed in Iberian, as well as to those reconstructed by Gorrochategui and Lakarra (2001) for Proto-Basque (Paulis 2008: 32–33). Additionally, various Iberian personal names are reminiscent of medieval Sardinian personal names of probable pre-Roman origin (Paulis 2008: 39). Paulis (2008: 48) interprets these structural similarities, as well as the handful of Sardinian-Basque lexical correspondences, as evidence for an Iberian origin of the pre-Roman language of Sardinia. In his view, the linguistic ancestors of Basque, Iberian, and the Sardinian substrate would have coexisted in the Iberian Peninsula ca. 10 000 – 9 000 years ago. This time depth is roughly twice that of the

<sup>17</sup> In general, Vasconists seem to be less optimistic about a genetic link, at whatever time-depth, between Basque and Iberian than Iberianists (e.g. Trask 1995: 77–80; cf. Orduña 2019: 221).

Indo-European language family, however, making it unlikely that superficial linguistic features as personal names would be recognizably preserved into the attested descendent languages. Moreover, this date appears to primarily be based on genetic insights (such as Vona & Calò 2006; Caramelli et al. 2007) that have since been surpassed due to the rapid developments in the study of ancient DNA (cf. Calò et al. 2021).

### 1.3.3 Proposed connections between the Sardinian substrate and attested languages

Apart from the unattested substrata of unknown affiliation (cf. § 1.3.2), various known languages or relatives thereof have been proposed to be spoken in Sardinia prior to the Roman conquest. These include Punic (§ 10.1), Berber (§ 10.2), Basque (§ 10.3), and Etruscan (§ 10.4). With the exception of Punic, for which the historical context is well-understood, each of these languages figures prominently in Mediterranean substrate comparisons (§ 1.3.2), as part of larger “currents” of linguistic influence. In more recent times, attempts have been made to explain all or most of the Sardinian substrate lexicon as coming from a language related to Berber (Argiolas 2012; 2020), Basque (Blasco Ferrer 2010a; 2010b; 2011; 2017), or Etruscan (Pittau 1995). These hypotheses are mutually exclusive. In § 10, the relevant evidence for each of these hypotheses is reviewed.

#### 1.3.3.1 Phoenico-Punic

Of all proposed connections between the Sardinian non-inherited lexicon and some non-Indo-European language, Punic is the most well-established. The first Phoenician settlements emerged in Sardinia in the eighth century BCE (Bondì 1987a; 1987b). These eventually came under Carthaginian influence from about 500 BCE (Bondì 1987c), until Sardinia was conquered by the Roman Republic in 238 BCE. Phoenician and Punic inscriptions in Sardinia are attested from the ninth century BCE until the second (or early third) century CE (Bondì 1987d; Adams 2003; 209):<sup>18</sup> a presence of more than a millennium with centuries of continuity into the Roman period. Given these facts, the presence of Phoenician or Punic elements in non-inherited Sardinian is expected. The first attempts at identifying these were made by Spano (1872). Unfortunately, Spano sees a Canaanite or Phoenician etymology behind almost any Sardinian word. His etymologies can almost invariably be dismissed (cf. Wagner 1957: 37). Wagner (1932: 212; 1997: 158–161; 1957: 104–109) instead proposes a Phoenico-Punic origin for a

<sup>18</sup> There may in fact be a Punic graffiti, in the Roman alphabet, on a hypogeum in San Salvatore di Cabras dated to the era of the Constantinian imperial dynasty — i.e. the 4<sup>th</sup> century CE (Meloni 1987: 290).



small number of lexical items,<sup>19</sup> hypothesizing that Phoenician and Punic linguistic influence was restricted to the plains of Campidano and Planargia (1997: 158).<sup>20</sup> The corpus of Phoenico-Punic etymologies in the Sardinian lexicon has been expanded by Paulis (1990). While Paulis accepts more evidence, he also rejects some of Wagner's earlier proposals at the same time. A difficulty in analyzing the material is that not all accepted punicisms in Sardinia have a clear Semitic etymology. In fact, Argiolas (2010: 23–24) has suggested some to be of Berber origin. The evidence for Punic loans in Sardinian is discussed in § 10.1.

### 1.3.3.2 Berber

Multiple ancient sources mention a migration of “Libyans” or “Africans” from North-Africa to Sardinia; e.g. Pausanias X: 17: 2 (Frazer 2012a: 523), Cicero XIX: 42, 45 (Cicero 1984: 19).<sup>21</sup> In linguistics too, various scholars have operated with an “African” or “Libyan” component within the Sardinian substrate, typically informed by lexical comparisons to Berber (e.g. Terracini 1927; 1936; Bertoldi 1928: 241; 1937b: 148, 151; Wagner 1932: 224; 1997: 263–263, 280; Serra 1954a; 1960; Hub-schmid 1953: 91; 1960a: 24, 40; 1965: 156; Alessio 1957: 263). Recently, Argiolas (2012; 2020) has argued for a Berber substrate in Sardinia. Given the possibility that some Berber linguistic material entered Sardinia in a Punic colonial context (Paulis 2008: 42; Argiolas 2010: 23–24), a key question is whether it is necessary to posit an independent North-African substrate, predating Punic presence in Sardinia. This, and the proposed Berber-Sardinian comparisons, are discussed in § 10.2.

On a practical note, there have been different conventions for the phonological representation of forms from the Berber languages. Wherever relevant, I have attempted to represent these in a uniform fashion. Wherever the notation used here differs from that used by the source cited, the original notation is given in a footnote. Forms found in sources that do not employ a linguistic transcription method, are cited between angle brackets (i.e. <...>). For the reconstruction of Proto-Berber, I adhere to the system used by Múrcia Sánchez (2011), which is the main source of the reconstructed Berber forms cited.

<sup>19</sup> See also Friedrich's (1957) mildly positive evaluation of Wagner's proposals.

<sup>20</sup> Wagner also discusses possible Phoenician origins of a range of toponyms (Wagner 1932: 208–212; 1997: 152–157; 1957: 38–43, 78–104), which are outside the scope of this study. Cf. also Paulis (1990: 625–639).

<sup>21</sup> Cf. Frazer (2012b: 320–321 [1898]) for a discussion.

## 1.3.3.3 Basque

Being the sole surviving pre-Indo-European language in Europe, it is no surprise that Basque is a major suspect when it comes to the identification of non-inherited linguistic material in the languages of Europe. While it is in itself reasonable to assume that the Basque language at some point in prehistory had relatives that were spoken in the wider region, it does not follow that all pre-Indo-European languages were related to Basque. In fact, even the Basque language's closest neighbor in antiquity, Iberian, appears to be genetically unrelated to it, despite being a pre-Indo-European language too (cf. Gorrochategui 1995; Trask 1995: 79).<sup>22</sup> For the Sardinian-Basque connections posited in the framework of the “Mediterranean substrate hypothesis”, see § 1.3.2.

During the last decades, a number of authors have discussed apparent connections between Basque and Sardinian. Santo Moreno (2000: 174) interprets the putative Sardinian-Basque lexical correspondences as pre-Roman *Wanderwörter*, rather than remnants of extinct language families. Bengtson (2017: 86–88) sees the pre-Roman Sardinian substrate as part of a larger Basque-Caucasian macro-family. Paulis (2008: 20) notes some possible similarities of medieval Sardinian personal names to Aquitanian and Iberian personal names.

In a recent account, the relation between Basque and the pre-Roman language(s) of Sardinia is explored highly concretely by Eduardo Blasco Ferrer. In a series of publications, he argues that the pre-Roman toponymic evidence on Sardinia demonstrates an ancient migration of Vasconic speakers to Sardinia (Blasco Ferrer 2010a; 2010b; 2011; 2017). Partly based on Wolf's (1998a) observations on recurring patterns in Barbaricine toponyms of pre-Roman origin, he divides synchronically opaque Sardinian toponyms into segments supposedly related to Basque lexical roots. For instance, the Ogliastran toponyms *Dolai*, *Dodolai* and *Iridolai*, analyzed as *\*dol-ai*, *\*do-dol-iai* and *\*iri-dol-ai*, in his view, all three contain a cognate of Bq. *odol* ‘blood’ (Blasco Ferrer 2011: 88).

Blasco Ferrer's methodology, like his conclusions, are deeply problematic. First and foremost, it is impossible to establish the *Benennungsmotive* of the hundreds of pre-Roman Sardinian toponyms. As a consequence, the meanings attributed to them are exclusively based on the proposed Basque etymologies and therefore speculative to the extreme. Additionally, our understanding of the derivational processes behind the formation of these toponyms is limited to the observation

<sup>22</sup> See Trask (1995) for an overview of the many languages that Basque has been proposed to be related to.

that some phonological sequences seem to reoccur. This essentially does not inform us on anything as long as there is no way to securely identify them as morphemes. As a result, Blasco Ferrer's (2011: 88) analysis of a toponym like *iri-dolai* as < \**iri-dol-ai* is entirely arbitrary. As such, the forms and meanings of the "morphemes" that are identified in Sardinian toponyms and compared to Basque are the outcome of speculative assumptions that are neither provable nor falsifiable on the basis of the evidence.

What is worse, the Proto-Basque phonological system as reconstructed by Michelena (1990) is dated to around two thousand years ago (Trask 1995: 67). This post-dates both the Phoenician and Roman arrivals on Sardinia, and also post-dates the arrival of "paleo-Basque" speakers on Sardinia between the 9<sup>th</sup> and 3<sup>rd</sup> millennium BCE, if we follow Blasco Ferrer (2010a: 162; 2017: 80–81). Even if we depart from Lakarra's (2009; 2016) internally reconstructed earlier phase of Proto-Basque, there are multiple millennia in which virtually no language change must have occurred, neither in the direct ancestor of Basque, nor in its hypothetical sister language on Sardinia.

In short, Blasco Ferrer's view of pre-Roman Sardinia as linguistically Vasconic can be considered unsubstantiated, due to its reliance on morphologically and semantically opaque toponyms.<sup>23</sup> This does not mean that a prehistoric migration from Iberia to Sardinia, or any other type of connection between pre-Roman Sardinian and Basque is ruled out; it simply does not follow from the material adduced by Blasco Ferrer. Any claim of prehistoric ties between Sardinia and Iberia must be supported by lexical evidence, including both form and meaning. Semantic scrutiny is the only tool we have to minimize the risk of chance correspondence between two or more phonologically similar forms (cf. § 2.1.2).<sup>24</sup> The proposed lexical correspondences between Sardinian and Basque are discussed in § 10.3.

#### 1.3.3.4 Etruscan

Like Basque, Etruscan was a non-Indo-European language spoken amidst Indo-European languages. Unlike Basque, Etruscan died out around the 1<sup>st</sup> century CE (Wallace 2008: 1). As for most of the apparently pre-Indo-European languages

<sup>23</sup> Cf. Wolf (2011) and Loporcaro (2013) for a more in-depth discussion of Blasco Ferrer's methodology.

<sup>24</sup> Although Blasco Ferrer's views do not seem to have gained widespread popularity in linguistic circles, they have unfortunately found their way into some genetic studies (e.g. Chiang et al. 2018: 1431).

that survived long enough to be attested, including Basque, Minoan, Tartessian etc., the linguistic affiliation of Etruscan has long puzzled scholars (Wallace 2008: 215–218). At this point, the only languages that are commonly accepted to be related to Etruscan are the Rhaetic language in the Alps, and the Lemnian language on the Greek island of Lemnos (Rix 1998: 60; Wallace 2008: 218–225). Regarding the eventual provenance of this “Tyrrhenian” language family, there are two main hypotheses. One is that the Etruscans as well as the Rhaetians originated in Anatolia, of which Lemnian would be a local vestige (Schachermeyr 1929; Beekes 1993; 2003; Van der Meer 2004; Kloekhorst 2022a). The other is that Etruscan and Rhaetic were native to Italy, and Lemnian a colonial offshoot of Etruscan (De Simone 1998: 409–410; Wallace 2008: 222; Oettinger 2010; Simon 2021: 245).<sup>25</sup> With regard to pre-Roman Sardinia, the idea of (one of) its languages being related to Etruscan is by no means absurd. Sardinia and Etruria are geographically in close proximity, and an Etruscan presence has been demonstrated in Corsica (Wallace 2008: 3). Still, any hypothesis of a Tyrrhenian language being spoken in pre-Roman Sardinia must be supported by compelling linguistic evidence.

The scholar that has argued in favor of a connection between the pre-Roman language(s) of Sardinia and Etruscan most fervently, is Pittau. Pittau (1984: 18–20; 1995; 1997: 29) argues that the pre-Roman language of Sardinia (“Nuragic”, in his definition), was a sister language to Etruscan, and as such part of the Tyrrhenian language family. Nuragic and Etruscan would have been introduced to Sardinia and Etruria respectively ca. 1 000 BCE from Anatolia (Pittau 1995: 147, 171–172). More specifically, the putative introduction of Tyrrhenian in these places, as well as in Corsica, the Balearic islands and the Iberian Peninsula (Pittau 1995: 148–157), is connected to a so-called “thalassocracy” and ensuing colonization by the Lydian people of Anatolia. Pittau’s hypothesis rests on a combination of historiographical, archeological and linguistic arguments. Here, I will focus on Pittau’s linguistic arguments.

First and foremost, Pittau’s (1995: 172; 1997: 28–33) connection between Etruscan and Lydian cannot be accepted. Even if one does not agree with the common view that Etruscan is not an Indo-European language (e.g. Wallace 2008: 1, 216–217), it is most definitely not an Anatolian language, which Lydian clearly is (Simon 2021; Kloekhorst 2022b). As a logical consequence, the “Nuragic” language posited by Pittau for pre-Roman Sardinia, cannot have been related to both

---

<sup>25</sup> The genetic findings of Posth et al. (2021) are more consistent with an autochthonous Italian origin of the Etruscans.

Etruscan and Lydian. To my knowledge, no Indo-European Anatolian presence has ever been proposed for Sardinia. The “Lydian” hypothesis will therefore not be pursued any further here.

A hypothetical connection between pre-Roman Sardinia and the Tyrrhenian language family is still on the table however, regardless of the ultimate origins of the latter. The evidence for this, adduced by Pittau (1995: 197–233), consists of four categories. These are: 1) comparisons of Sardinian forms of suspected pre-Roman origin with Etruscan words, 2) comparisons with Latin words of probable Etruscan origin, 3) comparisons with Tuscan words of obscure etymology for which an Etruscan origin is plausible, and 4) comparisons with dialectal Italian words of obscure etymology for which an Etruscan origin is plausible. For these categories, we must address a number of caveats. Firstly, even though Pittau (1995: 196) objects against the use of person names and toponyms in linguistic comparisons, due to their semantic opacity, he does make use of both of these himself in categories 1 and 2. He argues that onomastic and toponymic material can be used if the compared forms consist of a significant string of phonemes, diminishing the risk of “fortuitous homophony”. Yet, the issue of semantic opacity remains. We could find an eight-syllable-long place name in Sardinia and in Anatolia and it would be an enormous coincidence, but a coincidence nonetheless as long as we cannot verify that their meanings are comparable. These comparisons can therefore not be used to build a strong case for Tyrrhenian linguistic presence in Sardinia. Secondly, without direct evidence, an Etruscan origin of the Tuscan and dialectal Italian forms in categories 3 and 4 remains rather speculative, due to our scarce knowledge of the Etruscan lexicon and our limited knowledge of any other pre-Roman languages in the region that could have been the source of etymologically obscure words. The lexical and structural lines of evidence proposed to point to a connection between Etruscan and pre-Roman Sardinian are discussed in § 10.4.