DE GRUYTER
MOUTON

SUB-INDO-EUROPEAN EUROPE

PROBLEMS, METHODS, RESULTS

Edited by Guus Kroonen

TRENDS IN LINGUISTICS

Guus Kroonen (Ed.)
Sub-Indo-European Europe

Trends in Linguistics
Studies and Monographs

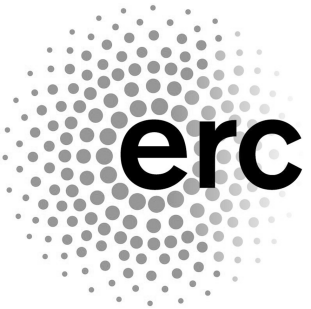
Volume 375

Sub-Indo-European Europe

Problems, Methods, Results

Edited by
Guus Kroonen

DE GRUYTER
MOUTON



This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement nº 716732).

ISBN 978-3-11-133705-0
e-ISBN (PDF) 978-3-11-133792-0
e-ISBN (EPUB) 978-3-11-133813-2
ISSN 1861-4302
DOI https://doi.org/10.1515/9783111337920

Library of Congress Control Number: 2024939066

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at http://dnb.dnb.de.

Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Foreword

This volume is a collection of thirteen chapters that together offer the state of the art of the study of sub-Indo-European Europe, i.e. the transitional state during which Indo-European- and non-Indo-European-speaking groups coexisted and interacted in Europe and adjacent areas. In their contributions, the authors review old and provide new evidence related to linguistic contacts that ensued from the Indo-Europeanization from the 3rd millennium BCE onward. In addition, they propose adjustments and refinements of the methodologies for studying linguistics substrates in general, thus offering new ways of interpreting data related to the reconstruction of situations of contact with extinct and marginally documented languages.

The genesis of this volume lies in a workshop held at Leiden University on August 30–31, 2021. It was organized by the ERC-funded Starting Grant project 'The Linguistic Roots of Europe's Agricultural Transition' (EUROLITHIC, grant no. 716732) in synergy with the 'Towards a New European Prehistory' project financed by Riksbankens Jubileumsfond (M16-0455:1). The workshop brought together an international selection of both junior and senior researchers with expertise in the main Indo-European subgroups of Western Eurasia. The focus was on the accumulation and interpretation of lexical evidence, to identify commonalities as well as particularities of the non-inherited components of the various Indo-European subgroups. An additional focus was the evaluation of the methodological principles developed over the past century for the detection of prehistoric loanwords.

Part I introduces the history of substrate methodology. In the introductory chapter, Guus Kroonen traces this methodology back to its roots, which he argues are inextricably entangled with the rise of the comparative-historical method itself. In the chapter, he additionally evaluates the various criteria that have been proposed for the identification of linguistic substrates.

Part II of this volume covers Northern and Eastern Europe, focusing on the Balto-Slavic languages. In chapter 1, Anthony Jakob revisits the Balto-Slavic evidence related to some "Avidic" words, i.e. bird names borrowed from unknown sources. Integrating parts of the research history of substrate studies, Jakob departs from Karel Oštir's work, which is largely obsolete, but at the same time invites the formulation of more strict methodological strategies for stratifying prehistoric loans. Ranko Matasović takes a different approach, rooted in semantics. Compiling a dataset of Slavic tree names, he assesses the proportion of Indo-European and non-Indo-European elements. Based on the resulting evidence, he shows that economically less important trees are less likely to be of Indo-European origin, thus reproducing an important semantic tendency.

https://doi.org/10.1515/9783111337920-202

Part III on Western and Central Europe has the Celtic languages as its main focus. Paulus van Sluis offers a comprehensive survey of the irregular vocabulary involving Celtic. By subsequently analyzing the resulting corpus, he shows how recurrent irregular features can be correlated with the geographic distribution of non-Indo-European vocabulary. Anders Richardt Jørgensen identifies a previously unknown Celtic suffix *-*anno*-, which he demonstrates regularly developed into -*enn* in Middle Breton. As the suffix predominantly occurs in etymologically obscure bird names, he considers the possibility that this suffix continues an element borrowed from an unknown language, at the same time acknowledging the risk of onomatopoetic mechanisms polluting the evidence for prehistoric loans. David Stifter offers a highly in-depth study of the potential layers of prehistoric loans in the Celtic languages, especially focusing on the Goidelic evidence. Candidates for such loans are scrutinized, as well as the criteria that can be used to detect them.

Part IV focuses on the languages of the Mediterranean. Andrew Wigman identifies three irregular lexical items with manifestations in Italic, but also in Germanic, Celtic and Greek. These are 'coot', 'fern' and 'shrew', items that fall within the usual semantic categories associated with prehistoric loans from local languages. Cid Swanenvleugel treats the complete evidence for the prefixes **kV-* and **ϑi-* in the pre-Roman lexicon of Sardinia, traditionally referred to as Paleo-Sardinian. This richly illustrated chapter offers a wealth of evidence on the distribution of these suffixes across the island and assesses the possibility that they arose as variants of the same Paleo-Sardinian element. Moving to the east of the Mediterranean, Lotte Meester and Guus Kroonen address problems related to the well-known Pre-Greek substrate. Meester offers a critical evaluation of the hypothesis that the non-Indo-European elements in Greek come from a single Pre-Greek substrate language. From a methodological perspective, she underlines the possibility that recurrent irregular sound correspondences do not reflect a phenomenon of the source language and instead are an artifact of the borrowing process. Kroonen revisits the evidence for the Pre-Greek suffix -νϑ-. By eliminating previously adduced variants without a nasal, he points to a partially complementary distribution between the suffix variants -ινϑ- and -υνϑ-.

Part V is the final part of the volume and extends the focus to Anatolia and the Caucasus. Rasmus Thorsø identifies a cluster of plausible European substrate loans, in which Armenian shows the reflex of a diphthong **ou* as opposed to a monophthong in other Indo-European branches. It follows that Armenian in its early prehistory must have been part of the circle of European Indo-European languages that were in contact with a single non-Indo-European stratum. The chapter by Zsolt Simon addresses a related question, namely whether Anatolian also contains manifestations of the substrate vocabulary found in the European Indo-

European branches. It offers extensive etymological discussions of existing proposals and shows that the majority of them cannot be maintained. Finally, Peter Schrijver returns to the tantalizing question of the linguistic affiliation of the source language that donated vocabulary items to the European Indo-European languages. He offers a new etymology for the word for 'camel'. In addition, he draws the attention to reconstructed morphological patterns in East Caucasian, which are highly reminiscent of European substrate words with the so-called *a*-prefix.

Although the authors of the present volume come with their own perspectives, start their explorations from different datasets and make use different approaches, they all share a common motivation, i.e. to explore new avenues into the linguistic landscape that predated the Indo-European dispersal. While providing definite answers to questions related to this topic may seem a difficult if not insurmountable task, accumulative advances in the methodology, including an explicit acknowledgment of its weaknesses, will hopefully lead to increasingly verifiable results and not least a more widely supported consensus on the interpretation of these results.

Contents

Language abbreviations

Akk.	Akkadian
Alb.	Albanian
Arab.	Arabic
Arm.	Armenian
Ass.	Assyrian
Att.	Attic
Av.	Avestan
Balt.	Baltic
Bav.	Bavarian
Bel.	Belarusian
Brb.	Berber
Bret.	Breton
Bsq.	Basque
Bulg.	Bulgarian
C(elt.)	Celtic
Camp.	Campidanese
Capp.	Cappadocian
Car.	Carian
Cat.	Catalan
Celtib.	Celtiberian
CLuw.	Cuneiform Luwian
Corn.	Cornish
Cpt.	Coptic
Cr.	Croatian
Cret.	Cretan
CS	Church Slavonic
Cypr.	Cypriot
Cz.	Czech
Da.	Danish
Dor.	Doric
Du.	Dutch
E	English
Eg.	Egyptian
Elfd.	Elfdalian
Etr.	Etruscan
Fal.	Faliscan
Far.	Faroese
Fi.	Finnish
Fr.-Prov.	Franco-Provençal
G	German
Gall.	Gallurese
Gallo-Lat.	Gallo-Latin
Gasc.	Gascon
Gaul.	Gaulish
Geo.	Georgian
Gm.	Germanic
Goid.	Goidelic
Gr.	Greek
Hatt.	Hattian/Hattic
Hebr.	Hebrew
Hitt.	Hittite
HLuw.	Hieroglyphic Luwian
Hung.	Hungarian
HVann.	Haut-Vannetais
Icel.	Icelandic
IE	Indo-European
Ion.	Ionic
Ir.	Irish
It.	Italian
Kash.	Kashubian
Khwar.	Khwarezmian
Kurd.	Kurdish
Lad.	Ladin
Lat.	Latin
Latv.	Latvian
LCorn.	Late Cornish
LG	Low German
Lith.	Lithuanian
LLat.	Late Latin
LSrb.	Lower Sorbian
Log.	Logudorese
Lomb.	Lombard
Luw.	Luwian
Lyc.	Lycian
Mac.	Macedonian
MBret.	Middle Breton
MCorn.	Middle Cornish
MDu.	Middle Dutch
ME	Middle English
MEg.	Middle Egyptian
MFr.	Middle French
MHG	Middle High German
MIr.	Middle Irish
MLat.	Middle Latin
MLG	Middle Low German
MoGr.	Modern Greek
MW	Middle Welsh
Myc.	Mycenaean

https://doi.org/10.1515/9783111337920-204

Nuor.	Nuorese
Nw.	Norwegian
OBret.	Old Breton
Occ.	Occitan
OCorn.	Old Cornish
OCS	Old Church Slavonic
OCz.	Old Czech
ODa.	Old Danish
ODu.	Old Dutch
OE	Old English
OEg.	Old Egyptian
OFr.	Old French
OFri.	Old Frisian
OGeo.	Old Georgian
OHG	Old High German
OIr.	Old Irish
OLat.	Old Latin
ON	Old Norse
OLFra.	Old Low Franconian
OOcc.	Old Occitan
OPic.	Old Picard
OPo.	Old Polish
OPr.	Old Prussian
ORu.	Old Russian
OS	Old Saxon
Osc.	Oscan
Oss.	Ossetic
OSw.	Old Swedish
OW	Old Welsh
P	Persian
PBalt.	Proto-Baltic
PBrit.	Proto-British
PBSl.	Proto-Balto-Slavic
PC(elt.)	Proto-Celtic
PGm.	Proto-Germanic
PGr.	Proto-Greek
Pic.	Picard
PIE	Proto-Indo-European
PIt.	Proto-Italic
Po.	Polish
Polab.	Polabian
Port.	Portuguese
Prov.	Provençal
PSem.	Proto-Semitic
PSl.	Proto-Slavic
Rht.-Rom.	Rhaeto-Romance
Rom.	Romanian
Ru.	Russian
RuCS	Russian Church Slavonic
Sass.	Sassarese
ScG	Scottish Gaelic
SCr.	Serbocroatian
Serb.	Serbian
SerbCS	Serbian Church Slavonic
Skt.	Sanskrit
Slav.	Slavic
Slk.	Slovak
Sln.	Slovene
Sogd.	Sogdian
Sp.	Spanish
Srd.	Sardinian
Sum.	Sumerian
Sw.	Swedish
ToA	Tocharian A
ToB	Tocharian B
Tsak.	Tsakonian
Ukr.	Ukrainian
USrb.	Upper Sorbian
Umb.	Umbrian
Vann.	Vannetais
VLat.	Vulgar Latin
W	Welsh
Wall.	Walloon
WFri.	West Frisian
Yid.	Yiddish

Part I: **Introduction**

Guus Kroonen

1 A methodological introduction to sub-Indo-European Europe

1 The rise of Indo-European substrate research

The quest for answers about the linguistic make-up of pre-Indo-European Europe is as old as the puzzle of Indo-European origins itself. Just as the Indo-Europeanization of a whole continent continues to perplex us today, millennia after its occurrence, the cataclysm it entailed for the pre-existing languages has stirred the imagination of generations of scholars. What languages or language families were spoken in Europe prior to their widespread extinction? Did they belong to a small number of large families or should we rather think of a patchwork of related and unrelated languages (Trask 1997: 364)? Did they belong to any of the known existing families, such as Caucasian, Basque and Semitic (Vennemann 2003), or were they related to extinct languages, such as Etruscan (Braun 1922), Minoan, Hattic or Sumerian (Schrijver 2011; 2018)? Should we actually expect to be able to identify any relationship to languages historically documented in the first place?

These questions cannot readily be answered. It is challenging enough to reconstruct the Indo-European proto-language itself, despite the availability of a universally accepted method and a wealth of well-documented descendant languages, historically spoken or still in use today. Yet these challenges pale in comparison to those related to reconstructing the pre-Indo-European linguistic landscape. Without direct sources, data is exceedingly more difficult to obtain, consisting only of those few features, phonological, lexical, morphological and syntactic, that happened to be absorbed by the Indo-European languages that survived into the dawn of history, and actually are objectively identifiable as foreign.

What is more—no full consensus exists on the methodology for studying unattested linguistic substrates except perhaps that it will never approach, let alone surpass, the gold standard of the historical-comparative method. Still, as our understanding of Proto-Indo-European advances, the features in the Indo-European languages that resist a native interpretation become increasingly difficult to ignore. Whatever data cannot be interpreted within an Indo-European framework must inevitably be interpreted otherwise. To do this, various strategies have been devised over the years by different groups of Indo-Europeanists, focusing on various aspects of language, including phonology, morphology, vocabulary and syntax.

For the current state-of-the-art, the period of the 1990s was a defining one. It saw a surge of publications proposing guidelines for the identification of various

https://doi.org/10.1515/9783111337920-001

traits of lost substrate languages. The initial impetus of this wave was mainly given by American scholars, including Markey (1989), Hamp (1990), Polomé (1986; 1989; 1990; 1992), Huld (1990) and Salmons (1991; 2004). Later, the resulting advances were integrated with parallel but largely independent insights from the Leiden School, including those of Kuiper (1995), Beekes (1996), Boutkan (1998; 2000; 2003), Schrijver (1997) and Lubotsky (2001), all of whom built on previous results obtained by Kuiper (1956) and Furnée (1972; 1979). Despite its divergent origins, this "Dutch-American strand" of substrate research (thus Bichlmeier 2016: 319) was characterized by a common focus on methodology. By canonizing a number of explicit criteria, these scholars strived to make the identification of prehistoric loans more systematic and ultimately more objective. To this aim, the following criteria were formulated:

- non-formal criteria
  - a lack of (consensus on) an Indo-European etymology
  - a local distribution
  - appurtenance to certain semantic fields
- formal criteria
  - non-native morphology
  - non-native phonology / phonotactics
  - irregular sound correspondences

More than two decades after the first focused formulations of these criteria, the question is as to how these advances should be evaluated and in what aspects they should be improved. An important limitation, as pointed out by the aforementioned Dutch and American scholars already in the 1990s, is that none of the above criteria are decisive in isolation, even if some are more probative than others.

To start, the first two non-formal criteria, while a prerequisite for assuming a non-Indo-European origin, cannot be used to substantiate it. When a word is restricted to two neighboring branches of Indo-European, an Indo-European origin cannot mechanically be assumed due the risk of prehistoric borrowing (Hirt 1905: 234). On the other hand, while a local geographic distribution is conspicuous, it is not in itself proof of borrowing. An important caveat applies: as long as the sound correspondences are regular, it is impossible to exclude loss in other Indo-European branches. Likewise, the possibility of lexical loss severely impairs the closely related criterion of the absence of (consensus on) an Indo-European etymology. As noted by Streitberg and Michels (1927: 51), "[d]er Mangel einer Deutung aus dem Indogermanischen ist so lange kein Beweis für die nicht-indogermanische Herkunft eines Wortes, als nicht dargetan ist, daß die Wortform selbst ein unindogermanisches Gepräge zeigt" (see also Bichlmeier 2016: 323). In other words, just as we should not

permit ourselves to mechanically assume an Indo-European origin for a word that is isolated to a single European branch, we cannot assign the same word to an unknown source without additional, positive evidence of borrowing.

The third non-formal criterion, that of appurtenance to certain semantic fields, such as local flora and fauna, local cuisine or region-specific cultural advances, has traditionally played a prominent role (cf. Meillet 1908; Bertoldi 1931: 94). Nonetheless, it is evidently the weakest of all (see Salmons 1991: 267). For instance, Meillet, followed by a slew of Mediterranean substraticists (Bertoldi 1942: 162; Alessio 1944: 108; Battisti 1960: 367), assigned a substrate origin to the pair Lat. *vīnum*, Gr. οἶνος 'wine' merely for belonging to the "non-Indo-European" semantic field of viticulture. However, as long as the formal reconstruction is unproblematic, we cannot exclude the possibility that an inherited Indo-European word for '(wild) grapevine' was repurposed to indicate 'wine' (Gorton 2017). When a language is subjected to external cultural influences, either through migrating to a new area or through the immigration of foreign speakers, there can be a correlation between the adoption of loanwords and cultural innovations. However, using such a cultural criterion can quickly turn circular, whereby any word that does not fit in a preconceived notion of Indo-European society or ecology is automatically analyzed as a borrowing.

These preliminary caveats already highlight several challenges at proving a non-native origin. This introduction will address the remaining methodological criteria in greater detail, as the questions they raise must be considered fundamental. Before moving on, however, it is important to realize that all the above criteria exclusively apply to lexical evidence. As mentioned above, during the initial phase of substrate linguistics the focus was placed not so much on lexical evidence but in fact on the phonology of the Indo-European branches, i.e. whether or not the daughter branches preserve the Indo-European *Lautstand*.

2 Early days: *Lautstand* and phonological shifts

When William Jones addressed the Asiatick Society 1786, he famously acknowledged the inclusion of Sanskrit in the Indo-European language family. Less well known is his characterization, in the same speech, of "the Germanick and the Celtick" as "blended with a very different idiom" (1788: 423; see also Polomé 1990: 268). Although the exact considerations of Jones remain elusive, there can be little doubt that he alluded to the shifted phonologies of these Indo-European branches vis-à-vis Greek, Latin and especially Sanskrit, which after all was perceived as the most "perfect" Indo-European language. Ever since, a shift in the phonology has

been widely acknowledged as an indicator of language contact, or more specifically, of the absorption of a different language, i.e. an "Idiom der Unterworfenen" (Nörrenberg 1900: 386).

Undoubtedly, the radically different phonology of Celtic, or at least Insular Celtic, was apparent to even the earliest students of the Indo-European languages and the question it raised about prehistoric language contact remained pertinent throughout the 19th and 20th centuries. This was phrased in clear terms by Kretschmer: "Da Spuren einer *voridg. Bevölkerung Britanniens* vorhanden sind, die vor Ankunft der Kelten dort schon bestand, so erklärt sich jener auffällige Charakter des neueren Keltisch wohl durch Einfluß der nicht-indogermanischen Sprache älterer Stämme auf den britischen Inseln" (1925: 29, emphasis in the original). In the 20th century, Pokorny famously sought to identify this substrate as Semitic (1927; 1928; 1930), not least for reasons of syntax. This school of thought has continued into the 21st century through the work of Wagner (1959), Gensler (1993), Jongeling (1995) and – more controversially – Vennemann (2003) and Mailhammer, who see a Semitic substrate across northwestern Europe, one that also affected Germanic (Mailhammer & Vennemann 2019).

The notion that Germanic is a mixed Indo-European dialect is a persistent one. Hirt saw a clear sign of "Sprachmischung" in the Germanic sound shifts (1899: 572; 1905: 196), a conclusion shared by Feist (1910; 1932) and Meillet (1930: x). The view prominently featured in the dangerously politicized debate on the Indo-European homeland of the late 19th and early 20th centuries. By then, the Germanic people had been proclaimed as the direct and least adulterated descendants of the Indo-European speech community, neither having moved from the homeland nor mixed with non-Indo-Europeans (Penka 1886; Wilser 1900; Kossinna 1911). Supporters of a non-local origin, including those favoring Schrader's steppe hypothesis, tried to counter this exactly by pointing to the sound shifts. Childe called Germanic "manifestly degenerate from a phonetic standpoint" (1926: 167). Braving the political Zeitgeist, the early Güntert (1934: 72) similarly attributed Grimm's and Verner's laws to "Wirkungen vorindogermanischer Sprechgewohnheiten", which he thought to demonstrate that "die Germanen aus Vermischung der nördlich vordringenden „Schnurkeramiker" mit dem steinzeitlichen Bauernadel entstanden sind" (but see Lincoln 2008 for the evolution of his views). The argument was reversed, however, by Nörrenberg (1900; followed by Wilser 1900: 147), who sought confirmation of the putative Scandinavian Indo-European homeland in the supposed impact of a Uralic substrate (cf. Karsten 1928: 125). While Germanic no longer plays a central role in debate on the Indo-European homeland, the notion that this branch was transformed by prehistoric language contact has not been abandoned, and in recent years some have again looked to analyze the "phonetic dislocations" of the Germanic sound shifts through the lens of Uralic (Wiik 1997; Kallio 1997; Schrijver 2014).

A third Indo-European branch with a strikingly divergent *Lautstand* is Armenian. Once Armenian had been identified as a separate branch of Indo-European (Hübschmann 1877), its markedly altered Indo-European character, most notably its sound shift, invited Kretschmer (1896: 121) to assume the influence of Urartian. On account of the striking similarities of the Old Georgian and Classical Armenian sound inventories, Deeters (1926; 1927) alternatively proposed Kartvelian as the primary substrate influencing Armenian. This view has garnered increasing support in recent scholarship (Schmidt 1992; Gippert 2005; Aslanov 2017).

The above explanations concerning the driving forces behind sound shifts were hypotheses that captivated the scholarly community of the time. However, demonstrating that a particular change was induced by a substrate is typically challenging, especially when the latter is not attested (cf. Schmitt 1936: 355–356). It is usually not feasible to identify substrates, whether they are Uralic, Semitic or Etruscan, on the basis of sound shifts alone. A statement of the type "[s]o ist auch die Aussprache des Lateinischen im Munde der Bewohner Etruriens, die ph, th, kh anstatt p, t, k artikulierten, etruskischem Einfluß zuzuschreiben" (Schrijnen 1921: 88) may at first seem reasonable, but cannot be verified. Typological comparisons often devolve into a "zweckloses Ratespiel" (Schmitt 1936: 362) that does little to improve our understanding of the prehistoric linguistic landscape. More fatally, sound change is not necessarily induced by language contact and can equally well result from internal drift (Schmitt 1936: 350; Huld 1990: 389; Ternes 1998: 279).

Similarly, the lack of a large-scale phonological shift does not necessarily demonstrate a lack of language contact. A case in point is Baltic. Baltic, in particular Old Prussian and Lithuanian, is often juxtaposed with Germanic and represented as an untouched form of Indo-European, i.e. an Indo-European dialect that was not affected by other languages in its prehistory (cf. Childe 1926: 167). The conservatism of the Baltic group has been a romantic trope at least since Poesche (1878: 123): "seine herrliche uralte Sprache hat sich rein erhalten, und lässt noch heute die ältesten Formen der arischen Zunge mit grosser Treue ertönen". According to a famous quote by Meillet (1913: 205), "[q]ui veut retrouver sur les lèvres des hommes un écho de ce qu'a pu être la langue commune indo-européenne, va écouter les paysans lituaniens d'aujourd'hui". Still today it is often claimed that Balto-Slavic has experienced "little or no non-I[ndo-]E[uropean] contact" (Nichols 1998: 254) or that "[i]n the course of 5,000 years, since Proto-Indo-European times, Lithuanian has not mixed with any other Indo-European or non-Indo-European language" (cf. Klimas 2002: 52–54). But Schleicher had in fact remarked more than a century earlier that Lithuanian is a language with a limp: "litauisch ist ein ungleich entwickelte

sprache, der treusten erhaltung des alten geht auffallender verlust zur seite" (1852: 10). And even Meillet himself noted that, although "[l]e lituanien est remarquable par son aspect d'antiquité indo-européenne", "le verbe, en particulier, a une structure toute nouvelle" (1922: 49). We may add that the lexicon in fact shows compelling evidence of prehistoric contact with other languages (Derksen 2000; Jakob, this volume).

In conclusion, despite its initially big impact on Indo-European studies, the research on substrates through phonological shifts did not leave a strong legacy in modern day practice. Instead, the attention has largely shifted to the identification of prehistoric loanwords, not least because lexical evidence has a greater potential for the recovery of the features of the (known or unknown) donor language.

3 Lexical criteria

For the accumulation of the lexical evidence on prehistoric language contact, several criteria have been formulated, as mentioned above. These criteria, too, have deep roots in the field and are closely linked to the principles of the comparative method.

3.1 Non-Indo-European morphology

As we have observed, Germanic and Celtic played a central role in the formulation of what late 19th and early 20th century Indo-Europeanists referred to as the "substrate theory". However, they are not the only Indo-European branches that were early on suspected of harboring influences from assimilated languages. Greek, spoken in a very different part of the Indo-European-speaking world, played a similarly formative role in the development of the methodology for studying substrates. In the case of Greek, however, the focus was placed on the identification of non-inherited morphology. Intriguingly, the roots of this focus predate even the rise of the comparative method itself. Pott (1853: 451) identified words ending in -ισσος and -ινθος as (non-Indo-European) "Pelasgian" at a very early stage in the development of the discipline (see Kroonen, this volume). He argued that these elements, which "ihrer räthselhaften Natur wegen etwas Fremdartiges zu haben scheinen", may be "Ueberrest aus der Sprache eines den Hellenen vorausgegangenen Geschlechts". Pott thus hinted at what would become a key principle of modern substrate linguistics, namely that prehistoric language

shifts can be betrayed by lexical borrowing, specifically loanwords adopted in a period when the language shift had not yet been completed.

The notion of a lost substrate is also promoted by Antoine Meillet in the highly influential paper "De quelques emprunts probables en grec et en latin" (1908). Similar to Kretschmer (esp. 1896: 401–409), Meillet acknowledged the foreign suffixes -ινθος and -ισσος as markers of non-Indo-European origin. Expanding upon the work by Kretschmer and Pott, Meillet inferred that isolated Latin and Greek comparanda, when they cannot be direct loans, may have been adopted by the two branches independently from a single "Mediterranean" source. This explanation marked an important step forward, not just because it offered a new model for understanding the non-Indo-European features shared by Latin and Greek, but also because it exposed the futility of attempting to trace every single word back to an Indo-European origin: "il est contraire à une saine méthode étymologique de vouloir expliquer tout le vocabulaire en grec et en latin par l'indo-européen" (Meillet 1908: 164).

Since then, numerous affixes have been proposed to be indicative of a non-Indo-European origin, in a wide variety of Indo-European languages, i.e. -αξ for Greek (Nehring 1925), *-ish-* for Albanian (Jokl 1923), *-ar-* (Terracini 1927: 139; Hubschmid 1960: 25) and *str-* (Bertoldi 1928: 234) for Romance, to mention some. The credibility of these proposals depends largely on the nature of the suffixes. The Greek suffix -ινθος represents a rather reliable indicator. It cannot be reconstructed as a PIE suffixal element, as there is no comparative data, and the form of the suffix in Greek precludes an Indo-European reconstruction. The alternation within Greek of the same root elements with and without the element -ινθ- is an additional argument for assuming that it was indeed a suffix in the language that transmitted it to Greek. The situation is rarely as clear-cut, however, particularly with simpler and more generic suffixes. For instance, Chantraine (1933: 278–279) demonstrates that the suffix -αξ, whatever its origin, is productive in Greek, as a result of which it cannot be mechanically interpreted as a substrate marker (see also Wigman, this volume). While Meillet's work evidently constituted a useful advancement, it is therefore clear that the identification of non-Indo-European morphological elements necessitates careful scrutiny on a case-by-case basis. In this volume, Anders Jørgensen carefully weighs the value of the Celtic suffix *-ann-* as a substrate marker and Cid Swanenvleugel does the same for the Sardinian prefixes **k(V)-* and **θi-*.

The complexity of the problem is neatly illustrated by the ambiguity surrounding the putative non-native suffix **-n-*. Kuiper (1995: 80) suspected the nasal of PGm. **baunō-* 'bean' to be a feature of the donor language of the word, its suffixal nature being revealed by its absence in Lat. *faba* and PSl. **bobъ*. Indeed, a nasal extension emerges in various other plausible prehistoric loans found across Europe (see Table 1). While it is conceivable that this suffix is non-native in origin,

especially considering the phonological irregularities of the comparanda, the feature itself is not necessarily probative. As Proto-Indo-European itself possessed a suffix *-no/eh₂-*, it is difficult to reject a scenario in which this element was productively added to a substratal element **bʰau-* in Germanic, so as to nativize it. Then again, since a nasal suffix reoccurs in several other demonstrably irregular cognate sets, across varying branches, Kuiper's analysis may still hold some validity. Interestingly, it has been observed that many of the cognate sets featuring the potentially non-native suffix **-n-* also feature an element **-e/is-* (Matasović 2013: 87).

Table 1: Plausible non-Indo-European terms exhibiting a suffixal alternation between *n* (in bold) and no *n*.

meaning	Greek	Albanian	Italic	Celtic	Germanic	Baltic	Slavic
alder			*als**n**o-		*alis/zō-	*alis**n**ā-	*olьxa
bean			*fabā-		*bau**n**ō-	?*babā-	*bobъ
fern	βλῆχ**ν**ον		*felik-		*brek**n**a-		
holly				*koles**n**o-	*hulisa-		
maple			*akro-		*ahur**n**a-		
oats			*awes**n**ā-			*avižā-	*ovьsъ
pine 1	πίτυς	*pishë*	*pi(t)s**n**o-				
pine 2				*gisusto-	*kiz**n**a-		

Indeed, the method for identifying non-Indo-European morphology, as pioneered by Pott, Kretschmer and Meillet, can be improved by combining multiple morphological features. Probably the strongest evidence for prehistoric interference comes from affixes that systematically co-occur with other non-native features. An *a*-prefix has been recognized as occurring relatively frequently in non-inherited vocabulary (Kretschmer 1932: 86–90; Battisti 1959: 155 fn. 1; Furnée 1972: 368–370). This element itself is short and generic, and productive nominal morphemes of a similar shape are present in numerous unrelated language families, including e.g. Berber or Abkhaz. This significantly raises the likelihood that not all occurrences of the *a*-prefix in the Indo-European language groups of Europe reflect one and the same phenomenon. However, the *a*-prefix has been observed to occasionally co-occur with an ablaut pattern that deviates from the Indo-European ablaut (Schrijver 1997). The presence of this additional feature tips the balance in favor of a non-Indo-European origin, as lexemes exhibiting it are more likely to belong to the same stratum, i.e. a single donor or set of related donors, than those that do not. While Kretschmer (1932: loc. cit.) previously linked the stratum of words containing the *a*-prefix to the isolated Hattic, Schrijver (2011: 246–249, 254) identified a potential parallel for the non-Indo-European ablaut in this language. Irrespective of

the linguistic affiliation of this stratum, the *a*-prefix appears to have been an element of it. That said, a maximally critical application of this categorization would exclude any *a*-prefixed words from the stratum if they do not exhibit ablaut (Table 2).

Table 2: Cognate sets potentially exhibiting an *a*-prefix in combination with non-Indo-European ablaut.

meaning	Greek	Albanian	Italic	Celtic	Germanic	Baltic	Slavic
blackbird			*mesVlā-	*mesal-kā-	*amslōn-		
ore			*raudo-		*arut-		
pear	ἀχράς, -άδος	*dardhë*					
swan					*albet-		*lebedь
turnip	ʔράφυς		*rāpo-	*arb-īn-	*rōb(j)ōn-	*rāpiā-	*rěpa

3.2 Irregular sound correspondences

As argued in the preceding section, non-Indo-European morphology is particularly diagnostic when it appears in lexical comparanda exhibiting other formal problems. In the aforementioned paper, Meillet acknowledged the Greek suffixes -ινθος and -ισσος as markers of non-Indo-European origin (cf. Kretschmer 1896). He furthermore noted that in at least one instance Gr. -ισσος corresponds to Lat. *-essus*, i.e. in Gr. κυπάρισσος vs. Lat. *cupressus* 'cypress tree', and that the comparison violates the regular sound correspondences deduced from inherited material. The employment of such irregular correlations marked the integration of the Neogrammarian doctrine of the regularity of sound change into the methodology for studying prehistoric loans from unknown sources. While the principle of the regularity of sound change allows to demonstrate cognacy through inheritance, the violation of this regularity conversely provides an objective approach for identifying loans. This was expressed very clearly by Cuny (1910: 158): "Mots existant à la fois en grec et en latin, mais qui ne sont pas exactement superposables d'après les formules phonétiques de ces langues": irregularly corresponding words reveal their non-inherited origin by suggesting incongruent proto-forms.

The advancement outlined here proved fundamental to the study of linguistic substrates and remains integral to the methodology today. The study of the internally irregular vocabulary in Greek spawned an entire strand of substrate research, as exemplified by the work of Furnée and Beekes. The methodology also raised new problems, however, not all of which have been satisfactorily addressed. While the detection of phonological irregularities represents a relatively strong indicator of borrowing, it suffers from an evident weakness. In absence of

regular sound correspondences, it becomes methodologically impossible to establish whether formally similar words share an etymological connection or are mere coincidental resemblances. The decision whether formally irregular lexical manifestations are valid comparanda, i.e. sufficiently similar to be compared, becomes intuitive and non-scientific in the sense that it cannot fully rely on the rigor of the Neogrammarian principles.

This methodological weakness can lead to contrastive approaches in cases where it is difficult to draw the line between regular and irregular lexical correspondences. For instance, those who derive Lat. *faba* and PSl. **bobъ* from a Proto-Indo-European base **bʰabʰ-* are forced to assume an irregular (i.e. *ad hoc*) development **-bn-* > **-u̯n-* in Germanic (Pokorny 1959: 106) or to dismiss the appurtenance of PGm. **baunō-* altogether (Kretschmer 1896: 146; Onions 1966: 83). Conversely, those who argue for a non-Indo-European origin of the cluster may additionally include Gr. φακός 'lentil' and Alb. *bathë* 'broad bean' < **bʰaḱ-o-* (Kuiper 1995: 80), despite these forms being formally more remote and – more importantly – mutually regular. The cherry picking of comparanda by the former is mirrored by a tendency to maximize the formal incongruity of an already irregular set of cognates by the latter.

Ultimately, the challenge lies in striking the right balance between lumping and splitting, between etymological ingenuity and restraint. On the one hand, when the phonologies of the donor and recipient languages are dissimilar, it is not uncommon for loanwords to undergo radical transformations. For instance, Basque *berun* 'lead' has been interpreted as an independent manifestation of the "Mediterranean" etymon represented by Gr. μόλυβδος, Hom. μόλιβος and Lat. *plumbum* (Bertoldi 1939: 96–97), suggesting a proto-form **b(e)lum* (cf. Hubschmid 1960: 33). In view of the known sound substitutions, however, it is impossible to exclude a much more recent loan from Romance **pluN* < Lat. *plumbum* (Tovar 1970: 272). Since the latter explanation has the least far-reaching consequences, it must be given preference over the former, despite the involved sound substitutions being more complex than those implied by the reconstruction of substratal **belum*. In fact, the example taken from Basque underlines the importance of supporting hypothetical sound substitutions with tangible parallels, even if this proves difficult because the donor language is hypothetical as well. Best practice mandates that the forms compared be as phonologically similar as possible, with the alternations themselves as limited as possible. From that perspective, it seems preferable to separate the manifestations of **bʰau-* ~ **bʰab(ʰ)u̯-* from those of **bʰaḱ-*.

Another possible strategy for mitigating the risk of excessive inclusivity, when comparing irregular lexical resemblances, is to increase not only formal but also semantic stringency. Unlike sound change, semantic change is not governed by regularities. Within Romance, a well-known example of extreme semantic drift is the

native etymon represented by It. *fegato*, Sp. *hígado*, Fr. *foie* 'liver', all of which developed from Lat. *fīcātum* 'stuffed with figs'. However, the unpredictability of semantic change should not be used as a license to accept any opaque semantic shift. In dealing with loans, we cannot afford the same liberty as with regularly corresponding cognates of inherited etyma. For instance, Furnée (1972: 132) connects Gr. δάφνη and Thess., Cypr. δαύχνα 'laurel', their co-occurrence convincingly demonstrating an irregular Proto-Greek doublet **dak^{wh}nā-* : **dauk^{h}nā-*. His additional inclusion of δαῦκος 'several types of umbelliferous plant', however, raises a semantic concern. While the formal irregularity of the first two comparanda may seem to sanction the inclusion of the third, it crosses a line in that it requires the additional assumption of an opaque semantic shift.

To prevent false positive matches, a maximally stringent approach is paramount to the study of non-inherited comparanda. The risk of chance resemblance is a genuine concern, as sufficiently demonstrated by Larry Trask in his comparison of Hungarian and Basque (1997: 413–414). In absence of regular sound correspondences, excessive leniency can easily create precedents for the lumping of formally and semantically distant elements, potentially culminating in the creation of what Hubschmied termed a "voridg. *kala-pala-bara*-Sprache" (1942: 118). With these caveats, the utilization of irregular sound correspondences, as suggested by Meillet and Cuny, is widely held to be the most objective approach for identifying prehistoric loans from known and unknown source languages. In this regard, it can reasonably be considered the methodological gold standard.

3.3 Non-native phonology and phonotactics

The detection of irregular sound correspondences, while a powerful tool, alongside the aforementioned methodological limitations, suffers from an important practical limitation: it requires the presence of lexical comparanda in at least two branches of the same language family. When a word is confined to a single branch, the criterion of irregular sound correspondences cannot be leveraged, and an alternative strategy must be adopted.

One such strategy relies on the criterion of non-native phonological and phonotactic features. Through the analysis of the phones and their placement in the word, i.e. the phonotactics, it is sometimes possible to reject an Indo-European origin for a word based solely on its root structure. Non-native phones are generally viewed as relatively reliable markers of a foreign origin. For instance, in English, the word *genre* /ʒɑnɹə/ can be identified as a loan even without knowledge of French phonology (Huld 1990: 392), as /ʒ/ is a rare phone in English and does not occur in inherited vocabulary. This principle has been successfully applied to

non-native comparanda in the modern Saami languages, which contain the loan phoneme /ʃ/ (Aikio 2004: 13). When it comes to the Celtic languages, words containing **p* cannot plausibly be analyzed as inherited from Proto-Celtic (see Stifter, this volume).

Cross-linguistically, the principle generally appears robust. However, like the other loanword criteria, the use of non-native phonotactics is not fully bulletproof, at least not in all circumstances. Salmons (2012: 28–29) exemplifies the limitations of the criterion through an investigation of the distribution of the phone /ʃ/ in English. He demonstrates that this phone originally did not occur after long vowels in this language. The gap was ultimately filled by loans from French, such as *gauche* and *brioche*, illustrating how the feature's phonotactics are indeed indicative of borrowing. However, the same phonotactics later appear in affective words, cf. *swoosh* and *smoosh*, which are evidently and in fact preeminently native. Consequently, it is essential to exclude affective parts of the vocabulary when identifying putative prehistoric loans. Still, the question on where to draw the line between "regular" and affective words is notoriously difficult to answer.

Additionally, the utilization of foreign phones and phonotactics to identify loans may be less convincing for reconstructed languages, not just because loanwords may become integrated into native phonology by successive waves of sound changes, but also due to a lack of consensus regarding the exact phonological configuration of reconstructed languages. For instance, the phonemes **a* (Kuryłowycz 1956: 194–195; Kuiper 1995; Beekes 1996: 216) and **b* (Hamp 1990: 298; Huld 1990: 392) have been suggested as intrusive phones confined to vocabulary restricted to Europe. However, their contested status in reconstructed Proto-Indo-European phonology means that these "will only convince part of the scholarly community" (Schrijver 1997: 295).

On the positive side, the problems outlined by Salmons and others do not imply that reaching a degree of consensus is entirely unattainable. The potential absence of Indo-European roots with two voiced stops or a combination of plain voiceless and voiced aspirated stops was noted from an early stage (Meillet 1912; see also Lubotsky 1998). In addition, it is widely acknowledged that Proto-Indo-European exclusively possessed monosyllabic morphemes, including both roots and suffixes (Saussure 1878). Consequently, the presence of a disyllabic root, i.e. a derivational base that resists further segmentation, serves as a relatively unambiguous indicator of a non-Indo-European origin (Huld 1990: 392–393).

There additionally appears to be relatively broad consensus regarding the absence of geminate consonants in Proto-Indo-European. Although geminates emerged regularly in many Indo-European branches, including Germanic and Celtic, it is unlikely that all the evidence for geminates in all branches can be accounted for by regular sound change (see now Stifter 2023 on Celtic). Notably, geminate consonants

occasionally appear in lexical items from across Europe, not least in those that also exhibit other substrate markers (see Table 3). A point in case is Gr. βέλεκκος (Arist., Hsch.) 'a kind of pulse', whose disyllabic root starts in a **b* (≠ **g*^{w}), ends in a geminate, and which semantically falls into the field of agriculture.

Table 3: Irregular comparanda in the European Indo-European languages exhibiting geminates.

meaning	Italic	Celtic	Germanic	Slavic
coot	*fulVkā-	*bo/ula/okk-	*balikōn-	
stick; crook		*bakko-	*pagjō-	
many; often		*menekki-	*managa-	*mъnogъ

4 Stratifying the corpus

Aside from the methodological difficulties involved in the collection of potential prehistoric loans, challenges surround their interpretation.

To begin with, no full agreement exists on the mechanisms underlying the formal irregularities. Furnée attributed the alternations displayed by the non-Indo-European elements in Greek to affective and expressive variation in the pre-Indo-European language of Greece (1972: 89–90). Conversely, Beekes treated the variation exclusively as a reflection of the mismatch between Greek and Pre-Greek sound inventories, using it to "reverse engineer" Pre-Greek phonology. However, some of the variation may alternatively be due to differences in the sound substitutions resulting from the diversification of Proto-Greek into its historical dialects: "Auch für die griechischen Dialekte darf man annehmen, daß sie im Keime bereits vorhanden waren, ehe das griechische Urvolk seine späteren Wohnsitze in der Balkanhalbinsel bezog" (Schrijnen 1921: 88). More importantly, neither Furnée nor Beekes fully considered the possibility of geographic or diachronic variation within Pre-Greek. Both of their approaches were built on the premise that Pre-Greek was a single, unified language, which is not a given (see Meester, this volume). It is in fact likely that Pre-Greek would have had dialects, even if it was formed by successive phases of homogenization or koineization. Thus, a model of language contact in Bronze Age Greece would have to factor in interactions between dialectally diversified Greek and dialectally diversified Pre-Greek, making the rules for the sound substitutions in Pre-Greek loans locally variable and seemingly erratic.

More generally, the presence of formal irregularities does not suffice to demonstrate that the identified loans were borrowed from a single source (cf. Šorgo 2020: 429). All loanwords, whatever their origin, will be captured by the methodology as long as they exhibit irregular reflexes in two or more Indo-European branches. The collection of loans identified by their irregular sound correspondences will therefore inevitably consist of a mixed bag containing lexical impositions from various prehistorically lost, unknown languages, including substrates, adstrates and superstrates, adopted through various mechanisms. As put by Hamp (1990: 301), "[w]e must be prepared to see an untidy set of intruding elements which cannot be claimed every time to be a fragment of evidence for our major subject, but it may simply be an intrusion from some other set of phenomena".

Distinguishing *Wanderwörter* from true substrate words represents a particularly significant challenge. While direct loans from attested languages can usually be identified with relative ease, as noted already by Meillet, early *Wanderwörter* may be impossible to distinguish from true substrate loans unless a large part of the borrowing chain is visible, which is not always the case. *Wanderwörter*, however, being associated with trade and the spread of new practices (cf. Haynie et al. 2014), constitute a distinct category (Beekes 1996: 215) and should be left out of consideration whenever possible (Polomé 1992: 51; Schuhmann 2016: 378).

In some cases, semantic arguments can be invoked to exclude prehistoric *Wanderwörter*. For instance, archaeological evidence shows that metallurgical practices related to lead spread after the Indo-European dispersal (cf. Winther Johannsen 2016). It follows that, if Proto-Germanic **blīwa-* is indeed connected to Gr. μόλυβδος, e.g. through Pre-PGm. **mlīwa-* (Kroonen 2013: 69), it is more likely to be a Bronze Age *Wanderwort* rather than a vestige of a local language absorbed by Germanic. However, it is only rarely feasible to invoke external evidence, such as from archaeology, to inform issues like this.

Identifying the mechanism or mechanisms underlying the irregular sound correspondences in sets of comparanda is inherently challenging. Using purely linguistic evidence, the problem can sometimes be resolved through the identification of regularities within the irregularities, i.e. through the detection of patterns within the irregularities. These patterns can be morphological in nature, such as the affixes described above, but they can also be phonological. Hester (1968), for instance, found clear phonotactic tendencies in the putative vocabulary of Pre-Greek. Across the European Indo-European branches, an alternation of *a* and *e* is found in multiple irregular cognate sets, the difference being distributed, in broad terms, along a north-south axis (see Table 4).

Naturally, the stratification of potential loans requires caution. There exists a risk of overinterpretation in the form of an unfounded tendency to attribute recurring irregularities, including the patterns they exhibit, to a single hypothetical

Table 4: Potentially non-native terms with northern *e*-vocalism (in bold) for southern **a*.

meaning	Greek	Albanian	Italic	Celtic	Germanic	Baltic	Slavic
hellebore	κάμ(μ)αρος				**h**e**merō-		**č**e**merъ
garlic				**kasno-			**č**e**snъ
bull	ταῦρος	*ter*	**tauro-	**tarwo-	**þ**e**ura-	**taura-	**turъ
boar			**apro-		**e**b**ura-	?**v**e**prja-	**v**e**prъ

substrate. Substantial evidence is required to demonstrate that recurrent patterns are manifestations of the same phenomenon. Regarding the European alternation of *a* and *e*, it is theoretically possible to postulate a single phone **æ* that was borrowed as either **e* or **a* (Šorgo 2020: 458). Reconstructing "substrate phones" risks introducing linguistic artifacts, however, especially when done in isolation. Although the distribution is imperfect, *e* tends to be more frequent in the northern Indo-European branches. This may alternatively suggest variation in the source language.

When stratifying potential loans, we should additionally be cautious of overly vague interpretations. At any rate, elevating vague patterns to a higher level, such as a "pre-Indo-European Sprachbund", i.e. "a collection of distinct, perhaps unrelated, languages, all of which share similar phonological, morphological and syntactic patterns and inventories" (Huld 1990: 390–391), does not contribute to an actual stratification. When a non-inherited alternation exhibits a spatially defined pattern, it is preferably interpreted as reflecting an underlying dialectal difference within the source. Such patterns emerge, for instance, from the distribution of Palaeo-Sardinian vocabulary across the Sardinian dialect continuum (see Swanenvleugel, this volume).

In essence, it is a valid undertaking to continue seeking regularities within the irregularities, of whatever nature, as an additional means of refining the methodology. Ultimately, however, the detection of recurrent morphology (see section 3.1), including prefixes and suffixes, may be the most reliable strategy for identifying strata within a corpus of irregular cognate sets, at least if the non-native origin of the involved morphology is sufficiently distinct.

5 A final word of caution

As argued in the present chapter, the methodology for identifying substrate words is demanding and requires acute awareness of its fundamental deficiencies. Each of the criteria that has been proposed suffers from shortcomings. It has therefore been argued that it is the combined evidence that decides the issue: "If the IE origin of a word is rendered suspicious by a number of criteria, it is usually the cumulative evidence rather than an individual criterion that tips the balance" (Schrijver 1997). Since combined weak evidence does not necessarily add up to strong evidence, however, even this claim is liable to criticism (see Simon, this volume). At any rate, the identification of substrate words requires an exhaustive evaluation of each of the individual criteria, in every instance, a task that is arduous and time-consuming.

A final word of caution is therefore in order. It has often proven tempting to abandon the methodological principles and bypass the etymological procedures outlined here by resorting to direct word-to-word comparisons between Indo-European and non-Indo-European languages (Baldi & Page 2006: 2187). Unfortunately, such direct, i.e. *ad hoc* lexical comparisons carry a substantial risk of producing false positive matches, the likelihood of chance resemblances being difficult to overstate (cf. Boutkan & Kossmann 1999). Through direct word-to-word comparisons, many grandiose substrate theories have been conceived, including those known by such appellations as Alarodian (Oštir 1921; 1930), Japhetitic (Marr 1920), Euskaro-Caucasian (Lafon 1949; cf. recently Bengtson 2017a; 2017b), Euro-African (Hubschmid 1953: 101), as well as Vennemann's Vasconic and Semitidic. As stated by Bichlmeier (2016: 326), "[s]o ingenuös diese Szenarien erdacht sind, fehlt ihnen schlicht eines: die Wahrscheinlichkeit". When identifying substrates, "[t]he essential task should be to refrain from diluting the concept by stretching it over the globe" (Craddock 1969: 42). The hallucinations of the substratomaniacs, to take a label from Rohlfs (1957), are notoriously alluring and the substratophobics cannot perhaps be blamed for hoping for someone to decree a definite end to all speculation (D. B. Y. 1950).

It is by all means advisable to analyze potential prehistoric loans, such as those in the European languages, using the accepted methodological principles before searching for comparanda outside Indo-European. And even then, the temptation to see parallels in large hypothetical "macro-substrates" may prove irresistible. To illustrate this, it is easy to find potential matches for multiple substrate items identified by scanning the European Indo-European branches for irregular lexis in the Northeast Caucasian languages (see Table 5).

Certainly, contact between some of the Indo-European languages and Northeast Caucasian (Schrijver 2018) cannot *a priori* be rejected and has been argued to

Table 5: Some similarities between phonotactically or phonologically irregular cognates sets in the European Indo-European languages and North-East Caucasian.

meaning	Greek	Armenian	Albanian	Italic	Celtic	Germanic	Slavic	North-East Caucasian
(wild) carrot						*murhō-	*mъrky	Lak *marx:*, *Dargwa marqːʷa* ‘root’
(thunder) mushroom	σπόγγος, σφόγγος	*sownk(n)*, *sowng(n)*		*fongo-		*swamba-		Avar *sːaḳ*, Tsez *ziḳu*, Udi *šaˤmk:al*
fig	σῦκον	*tˤowz*		*þīko-				Old Udi *tåxan*, Udi *toˤxaˤn*
pear 1	ἀχράς, ἄχερδος		*dardhë*					Lak *quˤrt*, Rutul *χˤir*, Archi *χˤert*
pear 2	ἄπιος			*piso-				Khinalug *bzɨ*
rye				*sekāl-				Rutul *siḳɨl*, Khinalug *sɨlg-li*
leek	πράσον			*porso-				Avar, Karata *per*, Archi *pil* (‘onion’)
spider	ἀράχνη			*ara(k)-sn-				Karata *χːarazan*
lightning	ἀστραπή, στεροπή				*s(t)rabi- (‘sulfur’)			Axwax *ʗa-piri*, Lak *ʗu-par*, Rutul *ʗaj-rap* (‘fire lighting’)

offer a realistic explanation for at least a couple of the so-called "Mediterranean" words (Matasović 2012: 8). However, given the time depth of the implied contact any serious comparisons should start out not from the attested Northeast Caucasian languages but from a Northeast Caucasian proto-language. For that, it would be important to reconstruct it, which happens to be "the one thing that has not been done" (Nichols 2003: 208). As a result, a danger once more lies in the risk of overinterpreting mere superficial resemblances, even with a limited corpus of carefully selected non-Indo-European words.

6 Outlook

It should be abundantly clear by now that the identification of linguistic substrates poses formidable challenges, even within a well-developed field such as Indo-European studies. The scarcity of the evidence means that achieving a complete reconstruction of the Pre-Indo-European linguistic landscape will always remain unattainable. At the same time, there is no point in denying that all the European Indo-European branches, as well as Armenian (cf. Thorsø, this volume) and perhaps even Indo-Iranian (see Kroonen, this volume), were impacted by at least one linguistically identifiable non-Indo-European stratum. The identification of a potential extra-linguistic proxy for this stratum poses a valid problem that deserves to be duly addressed.

Rather than evaluating the problem from a purely linguistic angle, exploring the interface with archaeology, with the help of linguistic palaeontological techniques, can provide potential clues. Semantic analysis reveals that the non-Indo-European vocabulary shared by the European languages exhibits a high number of words related to the so-called founder crops. This raises the question whether these elements were borrowed from a farming population (cf. Kroonen 2012a; 2012b; Iversen & Kroonen 2017). Would it be possible, to recall the caveat voiced by Bichlmeier, to offer a realistic scenario for the adoption of this vocabulary by the involved European Indo-European branches?

Quite possibly, interdisciplinary approaches will reveal new opportunities for finding answers. During the past years, several break-through archaeogenomic studies have shown that, prior to the Indo-European dispersal, Europe was inhabited by closely related populations of farmers who had colonized the continent from the East Mediterranean (Haak et al. 2005; Haak et al. 2010; Skoglund et al. 2012; Kılınç et al. 2016; Goldberg et al. 2017). Given their genetic homogeneity, it is not at all implausible that these farmers also introduced a new language (cf. Shennan 2018: 105), even if this language would have diversified considerably

at the time of the Indo-European dispersal. Future archaeogenomic studies will likely provide more detailed scenarios for the contacts between intrusive Indo-Europeans and preexisting European groups. In any case, the rapidly increasing biomolecular data volume on Europe's population history underlines the necessity of linguistic studies on prehistoric language contact, a necessity that has been apparent from the perspective of archaeology from the beginning. Almost a century later, we can therefore still applaud the determination shown by the early Güntert: "Die sog. „Substrattheorie" ist vorgeschichtlich gestützt, ja sie wird von der Vorgeschichte gefordert" (1934: 72).

References

Aikio, Ante. 2004. An essay on substrate studies and the origin of Saami. In Irma Hyvärinen, Petri Kallio & Jarmo Korhonen (eds.), *Etymologie, Entlehnungen und Entwicklungen: Festschrift für Jorma Koivulehto zum 70. Geburtstag*, 5–34. Helsinki: Société Néophilologique.

Alessio, Giovanni. 1944. Suggerimenti e nuove indagini sul problema del sostrato "mediterraneo." *Studi Etruschi* 18. 93–157.

Aslanov, Cyril. 2017. Armenian before Grabar: The emergence of the historically attested language in the shadow of the contact with non-Indo-European languages. In Lorenzo DiTommaso, Mathias Henze & William Adler (eds.), *The embroidered Bible: Studies in Biblical apocrypha and pseudepigrapha in honour of Michael E. Stone*, 79–95. Leiden/Boston: Brill.

Baldi, Philip & B. Richard Page. 2006. Europa Vasconica – Europa Semitica. *Lingua* 116(12). 2183–2220.

Battisti, Carlo. 1959. *Sostrati e parastrati nell'Italia preistorica*. Firenze: Felice Le Monnier.

Battisti, Carlo. 1960. Il sostrato mediterraneo nella fitonimia greco-latina. *Studi Etruschi* 28. 349–384.

Beekes, Robert S. P. 1996. Ancient European loanwords. *Historische Sprachforschung* 109. 215–236.

Bengtson, John D. 2017a. The anthropological context of Euskaro-Caucasian. *Iran and the Caucasus* 21. 75–91.

Bengtson, John D. 2017b. *Basque and its closest relatives: A new paradigm*. Cambridge, Massachusetts: Mother Tongue Press.

Bertoldi, Vittorio. 1928. Antichi filoni della toponomastica mediterranea incrociantisi nella Sardegna. *Revue de Linguistique Romane* 4. 222–250.

Bertoldi, Vittorio. 1931. Problèmes de substrat. *Bulletin de la Société Linguistique de Paris* 32. 93–183.

Bertoldi, Vittorio. 1939. Prerogative culturali mediterranee nel lessico greco e latino. *Archivio Glottologico Italiano* 31. 85–101.

Bertoldi, Vittorio. 1942. *Linguistica storica. Questioni di metodo*. Seconda edizione interamente riveduta e rinnovata. Naples: Dante Alighieri.

Bichlmeier, Harald. 2016. Zur Frage des nichtindogermanischen Substrats im Germanischen: ein kurzer Überblick. *Slovo a slovesnost, ročník* 77(4). 316–336.

Boutkan, Dirk H. F. 1998. On the form of North European substratum words in Germanic. *Historische Sprachforschung* 111. 102–133.

Boutkan, Dirk H. F. 2003. Lithuanian *šlākas*, Old Norse *slag*: Some features of North European substrate words exemplified by an alleged Indo-European etymon. In Alfred Bammesberger & Theo Vennemann (eds.), *Languages in Prehistoric Europe*, 245–252. Heidelberg: Carl Winter.

Boutkan, Dirk H. F. & Maarten G. Kossmann. 1999. Some Berber parallels of European substratum words. *The Journal of Indo-European Studies* 27(1/2). 87–100.

Boutkan, Dirk H. F. & Arend Quak (eds.). 2000. *Language contact: Substratum, superstratum, adstratum in Germanic languages*. Amsterdam: Rodopi.

Braun, Friedrich. 1922. *Die Urbevölkerung Europas und die Herkunft der Germanen*. Berlin: Verlag Kohlhammer.

Chantraine, Pierre. 1933. *La formation des noms en grec ancien*. Paris: Librairie ancienne Honoré Champion.

Childe, V. Gordon. 1926. *The Aryans: A study of Indo-European origins*. London: K. Paul, Trench, Trubner & Co.

Craddock, Jerry Russell. 1969. *Latin legacy versus substratum residue*. Berkeley, Los Angeles: University of California Press.

Cuny, Albert. 1910. Les mots du fonds préhellénique en grec, latin et sémitique occidental. *Revue des Études Anciennes* 12(2). 154–164.

D. B. Y. 1950. The Stalin-Marr philological controversy in the U.S.S.R. *The World Today* 6(8). 355–364.

Deeters, Gerhard. 1926. Armenisch und Südkaukasisch: ein Beitrag zur Frage der Sprachmischung. *Caucasica* 3. 37–82.

Deeters, Gerhard. 1927. Armenisch und Südkaukasisch: ein Beitrag zur Frage der Sprachmischung. *Caucasica* 4. 1–64.

Derksen, Rick. 2000. Old Icelandic *jarpi* 'hazel-grouse', *rjúpa* 'ptarmigan' and their Germanic and Balto-Slavic cognates. In Dirk H. F. Boutkan & Arend Quak (eds.), *Language contact: Substratum, superstratum, adstratum in Germanic Languages*, 97–105. Amsterdam: Rodopi.

Feist, Sigmund. 1910. Die germanische und die hochdeutsche Lautverschiebung sprachlich und ethnographisch betrachtet. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 36. 307–354.

Feist, Sigmund. 1932. The origin of the Germanic languages and the Indo-Europeanising of North Europe. *Language* 8(4). 245–254.

Furnée, Edzard J. 1972. *Die wichtigsten konsonantischen Erscheinungen des Vorgriechischen*. The Hague/ Paris: Mouton.

Furnée, Edzard J. 1979. *Vorgriechisches-Kartvelisches: Studien zum ostmediterranen Substrat nebst einem Versuch zu einer neuen pelasgischen Theorie*. Leuven-Louvain: Edition Peeters.

Gensler, Orin. 1993. *A typological evaluation of Celtic/Hamito-Semitic syntactic parallels*. Berkeley: University of California.

Gippert, Jost. 2005. Das Armenische – eine indogermanische Sprache im kaukasischen Areal. In Gerhard Meiser & Olav Hackstein (eds.), *Sprachkontakt und Sprachwandel: Akten der XI. Fachtagung der indogermanischen Gesellschaft, 17.–23. September 2000, Halle an der Saale*, 139–160. Wiesbaden: Reichert.

Goldberg, Amy, Torsten Günther, Noah A. Rosenberg & Mattias Jakobsson. 2017. Ancient X chromosomes reveal contrasting sex bias in Neolithic and Bronze Age Eurasian migrations. *Proceedings of the National Academy of Sciences* 114(10). 2657–2662.

Gorton, Luke. 2017. Revisiting Indo-European "wine". *The Journal of Indo-European Studies* 45(1/2). 1–26.

Güntert, Hermann. 1934. *Der Ursprung der Germanen*. Heidelberg: C. Winter.

Haak, Wolfgang, Oleg Balanovsky, Juan J. Sanchez, Sergey Koshel, Valery Zaporozhchenko, Christina J. Adler, Clio S. I. Der Sarkissian, et al. 2010. Ancient DNA from European early Neolithic farmers reveals their Near Eastern affinities. (Ed.) David Penny. *PLoS Biology* 8(11). e1000536.

Haak, Wolfgang, Peter Forster, Barbara Bramanti, Shuichi Matsumura, Guido Brandt, Marc Tänzer, Richard Villems, et al. 2005. Ancient DNA from the First European Farmers in 7500-Year-Old Neolithic Sites. *Science* 310(5750). 1016–1018.

Hamp, Eric P. 1990. The Pre-Indo-European language of Northern (Central) Europe. In John A. C. Greppin & Thomas L. Markey (eds.), *When worlds collide: the Indo-Europeans and the Pre-Indo-Europeans*, 291–309. Ann Arbor: Karoma Publishers.

Haynie, Hannah, Claire Bowern, Patience Epps, Jane Hill & Patrick McConvell. 2014. Wanderwörter in languages of the Americas and Australia. *Ampersand* 1. 1–18.

Hester, David A. 1968. Recent developments in mediterranean «substrate» studies. *Minos* 9(1). 219–235.

Hirt, Hermann. 1899. Die Urheimat der Germanen: Nachwort. *Neue Jahrbücher für das klassische Altertum, Geschichte und deutsche Litteratur* 3. 570–572.

Hirt, Hermann. 1905. *Die Indogermanen: ihre Verbreitung, ihre Urheimat und ihre Kultur*. Vol. 1. Strassburg: Karl J. Trübner.

Hübschmann, Heinrich. 1877. Ueber die stellung des armenischen im kreise der indogermanischen sprachen. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 23(1). 5–49.

Hubschmid, Johannes. 1953. *Sardische Studien* (Romanica Helvetica 41). Bern: A. Francke.

Hubschmid, Johannes. 1960. *Mediterrane substrate mit besonderer Berücksichtigung des Baskischen und der west-östlichen Sprachbeziehungen*. Bern: A. Francke.

Hubschmied, Johann Ulrich. 1942. Rätoromanisch. *Zeitschrift für Romanische Philologie* 62. 107–128.

Huld, Martin E. 1990. The linguistic typology of the Old European substrata in North Central Europe. *Journal of Indo-European Studies* 18(3/4). 389–423.

Iversen, Rune & Guus Kroonen. 2017. Talking Neolithic: Linguistic and archaeological perspectives on how Indo-European was implemented in southern Scandinavia. *American Journal of Archaeology* 121(4). 511–525.

Jokl, Norbert. 1923. *Linguistisch-kulturhistorische Untersuchungen aus dem Bereiche des Albanischen*. Berlin/Leipzig: Walter de Gruyter.

Jones, William. 1788. Third anniversary discourse: on the Hindus (delivered 2 February 1786). *Asiatick Researches* 1. 415–431.

Jongeling, Karel. 1995. Afro-Asiatic and Insular Celtic. *Near Eastern Languages and Literatures Foundation* 1. 135–165.

Kallio, Petri. 1997. Uralic Substrate Features in Germanic? *Journal de la Société Finno-Ougrienne* 87. 123–130.

Karsten, Torsten Evert. 1928. *Die Germanen: eine Einführung in die Geschichte ihrer Sprache und Kultur*. Leipzig: Walter de Gruyter.

Kılınç, Gülşah Merve, Ayça Omrak, Füsun Özer, Torsten Günther, Ali Metin Büyükkarakaya, Erhan Bıçakçı, Douglas Baird, et al. 2016. The demographic development of the first farmers in Anatolia. *Current Biology* 26(19). 2659–2666.

Klimas, Antanas. 2002. Lithuanian in the 21st century: Linguistic snippets and tidbits. (Trans.) Alfonsas Laučka. *Lituanus* 48(3). 52–80.

Kossinna, Gustaf. 1911. *Die Herkunft der Germanen: zur Methode der Siedlungsarchäologie*. Würzburg: C. Kabitzsch.

Kretschmer, Paul. 1896. *Einleitung in die Geschichte der griechischen Sprache*. Göttingen: Vandenhoeck & Ruprecht.

Kretschmer, Paul. 1925. *Die indogermanische Sprachwissenschaft: eine Einführung für die Schule*. Göttingen: Vandenhoeck & Ruprecht.

Kretschmer, Paul. 1932. Zur ältesten Sprachgeschichte Kleinasiens. *Glotta* 21(1/2). 76–100.
Kroonen, Guus. 2012a. Non-Indo-European root nouns in Germanic. In Riho Grünthal & Petri Kallio (eds.), *A linguistic map of prehistoric Northern Europe*, 239–260. Helsinki: Suomalais-Ugrilainen Seura.
Kroonen, Guus. 2012b. On the etymology of Greek ἄγλις and γέλγις 'garlic.' *Journal of Indo- European Studies* 40. 289–299.
Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic*. Leiden: Brill.
Kuiper, Franciscus B. J. 1956. The etymology of ἄνθρωπος. In Heinz Kronasser (ed.), *ΜΝΗΜΗΣ ΧΑΡΙΝ. Gedenkschrift Paul Kretschmer*, vol. 1, 211–226. Wien: Wiener Sprachgesellschaft.
Kuiper, Franciscus B. J. 1995. Gothic *bagms* and Old Icelandic *ylgr*. *North-Western European Language Evolution (NOWELE)* 25. 63–88.
Kuryłowycz, Jerzy. 1956. *L'apophonie en indo-européen*. Wroclaw: Zakład Imienia Ossolińskich.
Lafon, René. 1949. Sur les origines des Basques et de leur langue. *Les Cahiers d'Outre-Mer* 7. 193–207.
Lincoln, Bruce. 2008. Hermann Güntert in the 1930s. Heidelberg, politics, and the study of Germanic/Indogermanic religion. In Horst Junginger (ed.), *The study of religion under the impact of fascism*, 179–204. Leiden/Boston: Brill.
Lubotsky, Alexander M. 1998. Vedic roots of the type *TERD^{h}-. *Studia Indogermanica Lodziensia* 2. 75–81.
Lubotsky, Alexander M. 2001. The Indo-Iranian substratum. In Christian Carpelan & Petteri Koskikallio (eds.), *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations*, 301–317. Helsinki: Société Finno-Ougrienne.
Mailhammer, Robert & Theo Vennemann. 2019. *The Carthaginian North: Semitic influence on early Germanic: A linguistic and cultural study*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
Markey, Thomas L. 1989. The spread of agriculture in western Europe: Indo-European and (non-)pre-Indo-European linguistic evidence. In David R. Harris & Gordon C. Hillman (eds.), *Foraging and farming: the evolution of plant exploitation*, 585–606. London: Routledge.
Marr, Nikolai. 1920. *Yafeticheskii Kavkaz i Treth Etnicheskii Element v Sozidanii Sredizemnomorskoi Kultury*. Leipzig: Stamper's Typography.
Matasović, Ranko. 2012. Areal typology of Proto-Indo-European: The case for Caucasian connections. *Transactions of the Philological Society* 110(2). 283–310.
Matasović, Ranko. 2013. Substratum words in Balto-Slavic. *Filologija* 60. 75–102.
Meillet, Antoine. 1908. De quelques emprunts probables en grec et en latin. *Mémoires de la Société de Linguistique de Paris* 15. 161–164.
Meillet, Antoine. 1912. À propos de avestique zrazdā-. *Mémoires de la Société de linguistique de Paris* 18. 60–64.
Meillet, Antoine. 1913. La langue lithuanienne et la langue lette. *Les annales des nationalités* 2(5/6). 204–205.
Meillet, Antoine. 1922. *Introduction à l'étude comparative des langues indo-européennes*. Paris: Librairie Hachette.
Meillet, Antoine. 1930. *Caractères généraux des langues germaniques*. 4th edn. Paris: Klincksieck.
Nehring, Alfons. 1925. Griech. τίτοξ, τιτήνη und ein vorgriechisches k-Suffix. *Glotta* 14(1/2). 153–192.
Nichols, Johanna. 1998. The Eurasian spread zone and the Indo-European dispersal. In Roger Blench & Matthew Spriggs (eds.), *Archaeology and Language II*, 220–266. London: Routledge.
Nichols, Johanna. 2003. The Nakh-Daghestanian consonant correspondences. In Dee Ann Holisky & Kevin Tuite (eds.), *Current trends in Caucasian, East European and Inner Asian linguistics: Papers in honor of Howard I. Aronson*, 207–264. Amsterdam: John Benjamins Publishing Company.

Nörrenberg, Constantin. 1900. Was ist *NORD*? *Globus* 77(23). 371–389.
Onions, Charles Talbut. 1966. *The Oxford dictionary of English etymology*. Oxford.
Oštir, Karl. 1921. *Beiträge zur alarodischen Sprachwissenschaft I*. Wien; Leipzig: Beyers.
Oštir, Karl. 1930. *Drei vorslavisch-etruskische Vogelnamen*. Ljubljana: Znanstveno društvo.
Penka, Karl. 1886. *Die Herkunft der Arier. Neue Beiträge zur historischen Anthropologie der europäischen Völker*. Wien; Teschen: K. Prochaska.
Poesche, Theodor. 1878. *Die Arier, ein Beitrag zur historischen Anthropologie*. Jena: Hermann Costenoble.
Pokorny, Julius. 1927. Das nicht-indogermanische Substrat im Irischen. *Zeitschrift für celtische Philologie* 16. 95–144, 231–266, 363–394.
Pokorny, Julius. 1928. Das nicht-indogermanische Substrat im Irischen. *Zeitschrift für celtische Philologie* 17. 373–388.
Pokorny, Julius. 1930. Das nicht-indogermanische Substrat im Irischen. *Zeitschrift für celtische Philologie* 18. 233–248.
Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. 2 vols. Bern, München: A. Francke.
Polomé, Edgar C. 1986. The non-Indo-European component of the Germanic lexicon. In Annemarie Etter (ed.), *O-o-pe-ro-si, Festschrift für Ernst Risch zum 75. Geburtstag*, 661–672. Berlin/New York: De Gruyter.
Polomé, Edgar C. 1989. Substrate lexicon in Germanic. *North-West European Language Evolution* 14. 53–73.
Polomé, Edgar C. 1990. Types of linguistic evidence for early contact: Indo-Europeans and non-Indo-Europeans. In Thomas L. Markey & John A. C. Greppin (eds.), *When worlds collide: The Indo-Europeans and the Pre-Indo-Europeans*, 267–289. Ann Arbor: Karoma Publishers.
Polomé, Edgar C. 1992. Germanic, Northwest-Indo-European and pre-Indo-European substrates. In Rosina L. Lippi-Green (ed.), *Recent developments in Germanic linguistics*, 47–55. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Pott, August Friedrich. 1853. *Die Personennamen insbesondere die Familiennamen und ihre Entstehungsarten: auch unter Berücksichtigung der Ortsnamen*. Leipzig: F.A. Brockhaus.
Rohlfs, Gerhard. 1957. Zur Methodologie der romanischen Substratforschung (Substratomanie und Substratophobie). In *Syntactica et Stilistica. Festschrift für Gamillscheg*, 459–509. Tübingen: Niemeyer.
Salmons, Joe. 1991. Northwest Indo-European vocabulary and substrate phonology. In Roger Pearson (ed.), *Perspectives on Indo-European language, culture and religion: studies in Honor of Edgar C. Polomé*, vol. 2, 265–279. Washington, DC.: Institute for the Study of Man.
Salmons, Joe. 2004. How (non-)Indo-European is the Germanic Lexicon? And what does that mean? In Irma Hyvärinen, Petri Kallio & Jarmo Korhonen (eds.), *Etymologie, Entlehnungen und Entwicklungen: Festschrift für Jorma Koivulehto zum 70. Geburtstag*, 311–321. Société Néophilologique de Helsinki.
Salmons, Joe. 2012. *A history of German*. Oxford: Oxford University Press.
Saussure, Ferdinand D. de. 1878. *Cours de linguistique générale*. Paris: Payot.
Schleicher, August. 1852. *Formenlehre der kirchenslawischen Sprache*. Bonn: H. B. König.
Schmidt, Karl Horst. 1992. Kartvelisch und Armenisch. *Historische Sprachforschung* 105(2). 287–306.
Schmitt, Alfred. 1936. Die germanische Lautverschiebung und ihr Wert für die Frage nach der Heimat der Indogermanen. In Helmut Arntz (ed.), *Germanen und Indogermanen: Volkstum, Sprache, Heimat, Kultur; Festschrift für Herman Hirt*, vol. 2, 343–362. Heidelberg: Carl Winters Universitätsbuchhandlung.

Schrijnen, Jos. 1921. *Einführung in das Studium der indogermanischen Sprachwissenschaft, mit besonderer Berücksichtigung der klassischen und germanischen Sprachen*. Heidelberg: C. Winter.

Schrijver, Peter. 1997. Animal, vegetable and mineral: some Western European substratum words. In Alexander Lubotsky (ed.), *Sound law and analogy*, 293–314. Amsterdam: Rodopi.

Schrijver, Peter. 2011. La langue hattique et sa pertinence possible pour les contacts linguistiques préhistoriques en Europe occidentale. In Coline Ruiz-Darasse & Eugenio R. Luján Martínez (eds.), *Contacts linguistiques dans l'occident méditerranéen antique*, 241–255. Madrid: Casa de Velásquez.

Schrijver, Peter. 2014. *Language contact and the origins of the Germanic languages*. New York; London: Routledge.

Schrijver, Peter. 2018. Talking Neolithic: The case for Hatto-Minoan and its relationship to Sumerian. In Guus Kroonen, Bernard Comrie & James P. Mallory (eds.), *Talking Neolithic: Proceedings of the workshop on Indo-European origins held at the Max Planck Institute for Evolutionary Anthropology, Leipzig, December 2–3, 2013*, 336–374. Washington, D.C.: Institute for the Study of Man.

Schuhmann, Roland. 2016. Where is the substrate in the Germanic lexicon? In Bjarne Simmelkjær Sandgaard Hansen, Benedicte Nielsen Whitehead, Thomas Olander & Birgit Anette Olsen (eds.), *Etymology and the European Lexicon: Proceedings of the 14th Fachtagung der Indogermanischen Gesellschaft, 17–22 September 2012, Copenhagen*, 377–384. Wiesbaden: Reichert Verlag.

Shennan, Stephen. 2018. *The first farmers of Europe: An evolutionary perspective*. Cambridge University Press.

Skoglund, Pontus, Helena Malmström, Maanasa Raghavan, Jan Storå, Per Hall, Eske Willerslev, M. Thomas P. Gilbert, Anders Götherström & Mattias Jakobsson. 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336(6080). 466–469.

Šorgo, Aljoša. 2020. Characteristics of lexemes of a substratum origin in Proto-Germanic. In Romain Garnier (ed.), *Loanwords and substrata: proceedings of the colloquium held in Limoges, 5th-7th June 2018*, 427–472. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Stifter, David. 2023. The rise of gemination in Celtic. *Open Research Europe* 3. 1–24.

Streitberg, Wilhelm & Victor Michels. 1927. *Die Erforschung der indogermanischen Sprachen*, Band 2: *Germanisch, 1. Lieferung 1*. Berlin/Leipzig: Walter de Gruyter & Co.

Ternes, Elmar. 1998. 469. Keltisch und Romanisch. In Günter Holtus, Michael Metzeltin & Christian Schmitt (eds.), *Kontakt, Migration und Kunstsprachen*, 266–291. De Gruyter.

Terracini, Benvenuto. 1927. *Osservazioni sugli strati più antichi della toponomastica sarda*. Reggio Emilia: Office Grafiche Reggiane.

Tovar, Antonio. 1970. The Basque language and the Indo-European spread to the west. In George Cardona, Henry M. Hoenigswald & Alfred Senn (eds.), *Indo-European and Indo-Europeans*, 267–278. University of Pennsylvania Press.

Trask, R. Larry. 1997. *The history of Basque*. London: Routledge.

Vennemann, Theo. 2003. *Europa Vasconica – Europa Semitica*. Edited by Patrizia Noel Aziz Hanna. Berlin/New York: Mouton de Gruyter.

Wagner, Heinrich. 1959. *Das Verbum in den Sprachen der Britischen Inseln*. Tübingen: Max Niemeyer.

Wiik, Kalevi. 1997. The Uralic and Finno-Ugric phonetic substratum in Proto-Germanic. *Linguistica Uralica* 33(4). 258–280.

Wilser, Ludwig. 1900. Die "Kruger-Penkasche Hypothese": ein Beitrag zur Geschichte der arischen Frage. *Globus* 78. 144–147.

Winther Johannsen, Jens. 2016. Heavy metal: Lead in Bronze Age Scandinavia. *Fornvännen* 111(3). 153–161.

Part II: **Northeastern and Eastern Europe**

Anthony Jakob

2 Three pre-Balto-Slavic bird names, or: A more austere take on Oštir

To the strict Neogrammarian, a phonological "near miss" is as good as useless, and therefore better left unmentioned. The kinds of irregular comparisons which could point to independent loanwords from lost languages are ignored by those working within the canonical framework for the study of Indo-European. In the context of this volume, the exploration of works situated outside of the Neogrammarian framework can yield unexpected fruits. In this study, I will turn to the works of the Slovene linguist Karel Oštir.

Oštir was educated as a comparative linguist in Graz under the tutelage of Rudolf Meringer (Čop 1988: 4–5), and it was in the latter's journal *Wörter und Sachen* that he published several early papers revealing a flair for etymology with a particular interest in Armenian and Balto-Slavic (see Čop 1973: 15–26).[1] This appreciation of Neogrammarian theory is what gave Oštir's later forays into pre-Indo-European languages a qualitative advantage over those of his contemporary Nikolai Marr, whose ideologically-driven works stood in direct opposition to the established linguistic theory (e.g. Marr 1928; cf. Čop 1973: 30). Oštir sought to supplement, rather than to deny, the comparative method, and his work must therefore be viewed as part of the same tradition as Hübschmid (1950), Furnée (1972) and Beekes (2014).

What gives Oštir's work value for the study of European linguistic prehistory is the fact that the "Alarodian"[2] elements he identified included implied loanwords in the Indo-European languages of Europe. In other words, Oštir's work, and in particular his *Drei Vogelnamen* (1930), can be seen as one of the first attempts to catalogue potential substrate words in Europe. His significant methodo-

1 For the Balticist, the most important contribution is Oštir 1912: 214–215, which features three Baltic etymologies — Lith. *síena* 'wall' (to *siẽti* 'bind, tie up'), *žaltỹs* 'snake' (to *žãlias* 'green') and *mė́žti* 'heap manure' (to OCS *mazati* 'anoint') — all of which have been well received (see Mühlenbach and Endzelīns, 2: 622, 3: 858; Fraenkel: 444, 782, 1288). Also note Oštir's more extended contribution *Baltoslovanska metatonija* (1925).

2 Outside of his dissertation (*non vidi*, cf. Čop 1973: 31–34), Oštir rarely presented any kind of exposition of the so-called "Alarodian" language family. In name, and to a large extent in composition, Oštir's family can be identified with that proposed by Hommel (see e.g. Hommel 1884), consisting of "die vorindogermanischen Sprachen Kleinasiens, Nordsyriens, Armeniens und Elams" (Hommel 1904: 36), to which Oštir would add many more including Hamitic (Ancient Egyptian and Berber), Alteuropäisch (incl. Basque and Etruscan) and "Caucasian" in general.

https://doi.org/10.1515/9783111337920-002

logical contribution is the focus on "alternations" as evidence of borrowing from an unrecorded source.[3] Equally important is Oštir's almost unrelenting presentation of parallels for any suggested alternation: an isolated irregularity was for him not sufficient in itself unless it could be shown to form part of a greater pattern.

The obscurity of Oštir's work is certainly due, at least partly, to the dense and barely penetrable writing style: aside from the extended discussion of the word for 'fig' in his *Beiträge zur alarodischen Sprachwissenschaft* (1921: 1–33), most of his works in the Alarodian framework barely contain a single complete sentence, consisting almost entirely of condensed mathematical presentations of data (e.g. Oštir 1923, 1929, 1930; see Figure 1). That is not to say his work is full of gems — indeed, most of the comparisons are of questionable quality. Contemporary reviewers (cf. Schuchardt 1922: 21; Meillet 1922: 130; Uhlenbeck 1923) praised the author's knowledge and erudition, but pointed to his lack of criticism. Schuchardt puts it best: "er gibt sich keine Rechenschaft über die Grenzen der Erkenntnismöglichkeit" (1922: 21).

9 *ka*[*r̃*]*itz* ‚tempête orage' ∞ (*h*)*a*[*r̃*]*ize* ‚vent' (*h*)*atz*/*s* ‚haleine' mit [*r̃*] ∞ *r̃* in redupl. **xar̃-xanitze* > *arnase* > *asnase* ds. — zu [*r̃*]*i* : [*r̃*] in bask. *haize* : *hatz*/*s* vgl. bask. *haritz* ‚Eiche' (: *ez-kur* ‚gland') > iber. **ka*[*r*]*iss-* > span. *quejigo* ‚Bergeiche' vgall.-rom. **ca*[*r*]*ssan-* ‚Eiche' REW 1740 —; aus dem Kauk. hieher redupl. did. *ko-koni* ‚Wind' (< *r̃* ∞ *n* >) gru. *khari* kub. *xur*

Figure 1: Oštir's dense writing style (Oštir 1921: 38).

Yet the work is not void of interest, either.[4] In the following, I would like to revisit some of the irregular comparisons taken as Alarodian by Oštir and present them in a slightly different context. I will forego any discussion of the supposed *genetic* connections between the pre-Indo-European languages of Europe and other attested non-IE languages. Such theories, despite continued debate (e.g. Schrijver 2018: 361–363), are unlikely ever to be proven in a satisfactory, empirical manner purely due to the dearth of evidence. In the remainder of this study, I will assess

3 True, like Furnée, Oštir was inclined to assume such alternations were themselves present in the source language, rather than being artefacts of the loaning process. While Furnée (1972: 83–92) preferred to operate with expressive alternations within pre-Greek, Oštir (1921: 94–99) posited a system of consonant and vowel gradation, "wie im Uralischen".

4 Cf. "on trouvera sans doute à prendre dans ce monceau extraordinaire de rapprochements entre les langues les plus diverses", Meillet (1922: 130).

three "Alarodian" bird names in Balto-Slavic.[5] In doing so, I will follow a modified version of Oštir's methodology, cataloguing irregularities as potential indications of borrowing and seeking out parallels for any suggested alternation.

1 'Pigeon'

The comparison Lat. *columba* : OCS *golǫbъ* 'pigeon' (Oštir 1921: 49; 1923: 276; 1930: 39) is self-evident, and goes back to the start of Indo-European linguistics.[6] Yet it did not survive the Neogrammarian revolution, with the Slavic comparandum being dropped from the fourth edition of Fick's *Vergleichendes Wörterbuch* (cf. Stokes 1894: 92), and Prellwitz (1897: 102–104) establishing the new common opinion: the Latin and Slavic words are simply parallel formations to different roots (Brugmann 1906: 386). At the heart of this theory comes Skt. *vṛ́ṣan-* beside *vṛ̥ṣabhá-*, both occurring in the senses 'manly' and as a substantive 'bull'. If *vṛ̥ṣabhá-* reflects **urs-n-bʰo-*, then, according to Prellwitz, Lat. *columba* should also be derived from an *n*-stem, which he identified in Gr. κελαινός 'black, dark' (see also Persson 1912: 169–171; Pokorny 1959: 547–548; Batisti 2021: 206–207).

Except for the synonym *ṛ̥ṣabhá-* 'bull' (belonging with Av. *aršan-* 'manly'), other examples of this pattern of derivation in Indo-Aryan are quite uncertain. Skt. *śarabhá-*, a kind of game animal, continued in Dardic and Nuristani in the senses 'markhor, ibex, mountain goat' (CDIAL: 714), is supposedly connected to Lat. *cornū* 'horn' (Uhlenbeck 1898–1899: 304; Mayrhofer, 2: 616), but this is far from certain. Two words for 'donkey' — *rāsabhá-* and *gardabhá-* — are not well explained; at least the latter is probably not of Indo-European origin (see Mayrhofer, 1: 473; Pinault 2008: 393–394). A close parallel to *vṛ́ṣan-* beside *vṛ̥ṣabhá-* is nevertheless found in Gr. ἔλαφος 'deer', which beside OCS *elenь* 'deer', Arm. *ełn* 'doe, hind' has traditionally been derived from **h₁el-n-bʰo-* (Prellwitz 1897: 100; Osthoff 1901: 305–308; Meillet and Vaillant 1933: 102). At the same time, other examples of the supposed suffix **-n-bʰo-* remain rare, being limited to a few European bird-names: Lat. *palumbēs* 'wood pigeon', Arm. *salamb* 'francolin' and our words for 'pigeon'.

Despite the potential derivational parallel in the words for 'deer', the separation of Lat. *columba* and OCS *golǫbъ* feels artificial: the words mean exactly the same thing, and aside from the voicing of the initial stop, show an identical stem.

5 Although the title of this study is a play on the title of Oštir's *Drei vorslavisch-etruskische Vogelnamen* (1930), the three bird names discussed here do not coincide with Oštir's three.

6 Bopp 1833: 336; Pott 1861: 449; Miklosich 1865: 135; Fick 1871: 349. Of these, only Miklosich saw the need to address the Slavic *g-*, comparing Sln. *kâvran* : *gâvran* 'raven'.

Furthermore, the Slavic suffix is entirely unique, and in Latin it is paralleled only by another word for 'pigeon'. It is hardly plausible that Latin and Slavic would independently choose to employ this inherited suffix only in words for pigeon, and end up choosing almost identical roots. In this context, it is unsurprising that scholars have used the word in support of a number of competing theories, such as Indo-European *Konsonantenwechsel* (Hirt 1927: 334; Otrębski 1939: 156) or borrowing through an unidentified Indo-European language (Haas 1960: 34; Holzer 1989: 161–162). Further, some have suggested the word was borrowed from Latin into Slavic, either directly (Szemerényi 1967: 20–21) or through the mediation of an unknown language (Sobolevskij 1914: 441).

What unites all these theories is the assumption that Lat. *columba* is ultimately inherited. However, there is no solid indication of this. Prellwitz's root etymology is merely possible, but certainly not compelling.[7] The idea that the Latin and Slavic words could be independent loans from an unidentified source has been put forward only occasionally since Oštir (cf. Machek 1951: 103–104; Bezlaj, 1: 159; Kleyner 2015: 53–54; ERHJ, 1: 284).

The alternation $^{*}k \infty {}^{*}g^{(h)}$ can be supported by a number of parallels; in each case, not only do we find the same consonantal alternation, but also a comparable distribution:

i. Lith. *dal̃gis*, OPr. E *doalgis* ~ Lat. *falx -cis* 'scythe' (cf. Hirt 1927: 299; Alessio 1946: 165)[8]
ii. Latv. dial. *dradži* 'dregs of melted fat', OPr. E *dragios* PL. 'dregs' ~ Lat. *fracēs* F.PL. 'olive pomace'[9]

7 See Batisti 2021: 207 with lit. for other root etymologies, none of which are any more convincing. Walde and Hoffmann (1: 249) insist that the Slavic word must be native because of the colour term seen in Ru. *golubój* 'light blue', but the latter is instead a derivative of the word for 'pigeon' (Loewenthal 1901: 31–32; Machek 1951: 103; Herne 1954: 91).

8 For both Lat. *falx* and *fracēs*, one has assumed generalization of /k/ from the nominative singular (see Mikkola 1899: 74; Walde and Hofmann, 1: 539, respectively). For *fracēs*, this is hardly possible as the word is *plurale tantum*, *frax* only being recorded in glosses. Neutralization on the basis of the nominative would in any case be typologically unusual (see Niedermann 1918: 22; Decaux 1966). Both *falx* and *fracēs* imply IE $^{*}a$, which combined with the illegal root shape $^{*}D^{h}\_k$-, has encouraged others to seek a non-IE origin (cf. Ernout and Meillet 1951: 214, 251; Schrijver 1991: 486; de Vaan 2008: 200, 238).

9 OCS *drožďьję* 'dregs', Po. *drożdże* F.PL. 'yeast, leaven' are usually included here, but as these imply an underlying **drazg-* or **drazdj-*, it seems phonologically easier to compare OE *drœst*, *dærste* 'leaven, dregs' ?< $^{*}d^{h}$*rosd-*. Despite Meyer (1891: 72; also Demiraj 1997: 141), Alb. *dra* 'dregs of melted fat' cannot derive from **dragā*, as $^{*}g$ was not lost intervocalically (cf. Schumacher 2013: 240). A possibility would be to posit a preform **drasā-* < $^{*}d^{h}$*rHs-*, and compare OE *drōsna*, Du. *droesem* 'dregs, sediment' < $^{*}d^{h}$*roHs-*.

iii. Lith. *rugiaĩ*, Latv. *rudzi*, OPr. E *rugis*, Ru. *rož'*, Sln. *rȓž* ~ MW *ryc* 'rye'[10]
iv. Lith. *gulbė̃*, Latv. *gùlbis* 'swan', OPr. E *gulbis* ~ Kash. *kôłp*, SCr. dial. *kûp* 'swan' (Derksen 1999)[11]
v. Ru. *grab*, Cz. *habr* ~ Lat. *carpinus* 'hornbeam' (cf. Machek 1950b: 152; Matasović 2013: 84)[12]

The above examples all show a north-south divide, with $^{*}g^{(h)}$ regularly occurring in Baltic and $^{*}k$ in Italic and Celtic. While the distribution is slightly broken by Slavic, which shows conflicting reflexes, this could be explained by the intermediate geographical position of Slavic, which would permit contact with both southern and northern neighbours. As for Germanic, ON *rugr* 'rye' and *dregg* 'yeast, (PL.) dregs' tell us precious little, as Verner's law does not allow us to decide between $^{*}k$ and $^{*}g^{h}$

Oštir (1921: 49) adduced another curious comparandum from Coptic: Sahidic *croompe*, Bohairic *crompi*, Lycopolitan *crampe* 'pigeon' derive from a Late Egyptian (~ 12th c. BC) form *gr-(n)-p.t* */kʰV̆rámpV̆/ (cf. Allen 2020: 115).[13] The word is spelled as '*gr*-bird of the sky', and has been viewed by Egyptologists as a native formation. Peust (1999: 280) and Ivanov (2002) have suggested that Egyptian may be the source of the Indo-European words. Indeed, it seems quite unlikely that the similarity is coincidental. Vycichl (1990: 249), on the other hand, has argued that the Egyptian spelling is folk-etymological ("la colombe n'est pas un « oiseau

10 A non-Indo-European origin has often been suspected, although on the basis of non-linguistic evidence (e.g. Schrader 1901: 639; Hoops 1915–1916: 509–510; Charpentier 1930: 71; Polomé 1992: 70). MW *ryc* has generally been derived from OE *ryge* (Schrader 1901: 639; GPC, 3: 3136), but this is chronologically difficult. Welsh /k/ could hardly substitute the Old English spirant /j/ (the spirantization of /g/ has been dated to the continental Old English period, cf. Campbell 1959: 173). On the other hand, if the loan were of Proto-Celtic age, one would expect Celtic $^{*}g$ (> Welsh $^{**}Ø$). At face value, the Welsh data points to $^{*}rukio$- or $^{*}rukī$- (cf. Kroonen et al. 2022: 22), disagreeing with $^{*}g^{h}$ elsewhere.

11 Already Oštir 1930: 66. The rare and only lexicographically recorded SCr. *gûb* 'swan' is better disregarded (Vaillant 1929: 270; Sławski 1960: 40). The Sln. dial. *golbica*, which Bezlaj (1: 157) adduces in this connection, refers to the 'skylark', a tiny passerine bird which, aside from being a bird, has absolutely nothing in common with the swan.

12 The frequent derivation of the Latin word from *carpō* 'pluck' (supposedly < *'cut') based on the hornbeam's crenated leaves (Walde and Hofmann, 1: 171; Schrijver 1991: 430) is hardly logical, as the plant's leaves are neither sharp nor capable of cutting.

13 Allen actually reconstructs a final */-nipV̆/, but apparently only because the Egyptian genitive marker ⟨n⟩ is reconstructed as */ni/. This might be anachronistic, as spellings with ⟨m⟩ are already attested in Late Egyptian (Allen 2020: 115; see Erman and Grapow 1931: 181), suggesting that no vowel was present in at least some Late Egyptian varieties. Perhaps the spellings with ⟨n⟩ might be interpreted as archaicizing or folk-etymological.

du ciel » comme l'aigle ou le faucon"), and supported the earlier suggestion that we are dealing with a borrowing from an unknown source (cf. Worrel 1934: 67). It is in any case clear that Egyptian cannot be the direct source of the European words, due to both a mismatch in vocalism (Latin *-umb-* requires a labial vowel, cf. Leumann 1977: 81), and Egyptian *-r-* vs. European *-l-*.[14] Note that the latter alternation is paralleled in the Mediterranean by the example of Lat. *hirundō* ~ Gr. χελῑδών 'swallow'.[15]

A largely neglected piece of evidence could provide important additional support for the word's foreign origin. In Old English, we find the forms *culfre* ~ *culufre* 'dove'. While Skeat (1882: 146) saw this as a 'corrupted' Lat. *columba*, Pogatscher (1898: 97) suggested a diminutive **columbula* as the source (cf. OOcc. *colombla* 'dove', FEW, 2: 930). Either solution is phonologically problematic, as the loss of the nasal cannot be accounted for, whether we assume that *culfre* is a result of syncope (Campbell 1959: 159) or *culufre* the result of vowel epenthesis (Hogg 1992: 231–232).[16] I would therefore propose that we are dealing with another non-IE alternation. A close parallel can be found in RuCS *erębь*, Po. *jarząbek*, with a nasal vowel as against Lith. *jerubẽ* 'hazel grouse' (cf. Derksen 2000).

Alternations of the type **VNC* ∞ **V̌C* have long been identified as non-IE (cf. Kretschmer 1896: 403; Oštir 1930: 14; Kuiper 1956: 213–215; Furnée 1972: 275–291). Kuiper's Munda-inspired term "prenasalization" has become popular in Leiden (Beekes 1996: 223–226; Schrijver 2001: 420–421), but is probably best avoided, as other interpretations are possible, e.g. the alternation may reflect a diachronic or dialectal development within the putative source language, rather than a synchronic feature.

14 On the nature of Egyptian ⟨r⟩, see Peust (1999: 127–129).

15 And perhaps—if not a mere dissimilation—by Lat. *līlium* ~ Gr. λείριον 'lily'. The latter are frequently also connected with Cpt. *hrēre, hlēli* 'flower' < Eg. *ḥrr.t* */harīra.t/ (Worrel 1934: 67, Beekes 2010: 845; on the reconstructed vocalism, see Vycichl 1990: 94), but this etymology is suspect due to the absence of any reflection of the first syllable in the European languages, and the imprecise semantic match (cf. Vycichl 1983: 310).

16 Hogg's solution appears less likely: the epenthesis is typically Northumbrian, while the form *culufre* seems to have been more widely distributed (according to the data in the *Dictionary of Old English Web Corpus*). Furthermore, judging by the examples provided by the cited authors, the epenthesis typically occurs in a different environment: before word-final *_RC#* (where *C* is usually a velar) or before the clusters *-ht-* or *-gd-*.

2 ‘Swan’

Another example of such a nasal alternation cited by Oštir (1923: 309; 1930: 14, 68; 1930–1931: 14) is the Slavic word for ‘swan’. Two similar words for ‘swan’ are found in Slavic which cannot reflect a single preform: (1) Po. *łabędź*, Cz. *labuť*, SCr. *lȁbūd*, Sln. *labǫ́d*, and (2) Ru. *lébed’*, Ukr. *lébid’* (GEN.SG. *-edja*), CS **lebedь* (attested *debelь*, cf. Bulg. *lébed*). The two forms are almost in complementary distribution.[17] The second syllables clearly do not match, with the first group implying **-bǫdь* and the second **-bedь*.[18] The difference is difficult to account for in Indo-European terms, but is reminiscent of the alternation found in the two bird names discussed above. The second syllable ends with a voiceless stop before which a nasal is sometimes, and sometimes not, present (cf. Kroonen 2013: 20). Compare:

Sl. **GolomB-*	~	OE **gulubʰ-* ‘pigeon’
Sl. **i̯eremB-*	~	Balt. **i̯erubʰ-* ‘hazel grouse’
Sl. **albonD-*	~	Sl. **lebʰedʰ-* ‘swan’

There is also a mismatch between the initial syllables. While group (2) is now usually reconstructed **elbedь* (Bulaxovskij 1948: 118; ÈSSJa, 4: 19; Sławski SP, 6: 40), it is quite uncertain that this would yield the attested forms. Despite the frustrating lack of evidence for the outcome of **eRC-* in Slavic,[19] one would certainly expect a treatment parallel to **aRC-* (cf. Vaillant 1950: 160–161; Arumaa 1964: 148). It is therefore preferable to return to the older reconstruction **lebedь* (Miklosich 1886: 162; Osthoff 1898: 65–66). The first group, on the other hand, must be compared to ON *ǫlpt*,

17 Pleteršnik (1: 503) cites a rather doubtful looking Sln. *lebed* from the dictionaries of Jarnik and Janežič, which does not appear to be attested dialectally (Tijmen Pronk p.c.), while some other forms in South Slavic, e.g. Mac. *labed* and SCr. obs. *lȅbūt* (RJA, 5: 944) seem to show a confusion between the two forms.

18 Despite ÈSSJa (6: 19) and Nikolaev (2020: 39, fn. 6; cf. Zaliznjak 2019: 640), it seems incorrect to take the East Slavic forms from **lebędь*. All of the Old Russian evidence suggests *-bed-* (SDRJa, 2: 13–14), as does Ukr. *lébid’* (GEN.SG. *lébedja*). The adjective *lebjážij*, is by all appearances a late creation, replacing earlier *lebežii* in the 17th century (SRJa 11–17, 8: 183; cf. Bulaxovskij 1968: 103). It can be considered a hypercorrection due to the widespread merger of /’a/ and /e/ in unstressed syllables (DARJa, 1, Map 3).

19 The example of Sln. *réšek* (sic Erjavec 1879: 126) ‘prickly sow thistle’ (here also SCr. dial. *rèkeš* ‘eryngo’ among other variants; cf. RJA, 13: 860; Bezlaj 1977: 17) ~ Lith. *eršk̃etis*, Latv. *ẽršķis* ‘wild rose’ (Persson 1912: 841; Fraenkel 122–123; Andersen 1996: 140–141) is highly uncertain, as Sln. *-š-* does not regularly correspond to Baltic *-šk-*.

OHG *albiz* 'swan'.[20,21] One could only connect **lebedъ* and **albǫdъ* by assuming a doubtful *Schwebeablaut* (Andersen 1996: 124), but given that the second syllables also fail to correspond, such a last resort solution is unattractive.

Derksen (2000: 84) has suggested we account for the different Slavic reflexes by assuming a 'prefix' **a-* (thus **a-lb-* : **leb-*). Although this idea was not taken over in his dictionary (2008: 143), it does seem a plausible way to account for the two forms. The behaviour of this 'prefix' would be consistent with that of the supposed non-IE morpheme **a-* with concomitant stem reduction adduced by Schrijver (1997: 307–310). The closest parallel can be seen in another bird name:

– Lat. *merula* (< **mesal-*) ~ OE *ōsle*, OHG *amsala* (< **a-msl-*) 'blackbird'
– Ru. *lébed'* (< **lebʰedʰ-*) ~ ON *ǫlpt*, OHG *albiz* (< **a-lbʰed-*) 'swan'

Another excellent parallel is found in the relationship between Lat. *raudus* 'piece of copper or brass' and ODu. *arut*, OHG *aruz* 'ore' (Schrijver 1997: 308; Kroonen 2013: 37). As well as showing a similar correlation between the presence of (**a-* and a "reduced" stem (**raud-* : **a-rud-*), the variants also show a comparable geographical distribution. In conclusion, the words for 'swan' in Slavic and Germanic show several irregularities which make it highly probable we are dealing with a lexeme of foreign origin (cf. also Machek 1968: 316; Kroonen 2013: 20). As shown in the above discussion, all of the attested irregularities are paralleled in other European bird names.

3 'Oriole'

Oštir (1930: 101) has compared Lith. *volungẽ* 3ª 'oriole'[22] with the Slavic synonym Po. *wilga*, SCr. *vùga*, presenting it as an example of the alternation **ā* ∞ **ĭ* and nasal loss. The comparison certainly looks attractive, as was already recognized by

20 Osthoff reconstructed **lōbʰ-* for group (1) comparing the Hesychian gloss ἀλωφούς· λευκούς, which occurs alongside ἀλφούς· λευκούς. The former is most probably a mere transmission error (Beekes 2010: 77; Gippert 2017: 184–185), meaning that Osthoff's reconstruction has no real basis.

21 To account for ON *ǫlpt*, OHG *albiz* beside OE *ielfetu*, OHG *elbiz* one has often assumed two byforms, **albit-* and **albut-* (Noreen 1892: 93; Specht 1947: 114; de Vries 1962: 101; EWAhd, 1: 1033). A more cautious approach may be in order: since the *u*-umlaut in ON *ǫlpt* (GEN.SG. *alptar*) can be attributed to the analogical extension of *u*-umlaut to feminine consonant stems (cf. Noreen 1892: 183; Kroonen 2013: 26), the most straightforward solution would be to posit a root noun **albet-* (cf. somewhat similarly Orel 2003: 13).

22 Latv. *vãluôdze* 'oriole' suggests an underlying **-ang-* in the second syllable, which does not match the Lithuanian form. By way of a solution, ALEW 1469 suggests that the standard Latvian form is a hypercorrection based on the Latgalian continuant. Alternatively, we may be dealing with an East Lithuanian form with **anC* > **unC* which has been adopted into the other dialects

Miklosich (1865: 68; 1886: 379) — not only are the semantics perfect, the Baltic and Slavic words share a consonantal skeleton **v–l–g*. Yet since an Indo-European explanation for the correlation is not easy to come by, some recent works have rejected the relationship altogether (Smoczyński 2018: 1693; ALEW 1469). Treating the Baltic second syllable as a suffix (Endzelin 1924: 123, citing the river name *Bebrunga*) does not help to clarify the relationship with the Slavic form.[23]

Here probably belong Germanic forms like ME *wode-wale*, MHG *wite-wal* 'oriole' (ME *wode*, MHG *wite* 'wood'; Endzelin 1924: 123). A trace of the velar of the Balto-Slavic forms could be found in Swiss and Bavarian dial. *Wiedewalch* (since the 15th century, see Suolahti 1909: 170). In a bid to unite the material, Machek (1950a: 49–50) suggests the Germanic reconstruction **-walka*, but the loss of **-k-* elsewhere would be irregular. However, if we again operate with the non-IE alternation **g(ʰ) ~ *k*, we could posit **-walhō-* for Germanic. The loss of **h* in Low German and Dutch would be regular, cf. MDu., MLG *male* 'bag' (< **malhō-*, Kroonen 2013: 351). While it is more sporadic in Middle English and High German, the simplification of the cluster may have been supported by the word's unstressed position as the second element of a compound. This bird name therefore can be added to the examples of the **g(ʰ) ~ *k* alternation, discussed above. The material can be systematized as follows:

Baltic:	**u̯**	*â*	**l**	*an*	**g-**
Slavic:	**u̯**	*i*	**l**	Ø	**g-**
Germanic:	**u̯**	*a*	**l**	Ø	k-

A kind of "nasal loss" can be observed here, too, but the behaviour differs from that of the previous examples. Here, where the nasal is absent, the syllable is lost entirely. Furthermore, the initial syllable vocalism is highly irregular. If this comparison is to be substantiated, it would be desirable to find parallels for these alternations.

For the nasal loss, Oštir (1930: 68; cf. 1921: 55) cites a close parallel in OCS *golǫbъ* 'pigeon' : Lith. *gulĨbė* 'swan', an equation occasionally seen elsewhere (cf. Prellwitz 1897: 103; Fraenkel 175). On the other hand, the semantic connection between 'pigeon' and 'swan' is tenuous. Derksen (2015: 510) cites a more promising pair —

(note in this respect the rare South Aukštaitian *ulangėlė* cited in LKŽ, and cf. Derksen 2008: 386 on *ungurỹs* 'eel').

23 The etymological comparison (cf. Endzelin 1914: 126; Fraenkel: 1273–1274) with Av. *vārənjana-*, *vārəγna-* a bird of prey (cf. Sogd. *w'rγn'k*, Khwar. *w'rγnyk* 'falcon', Hintze 1994: 198–199) is semantically weak. Note that Endzelīns and followers operate with Bartholomae's non-specific translation "Name eines Vogels", which might explain their enthusiasm.

Lith. *jerumbẽ* : *ìrbė* ‘hazel-grouse’. However, it seems questionable that these two variants within Lithuanian should be projected back to a more distant prehistory. In fact, *jerumbẽ* is a very rare form;[24] how it is to be explained is rather unclear, but the limitation to some isolated dialects suggests a secondary development. Perhaps, in some areas, a certain role may have been played by Po. *jarząb(ek)*.[25] Lithuanian forms in *irb-* appear to be absent dialectally, and perhaps all the relevant data derives ultimately from Latvian.[26] Within Latvian, Latv. *ĩrbe* can be explained as the result of syncope from an older *(*)ierube* (cf. East Latvian *ìrube*), for which a convincing parallel can be found in *ilkss* : East Latvian *ielukši* ‘carriage pole’ (Endzelin 1923: 47).[27]

A possible parallel may be found in the equation between Lith. *balánda*, Latv. dial. *baluôda* ‘goosefoot, *Chenopodium*’ and OHG *melta* ‘orache, *Atriplex*’ (Machek 1950b: 149). The correlation between Baltic **balañd-* and Germanic **maldjō-* (cf. LÄGLOS, 2: 248; Kroonen 2013: 251), is quite similar to that between **vâlang-* and Germanic **walhō-* ‘oriole’. For the initial correlation **b-* ∞ **m-*, we might adduce Latv. *buřkãns* ‘carrot’ as against OHG *moraha* ‘edible root’, which have previously been interpreted as parallel loanwords from an unknown source (Kroonen 2013: 378; Matasović 2013: 88; Pronk and Pronk-Tiethoff 2018: 282).

For the vocalic alternation, Oštir (1930: 22) has adduced Lat. *taxus* ~ Ru. *tis*, Sln. *tîsa* ‘yew tree’. While the latter is probably indeed of non-IE origin (cf. Machek 1950b: 152; Sławski SEJP, 1: 103), the parallel is imperfect due to differences in vowel length. Perhaps more promising is the comparison between Lat. *grāmae*[28] and RuCS *groměždь*, *gremežďь* (SRJa 11–17, 4: 129) ‘rheum in the eye’,[29] which has been suggested in a non-IE context by de Vaan (2008: 270) and Mata-

24 Cited as a variant of *jerubẽ* in Juška (2: 684), and recurring in the form *arumbẽlė* in a daina (Palėvenė, LKŽ s.v. *arumbė*) and in the variant *vierumbẽlė* (Marcinkonys, LKŽ).

25 The form **jérumbė* (in dialect notation *jiêrǫ̂mbę*), recorded in Šateikiai (*Papildinimų kartoteka*) may be a direct loanword from Polish, showing /ė/ regularly for Slavic /a/ after a palatal.

26 Thus, *Ýrbenis* ‘Viburnum’ (Pabrėža 1834: 49) seems to be based on Latv. *irbenes* (cited there by the author). The bird name *ìrbė* ‘hazel grouse’ is only known from Šlapelis’ dictionary (*apud* LKŽ), while *vìrbė* ‘hazel grouse’ was perhaps popularized by Ivanauskas’ *Lietuvos paukščiai*. It does not seem to be attested prior to the early 20th century. Another Latvianism attributable to Ivanauskas is *lestė* ‘flounder’ (= Latv. *leste*, see LKŽ).

27 Compare also Lith. dial. *jérbė* which is surely the result of syncope (cf. the place name *Jeřbiškiai* < *Jerùbiškiai* cited in Zinkevičius 1966: 132).

28 Sic. TLL, 6: 2165. The word is rare, but the long vowel is metrically secured in Plautus. The derived adjective *grammō(n)sus* would therefore show the littera rule.

29 Further Slavic forms have an unclear initial **k-*: Sln. *krmẹ́žəlj*, in addition to which Sławski (SP, 8: 267) adduces SCr. Čak. dial. *kȓmež*, Kajk. dial. *kȓmeželj*. Both languages also attest a shorter form: Sln. *krmẹ́lj* (lexicographically recorded), SCr. dial. *kȓmelj* (haplology?).

sović (2013: 84). The reconstruction of the RCS form is uncertain, but **grъm-* would at least be a possibility (SDRJa, 1: 602; Berneker, 1: 360; Sławski SP, 8: 267).

Finally, we should note that part of the Slavic reflexes show an unclear initial *i-*, specifically East Slavic (Ru. *ívolga*, Bel. *ívalha*), as well as Bulg. *avlíga* (which is most easily explained by metathesis from older *(*)ivlъga*).[30] Although Po. *wilga* and Slk. *vlha* may also reflect **jьvьlga* with **jь-* > **Ø-*, SCr. *vȕga* and Sln. *vólga* 'oriole' cannot reflect such a preform (*pace* ÈSSJa, 8: 251–252), as initial **jь-* is always preserved in South Slavic (cf. Derksen 2003). This "prefix" cannot easily be explained: it is hardly, with Trubačev (1972: 19–20), an irregular reduction of the prefix **jьz-* (**z* would not be lost before **v*); neither is the parallel in Ru. *zubr* 'bison' beside *izjúbr'* 'Manchurian wapiti' (Kroonen 2012: 254; 2013: 571) watertight.[31]

To conclude, the Germanic, Baltic and Slavic words for 'oriole' strongly resemble one another. Unlike with the words for 'swan' and 'pigeon', the irregularities do not all have reliable parallels in other non-IE vocabulary, but it nevertheless seems fairly clear that we are dealing with a word of non-IE provenance: the words can hardly be separated from one another, yet do not reflect common proto-forms (cf. Machek 1968: 694; Matasović 2013: 87).

4 Similar bird names

The three bird names discussed in this study have a similar structure, with a disyllabic stem of the shape **CVCVNC*. To these we can also add the word for 'grouse', mentioned in the course of the discussion. The data may be presented as

30 This CS form is attested among a list of birds in the Hexameron of John the Exarch; however, it is not entirely certain how it is to be read. The actual manuscript has "косыжє · йсоіє · йвлъгы · йжлъны · щурыжє". Since the sequences ⟨йсоіє⟩ 'and jays' and ⟨йжлъны⟩ 'and woodpeckers' clearly both contain the word *i* 'and', it is natural to suspect that ⟨йвлъгы⟩ does, too (thus Miklosich 1865: 68, and thence the CS form *vlъga* usually encountered in the literature, e.g. ÈSSJa, 8: 251). Aitzetmüller (1958: 38), on the other hand, reads "*ivlъga*" here, citing a variant ⟨и ивлъга⟩ and the modern Bulgarian evidence. This theory is supported by Bulg. *ívolga* attested in Gerov (2: 171; not a Russian loan, but a dialectal form with **lъ* > /ol/ like others recorded in this source, e.g. *mólzǫ* 'to milk', *mórkovъ* 'carrot', Gerov, 3: 78, 82).

31 A couple of other words in Siberian dialects show an epenthetic /i-/ before /z-/, cf. Sib. *izábol'* 'indeed' (SRNG XII: 84) = dial. *zábyl'* (Anikin 2003: 201), Sib. *izúfr'* = *zuf'* 'a kind of woolen fabric', of Turkic origin (~ Turkish *sof* 'woolen fabric', Anikin 2000: 220). This is probably to be explained by assuming the interference of a substrate in which initial /z-/ is not permitted, cf. Khakas *izep* 'pocket' ← Ru. dial. *zep'* (itself of Turkic origin, cf. Anikin 2003: 216), Yakut dial. *ïhȳr* 'fat' ← Ru. *žir* (Anikin 2003: 199).

in Table 1, with quasi-IE reconstructions. Here, I have highlighted forms lacking nasals, as well as those with an apparent prefixal element:

Table 1: Irregular alternations in bird names.

	Baltic	Slavic	Germanic	Elsewhere
'pigeon'		*g$^{(h)}$olomb$^{(h)}$-	*gulub^{h}-r-	Lat. *kolomb^{h}-
				Eg. *k^{h}Vramp-
'swan'		***a-lband^{h}-**	***a-lb^{h}ed-**	
		*leb^{h}ed^{h}		
'oriole'	*u̯aHlang-	*u̯lg-	*u̯alk-	
		***i-u̯lg-**		
+ 'grouse'	*i̯erub^{h}-	*i̯eremb^{h}-		

A couple of other European bird names can be noted with a similar structure. The first has already been mentioned briefly on p. 5 — the Mediterranean word for 'swallow'. It would seem certainly ill-advised to separate Lat. *hirundō* from Gr. χελῑδών 'swallow' (cf. Curtius 1862: 167; Chantraine, 4: 1253), or from Alb. *dallëndyshe* 'swallow' (cf. Meyer 1891: 59),[32] although they cannot go back to a common proto-form. There is a disagreement both in terms of vocalism and between *-r-* and *-l-*.[33] The variant without the nasal in Gr. χελῑδών strongly recalls the forms lacking a nasal highlighted in the above table.

In Latin, we have seen *palumbēs* 'wood pigeon'. Already Fick (1871: 441) has noted the similarity with Lith. *balañdis*, Latv. *baluôdis* 'pigeon'. Could this be another non-IE form? The correlation between Latin *p-* and Baltic *b-* has already been noted in two lemmata also exhibiting a parallel *k ∞ *$g^{(h)}$, namely the words for 'swan' (Lith. *gulbė̃* ~ Kash. *kôłp*) and 'hornbeam' (Ru. *grab* ~ Lat. *carpinus*). Both words for 'pigeon' are again traditionally explained as derivatives of colour terms (cf. Schulze 1910: 799–800). For Baltic, Karaliūnas (1993: 110) assumes an original colour adjective **balandas* 'whitish' to the root of *bálti* 'whiten' (cf. Fraen-

32 Alb. *d-* regularly corresponds to Lat. *h-* (Alb. *dimër* ~ Lat. *hiems* 'winter'). It must be admitted that the alternative comparison with the Illyrian tribal name Ταυλάντιοι, reported by Hecateus of Miletus to have neighboured the Χελιδόνιοι, is tempting (Çabej 1987: 158).

33 A solution is attempted by André (1967: 93–94), who tentatively suggests dissimilation from an original *$ǵ^{h}$eni/und-*, but this still leaves many things unanswered, such as the loss of the second nasal in Greek and the disagreement in second-syllable vocalism.

kel: 31), but an adjective suffix *-anda-* is not paralleled (Skardžius 1941: 101). Lat. *palumbēs* is generally taken with *palleō* 'be pale' (Prellwitz 1897: 102; Walde 1906: 443), cf. Gr. πέλεια 'wild pigeon' : πελιός 'blue, dark'. This is semantically unproblematic,[34] but morphologically difficult. Lockwood (1990: 262–263) assumes an original **palēs* took over its suffix from *columba*. This theory might be more attractive were we to start from an original **palond-* corresponding, aside from the initial stop, to Lith. *balañdis*. While this is possible, the fact that a contamination must still be assumed makes the advantage over the traditional etymology less evident.

5 Conclusions

Although starting out with three bird names, the search for parallels has led us to examine several other words of probable non-Indo-European origin. The similar structure of the bird names (Po. *jarząbek*, *gołąb*, *łabędź,* Lith. *balañdis*, Lat. *hirundō*; Lith. *volungė̃*) is alone striking, and probably suggests that they have something to do with each other. In addition, the following recurring irregularities were identified:

– voicing alternations (*gołąb*, *volungė̃*, ?*balañdis*)
– "nasal loss" (*jarząbek*, *gołąb*, *łabędź*, *hirundō*, *volungė̃*)
– "a-prefixation" (*gołąb*, *łabędź*, cf. also Lat. *merula*)

The fact that these alternations, all of which have also been identified in other possible Indo-European substrate words, are found in the same group of words, may suggest that all of the words originated in related languages, or perhaps that they were mediated by related languages. The idea that these alternations should be reflections of developments within the source language (or mediation by another unattested language) is implied by the way in which the variants are distributed. Where a voicing alternation is found, Baltic always attests a voiced stop, while Italo-Celtic always shows voiceless ones. Where a "nasal" alternation is found, Germanic never features a nasal, while Balto-Slavic vacillates between both variants. "Prefixation with stem reduction" seems particularly common in Germanic, and particularly rare in Italic.

It is curious that two of the bird names appear to show two irreconcilable variants within Slavic. This is remarkable, since the split of the Slavic language

34 At least if we take the Greek meaning as original: cf. Ru. *sizják* 'feral pigeon' < *sízyj* 'dark bluish-grey' (Dal' 1882: 187); Oss. *æxsīnæg* 'wild pigeon' < *æxsīn* 'dark grey' (Abaev 1958: 220–221).

family was relatively late, datable to the first millennium CE. The existence of multiple variants in Slavic would appear to suggest that multiple non-IE languages were still present in Europe in the Common Era. This hypothesis would be supported by the very similar distribution of the two variants within Slavic. Both **lebedь* and **jьvьlga* are largely limited to East Slavic and Bulgarian, thus to the easternmost part of the Slavic territory. At face value, this would imply that the Slavs encountered different substrates in Central and Eastern Europe and absorbed these words independently into differentiated dialects.

While it is clear that Oštir's ideas had some merit, his results must be taken with a very large pinch of salt. His lack of criticism led to the inclusion of an enormous amount of probably irrelevant data. Therefore, while some of Oštir's methodological concepts can be of use to us in the modern era, a more austere approach is necessary. I would suggest the following modifications:

– Limit the number of alternations. Oštir (1930: 56–57) compares OHG *obaz* 'fruit' with *apful* 'apple' assuming $^{*}a \infty {}^{*}u$, $^{*}b \infty {}^{*}\dot{b}$ and $^{*}l \infty {}^{*}t$. In other words, none of the segments in the two words actually match, and it seems obvious that this methodology, taken far enough, would allow essentially any two words to be equated.
– Pay attention to geographical patterns. These may not always be found, but as in the case of $^{*}k \infty {}^{*}g^{(h)}$, a particular pattern in the reflexes certainly decreases the chance that the parallels emerged by chance or that the individual cases represent unrelated phenomena.

In conclusion, it is always possible to make a methodology more scientific by limiting the scope for arbitrary speculation and the possibility of false positives. While we will never have a method as robust as the comparative method for identifying non-Indo-European loanwords in the languages of Europe, that does not mean that we should abandon such a branch of research without attempting to refine our methods. Indeed, the fact that there is no generally acknowledged methodology for dealing with semantics equally does not mean that we should reject any proposed semantic shift. Rather, we should approach such aspects with caution, and strive to make our methods as scientific as possible.

References

Abaev, Vasilij. 1958. *Istoriko-ètimologičeskij slovar' osetinskogo jazyka* [Historical and etymological dictionary of Ossetic]. Vol. 1: *A–K'*. Moscow/Leningrad: Nauka.

Aitzetmüller, Rudolf. 1958. *Das Hexaemeron des Exarchen Johannes*. Vol. 1. Graz: Akademie.

Alessio, Giovanni. 1946. Un' oasi linguistica preindoeuropea nella regione baltica? [A pre-Indo-European linguistic oasis in the Baltic region?]. *Studi Etruschi* 19. 141–176.
ALEW = Wolfgang Hock (ed.). 2019. *Altlitauisches etymologisches Wörterbuch*. Version 1.1. Berlin: Humboldt-Universität.
Allen, James. 2020. *Ancient Egyptian Phonology*. Cambridge: Cambridge University Press.
Andersen, Henning. 1996. *Reconstructing prehistorical dialects: Initial vowels in Slavic and Baltic*. Berlin/New York: Mouton de Gruyter.
André, Jacques. 1967. *Les noms d'oiseaux en latin*. Paris: Klincksieck.
Anikin, Aleksandr. 2000. *Ètimologičeskij slovar' russkix dialektov Sibiri* [Etymological dictionary of the Russian dialects of Siberia]. 2nd edn. Moscow/Novosibirsk: Nauka.
Anikin, Aleksandr. 2003. *Ètimologičeskij slovar' zaimstvovanij v jazykax Sibiri* [Etymological dictionary of borrowings in Siberian languages]. Novosibirsk: Nauka.
Arumaa, Peeter. 1964. *Urslavische Grammatik*. Vol. 1: *Einleitung. Lautlehre*. Heidelberg: Winter.
Batisti, Roberto. 2021. Birds of a feather? Latin *columba* 'pigeon, dove' and Greek κόλυμβος 'grebe'. *Indogermanische Forschungen* 126. 205–228.
Beekes, Robert S. P. 1996. Ancient European Loanwords. *Historische Sprachforschung* 109. 215–236.
Beekes, Robert S. P. 2010. *Etymological Dictionary of Greek*. Leiden/Boston: Brill.
Beekes, Robert S. P. 2014. Stefan Norbruis (ed.), *Pre-Greek. Phonology, Morphology, Lexicon*. Leiden/Boston: Brill.
Berneker = Berneker, Erich. 1908–1914. *Slavisches etymologisches Wörterbuch*. 2 vols. Heidelberg: Winter.
Bezlaj = Bezlaj, France. 1977–2007. *Etimološki slovar slovenskega jezika* [Etymological dictionary of the Slovene language]. 5 vols. Ljubljana: Mladinka knjiga/Založba ZRC.
Bezlaj, France. 1977. Slovensko-baltijskie leksičeskie paralleli [Slovene-Baltic lexical parallels]. *Baltistica* 2 *Priedas*. 15–19.
Bopp, Franz. 1833. *Vergleichende Grammatik des Sanskrit, Zend, Griechischen, Lateinischen, Litthauischen, Gothischen und Deutschen*. Berlin: Akademie der Wissenschaften.
Brugmann, Karl. 1906. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*, 2nd edn. Vol. 2: *Vergleichende Laut-, Stammbildungs- und Flexionslehre*. Strasbourg: Trübner.
Bulaxovskij, Leonid. 1948. Obščeslavjanskie nazvanija ptic [Common Slavic bird names]. *Izvestija akademii nauk SSSR. Otdelenie literatury i jazyka* 7/2. 97–124.
Bulaxovskij, Leonid. 1968. Morfologičeskaja problematika russkix naimenovanij ptic [A morphological inquiry into Russian bird names]. *Voprosy jazykoznanija* 1968/4. 100–106.
Çabej, Eqrem. 1987. *Studime Etimologjike në fushë të Shqipes* [Etymological studies in the Albanian language]. Vol. 3. Tirana: Akademia e Shkencave.
Campbell, Alistair. 1959. *Old English Grammar*. London: Oxford University Press.
CDIAL = Turner, Ralph L. [1966] 1973. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.
Chantraine = Chantraine, Pierre. 1968–1980. *Dictionnaire étymologique de la langue grecque*. 4 vols. Paris: Klincksieck.
Charpentier, Jarl. 1930. Der Name des Roggens. *Arkiv för nordisk filologi* 46. 63–73.
Čop, Bojan 1973. Oštirs sprachwissenschaftliche Ideenwelt. *Linguistica* 13. 13–96.
Čop, Bojan 1988. Hundert Jahre eines bekannten slowenischen Sprachwissenschaftlers, Professor dr. Karel Oštir. *Linguistica* 29. 3–12.
Curtius, Georg. 1862. *Grundzüge der griechischen Etymologie*. Leipzig: Teubner.
Dal', Vladimir. 1882. *Tolkovyj slovar' živago velikorusskago jazyka* [Explanatory dictionary of the living Great Russian language]. 2nd edn. Vol. 4. Saint Petersburg/Moscow: Wolf.

DARJa = Ruben Avanesov and Sof'ja Bromlej (eds.). 1986. *Dialektologičeskij atlas russkogo jazyka* [Dialect atlas of the Russian language]. Vol. 1: *Fonetika* [Phonetics]. Moscow: Nauka.

Decaux, Etienne. 1966. La neutralisation n'est pas un changement structurel. *Bulletin de la Société de linguistique de Paris* 61. 57–81.

Demiraj, Bardhyl. 1997. *Albanische Etymologien. Untersuchungen zum albanischen Erbwortschatz*. Amsterdam/Atlanta: Rodopi.

Derksen, Rick. 1999. North European **KulP-* 'beak, snout'. *Baltistica* 34. 69–76.

Derksen, Rick. 2000. Old Icelandic *jarpi* 'hazel-grouse', *rjúpa* 'ptarmigan' and their Germanic and Balto-Slavic cognates. In Dirk Boutkan & Arend Quak (eds.), *Language Contact. Substratum, superstratum, adstratum in Germanic Languages*, 75–78. Amsterdam/Atlanta: Rodopi.

Derksen, Rick. 2003. Slavic **jь-*. In Jos Schaeken, Peter Houtzagers & Janneke Kalsbeek (eds.), *Dutch Contributions to the Thirteenth International Congress of Slavists*, *Ljubljana*: *Linguistics*, 97–105. Amsterdam/Atlanta: Rodopi.

Derksen, Rick. 2008. *Etymological Dictionary of the Slavic Inherited Lexicon*. Leiden/Boston: Brill.

Derksen, Rick. 2015. *Etymological Dictionary of the Baltic Inherited Lexicon*. Leiden/Boston: Brill.

Endzelin, Jan. 1914. Miszellen. *Indogermanische Forschungen* 33. 119–127.

Endzelin, Jan. 1923. *Lettische Grammatik*. Heidelberg: Winter.

Endzelin, Jan. 1924. Germanisch-baltische Miszellen. *Zeitschrift für vergleichende Sprachforschung* 52. 110–128.

ERHJ = Ranko Matasović, Dubravka Ivšić Majić and Tijmen Pronk (eds.). 2016–2021. *Etimološki rječnik hrvatskoga jezika* [Etymological dictionary of the Croatian language]. 2 vols. Zagreb: Institut za hrvatski jezik i jezikoslovlje.

Erjavec, Fran. 1879. Iz pótne torbe [From the travel bag]. *Letopis Matice Slovenske* 1879. 118–223.

Erman, Adolf & Hermann Grapow. 1931. *Wörterbuch der aegyptischen Sprache*. Vol. 5. Leipzig: Hinrichs.

Ernout, Alfred & Antoine Meillet (1951). *Dictionnaire étymologique de la langue latine*. 3rd edn. Paris: Klincksieck.

ÈSSJa = Oleg Trubačev and Andrej Žuravlev (eds.). 1971–. *Ètimologičeskij slovar' slavjanskix jazykov* [Etymological dictionary of the Slavic languages]. 41 vols. Moscow: Nauka.

EWAhd = Rosemarie Lühr et al. (eds.), *Etymologisches Wörterbuch des Althochdeutschen*. 7 vols. Göttingen/Zürich: Vandenhoeck & Ruprecht.

Fick, August. 1871. *Vergleichendes Wörterbuch der indogermanischen Sprachen*. 2nd edn. Göttingen: Vandenhoeck & Ruprecht.

Fraenkel = Fraenkel, Ernst. 1955–1966. *Litauisches etymologisches Worterbuch*. Heidelberg–Göttingen: Winter, Vandenhoeck & Ruprecht.

Furnée, Edzard J. 1972. *Die wichtigsten konsonantischen Erscheinungen des Vorgriechischen*. The Hague/Paris: Mouton.

Gerov = Gerov, Naiden. 1895–1904. *Rěčnik na Blăgarskyj jazyk* [Dictionary of the Bulgarian language]. 5 vols. Plovdiv: Săglasie.

Gippert, Jost. 2017. Armeno-Albanica II: Exchanging doves. In Bjarne S. S. Hansen, Adam Hyllested, Anders R. Jørgensen & Guus Kroonen (eds.), *Usque ad radices*, 179–192. København: Museum Tusculanum.

GPC = Gareth A. Bevan and Patrick J. Donovan (eds.). 1950–2002. *Geiriadur Prifysgol Cymru: A Dictionary of the Welsh Language*. 4 vols. Cardiff: Gwasg Prifysgol Cymru.

Haas, Otto. 1960. *Das frühitalische Element*: *Versuch über die Sprache der ersten Indogermanen Italiens*. Vienna: Notring.

Herne, Gunnar. 1954. *Die slavischen Farbenbennenungen. Eine semasiologisch-etymologische Untersuchung*. Uppsala: Almqvist & Wiksell.

Hintze, Almut. 1994. *Der Zamyād-Yašt. Edition, Übersetzung, Kommentar*. Wiesbaden: Reichert.
Hirt, Hermann. 1927. *Indogermanische Grammatik*. Vol. 1. Heidelberg: Winter.
Hogg, Richard M. 1992. *A Grammar of Old English*. Vol. 1: *Phonology*. Oxford: Blackwell.
Holzer, Georg. 1989. *Entlehnungen aus einer bisher unbekannten indogermanischen Sprache im Urslavischen und Urbaltischen*. Wien: Akademie der Wissenschaften.
Hommel, Fritz. 1884. Die sumero-akkadische Sprache und ihre Verwandtschaftsverhältnisse. *Zeitschrift für Keilschriftforschung* 1. 161–178, 195–221, 323–342.
Hommel, Fritz. 1904. *Geschichte des alten Morgenlandes*. 3rd edn. Leipzig: Göschen.
Hoops, Johannes. 1911. *Reallexikon der Germanischen Altertumskunde*. Vol. 1. Strasbourg: Trübner.
Hübschmid, Johannes. 1950. *Sardische Studien*. Bern: Francke.
Ivanov, Vjačeslav. 2002. Rannie koptskie zaimstvovanija v slavjanskom [Early Coptic loanwords in Slavic]. In Tat'jana Nikolaeva (ed.), *Slavjanskaja jazykovaja i ètnojazykovaja sistemy v kontakte s neslavjanskim okruženiem* [Slavic linguistic and ethnolinguistic systems in contact with non-Slavic surroundings], 55–60. Moscow: Jazyki slavjanskoj kul'tury.
Juška = Juškevič, Anton. 1898–1904. *Litovskij slovar' s" tolkovaniem" slov" na russkom" i pol'skom" jazykax"* [Lithuanian dictionary with explanations in Russian and Polish]. 3 issues. St. Petersburg: Imperial Academy of Sciences.
Karaliūnas, Simas. 1993. Some remarks on the names for pigeon in Lithuanian. *Linguistica Baltica* 2, 109–113.
Kleyner, Svetlana D. 2005. From Yellow to Blue – or Not? In Dafydd Johnston, Elena Parina & Maxim Fomin (eds.), '*Yn llawen iawn, yn llawn iaith*', 47–52. Aberystwyth: Centre for Advanced Welsh and Celtic Studies.
Kretschmer, Paul. 1896. *Einleitung in die Geschichte der griechischen Sprache*. Göttingen: Vandenhoeck & Ruprecht.
Kroonen, Guus. 2012. Non-Indo-European root nouns in Germanic: evidence in support of the Agricultural Substrate Hypothesis. In Riho Grünthal & Petri Kallio (eds.), *A linguistic map of prehistoric Northern Europe*, 239–260. Helsinki: Suomalais-Ugrilainen Seura.
Kroonen, Guus. 2013. *Etymological Dictionary of Proto-Germanic*. Leiden/Boston: Brill.
Kuiper, Frans B. J. 1956. The Etymology of ἄνθρωπος. In Heinz Kronasser (ed.), Μνήμης χάριν: *Gedenkschrift Paul Kretschmer*. Vienna: Sprachgesellschaft.
LÄGLOS = Andries D. Kylstra, Sirkka-Liisa Hahmo, Tette Hofstra and Osmo Nikkilä (eds.). 1991–2012. *Lexikon der Alteren Germanischen Lehnworter in den Ostseefinnischen Sprachen*. 4 vols. Amsterdam: Rodopi.
Leumann, Manu. 1977. *Lateinische Grammatik*. Vol. 1: *Lateinische Laut- und Formenlehre*. Munich: Beck.
LKŽ = Juozas Balčikonis and Vytautas Vitkauskas (eds.). 1941–2002. *Lietuvių kalbos žodynas* [Dictionary oft he Lithuanian language]. 20 vols. Vilnius: Mokslas, Lietuvių kalbos instituto leidykla.
Lockwood, William B. 1990. Latin *columba*, *palumbēs*, Greek *κόλυμβος*. *Historische Sprachforschung* 103. 261–263.
Loewenthal, Wilhelm 1901. *Die slavischen Farbenbezeichnungen*. Leipzig: Preis.
Machek, Václav 1950a. Einige slavische Vogelnamen. *Zeitschrift für Slavische Philologie* 20. 29–51.
Machek, Václav 1950b. Quelques noms slaves de plantes. *Lingua Posnaniensis* 2. 145–161.
Machek, Václav 1951. Quelques mots slavo-germaniques. *Slavia* 20. 200–218.
Machek, Václav 1968. *Etymologický slovník jazyka českého* [Etymological dictionary of Czech]. Prague: Academia.
Marr, Nikolaj. 1928. *Jafetičeskaja teorija* [The Japhetic theory]. Baku: Azerbaidžanskoe gosudarstvennoe izdatel'stvo.
Matasović, Ranko. 2013. Substratum words in Balto-Slavic. *Filologija* 60. 75–102.

Mayrhofer = Mayrhofer, Manfred. 1992–2001. *Etymologisches Wörterbuch des Altindoarischen*. 3 vols. Heidelberg: Winter.
Meillet, Antoine. 1922. [Review] K. Oštir. – Beiträge zur alarodischen Sprachwissenschaft. *Bulletin de la Société de linguistique* 22. 128–130.
Meillet, Antoine & André Vaillant. 1933. Le slave commun **jarębь* et le suffixe en *-b-*. *Revue des études slaves* 13. 101–102
Meyer, Gustav. 1891. *Etymologisches Wörterbuch der albanesischen Sprache*, Strasbourg: Trübner.
Mikkola, Jooseppi. 1899. Baltische etymologien II. *Beiträge zur Kunde der indogermanischen Sprache* 25. 73–76.
Miklosich, Franz von. 1865. *Lexicon palaeoslovenico-graeco-latinum*. Vienna: Braumüller.
Miklosich, Franz von. 1886. *Etymologisches Wörterbuch der Slavischen Sprachen*. Vienna: Braumüller.
Mühlenbach & Endzelīns = Karl Mühlenbach & Jan Endzelin (eds.), *Lettisch-deutsches Wörterbuch*. 4 vols. (1923–1925). Rīga: Latvju grāmata.
Niedermann, Max. 1918. *Essais d'étymologie et de critique verbale latines*. Neuchâtel: Attinger Frères.
Nikolaev, Sergej. 2020. «*Slovo o polku Igoreve*»: *Rekonstrukcija stixotvornogo teksta* ["The Tale of Igor's Campaign": A reconstruction of a poetic text]. Moscow/Saint Petersburg: Nestor-Istorija.
Noreen, Adolf. 1892. *Altisländische und altnorwegische Grammatik*. 2nd edn. Halle: Niemeyer.
Orel, Vladimir. 2003. *A Handbook of Germanic Etymology*. Leiden/Boston: Brill.
Osthoff, Hermann. 1898. Griechische und lateinische Wortdeutungen III. *Indogermanische Forschungen* 8, 1–68.
Osthoff, Hermann. 1901. *Etymologische Parerga* I. Leipzig: Hirzel.
Oštir, Karl. 1912. Etymologische und grammatische Versuche. *Wörter und Sachen* 4, 211–218.
Oštir, Karl. 1921. *Beiträge zur alarodischen Sprachwissenschaft*. Vienna/Leipzig: Beyer.
Oštir, Karel. 1923. Alarodica I. *Razprave Znanstvenega društva v Ljubljani* 1, 273–310.
Oštir, Karel. 1925. Baltoslovanska metatonija [Balto-Slavic metatony]. *Razprave Znanstvenega društva v Ljubljani* 2, 55–120.
Oštir, Karel. 1929. Japodi [Iapydes]. *Etnolog* 3, 87–113.
Oštir, Karl. 1930. *Drei vorslavisch-etruskische Vogelnamen*. Ljubljana: Znanstveno društvo.
Oštir, Karel. 1930–1931. Predslovansko **sębьrъ* „zadrugar" [Pre-Slavic **sębьrъ* 'cooperative worker']. *Etnolog* 4, 1–29.
Otrębski, Jan. 1939. *Indogermanische Forschungen*. Vilnius: Towarzystwo przyjaciół nauk.
Persson, Per. 1912. *Beiträge zur indogermanischen Wortforschung*. Uppsala: Akademiska bokhandeln.
Pabrėža, Jurgis A. 1834. [Manuscript]. *Waardaa tayslyynee Augimiu atsyrądąntiu Augmyniićźioo źemaytyyniee* [The systematic names of plants found in Žemaitian botany]. Kretinga. https://www.epaveldas.lt/preview?id=C10000689960 (accessed 9 November 2023).
Peust, Carsten. 1999. *Egyptian Phonology*: *An Introduction to the Phonology of a Dead Language*. Göttingen: Peust & Gutschmidt.
Peyrot, Michaël. 2019. Tocharian Agricultural Terminology: Between Inheritance and Language Contact. In Guus Kroonen, James P. Mallory & Bernard Comrie (eds.), *Talking Neolithic* [= *Journal of Indo-European Studies Monograph* 65], 242–277. Washington DC: Institute for the Study of Man.
Pinault, Georges-Jean. 2008. *Chrestomathie tokharienne*. Leuven–Paris: Peeters.
Pleteršnik = Pleteršnik, Maks. 1894–1895. *Slovensko-nemški slovar* [Slovenian-German dictionary]. 2 vols. Ljubljana: Knezoškofijstvo.
Pogatscher, Alois. 1898. Altenglisch *ƀr* aus *mr*. In Jakob Schipper (ed.), *Festschrift zum VIII. allgemeinen deutschen Neuphilologentage in Wien Pfingsten 1898*, 97–107. Vienna/Leipzig: Braumüller,
Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern–München: Francke.

Polomé, Edgar C. 1972. Germanic and the other Indo-European Languages. In Frans van Coetsem & Herbert L. Kufner (eds.), *Toward a grammar of Proto-Germanic*, 43–69. Tubingen: Niemeyer.
Pott, August Friedrich. 1861. *Etymologische Forschungen auf dem Gebiete der indo-germanischen Sprachen*. *Zweiten Theiles erste Abtheilung*. 2nd edn. Lemgo/Detmold: Meyer.
Prellwitz, Walther. 1897. Studien zur indogermanischen etymologie und wortbildung [*sic*.]. *Beiträge zur Kunde der indogermanischen Sprachen* 22, 76–114.
Pronk, Tijmen and Pronk-Tiethoff, Saskia. 2018. Balto-Slavic Agricultural Terminology. In Guus Kroonen, James P. Mallory & Bernard Comrie (eds.), *Talking Neolithic*, 278–314 (= *Journal of Indo-European Studies Monograph* 65). Washington DC: Institute for the Study of Man.
RJA = Đuro Daničić, Pero Budmani et al. (eds.). 1880–1976. *Rječnik hrvatskoga ili srpskoga jezika* [Dictionary of the Croatian or Serbian language]. 23 vols. Zagreb: Jugoslavenska akademija znanosti i umjetnosti.
Schrader, Otto. 1901. *Reallexikon der indogermanischen Altertumskunde*. Strassburg: Trübner.
Schrijver, Peter. 1997. Animal, vegetable and mineral: some Western European substratum words. In Alexander Lubotsky (ed.), *Sound Law and Analogy*, 293–316. Amsterdam/Atlanta: Rodopi.
Schrijver, Peter. 2001. Lost Languages in Northern Europe. In Christian Carpelan, Asko Parpola & Petteri Koskikallio (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations*, 417–425. Helsinki: Suomalais-Ugrilainen Seura.
Schrijver, Peter. 2018. Talking Neolithic. The Case for Hatto-Minoan and its Relationship to Sumerian. In Guus Kroonen, James P. Mallory & Bernard Comrie (eds.), *Talking Neolithic*, 336–374 (= *Journal of Indo-European Studies Monograph* 65). Washington DC: Institute for the Study of Man.
Schuchardt, Hugo. 1922. Heimisches und fremdes Sprachgut. *Revista Internacional de los Estudios Vascos* 13. 70–82.
Schulze, Wilhelm. 1910. Etymologisches. *Sitzungsberichte der königlich preussischen Akademie der Wissenschaften* 1910. 787–808.
Schumacher, Stefan. 2013. Historische Phonologie. In Stefan Schumacher & Joachim Matzinger. *Die Verben des Altalabanischen*, 205–276. Wiesbaden: Harrassowitz.
SDRJa = Izmail Sreznevskij. 1893–1912. *Materialy dlja slovarja drevne-russkago jazyka po pis'mennym" pamjatnikam"* [Materials for a dictionary of the Old Russian language based on written documents]. St. Petersburg: Imperatorskaja akademija nauk.
Skardžius, Pranas. 1941. *Lietuvių kalbos žodžių daryba* [Lithuanian word formation]. Vilnius: Lietuvos mokslų akademija.
Skeat, Walter W. 1882. *An Etymological Dictionary of the English Language*. Oxford: Clarendon Press.
Sławski, Franciszek. 1960. Dubrownicka nazwa łabędzia [A Dubrovnik term for the swan]. *Rocznik sławistyczny* 21. 37–40.
Sławski SEJP = Franciszek Sławski (ed.). 1952–1982. *Słownik etymologiczny języka polskiego* [Etymological dictionary of the Polish language]. 5 vols. Kraków: Towarzystwo miłosników języka polskiego.
Sławski SP = Franciszek Sławski (ed.). 1974–2001. *Słownik prasłowiański* [Proto-Slavic dictionary]. 8 vols. Wrocław: Ossolineum.
Smoczyński, Wojciech 2018. *Lithuanian Etymological Dictionary*. Berlin: Peter Lang.
Sobolevskij, Aleksěj. 1914. Něskol'ko zamětok" po slavjanskim" vocalizmu i leksikě [A few notes about Slavic vocalism and lexicon]. *Russkij filologičeskij věstnik"* 64. 178–179.
Specht, Franz. 1947. *Der Ursprung der Indogermanischen Deklination*. Göttingen: Vandenhoeck & Ruprecht.
SRJa 11–17 = Stepan Barxudarov et al. (ed.). 1975–. *Slovar' russkogo jazyka XI–XVII vv*. [Dictionary of the Russian language, 11th–17th c.]. 31 vols. Moscow/St. Peterburg: Nauka/Nestor-Istorija.

SRNG = Fedot Filin, Fedor Sorokoletov and Sergej Myznikov (eds.). 1966–. *Slovar' russkix narodnyx govorov* [Dictionary of Russian dialects]. Leningrad/St. Petersburg: Nauka.
Stokes, Whitley. 1894. *Urkeltischer Sprachschatz* [= *Vergleichendes Wörterbuch der indogermanischen Sprachen*. 4th edn. Vol. 2.]. Göttingen: Vandenhoeck & Ruprecht.
Suolahti, Hugo. 1909. *Die deutschen Vogelnamen*. Strasbourg: Trübner.
Szemerényi, Oswald. 1967. Slavjanskaja ètimologija na indoevropejskom fone [Slavic etymology in an Indo-European context]. *Voprosy jazykoznanija* 1967(4). 3–24.
TLL = *Thesaurus linguae latinae*. 1900–. Leipzig: Teubner/Berlin: De Gruyter.
Trubačev, Oleg 1972. Zametki po ètimologii i sravnitel'noj grammatike [Notes on etymology and comparative grammar]. *Ètimologija 1970*. 3–20.
Uhlenbeck, Christianus C. 1898–1899. *Kurzgefasstes etymologisches Wörterbuch der altindischen Sprache*. Amsterdam: Müller.
Uhlenbeck, Christianus C. 1923. [Review] K. Oštir, Beiträge zur alarodischen Sprachwissenschaft. *Revue internationale des études basques* 14. 365.
Vaillant, André. 1929. Deux emprunts slaves au roman de Dalmatie. *Revue des études slaves* 9. 270–272.
Vaillant, André. 1950. *Grammaire comparée des langues slaves*. Vol. 1: *Phonétique*. Lyon: IAC.
de Vries, Jan W. 1962. *Altnordisches etymologisches Wörterbuch*. Leiden: Brill.
Vycichl, Werner 1983. *Dictionnaire étymologique de la langue copte*. Leuven: Peeters.
Vycichl, Werner 1990. *La vocalisation de la langue égyptienne*. Vol. 1: *La phonétique*. Cairo: Insitut Français d'archéologie orientale.
Walde, Alois 1906. *Lateinisches etymologisches Wörterbuch*. Heidelberg: Winter.
Walde and Hofmann = Johann B. Hofmann (ed.). 1938–1954. Walde, Alois. *Lateinisches etymologisches Wörterbuch*. 3rd edn. 2 vols. Heidelberg: Winter.
Worrel, William H. 1934. *Coptic Sounds*. Ann Arbor: University of Michigan.
Zaliznjak, Andrej 2019. *Drevnerusskoe udarenie: Obščie svedenija i slovar'* [Old Russian accentuation: general information and dictionary]. 3rd edn. Moscow: Jazyki slavjanskix kul'tur.
Zinkevičius, Zigmas. 1966. *Lietuvių dialektologija* [Lithuanian dialectology]. Vilnius: Mintis.

Ranko Matasović

3 Proto-Slavic forest tree names: Substratum or Proto-Indo-European origin?

1 Introduction

There are three groups of Slavic forest tree names: the first group comprises those words that have a viable Indo-European etymology, in the sense that they are regularly derivable from PIE roots attested in both European and Asian branches of the family: **asenъ* 'ash, *Fraxinus*', **berstъ* 'elm, *Ulmus*', **berza* 'birch, *Betula*', **dǫbъ* 'oak, *Quercus*', **edla* 'silver fir, *Abies alba*', **glogъ* 'hawthorn, *Cornus sanguinea*', **xъbъtъ*, **xъbzъ* 'elder, *Sambucus ebulus*', **iva* 'willow, *Salix*', **lipa* 'lime, linden, *Tilia*', **vьrba* 'willow, *Salix*' (**buky* 'beech, *Fagus*' and **avorъ* 'sycamore maple, *Acer pseudoplatanus*' also belong to this group, in the sense that these words are originally Indo-European, but in Slavic they are known Germanic loanwords).

The second group comprises tree names that have possible cognates only in IE languages spoken in Europe (sometimes including Armenian), cannot be derived from any known PIE root, and the reconstruction of the common prototype is doubtful, often because sound correspondences between cognates are irregular: **grabrъ* 'hornbeam, *Carpinus betulus*', **jьlmъ* 'elm, *Ulmus*', **klenъ*, **klьnъ* 'maple, *Acer*', **orkyta* 'willow, *Salix*', **olьxa*, **elьša* 'alder, *Alnus*', **osa*, **asika* 'aspen, *Populus tremula*', **tisъ*, **tisa* 'yew, *Taxus baccata*', **topolь* 'poplar, *Populus*' (**avorъ* 'sycamore maple, *Acer pseudoplatanus*' may also belong to this group, but it might also be a Germanic loanword in Proto-Slavic).

Finally, the third group comprises those Proto-Slavic forest tree names for which no cognates in other branches have been found: **berka* 'service tree, *Sorbus torminalis*', **buzъ*, **bъzъ* 'elder, *Sambucus nigra*', **smerka* 'spruce, *Abies picea*', and **sosna* 'Scots pine, *Pinus sylvestris*'. Now, although "not having a viable PIE etymology" should not be interpreted as meaning the same thing as "having a substratum origin", one has to count with the possibility that some Proto-

Acknowledgements: I would like to thank professor Ante Aikio for sharing with me some of his data on Uralic tree names, and also to express my gratitude to my co-authors who worked with me on the *Etymological Dictionary of Croatian* (ERHJ I and II) during the past ten years – Tijmen Pronk, Dubravka Ivšić Majić and Dunja Brozović Rončević. Several of the etymologies discussed in this study draw on the materials collected during the work on that dictionary.

https://doi.org/10.1515/9783111337920-003

Slavic forest tree names belonging to the latter two groups have been borrowed from a pre-Indo-European language (or several languages) in Europe. In the present study this possibility will be investigated, and it will be shown that a few Slavic forest tree names indeed show some features indicative of substratum origin, such as a-vocalism of the root (which cannot be reduced to PIE *h_2e), non-IE Ablaut patterns (e.g. *i, *u alternating with *e, *a, and *o), alternations of voiced and voiceless stops, root shapes that contradict PIE phonotactic rules, etc. (Kuiper 1995, Iversen and Kroonen 2017, Matasović 2020). Moreover, a few new etymologies of Proto-Slavic tree names will be proposed.

2 Proto-Slavic forest tree names probably inherited from PIE

The following Proto-Slavic forest tree names have a probable PIE etymology:

1. *$asenь$, *$esenь$, *$asenъ$ 'ash, *Fraxinus*' (Sln. *jésen* ⟨Gen. *jesẹ́na*⟩, *jásen* ⟨Gen. *jasẹ́na*⟩, Bulg. *ósen*, Ru. *jásen'*, Po. *jesion*, dial. *jasień*, Cz. *jasaň*, *jasan*, Cz. dial. *jeseň*, ERHJ I: 397, ESSJa I: 79f.). The forms starting with *es- may have been influenced by the short vowel in the word for 'aspen' (PSl. *$osina$, *$esina$, no. 19 below), or may correspond to the forms with the short vowel found in the word for 'ash' in other IE languages. The Slavic words for 'ash' are cognate with Lith. *úosis*, Latv. *uôsis*, OPr. *woasis* < *$ōs$- (? < *h_3eHs-), Lat. *ornus* 'kind of ash tree' < *$osVno$-, MIr. *onn* < PCelt. *$osno$-, W (singulative) *onnen*, pl. *onn* < *$osnā$, OHG *asc* < *ask- < *osk-, Arm. *hacᶜi* 'ash-tree', Alb. *ah* 'beech' < *$osko$-, perhaps also Gr. ὀξύα 'beech' (from *$osko$- by metathesis?). It is in principle possible to reconstruct a PIE root *h_3eHs-/*h_3Hs- with different suffixes (*$-(V)n$-, *$-i$-, *$-k$-), but a possibility also remains that this word was borrowed into European languages from some pre-IE substratum. The distribution of ash was more limited before the Subatlantic period (from 500 BCE onward), but, on the whole, several *Fraxinus* species are found widely across Eurasia, as well as in North America and North Africa.[1] Ashwood is tough and resilient, and traditionally used to make bows, tool handles, and other utensils.

1 Unless otherwise stated, the data about the present-day and earlier distributions of forest tree species discussed in this study are taken from EAFTS. This source can be accessed on the internet fre of charge and contains useful maps showing the distribution of various tree species.

2. **bêrstъ* 'elm, *Ulmus*' (a. p. c) (Sln. *brẹ̑st*, Ru. *bérest*, Cz. *bĕst*, Po. *brzost*, ERHJ I: 85, ESSJa: 1: 189f.); probably derived from the PIE root **b^{h}erHǵ-* 'bright', since the bark of the elm is bright (cf. OHG *beraht* 'bright', W *berth* 'beautiful', EDPC 63, IEW 139). Elm trees are common everywhere in Europe except on its westernmost fringes.

3. **bérza* 'birch, *Betula alba*' (a. p. a) (CS *brěza*, Sln. *brẹ́za*, Ru. *berëza*, Cz. *bříza*, Po. *brzoza*, ERHJ I: 83, ESSJa I: 201ff.) < PIE **b^{h}erHǵ-* 'birch' (Lith. *béržas*, OHG *birihha*, Skt. *bhūrjá-* 'a kind of birch'). PIE probably had a root-noun, Nom. sg. **b^{h}erHǵs*, Gen. sg. **b^{h}rHǵos*. The PIE root is probably the same as in **bêrstъ* 'elm, *Ulmus*' (see above). Birch has had a wide distribution in Europe during the last three millennia, being native in the vast area from the Atlantic to the Urals.

4. **dǫbъ* 'oak, *Quercus robur*' (a. p. c) (OCS *dǫbъ*, Sln. *dọ̑b*, Ru. *dub*, Cz. *dub*, Po. *dąb*, ERHJ I: 204, ESSJa V: 95ff.). The comparison of PSl. **dǫbràva*, **dǫbròva* 'marshy woodland' (Sln. *dobrâva*, Ru. *dubróva*, Po. *dąbrowa*) with Lith. *dumbravà* 'puddle, muddy part of a meadow or road', Latv. *dubra* 'marshland' shows that the root is probably PIE **d^{h}ewb^h-* 'deep' (OCS *dъno* 'bottom', Go. *diups* 'deep', Lith. *dubùs*); PSl. **dǫbъ* is derivable from **d^{h}ub^h-no-* by metathesis (**d^{h}umb^{h}o-* < **d^{h}unb^{h}o-*). Much less likely is a derivation from the root **demh$_2$-* 'build' (Go. *ga-timan*, Lat. *domus* 'house', cf. Snoj 2003: 114f.), since **-b^h-* > **-b-* is not a suffix in Slavic.[2] PSl. **dǫbъ* probably originally denoted the lowland sub-species of oak (such as *Quercus robur* in contradistinction to *Quercus petraea* 'sessile oak'), which are often found in marshlands and on wet soils. The oak was the dominant tree species in Western Eurasia during the warmer Atlantic Period (5500–3000 BC), but during the cooler Subboreal period (3000–500 BC) it was partially replaced by the spread of beech (and fir at higher altitudes).[3] Oak wood is extremely hard and durable and has many uses: as timber, for making furniture, in shipbuilding, etc.

5. **èdlь*, **edlà* ⟨a. p. *c*⟩, **èdlъka* 'silver fir, *Abies alba*' (Sln. *jẹ̑l*, *jẹ̑la*, *jẹ́lka*, Ru. *el'*, *ëlka* 'spruce',[4] Po. *jodła*, *jedla*, Cz. *jedl*, ERHJ I: 404f., ESSJa VI: 14f.). Cognates of this word are found in Baltic: Lith. *ẽglė* 'spruce, fir', Latv. *egle*, OPr. *addle*; the compar-

2 However, in principle it would be possible to assume that the stem **dǫb-* was abstracted from derivatives in **-ro-* such as **dǫbràva* 'woody marshland', where **-b-* developed regularly between **-m-* and **-r-*. The semantic connection between 'to build' and 'oak' would lie in the fact that oak is excellent as timber.

3 Ante Aikio (p. c.) draws my attention to the fact that there is a Western Uralic (Finnic-Mordvin-Mari) word for 'oak' that can be reconstructed as **tammi-*, but I do not see it as similar enough to PSl. **dǫbъ* to posit a common source.

4 In Russian, the inherited word for 'fir' was replaced by a German loanword *píxta* (from G *Fichte*).

ison with Lat. *ebulus*, *ebulum* 'dwarf elder, *Sambucus ebulus*' allows us to reconstruct PIE *h_1ed^h-l- (EDL 185).[5] Note, however, that elder is a very different plant from fir and spruce, so we have to count with the possibility that the comparison is due to chance similarity (in which case the Balto-Slavic words are isolated within Indo-European). Gaul. *odocos* 'elder' may have been borrowed as OHG *attuh*, *attah* 'dwarf-elder, danewort' (EDL 185, DELG 238), but it is also possible that both the Germanic and the Gaulish word (which is attested in a late and possibly corrupt source, Marcellus of Bordeaux, *Med. Lib.* 7.13) were actually borrowed, via Lat. *acte* 'elder', from Gr. ἀκτέα 'elder-tree, *Sambucus nigra*' (of unknown origin, EDG 58). Fir was generally absent from Central and East Europe in the 4th millennium BC, but it spread during the Subboreal period (3000–500 BC) to the east (Cooper 2010: 42).

6. *glogъ (a. p. b) 'hawthorn, dogwood, *Cornus sanguinea*' (Cr. *glȍg*, Sln. *glȍg*, Ru. *glog*, Po. *głóg*, ERHJ I: 275, ESSJa VI: 166f.). Usually connected with Gr. γλωχίς 'projecting point, end of a yoke-strap', γλῶσσα 'tongue', since sticks made of hawthorn are very sharp (IEW 402). Note that the zero-grade *glg^h- (without a laryngeal) must be posited because of Gr. Ion. γλάσσα 'tongue' (EDG 278). This points to PIE *$glōg^h$-/*$glog^h$- and the Slavic forms can be derived from the root with the o-grade. Another possibility is to derive PSl. *glogъ from *dlogъ < *$dlog^h$- 'split' (OIr. *as-dloing*, ON *telgja*), Lith. *dal̃gis* 'scythe', with "Schwebeablaut" (*$delg^h$-/*$dlog^h$-). This is not problematic, as the new full grade *$delg^h$- (> ON *telgja*) and *$dolg^h$- (< Lith. *dal̃gis*) may have been formed on the basis of the zero-grade *dlg^h- which yielded *tulg- in Germanic and *tilg- or *tulg- in Balto-Slavic (the zero-grade is attested in Lith. *dilgùs* 'sharp', and perhaps in G dial. *Zungen-zolch* 'tongue', if the original meaning was 'tongue-tip', EDPG 525). The same root is attested (with metathesis) in OCS *glodati* (< *dlogati) 'gnaw'. Note that there is no reason why Gr. γλωχίς and γλῶσσα could not be from this root (this is also claimed in EDPG 525), with the probably regular assimilation *dl- > gl- as in γλυκύς 'sweet' < *dlukus (cf. Lat. *dulcis*). *Cornus sanguinea* has a wide distribution in Europe – it is common everywhere except in Southern Spain and Northern Scandinavia. Its wood is traditionally used to make sharp sticks, skewers and arrows.

7. *xъbъtъ, *xъbъzъ 'elder, *Sambucus ebulus* (Cr. *àbad*, Sln. *hebed*, Sln. dial. *həbât*, Ru. dial. *xobóta*, OPo. *chebd*, ERHJ I: 1, ESSJa I: VIII: 136f.). This word is probably derived from the root *$skewb^h$- 'quick' (Lith. *skubùs*, Latv. *skubrs*, perhaps also

5 Derivation from a PIE root *$(h_1)ed^h$- 'to stick' has been proposed (IEW 289f., Snoj 2003: 238), but reflexes of this alleged root would otherwise be attested only in Baltic (Lith. *adýti* 'to embroider', Latv. *adît* 'to weave', Lith. *ãdata* 'needle').

Go. *af-skiuban* 'push away' EDBIL 410); the semantic motivation lies in the fact that elder grows very quickly. The forms derivable from **xъbzъ* (Ru. dial. *xobz*, *xabz*, Cz. dial. *chebz*, *chabzda*) were analogically influenced by **bъzъ* 'elder' (on which see below). *Sambucus ebulus* has a very wide distribution in Europe (from the Atlantic coast to the Urals).

8. **íva* (a. p. a) 'willow, *Salix caprea*' (Sln. *íva*, Ru. *íva*, Po. *iwa*, Cz. *jíva*, ERHJ I: 372, ESSJa VIII: 248f.) < PIE **Hey(H)-weh$_2$* 'some tree with berries' (Lith. *ievà* 'bird-cherry', Latv. *iẽva* 'bird-cherry', OHG *īwa*, G *Eibe* 'yew', Gr. ὄα, ὄη, οἴη 'elderberry-tree', Lat. *ūva* 'bunch of grapes', Arm. *aygi* 'vine', OIr. *eó* 'yew', Hitt. ^{GIŠ}*eyan-* 'a kind of evergreen tree', IEW 297). In Slavic, the meaning changed from 'some tree with berries' to 'willow'. Note, however, that willows do not actually have berries – rather, they have yellow catkins. Of the two types of willow in Europe, **iva* probably originally denoted *Salix caprea* rather than *Salix alba*, whose catkins are less berry-like. Various sub-species of willow (*Salix*) are common everywhere in Europe (with the partial exception of Scandinavia).

9. **lípa* (a. p. a) 'lime, linden, *Tilia*' (Cr. *lȉpa*, Sln. *lípa*, Ru. *lípa*, Cz. *lípa*, Po. *lipa*, ERHJ I: 556f., ESSJa XV: 114ff.); Baltic cognates include Lith. *líepa* 'lime', OPr. *lipe* 'id.' and Latv. *liẽpa* 'id.'. These words are probably derived from the root **leyp-* 'glue' (OCS *lěpiti*, Cr. *lijépiti*, etc., cf. also Skt. *lepayati* 'smears', Gr. λίπος 'fat', IEW 670f.), and the acute root points to the Balto-Slavic lengthened grade (**lḗypā*); the semantic motivation is paralleled in E *lime tree* (from *lime* 'glue, sticky substance used to trap small birds')[6] and Ru. *víšnja*, Slk. *višňa* 'cherry' from the root of Lat. *viscum* 'mistletoe; bird lime', Gr. ἰξός 'mistletoe; bird-lime' (< **wik-sk-*), Markova 2008: 40.[7] The species *Tilia cordata* is common everywhere in Europe except in Southern Spain and Northern Scandinavia. Its wood was traditionally used for making shields.

10. **vьrba* 'willow, *Salix*' (OCS *vrъbije*, Cr. *vŕba*, Sln. *vŕba*, Ru. *vérba*, dial. *verbá*, Cz. *vrba*, ERHJ II: 752, Snoj 2003: 833) can be derived from PIE **werb^h-* 'twig, stick' (Lith. *vir̃bas* 'twig', Latv. *vir̃bs* 'stick', Lat. *verber* 'stick, whip', maybe also Gr. ῥάμ-

6 It is possible that *lime* is actually related to G *Linde* with *-m-* instead of *-n-* under the influence of the word for birdlime; however, this does not disprove the semantic connection between the word for a lime tree and words denoting glue or some sticky substance.

7 Similar words for 'alder' exist in West Uralic languages, with Finnic reflexes pointing to **leppä*, (Western) Saami to **lejpä*, and Mordvin to **lippä* or **lüppä* (Ante Aikio, p. c.). These could be either prehistoric borrowings from extinct Baltic dialects in which the meaning changed from 'lime' to 'alder' (an unlikely, but possible change), or chance similarities, but one cannot exclude the possibility that Balto-Slavic and West Uralic words have a common substratum source.

νος ‘bush *Rhamnus*’ < **wrb^{h}no-*). Perhaps this was the original word for *Salix alba* rather than *Salix caprea* (see **iva* above).

There are also two Slavic forest tree names that were borrowed from Germanic in the prehistoric period (and the Germanic words have viable PIE etymologies):

11. **búkъ*, **búky* (Gen. sg. **búkъve* a. p. a) ‘beech, *Fagus sylvatica*’ (Cr. *bȕkva*, *bȕk*, Sln. *búkev*, *bûkva*, Ru. *buk*, OCz. *bukev*, Cz. *buk*, Po. *bukiew*, ERHJ I: 95, ESSJa III: 90f.) probably borrowed from Go. *boka* (Pronk-Tiethoff 2013: 79–82; cf. G *Buche*, E *beech*) < PIE **b^{h}eh$_2$ǵo-* (Lat. *fāgus*, Gr. φηγός ‘kind of oak, acorn’). The present-day distribution of beech in Europe (from the Atlantic to the Ukrainian steppe) is the result of a recent spread (especially in the Subatlantic period after 450 BCE, with the peak around 700 CE). Earlier, *Fagus sylvatica* had a more limited distribution in Central Europe.

12. **ávorъ* (a. p. a) ‘sycamore maple, *Acer pseudoplatanus*’ (CS *avorъ* (Miklošić), Cr. *jȁvōr*, Sln. *jávor*, Ru. *jávor*, Po. *jawor*, ERHJ I: 399, ESSJa I: 96f.), probably from OHG **āhor* (> G dial. *Acher*, *Are*, *Ohr*, *Ure*, *Ere*, OHG *ahorn*, ODa. *ær*) < PIE **h$_2$eḱr-* (Lat. *acer*, perhaps Gr. ἄκαστος).[8] It is formally possible, but of course rather speculative, to derive PIE **h$_2$eḱr-* from the root **h$_2$eḱ-* ‘sharp’ (Lat. *acus* ‘needle’, OCS *ostrъ* ‘sharp’, IEW 20, Snoj 2003: 236f.) because of the sharpness of its leaves. The species *Acer pseudoplatanus* is native chiefly in Central Europe, from France to the borders of the Ukrainian steppe. Maple wood has limited use as it is not very hard, but it was traditionally used in making bowls and kitchen utensils, as well as for fuel.

3 Possible substratum words for forest tree names in Proto-Slavic

The following tree names do not have a viable PIE etymology in the sense mentioned above; they are not attested in the non-European branches of the family, they cannot be connected with any PIE root (with reflexes in other semantic fields

8 According to Pronk-Tiethoff (2013: 185f.) it is not altogether clear that **avorъ* ‘maple’ is a Germanic loanword. Germanic **-h-* is usually preserved in Slavic (though note the Montenegrin dialectal form *ahor* ‘maple’ and the oronym *Jȁhorina* in Bosnia where *-h-* has been preserved), and Germanic **a* is reflected as Slavic **o* in old loanwords. Pronk-Tiethoff concludes that both the Germanic and Slavic words (together with Lat. *acer* ‘maple’) could be loanwords from some non-IE substratum.

besides tree names), and all of them have one or more features typical of pre-IE substratum words: a-vocalism of the root (which cannot be reduced to PIE *h_2e), non-IE Ablaut patterns (e.g. **i*, **u* alternating with **e*, **a*, and **o*), alternations of voiced and voiceless stops, root shapes that contradict PIE phonotactic rules (see Kuiper 1995, Schrijver 1997, Iversen & Kroonen 2017, Matasović 2020: 335):

13. **grabrъ* 'hornbeam, *Carpinus betulus*' (Cr. *grȁb*, Bulg. *gábər*, Serb. *gàbar*, *gràbar*, Sln. *gâber*, *grâber*, Ru. *grab*, Po. *grab*, Cz. *habr*, *hrabr*, Slk. *hrab*, LSrb. *grab*, USrb. *hrab*, ERHJ I: 291, ESSJa VII: 99). In several languages there was a dissimilation of two rhotics, whereby either the first one or the second one was lost. The connection, often found in dictionaries, of this Slavic word for 'hornbeam' with Ancient Macedonian γράβιον 'some tree' and Umbr. *Grabovius* (an epithet of Jupiter) is extremely speculative and may rest on no more than chance similarity. The connection with Lat. *carpinus* is formally very difficult – the Latin noun is derived from PIE **(s)kerp-* 'pluck, cut' (Lith. *kerpù*, *kir̃pti* 'chop, cut', *skir̃pstas* 'elm', Lat. *carpō* 'pluck', maybe also Hitt. *karpina-* 'kind of fruit tree'); the semantic motivation is in the serrated leaves of hornbeams (EDL 94). Rather, PSl. **grabrъ* is probably related to Lith. *skrõblas*, *skrúoblas* 'hornbeam' (perhaps by dissimilation from **skrobris*) and Latv. *skābardis*, *skabarda* 'red beech', OPr. **scoberwis* (written *stoberwis* in Elbing Vocabulary, cf. Smoczyński 2007: 568). This points to a root **(s)greh$_2$b^h-* or **sgrob-* (with the acute as the result of Winter's law). OPr. *wosigrabis* '*Evonymus Europaeus*', lit. 'goat-hornbeam' may either be related to the Slavic words, or it is a Slavic loanword. The shape of the root that has to be posited for Balto-Slavic (**(s)g$^{(h)}$rob-r-*) looks distinctly non-Indo-European. If Alb. *shkozë* 'hornbeam, *Carpinus betulus*' is related, it points to a yet different variant of the root, **skēb^h-* (*-zë* is a common collective suffix, Demiraj 1997: 362). Hornbeam was absent from most of Eastern Europe (and the European parts of Russia) during the Boreal period, but it spread eastward and northward during later periods, especially during the Subatlantic (fom 500 BCE until present). It is of limited commercial value – although the wood is hard, it is difficult to work, because it has cross-grains and is not flexible. It makes excellent fuel, though.

14. **jьlmъ* 'elm, *Ulmus*' (Ru. *íl'm*, *ílem*, *íl'ma,* Po. *ilm*, *ilem*, LSrb. *lom*, Polab. *jelm*, Vasmer I: 478); in other languages we find the following cognates: Lat. *ulmus* 'elm' < **elmo-*, MIr. *lem* 'elm' < **limo-*, W *llwyf* (pl.) < **leymo-* and perhaps Gaul. tribal name *Lemo-vices* (whence the placename *Limoges*), ON *almr* < **olmo-*, OHG *elme*, *elm*, *elmo* < **elmo-* (G *Ulme* was borrowed from Lat. *ulmus*). The ablauting PIE paradigm EDPC (**h$_1$leyōm*/**h$_1$lim-os*) posited by EDPC (237) is not really persuasive. It is unlikely that the Slavic words were borrowed from Germanic, since the element *ilem-* is common in placenames (Ru. *Ilemno*, *Ilemka*, etc., Vasmer I: 478), and the vocalisms of the Slavic and Germanic words do not match. The irregular vowel al-

ternations point to a non-IE origin (Schrijver 1997).[9] Elms are easy to work and have many uses (for furniture, etc.), and are very resistant to water-decay. They are common in mixed woods with oaks and limes and their distribution was somewhat wider before the onset of the Subatlantic period (from ca. 500 BCE). The species *Ulmus glabra* has the widest distribution today, roughly from central France to the Urals.

15. **jálovьcь* 'juniper, *Juniperus communis*' (Ru. *jálovec*, Cz. *jalovec*, Po. *jałowiec*), perhaps from the same root as PSl. **jalovъ* 'sterile, barren' (ORu. *jalovъ*, CS *jalovъ*, Cr. *jàlov*, Vasmer III: 488). The semantic connection is found in the fact that juniper has dioecious flowers (each sporophyte plant has only one kind of spore-producing organ, all of whose spores give rise either to male gametophytes, which produce only male gametes (sperm), or to female gametophytes, which produce only female gametes); less likely appears the derivation from the root **h₁el-* 'reddish brown' (Snoj 2003: 239, IEW 302f., see **olьxa*, **elьša* below), since the lengthened grade would be unexpected. Young juniper trees are indeed reddish brown, while older specimens tend to be dark greyish.

A different word for 'juniper', **brinъ*, can be reconstructed on the basis of Cr. *brȉnje* 'juniper berries', Sln. *brin* 'juniper', Cz. *břím* 'larch, *Larix decidua*', dial. *břin* (Silesia) (Skok I: 210f., Machek 73). It obviously has a more limited distribution in Slavic, but it has no generally accepted etymology. The derivation from the root of the verb **briti* 'cut' (Cr. *brȉti* 'shave', Ru. *brít'*, Sln. *briti*) is possible, but uncompelling. The semantic connection would lie in the fact that juniper leaves are needle-like, and the formation would be as in Cr. *klȉn* 'wedge' from *klàti* 'cut, slaughter' < PSl. **kólti*. This shrub has the widest distribution of all conifers; it is found at lower as well as at higher altitudes across Europe and has many uses (in medicine as well as for making utensils).

16. **klenъ*, **klьnъ* 'maple, *Acer campestre*' (Cr. *klȅn*, Montenegrin *kûn*, Sln. *klẽn*, Ru. *klën*, Po. *klon*, ERHJ I: 447, ESSJa X: 194f.); cognates include OE *hlyn* 'maple', ON *hlynr*, Sw. *lönn* < PGm. **hluni-* (EDPG 232) and Lith. *klẽvas* 'maple tree', Latv. *kļavs* < **klyowo-*). These words for 'maple' might be formally compared with the word for 'holly' in West European languages: OE *holegn* 'holly, *Ilex aquifolium*' < PGm. **hulena-*, OHG *hulis-boum* (G *Hulst*) < PGm. **hul-is-*, ON *hulfr* < PGm. **hul-f-*(?), OIr. *cuilenn* 'holly tree', W *celyn* < PCelt. **kulisno-* or **kolisno-* (EDPC 213); similar

9 The similarity with the Latin word for holm-oak (*Quercus ilex*), *īlex*, is probably accidental. Lat. *īlex* has been derived from a PIE root **HleyHl-* and compared with Ru. *il* 'silt, clay', Latv. *īls* 'very dark, black' because of the dark colour of the bark of holm-oaks (Matasović 2016: 703).

words for 'holly' exist also in Arm. *kostłi*,[10] Srd. *golóstri* and Bsq. *gorosti*. It should be noted, though, that maple and holly are very dissimilar trees – maples are tall and strong, hollies are lower and their wood is not used as timber; maples lack the berries characteristic of hollies, and the leaves of the two trees are dissimilar. However that may be, the alternation between **kleno-*, **klunV-* and **klyowo-* in the word for 'maple' does not look Indo-European. A similar suffix alternation (**R/*uR*) is found in the word for 'hazel, *Corylus avellana*', cf. Lat. *corulus* 'hazel-tree', Lith. *kasùlas* 'hunter's spear, shaft', Lith. dial. (South Aukštaitian) *kasulà* 'plough shaft' < **kosulo-* vs. OIr. *coll* 'hazel', W *coll*, OHG *hasal*, ON *hasl* < **koslo-* (EDL 138f., EDPG 213).

17. **orkýta* (a. p. a) 'willow, *Salix*' (Cr. *ràkita*, Sln. *rakîta*, Ru. *rakíta*, *rokíta*,[11] Po. *rokita*, ERHJ II: 349); a Baltic cognate is probably Latv. *ērcis* 'juniper', although junipers and willows are not similar; possible parallels in other IE languages are Gr. ἄρκευθος 'juniper' and Lat. *arcus* 'bow' (bows are often made of willow-tree; cf. also E *arrow* < OE *earh* and its Germanic cognates pointing to PGm. **arhwō-*); since Gr. -θ- cannot correspond to Balto-Slavic **-t-*, the word is probably of substratum origin (similarly in EDG 132). For the distribution of willow trees, see **iva*.

18. **olьxa*, **elьša* 'alder, *Alnus*' (Cr. dial. *jóha*, Sln. *jélša*, Ru. dial. *ëlxa*, Po. *olcha*, *olsza*, LSrb. *wolša*, Cz. *olše*, ERHJ I: 408, ESSJa VI: 23f.); the Slavic forms point to a prototype **ol-is-* (or **al-is-*) and **el-is-*; cognates include Lith. *alk̃snis* < **alsni-*, OHG *elira* (G *Erle*), ON *alr* < **al-is-*, Lat. *alnus* < **al-i(s)no-* (by syncope). A derivation from the PIE root **h₁el-* 'reddish' (OHG *elo* 'yellow, fawn-coloured', perhaps also in Gr. ἔλαφος 'deer', CS *jelenь* 'deer', ON *elgr* 'elk'), allegedly because this tree turns reddish when peeled of its bark (Snoj 2003: 239), seems semantically improbable to me, besides being incompatible with the vocalism of Lat. *alnus*. Alder is very widespread in Eurasia (except in Southern Europe) and it is well attested in the steppes during the Atlantic period (5500–3000 BCE). The Slavic term probably referred originally to *Alnus glutinosa* (common alder) rather than to the other species (e.g. *Alnus cordata*), because it has the widest distribution in Europe, especially since the onset of the Subatlantic period (from 500 BCE onward). Its wood is soft and porous, but durable if kept under water, so it was often used for shipbuilding.

10 This Armenian word also means 'holm oak, *ilex*'; Martirosyan (EDAIL 371f.) does not connect it with the West European words for 'holly', but rather derives it from *kostł* 'twigs smeared with bird-lime to entangle birds' (of obscure origin).

11 Ru. *šeljúga* 'red willow, *Salix rubra*' was probably borrowed through Ukr. *šel'úha* from Alb. *shelg* 'willow' < Lat. *salix*. It is a "Wanderwort" spread by migrant, nomadic cattle-herders from the Balkans to the Carpathian Mountains.

19. **osa*, **osina*, **esina*, **asika* 'aspen, *Populus tremula*' (Cr. *jàsika*, Sln. *jasika*, *jesika*, Ru. *osína*, Cz. *osina*, Po. *osa*, *osina*, USrb. *wosa*, *wosyna*, Slk. *osika*, Vasmer II: 282); all of these forms are derivable from two prototypes: **opsā* and **ōpsīkā*,[12] which are presumably cognate with Lith. *ãpušė*, *apušẽ*, *apušìs* 'aspen' < **op-us-* (by analogy with *pušìs* 'pine'), Latv. *apse*, OPr. *abse*, OE *æspe æpse* 'aspen', OS *aspa* (with a metathesis, cf. EDPG 39). Arm. *opʿi* 'white poplar' might also be related (if from **osp-*, with a metathesis as in Germanic); several words for 'aspen' in Altaic languages have been compared, e.g. Tatar *awsak* (Tobolsk), Chuvash *ăvăs*, cf. also Fi. *haapa* 'aspen' < **šapa-*, so we might be dealing with a Eurasian 'Wanderwort'. The wood is not dense or strong and has limited commercial value. It has a wide distribution across Eurasia.

20. **tísъ*, **tísa* (a. p. a) 'yew, *Taxus baccata*' (RuCS *tisa* 'cedar', Sln. *tîsa*, Cr. *tȉsa*, Ru. *tis*, Cz. *tis*, USrb. *ćis*, Po. *cis*, ERHJ II: 215). Lat. *taxus* 'yew' seems too similar to be unrelated. These words are often connected to Gr. τόξον 'bow' and Persian *taxš* 'bow', since bows are often made of yew-tree. Lat. *taxus* could, in principle, be derived from **tk^{w}so-*, with a secondary *-a-* (EDL 607). However, comparing these words with PSl. **tisъ*, **tisa* does not allow us to reconstruct a common prototype. PIE **tok^{w}s-* and **tak^{w}s-* would yield PSl. **tox-*, and even the reconstruction **teyk^{w}s-* would not do (we would expect PSl. **tix-*). However, since stretchability is the prime characteristic of yew-wood, one is tempted to connect PSl. **tisa*, **tisъ* with Lith. *tiẽstis* 'stretch', *tiẽsti* 'straighten'. PSl. **tisъ* would then be derivable from **tēys-so-*, with the lengthened grade as in **lípa* 'linden',[13] Lith. *líepa* < **lēyp-*, from the root **leyp-* 'glue, be sticky' (see above), and with the suffix **-so-* as in PSl. **běsъ* 'anger', < **b^{h}oyd-so-* (Lat. *foedus* 'ugly', cf. Lith. *baisùs* 'terrible' < **b^{h}oyd-su-*). We would also have to assume that Proto-Slavic (or Proto-Balto-Slavic) had a geminate **-ss-* at the time when the RUKI-rule operated, and that the geminate was not affected by the otherwise regular change of **s* to **š* (> PSl. **x*) after **-ey-* and **-ēy-*.[14] The original meaning of **tēys-so-* (> PSl. **tísъ*) would have been 'stretchable (tree or wood)'. If this etymology is correct, PSl. **tisъ*, **tisa*

12 The long **ō* in **ōpsīkā* may have been due to the analogy with the word for 'ash' (PSl. **asenъ* < **ōs-en-*, no. 1 above).

13 The acute, probably induced on a secondary lengthened grade of the same root (**tēyso-*) is found on another derivative from the same root, PSl. **tíxъ* 'quiet' (OCS *tixъ*, Ru. *tíxyj*, Cr. *tȉh*, Po. *cichy*). The meaning 'silent, quiet' developed from 'flat' (cf. It. *piano* 'quiet < 'flat'), and this in turn from 'stretched'.

14 The RUKI-rule did not operate before clusters of two consonants in Proto-Slavic, cf. OCS *prъstъ* 'finger' vs. Lith. *piřštas* 'id.' < **prsto-*.

'yew' is not a substratum word.[15] The species *Taxus baccata* is native mostly to Central Europe and parts of Northern Spain, but it is also found in the Caucasus and the neighbouring areas.

21. **topòlь* 'poplar, *Populus*' (Sln. *tópol*, CS *topolь*, Ru. *tópol'*, Cz. *topol*, ERHJ II: 623). A possible Baltic cognate is found in Lith. *túopa* 'poplar', which points to a root **toHp-* or **tōp-* (but it could also be a Slavic loanword, see Gliwa 2008), while the Slavic forms are derivable from **top-ol-* or **tap-ol-*. Lat. *pōpulus* 'poplar' can also be from **toHp-* or **tōp-* (with the assimilation **t. .p* > **p. . .p*, EDL 620). It is less likely that the Latin form is from original **poHp-* and that the Balto-Slavic words were derived by dissimilation, since independent dissimilation would have to be posited for Slavic and Lithuanian.[16] Alb. *plep* 'poplar' is probably from VLat. **plōpulus* (cf. also Rom. *plop*), and OHG *papil(boum)* > G *Pappel* must be from MLat. variant *papulus*, with unexplained short *-a-*.[17] Furthermore, Gr. ἀπελλόν 'black poplar' (Hes.), if related, might point to a substratum stem **apel-* (with the prefix **a-* identified by Schrijver). Gr. πτελέα 'elm' (Myc. *pte-re-wa*) is probably unrelated (it is the likely source of Lat. *tilia* 'lime-tree', or both words were borrowed from some common source, see EDL 620). Since the bark of the (young) poplar tree can be yellowish to dark grey, the name of the tree has been connected with the following words: OHG *falo* 'pale', Lat. *palleō* 'be pale', Lith. *palvas* 'light yellow', but this does not explain the a-vocalism implied by Lat. *palleō* (the other words might, in principle, be from PIE **pol-*). The stems that must be posited to account for the attested forms (**tōp-(-ol-)*, **top-ol-*, **tap-ol-*, perhaps also **a-pel-*) do not look Indo-European, so we may be dealing with a substratum word for 'poplar'. The wood of this tree is not dense and has limited commercial value. The most common species of poplar, *Populus tremula*, has a very wide distribution in Western Eurasia.

15 Alb. *tis* 'yew' is a loanword from South Slavic.

16 For the same reason I consider it unlikely that Slavic **topolь* was borrowed from MLat. *papulus* (the source of G *Pappel*), which is discussed by Snoj 2003: 773.

17 The OHG word is also attested with a different vocalism as *popilboum*, cf. also MLG *poppele*, which might represent a variant **poppulus* (with short **-o-* by "Littera-rule"). Kluge (610) speculates that G *Vielbaum* 'black poplar' (cf. the Old High German placename *Vilbom*) could contain the reflex of an etymological cognate of Lat. *pōpulus* (perhaps *Viel-* could go back to **felV-* < **tpelV-*).

4 Some unresolved problems

For the following words no persuasive PIE etymology has been offered so far:

22. **berka* 'wild service tree, *Sorbus torminalis*' (Cr. *brèkinja*, Slk. *brekyňa*, Sln. *brẹ́ka*, Ru. *berëka*, *bérek*, Po. *brzęk*, ERHJ I: 83, ESSJa II: 194f.). The Polish form with the nasal vowel is unexplained; no convincing etymology has been proposed so far (the connection with PSl. **brękn̥ti* 'swell' > Cr. *bréknuti* is gratuitous, as there is no semantic connection). A similar tree name is found in Alb. *bërshen* 'yew, *Taxus baccata*', which is very dissimilar, but has similar berries.[18] The wood is of excellent quality, very dense and with good bending-strength. This tree has a somewhat limited distribution in Europe, being native to its western and southern parts (France, Italy and the Balkans), from where it spread eastwards during the last millennium.

23. **bъzъ*, **buzъ*[19] 'elder, *Sambucus*' (Cr. *baz*, *zóva* < **bъzova*, Sln. *bèz*, Ru. *buziná*, dial. *buzá*, *boz*, Ukr. *boz* <Gen. *bzu*>, LSrb. *bez*, *baz*, Po. *bez* <Gen. *bzu*>, ERHJ I: 47, ESSJa III: 144ff.). A derivation from PIE **b^{h}eh$_2$ǵo-* 'beech' (Lat. *fāgus*, OHG *buohha*, etc.), which is often found in the literature, has little to recommend itself: the vocalism of the two words is incompatible, and elder and beech are very different trees, not only by their appearance, but also with respect to the qualities of their wood.[20] Since elder berries are edible, one is tempted to derive PSl. **bъzъ*/**buzъ* from the PIE root **b^{h}ew(ǵ)^h-* 'enjoy' (Skt. *bhunákti*, Lat. *fungor*, Alb. *bungë* 'a kind of oak (with edible fruit, acorn)', but this is, of course, very speculative (and the Sanskrit forms point rather to the root ending in a plain velar rather than PIE **ǵ^h*). Oguibénine (2016: 39) notes the possibility that PSl. **bъzъ* is from the same root as PGm. **bukka-* 'goat' (OHG *boc*, ON *bukkr*, OE *bucca*), P *buz* 'goat', and perhaps Arm. *buc* 'lamb' < PIE **b^{h}uǵo-* (IEW 174), pointing to a similar metaphor, e.g., in Lat. *caprifolium* 'honeysuckle' (from *caper* 'goat'). This is unlikely, however, because we would expect PIE **u* to be lengthened in Slavic before **ǵ* by Winter's law. Therefore, it is best to consider PSl. **bъzъ*, **buzъ* as a word without a PIE etymology.

---

18 Demiraj 1997: 98f. rejects the connection of Alb. *bërshe* (Gheg. *bërshën*) and OIr. *ibar* 'yew-tree', but does not propose an alternative etymology. He reconstructs the Proto-Albanian form as **bVrš-* or **brVš-*, which is formally difficult to square with the Slavic word. The assumption that PSl. **berka* was borrowed from PGm. **berkō-* 'birch' (OE *beorc*, Du. *berk*, etc., cf. EDPG 61) is very improbable, since birches are very unlike service trees; the word would have to be borrowed from West (or North) Germanic, where PIE **e* and **i* do not merge as **i* as they do in Gothic.

19 The form with **u* may be secondary, since it is limited to East Slavic and it may be due to the analogy with the word for beech, PSl. **bukъ* (no. 11 above).

20 The beech wood is very hard, while that of the elder is very soft (the stems of the elder can be easily hollowed and are often used as tubes).

24. **smérka* (a. p. a) 'spruce, *Abies picea*' (CS *smrěka*, Sln. *smrẹ́ka*, Ukr. *smeréka*, Cz. dial. *smrek*, ERHJ I: 482). No convincing etymology has been proposed so far. Arm. *mayr* 'cedar, pine', perhaps also 'juniper' (Friedrich 1970: 151)[21] is probably unrelated (EDAIL 448), and a borrowing from Celtic **smriko-* (Oguibénine 2016: 282) < **sm-prḱo-* (allegedly related to Av. *hapərəsī* 'juniper') is too speculative, as the Celtic word is actually unattested. However, it is possible to connect this word with PGm. **furhō* 'spruce' (OHG *for(a)ha*, OS *furia*, OE *furh*, Kluge s. v. *Föhre*) and It. dial. (Trent.) *porca* 'spruce'. One has to start from PSl. **perka* which has been re-shaped as **smerka* by analogy with the verb **smьrděti*, **smьrdǫ* 'stink' (of all the coniferous trees, the spruce has the strongest odour). The question remains whether these words are related to Lat. *quercus* 'oak', perhaps also to OIr. *ceirt* 'apple-tree', MW *perth* 'bush, hedge, thicket' < PCelt. **k^{w}erx-t-* (note that this etymology is doubted by EDPC 178). Lat. *quercus* may be derivable from an u-stem **perku-*, gen. sg. **perkwos*, with the subsequent dissimilation **perkw-* > **k^{w}erkw-*, but the semantic connection is problematic, as fir trees are very much unlike oaks. The similarity may lie in the fact that spruces are the largest, tallest trees at higher altitudes, just as oak trees tend to be the most imposing tree species at lower altitudes. Spruce has very solid wood for timber (just as oak does), but it is not very durable. Spruce was absent from the steppe region in the 4th millennium B.C. (Cooper 2010: 42), but it subsequently spread, along with pine. Today, the species *Abies picea* has a fairly northern distribution in Europe (Scandinavia, Central and Northern Russia), but it is also found in the higher altitudes in the Alps.

25. **sosna* 'Scots pine, *Pinus sylvestris*' (Ru. *sosná*, USrb. *sosna*, Po. *sosna*, Polab. *süsnó*, Vasmer II: 701f.); this word does not have a convincing PIE etymology. The derivation from the root **ḱeh$_1$s-* 'grey' (Lat. *cānus* 'gray (of hair)' < **ḱh$_1$sno-*, OPr. *sasins* 'hare', OHG *haso* 'hare', OE *hasu* 'grey-brown', Skt. *śaśá-* 'grey') is semantically unmotivated and formally difficult (**ḱh$_1$s-* would probably yield PSl. **ss-*> s-, while **ḱh$_1$es-* would yield **xes-* > **šes-* or **ses-*). The derivation from **sopsna* < **sapsnā* (Trubačev, apud Vasmer II: 701) links this word to ON *safi* 'sap of a tree', OHG *saf*, perhaps also Arm. *ham* (if from **sapmo-*), Lat. *sapor* 'taste' (whence *sapīnus* 'pine or fir tree' > Fr. *sapin*), a derivative of *sapa* 'new wine boiled down to a proportion of its original volume' (EDL 538). Although formally better (if one accepts PIE **a*), this is semantically still a difficult etymology. On the other hand, exact parallels to this Slavic word for 'pine' exist in Saami, which points to bor-

21 Other words connected with PSl. **smerka* and Arm. *mayr* in the literature (on which see EDAIL 448), e.g. Hung. *mór* 'spruce', MEg. *mrw* 'Lebanese cedar' are probably chance similarities. The derivation of the Slavic words for 'spruce' from the root of Ru. *smerkat'sja* 'get dark' (Markova 2008: 43) is both formally and semantically unconvincing.

rowing from some common, non-IE and non-Uralic source: compare Saami *saasne* 'rotten tree', Skolt Saami *šošnn* 'dead pine tree', words that do not have a Uralic etymology (Aikio 2012).[22] The Scots pine was the dominant conifer in Central and Northern Europe during the Boreal period; its expansion to Western and Central Europe is rather recent. Pine wood is very strong and one of the most commercially important kinds of wood in the Nordic countries. It was used as timber and for making utensils.

5 Conclusion

In this study, we have argued that there are a number of probable non-IE loanwords in the corpus of Proto-Slavic forest tree names: **klenъ*, **klъnъ* 'maple', **olьxa*, **elьša* 'alder', **grabrъ* 'hornbeam', ?**osa*, **asika* 'aspen', **topolь* 'poplar', **orkyta* 'willow' and **jьlmъ* 'elm'. Three other words do not have a clear PIE etymology and could, in principle, also have been borrowed from some non-IE substratum: **bъzъ*, **buzъ* 'elder', **berka* 'service tree' and **sosna* 'pine'.[23]

It can be observed that names of economically less important trees are more often among the possible or certain loanwords from some substratum source (alder, hornbeam, aspen, poplar, elder, service tree); this is in line with the fact, already well established in substratum studies, that the language of the socially dominant group is more likely to borrow words for items that are not traded, or otherwise economically exploited, from the language of the socially lower group; in comparison, lexemes for economically important items are less likely to be borrowed from a substratum language.[24]

---

22 According to Aikio, these Saami words look like borrowings from different, but related sources: "Regarding the Saami words for 'dry/dead/rotten pine-tree': there are three phonological variants, which show a nearly complementary distribution. Eastern Saami (Inari, Skolt and Kildin Saami) have the form **šošnę*, but Western Saami languages show two forms with the sibilant **s* instead: the more widespread one is **suosnę* ~ **suosnō*, found throughout Western Saami, but in South Saami there is also a parallel form **sāsnę*" (Ante Aikio, p. c.).

23 For **asenь*, **esenь*, **asenъ* 'ash' a substratum origin is possible, but a viable PIE etymology exists; similarly, substratum origin is possible for **smerka* 'spruce' and **tisъ*, **tisa* 'yew', although we have proposed new PIE etymologies for these words.

24 Cf. the French words of Gaulish origin, which very often denote animals and plants not raised or grown for food, e.g. *alouette* 'lark', *vanneau* 'lapwing', *belette* 'weasel', *bruyère* 'heath', *gerzeau* 'corncockle', etc. Similarly to the Gauls with respect to Romans, the Neolithic farmers from which the substratum words for forest trees were probably borrowed into Proto-Slavic would presumably have been a socially lower group in comparison to the speakers of Proto-Slavic (or its ancestor language).

Finally, we have also seen that names of trees with more limited geographical spread before the Subatlantic period (beech, pine, hornbeam, and service-tree) were more often of substratum origin, which is quite understandable, given that the homeland of the Slavs (and possibly speakers of Proto-Balto-Slavic) before 500 A.D. was restricted to parts of Central and Eastern Europe.

References

Aikio, Ante. 2012. An Essay on Saami Linguistic Prehistory. In Riho Grünthal & Petri Kallio (eds.), *A Linguistic Map of Prehistoric Northern Europe*, 63–117. (Mémoires de la Société Finno-Ougrienne 266). Helsinki: Suomalais-Ugrilainen Seura.

Cooper, Brian. 2010. Russian words for forest trees: a lexicological and etymological study. *Australian Slavonic and East European Studies*. 24(1–2). 41–71.

DELG = Xavier Delamarre. 2003. *Dictionnaire étymologique de la langue gauloise*. Paris: Errance.

Demiraj, Bardhyl. 1997. *Albanische Etymologien*. Amsterdam: Rodopi.

EAFTS = San-Miguel-Ayanz, Juan,Daniele de Rigo, Giovanni Caudullo, Tracy Houston Durrant, & Achille Mauri (eds.). 2016. *European Atlas of Forest Tree Species*. Bruxelles: The European Commission, https://forest.jrc.ec.europa.eu/en/european-atlas/. (accessed March 2, 2022).

EDAIL = Martirosyan, Hrach. 2009. *Etymological Dictionary of the Armenian Inherited Lexicon*. Leiden: Brill.

EDG = Beekes, Robert S. P. 2010. *Etymological Dictionary of Greek*. Leiden: Brill.

EDL = De Vaan, Michiel. 2007. *Etymological Dictionary of Latin and the Other Italic Languages*. Leiden: Brill.

EDPC = Matasović, Ranko. 2009. *Etymological Dictionary of Proto-Celtic*. Leiden: Brill.

EDPG = Kroonen, Guus. 1913. *Etymological Dictionary of Proto-Germanic*. Leiden: Brill.

ERHJ I = Matasović, Ranko,Tijmen Pronk, Dubravka Ivšić & Dunja Brozović Rončević. 2016. *Etimološki rječnik hrvatskoga jezika*, *I: A-NJ* [Etymological dictionary of the Croatian language, I: A-NJ] Zagreb: Institut za hrvatski jezik i jezikoslovlje.

ERHJ II = Matasović, Ranko,Dubravka Ivšić Majić an& Tijmen Pronk. 2021. *Etimološki rječnik hrvatskoga jezika*, *II: O-Ž* [Etymological dictionary of the Croatian language, II: O-Ž] .Zagreb: Institut za hrvatski jezik i jezikoslovlje.

ESSJa = Trubačev, Oleg N. (ed.). 1974 – *Ètimologičeskij slovar' slavjanskix jazykov* [Etymological dictionary of the Slavic languages]. Moscow: Nauka.

Friedrich, Paul. 1970. *Proto-Indo-European Trees. The Arboreal System of a Prehistoric People*. Chicago: The University of Chicago Press.

Gliwa, Bernd. 2008. Ist litauisch *túopa* 'Pappel' wirklich ein Erbwort? *Baltistica* 43(2). 239–244.

IEW = Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern: Francke.

Iversen, Rune & Guus Kroonen 2017. Talking Neolithic: Linguistic and Archeological Perspectives on how Neolithic was Implemented in Southern Scandinavia. *American Journal of Archaeology*. 4 (1219). 511–525.

Kluge = Kluge, Franz. 1999. *Etymologisches Wörterbuch der deutschen Sprache* (23. Auflage bearbeitet von Elmar Seebold). Berlin/New York: De Gruyter.

Kuiper, Franciscus B. J. Gothic *bagms* and Old Icelandic *ylgr*. *North-Western European Language Evolution* 25. 63–88.

Markova, Elena M. 2008. Praslavjanskie nazvanija derev'ev kak otraženie fragmenta jazykovoj kartiny mira slavjan [Proto-Slavic names of trees as a reflex of a fragment of the Slavic world map]. *Acta Linguistica* 2(1). 37–45.

Matasović, Ranko. 2016. Latin tree and plant names with the suffix *-ix*, *-ex*. In Nikolai N. Kazanskij (ed.), *Indoevropejskoe jazykoznanie i klassičeskaja filologija XX (2). Materialy čtenij, posvjaščennyx pamjati professora Iosifa Moiseeviča Tronskogo, 20–22 ijunja 2016 g*. [Indo-European Linguistics and Classical Philology XX (2). Proceedings of the conference dedicated to the memory of Professor Iosif Moiseevich Tronskyj, June 20–22, 2016], 700–709. Sankt Peterburg: Nauka.

Matasović, Ranko. 2020. Language of the bird names and the pre-Indo-European substratum. In Romain Garnier (ed.), *Loanwords and Substrata*, 331–344. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Machek = Machek, Václav. 1966. *Etymologický slovník jazyka českého* [Etymological dictionary of the Czech Language]. Prague: Lidové noviny.

Oguibénine, Boris. 2016. *L'héritage du lexique indo-européen dans le vocabulaire russe*. Paris: Institut d' études slaves.

Pronk-Tiethoff, Saskia. 2013. *The Germanic loanwords in Proto-Slavic*. Amsterdam: Rodopi.

Schrijver, Peter. 1997. Animal, vegetable and mineral: some Western European substratum words. In Alexander Lubotsky (ed.), *Sound Law and Analogy*, 293–316. Amsterdam/Atlanta: Rodopi.

Skok = Skok, Petar. 1971–1974. *Etimologijski rječnik hrvatskoga ili srpskoga jezika* [Etymological dictionary of the Croatian or Serbian language]. Zagreb: Jugoslavenska akademija znanosti i umjetnosti.

Smoczyński, Wojciech. 2007. *Słownik etymologiczny języka litewskiego* [Etymological dictionary of Lithuanian]. Vilnius: Uniwersytet Wileński.

Snoj, Marko. 2003. *Slovenski etimološki slovar* [Etymological dictionary of Slovene]. Ljubljana: Modrijan.

Vasmer = Vasmer, Max. 1959. *Russisches etymologisches Wörterbuch*. Heidelberg: Winter.

Part III: **Western and Central Europe**

Paulus S. van Sluis

4 Substrate alternations in Celtic

1 Introduction

Many words in Celtic languages lack an etymology in Proto-Indo-European, the ancestor of Celtic. This fact has led to suspicions that prehistoric Celtic heavily borrowed from a non-IE substrate language (see introduction, this volume). More puzzling still is that most of these words do not appear to come from any known language, not even languages for which substratum influence has been suspected on the basis of morphosyntactic evidence (EDPC 441–444; Matasović 2012: 156–159).[1] Still, every word in every language has an etymology, raising the question where these words come from. Tracing back from the earliest attested Celtic languages to the latest insights on Indo-European origins provides some bounds as to when, where and how such languages may have lent some of their words to Celtic. The *communis opinio* among Indo-Europeanists is that Proto-Indo-European was originally spoken by pastoral nomads associated with the Yamnaya culture on the Pontic-Caspian steppe in the late fourth and early third millennium BCE (Anthony 2010). Speakers of the earliest attested Celtic languages lived in west-central Europe and the Atlantic coast in the first millennium BCE, and lived more sedentary lives. What happened between these two points, and what other languages did these Indo-Europeans encounter on their way?

Evidence from palaeogenomics shows that the beginning of the third millennium BCE saw a significant influx of steppe-related genes into central Europe, leading to admixture with the Neolithic farmers who had settled Europe some millennia prior. Populations with this ancestry are found across the Corded Ware horizon (Allentoft et al. 2015; Haak et al. 2015). Western Europe saw the emergence of

1 Parts of the Celtic lexicon have variously been connected to Basque (McCone 2005), Afro-Asiatic (Vennemann 2003), Finnic (Hyllested 2016), and (para-)Altaic (Mikhailova 2020), but these proposals have failed to gather widespread support. At any rate, it is best to first compile a list of lexemes and phonemic features that cannot be explained as inherited, and only after compilation of such a corpus should one compare known non-IE languages to judge the explanatory power of assuming an known language as a Celtic substrate (McCone 2005: 406).

Acknowledgements: The author would like to thank Anthony Jakob, Anders Jørgensen, Guus Kroonen, Axel Palmér, David Stifter, Cid Swanenvleugel, Rasmus Thorsø and Andrew Wigman and all the participants at the workshop for their insightful comments on this study, and for the opportunity to see preliminary versions of their own chapters. This study has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant n° 716732).

https://doi.org/10.1515/9783111337920-004

the Bell Beaker phenomenon in the early third millennium BCE, and while Bell-Beaker-associated individuals in Iberia primarily retained ancestry of Neolithic farmers, western central Europe saw Bell-Beaker-associated individuals with high levels of steppe-related ancestry. On the British Isles, the introduction of the Bell Beaker phenomenon came with the introduction of steppe-related ancestry amounting to a population turnover to the tune of 90% in Britain (Olalde et al. 2018) and 15–30% in Ireland (Cassidy et al. 2016). Iberia at the end of the third millennium BCE saw a replacement of about 40% by people with steppe ancestry (Olalde et al. 2019). Many of these partial population replacements entailed a sex bias, whereby steppe ancestry predominantly appeared along the male line. Later, the Middle to Late Bronze Age saw a resurgence of Neolithic farmer ancestry in Britain, signalling the incorporation of migrants from what is now France (Patterson et al. 2022).

The Indo-Europeanization of Europe therefore appears to have been a protracted process with several points of contact between people with steppe-derived ancestry and the indigenous farmers and, in some places, hunter-gatherers of Europe. Episodes of mixed families with steppe-derived fathers and local mothers alternate with the adoption of new cultural packages across genetic boundaries (Sjögren et al. 2020). Conversely, a lack of admixture in Scotland is observed in the wake of the Middle to Late Bronze Age genetic turnover in Britain, suggesting that contact between Celtic speakers and now-unknown non-Celtic speakers could persist well into later prehistory (Patterson et al. 2022). In Orkney, Neolithic farmer lineages survived well into the Iron Age along the male line, and influx of Steppe-derived lineages mainly occurred along the female line (Dulias et al. 2022).

These contact events described in the genetic literature must have entailed linguistic contact as well, including the borrowing of words from now-extinct languages spoken in Neolithic Europe. Some of these episodes of contact, such as the Corded Ware horizon, seem relevant for IE branches beyond Celtic, and one may expect that loanwords at this stage are also borrowed into other IE languages of Europe. An episode such as the arrival of steppe-related ancestry in Ireland, on the other hand, could hardly leave its mark on surviving IE languages other than Celtic, and language contact with non-IE may have persisted well beyond Proto-Celtic. This study charts what words may have been borrowed into Celtic in these episodes of contact.

2 Methodology

In western Europe we are faced with a gap spanning thousands of years between its Indo-Europeanization and its first written records, and many languages spoken by the pre-Indo-European population that came into contact with IE and

Celtic speakers have not entered the historical record. The lexicon of these languages only survives as loanwords into IE and its daughter languages.[2] The first step in establishing loanwords from these substrate languages is therefore to find a methodology whereby loanwords can be positively identified even when the donor language is unknown.

One linguistic technique which allows us to positively identify a word as borrowed entails turning the comparative method on its head. The comparative method allows one to establish cognacy between words by establishing a pattern of regular sound correspondences. This, in turn, should produce a single reconstructed form in their shared proto-language. Conversely, when a set of seemingly related words resists reconstruction to a single form in the proto-language, the word appears borrowed into two or more subgroups after the break-up of the proto-language (Iversen and Kroonen 2017: 517).

Systematically submitting words to the test of whether their cognates can be reconstructed to a single proto-form therefore allows one to build a corpus of putatively borrowed forms (also known as pseudo-cognates) by isolating the words requiring multiple reconstructions. In those cases, their irregular reconstructions may suggest prehistoric borrowing from or via a third language, even if this language is not known directly. This study presents a collection of such words, i.e. words attested in Celtic with cognates that do not allow for an identical reconstruction to their shared proto-language, and whose formal differences cannot be accounted for by IE morphology, mutual borrowing, or language-internal processes such as analogy and onomatopoeia; this study collects irregular cognates containing an irregularity between Celtic and other branches of Indo-European as well as between the Goidelic and Brittonic branches of Celtic.[3]

There is one caveat in this methodology of finding irregularities within a cognate set. The comparative method employs the principle of the exceptionlessness of sound change to establish cognates, and this principle allows one to distinguish cognates from chance resemblances or borrowings. Thus, when irregularities are found, a word may conceivably be borrowed, but it is also possible that words being compared are simply unrelated chance resemblances. Even if the words compared are somehow related, irregularities between cognates do not exclusively originate from borrowing from a third language. Numerous processes such as sound symbolism, taboo deformation or obscure morphological processes may introduce irregularities in other words as well.

2 And possibly in other surviving languages in northern and western Europe, i.e. Basque and Uralic.

3 Irregularities only within Goidelic or Brittonic are not included, because such irregularities would fall outside the temporal scope of this study.

2.1 Substrate alternations

What is necessary, then, is a methodology to distinguish irregularities due to third-language borrowing from chance resemblances or language-internal irregularities. Schrijver (1997) describes such a methodology employing recurring patterns in how reconstructions are irregular. The borrowings that constitute the corpus of this study are identified by this methodology.[4] Schrijver demonstrates its working by drawing on a parallel case where the donor language is known: borrowings from Arabic into early Romance. These borrowings exhibit a phenomenon whereby the same word is sometimes found both with and without a prefix *ar-/al-*, cf. Cat. *carxofa*, It. *carciofo* beside It. *articiocco*, Fr. *artichaut*, and Sp. *alcachofa* 'artichoke', and Sp. *berenjena*, Port. *berinjela* beside Cat. *albergínia* and Fr. *aubergine* 'aubergine'. The co-occurence of forms with and without *ar-/al-* is unresolveable to a single proto-Romance reconstruction and has no basis in Latin morphology, but is easily explained within Arabic, where *al-* serves as the definite article. The many Arabic borrowings into Romance show an irregularity that cannot be projected back to Proto-Romance, but the inconsistently prefixed *al-* is itself consistent in shape. This consistent manner in which the borrowings are inconsistent across the target languages reflects a single grammatical feature of a single language.

Other alternations do not reflect morphological processes in the source language, but rather reflect different approximations of a foreign phoneme. If so, alternations still provide evidence of borrowing, but they do not necessarily allow for the demarcation of source languages. In Dutch, a phoneme /g/ is absent from its native lexicon. When English [g] is borrowed, it may be variously nativized as /k/ or as /x/ (van Bezooijen and Gerritsen 1994: 153), and words like *drugs* and *goal* in Dutch may be heard with either realization, with a preference for /k/ in the north and /x/ in the south. This alternation cannot be the result of Dutch-internal grammatical processes, and its existence provides evidence that Dutch *drugs* and *goal* are borrowings. The intra-Dutch alternation ultimately has its source in a language with a phoneme /g/, but this phoneme is common across languages, and words with /g/ in other languages may yield the same alternation when borrowed into Dutch. Based on Dutch evidence alone, the alternation found in Dutch *drugs* and *goal* suggests borrowing from a language with /g/, but there is no need to posit that only one language with /g/ stood in contact with Dutch or that both words are borrowed from the same language. Similar alternations be-

4 Earlier works employing these substrate alternations include DveV, Furnée (1972) and Kuiper (1995).

tween two phonemes in the Celtic lexicon may betray prehistoric borrowing, without necessarily allowing words exhibiting the alternation to be connected to a single source.

Celtic words for 'thrush' provide examples of how irregularities may come from both source-language alternations and from different approximations of a foreign phoneme.[5] Irregularity in vocalism between Ir. *troisc* and W *tresglen*, Bret. *draskl* is the result of a morphological alternation in the source language. Irish *troisc* is borrowed from OE *þrysce* 'thrush', while W *tresglen* and Bret. *draskl* are borrowed from OE *þræsce* 'thrush'. The irregularity within Celtic reflects a Germanic morphological phenomenon, just like how the irregularity between Cat. *carxofa* and Sp. *alcachofa* reflects an Arabic morphological phenomenon. Irregularities between W *tresglen* and Bret. *draskl*, on the other hand, can be analyzed as the result of differing approximations of a foreign phoneme. W *tresglen* and Bret. *draskl* resist reconstruction into a single shared proto-Brittonic form, differing between *t-* and *d-* in their onsets and between *e* and *a* in their stem vowels. The donor form *þræsce* has an onset in [θ] not found word-initially in early Brittonic except as the result of spirantization, so the various onsets in *t-* and *d-* can be accounted for as different nativizations of this foreign sound. The vowel [æ] in *þræsce* is not found in early Brittonic at all, and different attempts by Brittonic speakers to render it yielded an alternation between *e* and *a*, much like how different attempts by Dutch speakers to render [g] in *drugs* yields an alternation between /k/ and /x/.

Drawing on these parallels, one may expect substrate borrowings in Celtic to exhibit alternations whereby an irregularity is similarly found within an etymon recurring across etyma. Knowledge of these alternations may help in identifying more substrate words and sometimes connecting such words to a single source language.

2.2 Further identification criteria

Substrate alternations may be most securely identified when words with an alternation are also suspected of being borrowings on other grounds. The following additional criteria laid out by Schrijver (1997) may help in identifying substrate borrowings. A comprehensive overview of these criteria for Celtic and an overview of words identifiable as substrate borrowings on their basis is given by

5 Stifter (this volume) discusses the further etymology and potential substrate status of OE *þrysce* and *þræsce*.

Stifter (this volume). None of the criteria below provide compelling evidence of a substrate origin by themselves, however (Simon, this volume). Words that satisfy one or more of the criteria below are only included in the corpus of this study if they also show an irregularity in reconstructions.

Limited geographical distribution within IE may suggest borrowing. An inherited PIE lexeme may easily be found from Ireland to India, but a word borrowed from a substrate language would not be expected to be found beyond the borders of this substrate language. This criterion cannot establish a word's IE or non-IE origin by itself; doing so would constitute an argument from silence. The absence of a word from far-off IE languages may well be because those languages lost it. Conversely, non-IE *Wanderwörter* are known to have spread over wide areas after the Indo-Europeanization of Europe.[6]

Remarkable word formation may also suggest borrowing. Some words require a reconstruction that violates the grammar of PIE. Nouns inherited from PIE in principle contain a closed class of morphological transformations, so that a word may well be borrowed when it has undergone morphological transformations not inherited from PIE. This criterion carries little force by itself, because a borrowed affix can be added to an inherited root, a possibility Schrijver (1997: 295) illustrates using English native roots with Romance suffixes of the type *leak-age*, *teach-er*.

Some reconstructions to PIE require a phoneme (sequence) that was rare or absent from PIE, such as $^{*}a$ or $^{*}b$. The shape of PIE roots is moreover subject to a number of phonological constraints: a root is always monosyllabic, may not begin and end in identical consonants and roots of the shape $^{**}TeD^{h}$-, $^{**}DeD$-, $^{**}D^{h}eT$- are not allowed (Hamp 1990: 298). Words with different shapes immediately become suspected borrowings. This criterion, however, carries a risk of circularity: if words that are to be reconstructed to PIE $^{*}a$ and $^{*}b$ or to an illegal root shape are considered borrowed for the sole reason that they contain these phonemes, then inherited material with these phonemes could never be found.

Another method of finding borrowed words is to collect etymologically obscure words containing a phoneme or phoneme cluster that cannot be inherited. PIE $^{*}p$- disappeared in Celtic, and no inherited Irish words start with /p/. So Irish words such as *partán* 'lobster' and *parn* 'whale' must be borrowed, and because these words cannot be shown to be borrowed from a known neighbouring lan-

6 An extreme case of limited geographical distribution concerns words isolated to a single language or subfamily. When these words are regularly reconstructible and no other positive criteria attest to their borrowed status, they may conceivably be lexical archaisms. Examples of such *Sondergleichungen* in Celtic are PC $^{*}banu̯o$- 'pig' and $^{*}k^{w}ezdi$- 'thing, piece'; more cases are given by EDPC (441–4), Matasović (2012), and Mikhailova (2020).

guage, they appear to be borrowed from an unknown substratum language (Schrijver 2000b; 2005). Intervocalic /f/ in Irish may have a similar role in identifying borrowings (Mac Eoin 2007). This method is reliable in identifying recent borrowings, but the identification of necessarily borrowed phonemes and phoneme clusters becomes less secure for earlier stages. Geminated stops are possible examples of phonemes reconstructible to Proto-Celtic, but without a PIE source.

Semantics may also help in finding substrate borrowings. Economically and culturally unimportant animals and plants often have a substrate origin (Schrijver 1997: 295). PIE speakers were likely to be pastoral nomads, and the European population of the late Neolithic mainly consisted of sedentary farmers, so one may expect IE vocabulary relating to a pastoral nomad lifestyle to be inherited, and vocabulary relating to a sedentary farmer lifestyle to be of borrowed (Iversen and Kroonen 2017). This argument has many counter-examples, however: a word for an animal as small and unimportant as a hornet is positively reconstructible to PIE. Moreover, assuming the etymology of a word on the basis of its semantics can lead to circular reasoning. This study aims to establish how the Celtic substrate lexicon may aid in reconstructing how Celtic speakers interacted with the substratum population. The usage of any semantic criteria in establishing the substrate lexicon would entail a presupposition regarding the nature of this contact.

2.3 Stratification

In order to establish where and when Celtic acquired its substrate lexicon, one needs criteria with which the borrowing of individual words can be dated. These are laid out here.

The first and foremost dating criterion is the relative chronology of sound changes in the target languages. When a word becomes more similar to its comparanda after reconstruction to a stage before a particular sound change, then the word appears borrowed before the operation of this sound change. A concrete example is found in GOAT/SHEEP discussed in the next section, where OIr. *cáera* is best reconciled with its comparanda containing *p* by assuming a pre-form containing **p*. The same reasoning holds for W *celwrn* (VESSEL), where **p* is lost while present in e.g. Lat. *calpar*. Both words must be borrowed into Celtic before loss of PIE **p*. A much later date may be assigned to e.g. OIr. *partán*, which can only have been borrowed after Irish regained /p/. The limitation of this criterion is that some words cannot be dated because they did not undergo dateable prehistoric sound changes.

Another criterion to date borrowing is by means of the distribution of irregularities among languages attesting to a word. When all comparanda within a branch of IE can be reconciled to a single reconstruction to its proto-language,

then borrowing appears to predate the proto-language. When irregularities are found within a branch, borrowing seems to postdate this proto-language. This principle may be illustrated by contrasting HOLLY with ELM. In HOLLY, OIr. *cuilenn*, W *celyn* and Bret. *kelenn* can be reconciled to a single Proto-Celtic reconstruction, implying that the word was borrowed before Proto-Celtic broke up. In ELM, OIr. *lem* and W *llwyf* cannot be reconciled to a single Proto-Celtic reconstruction, implying that the word entered Celtic after Proto-Celtic. This criterion can be misleading, however, when a word is borrowed multiple times in a single language. Both GOAT/SHEEP and CUP/HEAD appear independently borrowed into Celtic more than once, but in both cases there are Celtic reflexes showing IE treatment of **p*, implying both words were borrowed into Celtic before Proto-Celtic.

A somewhat weaker criterion to date borrowing is by means of its geographical distribution. A word with a pan-European distribution is likely to be borrowed earlier than a word restricted to western Europe. This criterion suggests that OIr. *sraib* (LIGHTNING/SULPHUR) has a longer pedigree in Celtic than MIr. *coirce* (OATS). OIr. *sraib* has comparanda in Greek, suggesting that it entered a language ancestral to Irish at a stage when it was geographically closer to the ancestor of Greek. Such a stage must have been rather soon after the Indo-Europeanization of Europe. MIr. *coirce* has cognates in Germanic, and because Celtic and Germanic are known to have been neighbouring languages in later prehistory, this word, with a northwest European distribution, can be a later borrowing *in situ*. However, this criterion can be misleading in the case of widespread *Wanderwörter*, and lexical loss may obscure the original distribution of a now-restricted word.

It is also possible to date the borrowing of a word with reference to its meaning. A word may be assumed to be borrowed after the invention or adoption of its corresponding concept. RYE and OATS refer to crops that were domesticated only in the later Bronze Age leading into the Iron Age (Stika and Heiss 2013: 362), so borrowing of the words must postdate this period if it is assumed that borrowing postdates domestication. Similarly SILVER and LEAD can hardly have been borrowed in a period when speakers were unfamiliar with these metals.

Finally, one may iteratively stratify substrate words with reference to their substrate alternations. The borrowing of most words can be dated with some of the aforementioned criteria, and consequently the borrowing of substrate alternations found in these words can be dated. If a substrate alternation consistently appears borrowed in a particular stratum, then this feature may be used to date the borrowing of otherwise undateable substrate words. One of the aims of this study is to establish the date of the various substrate alternations for this purpose. The results are found in Section 4.13.

3 The corpus

In each entry, a set of Indo-European or Celtic pseudo-cognates is given with reconstructions into quasi-Proto-Indo-European (qPIE) or quasi-Proto-Celtic (qPC). These reconstructions show how reconstructing the forms into their respective proto-languages make them look more similar to their pseudo-cognates. At the same time, differences in their reconstructions show that no single form may be reconstructed for the proto-language, so that shared inheritance may be excluded and that the word is borrowed into different languages after the dissolution of the proto-language. Reconstructions into quasi-proto-languages do not include all formally possible reconstructions, but only those that increase the similarity between pseudo-cognates. For instance in COOT, PGm. **balikōn-* allows for a reconstruction with either qPIE **a* or **o* in the first syllable, but **a* is incompatible with and/or further removed from its extra-Germanic comparanda than **o*, and therefore only the latter reconstruction is given. Reconstructions into qPIE are written without laryngeals, so what could be reconstructed as PIE **h₂e* is given as **a* unless there are reasons to do otherwise. Reconstructions into both qPIE and qPC may contain features that are (nearly) unattested in the proto-language. For example, qPIE reconstructions may contain geminated consonants, **a*, **b*, and disyllabic roots, even though these features were rare to non-existent in PIE. Reconstructions into Proto-Goidelic containing **p* are similarly anachronistic. The fact that many reconstructions require features not found in PIE or PC again attests to the borrowed nature of such words. Reconstructions into qPIE do not contain features that were added after borrowing as adaptation to target-language morphology such as case endings.[7] A question mark before a reconstructed form indicates that its appurtenance to the other words under the lemma is doubtful.

3.1 Badger

qPC **tazgo-* > MIr. PN *Tadg* (name of king with badger totem)
qPC **task(i)o-* > Gaul. PN *Tascos*, Ancient Brittonic PN *Tasciouanos*, OW PN *Teuh(u)ant*[8]

qPIE **taKs-* > PGm. **þahsu-* > MDu. *das*, MHG *dahs* 'badger'

7 GARLIC constitutes an exception where comparison between Celtic *u*-stem flexion with stems ending in **-us* in other languages suggests that this flexion reflects a feature of the donor language.

8 For the appurtenance of OW *Teuh(u)ant* to this name, see Koch (1992).

Goidelic **zg* against Gallo-Brittonic **sk* can be taken at face value as a voicing alternation, or they may be resolved within Celtic by assuming a formation **tazg-sko-* for the latter (Stifter, p.c.).[9] A further irregularity in the form of metathesis may be observed between Germanic **Ks* and Celtic **sK*. One may surmise from these irregularities that this word was borrowed independently into Goidelic, Brittonic and Germanic from a third language after the breakup of Proto-Celtic and before the Germanic consonant shifts (van Sluis, Jørgensen and Kroonen 2023). Outside Indo-European, there is Bsq. *aizkon* 'badger' whose source is ambiguous. It shows loss of initial **t-* when compared to the remaining comparanda, and initial voiceless stops were regularly lost early on in the prehistory of Basque (cf. Trask 1997: 180ff; Koch 2020: 15). It is therefore possible that *aizkon* is an early borrowing in unclear direction with western IE or the European substrate (Trask 1997: 136).

3.2 Bat

qPC **eittVlon-* > OIr. *íatlu*, MoIr. *ialtóg*, ScG *ialtag* 'bat'
qPC **eiss/t(V)loi/un-b/mo-* > W *ystlum* 'bat'

Matasović (2012: 157n) notes that a word for 'bat' without an IE etymology is shared between Brittonic and Goidelic. When reconstructed, two irregularities may be observed: W *-st-* reconstructs to **-st-*, but Ir. *-t-* reconstructs to **-tt-*;[10] the Irish *n*-stem ending reconstructs to PC nom. **-ū*, gen. *-nos*, while W *-um* suggests PC **oi/um/bV(C)*. The irregularity **st* ~ **tt* can be interpreted as an approximation of a foreign phoneme, perhaps [tt^{s}]. The vocalism of Irish *n*-stem inflexion against W *-um* is best reconciled to a foreign [ū] or [ō] that was borrowed after PC **ū* > **ī* in Brittonic, while its consonantism can be analyzed as a geminate-simplex alternation whereby a substratal [m] became Goidelic **-n* in absolute auslaut, while it was borrowed as a geminate and thematicized in Brittonic, resisting the effects of lenition and apocope. For the onset, a single reconstruction may be offered: assuming that W *y-* is not prothetic, it can be reconciled with OIr. *ía-* to PC **ei* (later **ē*), because it became W *y* in unstressed heavy syllables after syncope, cf. PC

9 EDPC (372) suggests that the voiced cluster in Goidelic is from dissimilation with the initial consonant, but as a language-internal explanation this is *ad hoc*.

10 David Stifter (p.c.) suggests that the *t* in OIr. *íatlu* can also stand for /d/, implying PGoid. **dd* or **nt*. The voicing distinction is neutralized in modern forms with metathesized *-lt-*, cf. OIr. *rét-glu* 'star' (with /d/) > MoIr. *réalt*. Of course, such a reconstruction removes *íatlu* even further from W *ystlum*.

**bleidaniiās* > W *blynedd* (SBCHP 251). A substratal **ētt^{s}(V)lūm(b)* may thus be adduced.[11]

3.3 Bee

qPIE **b(ʰ)ek-* > PC **beko-* > OIr. *bech* 'bee', W *beg-egyr* 'drone'
qPIE **b(ʰ)i(t)-* > PC **bitamon-* > OIr. *bethamain* (nom. pl.), W *bydaf* 'beehive'

qPIE **bʰi-on-* > PGm. **bīōn-* > ON *býr*, OE *bēo*, OHG *bīa* 'bee'
qPIE **bʰit-* > Lith. *bìtė*, *bìtis*, *bitìs*, Latv. *bite*, OPr. *bitte* 'bee'
qPIE **bʰi/ek-eleh₂-* > PSl. **bъčelà* > OCS *bъčela*, *bъčela*, Ru. *pčelá*, *bčelá*, Cz. *včela*, SCr. *pčèla* 'bee'
?qPIE **a-pi-* > Lat. *apis* 'bee'

An element **bʰi/e-* may be reconstructed, with various expansions in **k* and **t*.[12] Outside IE, it may be compared to Egyptian *bj.t* 'bee', and expansions in **t* can be explained as continuing the Egyptian feminine suffix *.t* along with the stem. However, direct borrowing from Egyptian is difficult to substantiate on geographical grounds, requires borrowing from two forms both with and without the suffix, and it cannot account for the remainder of the alternations between the IE languages. The appurtenance of Lat. *apis* is uncertain. It can be adduced with the assumption of a voicing alternation **bʰ ~ *p* as well as the existence of an *a*-prefix in *apis*. One must therefore adduce not one but two alternations not otherwise found in this lexeme to accept the comparandum, and there is little phonological material to further substantiate the comparison. A putative Iranian **baina-* 'fly, bee' is unrelated (Witczak 2005: 208), as are Lat. *fūcus* 'drone, gadfly', OE *bēaw*, LG *bau* 'gadfly' (van Sluis 2022).

3.4 Beer

qPIE **kurmi-* > PC **kurmi-* > OIr. *cuirm*, W *cwrw*, OCorn. *coruf* gl. *ceruisia*, Gaul. *curmi* 'beer'
qPIE **krem-* > Lat. *cremor* 'thick broth'
?qPIE **ker(e)u̯-* > Lat. *cervisia*, *cerevisia* 'beer'

11 Its phonological structure (C)VC(C)VNC- may be compared to other substrate words in Germanic and Balto-Slavic **KolVmb-* 'pigeon', **jerVmb-* 'hazel-grouse', **albVnd-* ~ **lebVnd-* 'swan' and others (Jakob, this volume).
12 For the Proto-Celtic **e*-vocalism, see LEIA (B-24) and van Sluis (2022).

McCone (2005: 404–5) adduces these comparanda as non-IE borrowings and reconstructs **kVr(V)-*, but a more precise reconstruction seems to be **kVreu̯/m-*.[13] EDPC (216) and EDLI (142) reconstruct PC **kormi-*, but this reconstruction is to be rejected in favour of **kurmi-*: Gaul. *curmi* suggests the latter, and the parallel of PC **mori-* > W *môr* 'sea' suggests that short **i* did not cause *i*-affection of **o* (SBCHP 268). The *u*-vocalism in PC **kurmi-* precludes any PIE qualitative ablaut with Lat. *cremor* 'thick broth'. Adducing the Lat. *cremor* to PC **kurmi-* introduces an alternation in vocalism suggesting a shared substrate origin. Although Lat. *cerevisia*, *cervesia* is said by Pliny to be from Gaulish, it is not actually attested in a Celtic language. Still it must be borrowed somehow into Latin after intervocalic **s* became *r*. DLG (133) suggests that *cer(e)visia* may be from the same source as *cremor* with lenition of *m*, but *m* is not in a position to become *v* here in Gaulish. Thus adducing *cer(e)visia* to **kurmi-* and *cremor* introduces an alternation between **m* and **u̯* as well as in vocalism. Even then, the ending *-isia* in *cer(e)visia* is unetymologized under this account, and it is for this reason that its appurtenance remains doubtful.

3.5 Berry

qPIE **bak-* > PC **bak-* > W *bagad* 'cluster (of flowers, berries), OBret. *bacat* gl. *bacca* 'berry', LCorn. *bagaz* (*eithin*) 'bush (of gorse)', W *bagwy* 'cluster'

qPIE **bāk-* (~**bakk-*) > Lat. *bāca* 'berry', It. *bacca*

A connection between Celtic and Latin is rather likely, but the Celtic short vowel does not match the Latin long vowel, so a scenario of shared inheritance or mutual borrowing can be excluded. A reconstruction to PIE at any rate requires the marginal phonemes **a* and **b*, implying that the word is not inherited. ScG *bagaid* 'cluster' is most likely borrowed from a British Celtic dialect; this is confirmed by its meaning 'cluster' otherwise only found in Welsh. There are several productive suffixes with the shape *-ad* in Brittonic, but the suffix *-wy* is more obscure; the presence of both suffixes shows that the stem is **bak-*. Latin *bāca*, with a long vowel and a simplex consonant, alternates with Proto-Romance **bacca* with a short vowel and a geminate. This alternation may be the result of the *littera* rule

13 This reconstruction is hardly reconcilable to McCone's idea that Bsq. *gari* 'wheat' is related; all the more so because compounds such as *galburu* 'ear of wheat', containing *buru* 'head' show that the *r* in *gari* regularly descends from **l* in intervocalic position (Michelena 1990: 311–318).

whereby long vowels followed by a simplex consonant become short vowels followed by a geminate consonant, although Weiss (2010) argues that this rule only operated on high vowels. Outside IE, Proto-Berber **bqā-* 'blackberry, mulberry' < ?**βqā* may be adduced (Boutkan and Kossmann 1999: 88; EDLI 67).

3.6 Blackbird

qPIE **mes(a)l-* > PC **mesal-skā* > W *mwyalch*, OCorn. *moelh*, Bret. *moualc'h* 'blackbird'

qPIE **mes(V)l-* > Lat. *merula* 'blackbird'
qPIE **a-msl-* > PGm. **amslōn-* > OE *ōsle*, OHG *amsala* 'blackbird'

A Germanic *a*-prefixed and syncopated in **a-msl-* contrasts with an Italo-Celtic prefixless and non-syncopated **mesl-* (EDLI 375, EDPC 268, EDPG 25). OIr. *smolach*, MoIr. *smólach*, *smaol*, ScG *smeòrach* 'thrush' appear borrowed from Brittonic, if an Old British **moialx* was adopted as *moilax* (Schrijver 1997: 307–308). The Celtic suffix **-sk-* denoting animals frequently occurs in words of unknown etymology (Stifter 2023b), but may not by itself suggest a substrate origin, and is found with inherited roots if OIr. *loscann* 'toad, frog' < PC **ɸlu-sk-* is from PIE **pleu-* 'to float, swim' (Marstrander 1908).

3.7 Bread

qPC **baregi-* > W, Bret., Corn. *bara* 'bread'
qPC **bareginā* > OIr. *bairgen* 'bread'

qPIE **bʰars-* > PIt. **fars-* > Lat. *far*, gen. *farris* 'grain, spelt; grits', Umb. *far* 'flour, meal'
qPIE **bʰar(e/i)s-* > PGm. **bar(i)z-* > Go. *bariz-eins* 'barley-', ON *barr* 'grain, barley', OE *bere* 'barley'
qPIE **bʰars-* > PSl. **borš-ьno-* > OCS *brašьno* 'food', Ru. dial. *bórošno* 'rye-flour', Sln. *brášno*, *brašnọ̑*, SCr. *brȁšno* 'flour, food'

The alternation between an **n* in OIr. *bairgen* and the lack thereof in Brittonic *bara* can point to an inherited adjectivizing **-no/ā-*. However, the meaning 'bread' is identical in both forms, so that the semantically empty *n*-suffix found in

plant-names can also be the source of this **n*. The alternation between Brittonic **-a-* and Goidelic **-e/i-* in the second syllable is reconcilable to **e* under the assumption of a Brittonic **e* > **a* before **ge*, **gi* (SBCHP 134–141). Outside Celtic, a likely non-IE **bʰar(e)s-* referring to various cereals or flour, can be adduced (IEW 359–366).

3.8 Bull

qPIE **tarṷo-* > PC **tarṷo-* > OIr. *tarb*, W *tarw*, Bret. *tarv*, MCorn. *tarow* 'bull'

qPIE **tauro-* > Lat. *taurus*, Umb. *turuf* (acc.pl.), Osc. ταυρομ (acc.sg.) 'bull'
qPIE **tauro-* > Gr. ταῦρος 'bull'
qPIE **(s)teuro-* > **steura-* ~ **þeura-* > Go. *stiur*, OE *stēor*, OHG *stior*, ON *þjórr* (castrated) bull'
qPIE **tauro-* > PBSl. **tauró-* > Lith. *taũras*, OCS *turъ* 'aurochs'
qPIE **ta(u)ro-* > Alb. *ter*[14] 'young bull'

Celtic **tarṷo-* shows metathesis with its comparanda **(s)tVuro-*.[15] The intra-Germanic alternation between forms with and without initial **s-* is explained by EDPG (478) as different sound-substitutions of a non-IE consonant, e.g. [θ]. The first element of Etruscan *θevrumines* 'minotaur' provides a suitable parallel to PGm. **steura-* ~ **þeura-*. Av. *staora-* 'cattle' is unrelated, as it seems derived from PIE **steh₂uro-* 'big' (EDPG 478), a root connection incompatible with Germanic and Albanian. A PIE reconstruction would require PIE **a*, and not a laryngeal. The only laryngeal compatible with the Germanic and Albanian *e*-vocalism (**h₁*) is incompatible with Greek. The word ultimately appears to be a *Wanderwort* also found outside Indo-European in Etr. *θevrumines* 'minotaur' and Proto-Semitic **ṯawr* > Akk. *šūru*, Arab. *ṯawr*, Hebr. *šôr* 'steer'. Its date of borrowing into most IE branches must be early in view of its regular reconstruction to Proto-Celtic, Proto-Italic, et cetera. Only Germanic preserves an alternation within the branch, but this may be a secondary *s*-mobile.

14 The *e*-vocalism is the result of later umlaut (Demiraj 1997: 384).
15 Conversely, Lat. *taurus* has unexpectedly not undergone the metathesis *-aur-* > *-arṷ-* found in *parvum* and *nervus* (EDLI 607), but this metathesis may not have been exceptionless, as shown by Lat. *īnstaurō* 'to repeat, restore' < PIE **st(e)h₂uro-* 'big' (EDLI 305). If it was exceptionless, *taurus* appears to be borrowed into Italic after this metathesis operated.

3.9 Clover

qPIE **semh₁r-* > PC **semm(V)rā/i-* > OIr. *semar*, *semair* 'clover, shamrock'
qPIE **sm(e)l-sti-* > PC **(s)me/al(s)ti-on-* > W *meillion*, Bret. *melchon* 'clover'
qPIE **u̯i-sum(a)r-* > PC **u̯i-sumaro-* > Gaul. *uisumarus* 'clover'

qPIE **smeh₁r-* > PGm. **smēran-* > Icel. *smári* 'clover', PGm. **smērjōn-* > Icel., Far. *smæra*, Nw., Da. *smære*, Sw. dial. *smäre* 'clover'

The Germano-Celtic comparison is difficult to maintain in IE terms (LEIA S-83, Schrijver 1997: 304): variation in vowel placement between OIr. *semar* and Icel. *smári* can only be accounted for in IE terms by assuming a base **smh₁-r-* with placement of the vowel on either side of the **m*, but even then this root is otherwise unattested. W *meillion*, Bret. *melchon*[16] 'clover' are best accounted for through a reconstruction **(s)mel(s)tion-*.[17] This suggests presence of the substratal **-(s)ti-*suffix. Gaul. *uisumarus* 'clover' has been derived from PC **u̯isu-* 'poison' and **māro-* 'great', supposedly because clover was used as an antidote (EDPC 424–425). However, it is unexpected to call an antidote 'great poison', and the precise semantic match with OIr. *semar* suggests a segmentation **ui-sumarus*, with a **u̯i-*prefix, cf. PGm. **wisund-* ~ Lith. *stumbras* 'bison' (DveV 26; Kroonen 2012a: 252–255). The prefix appears to be late and geographically restricted, because the alternation is found within Celtic and the prefix is otherwise restricted to Northern Europe.[18]

3.10 Coot

qPIE **bʰo/ula/okk-* > PC **bo/ula/okk-* > ScG *bolachdan* 'coot'

qPIE **bʰulik-* > Lat. *fulica* 'coot'

16 An OBret. *multion* is found glossing *qaleta*; here the glossed text is corrupt for *lucinnia calta*. Another hand writes in the margin: *lucinnia .calta.fructus niger*. Latin *caltha* probably refers to *Calendula officinalis*, but the text is corrupt and the glossator may have misinterpreted the Latin, so the meaning and appurtenance of OBret. *multion* cannot be verified (Fleuriot 1964: 261).
17 The reconstruction **smeli̯on-* suggested by Koch (2020: 73) is to be rejected, because Bret. *-lch-* suggests **-l(s)tiV-*, cf. OBret. *guiltiat* 'tonsure' > MoB *gwilchad*. W *meillion* for expected **meilltion* is paralleled by e.g. *eillt* 'unfree subject', but pl. *eillon*.
18 There is otherwise no evidence for the **u̯i-*prefix in Celtic. Perhaps the language with this prefix was in contact with Continental Celtic but not Insular Celtic; its rarity may then be the result of the poor transmission of Continental Celtic.

qPIE *b^holig- > PGm. *balikōn- > OS *beliko*, OHG *belihha*, MHG *belche*, G *Bölch*, *Belche* 'coot'

ScG *bolachdan* appears formed with the diminutive suffix *-an* < OIr. *-án*, but the stem *bolachd-* resists analysis.[19] ScG *bolachdan* is documented from the Gaelic of the southern Hebridian island Colonsay (McNeill 1910: 17); here, the reflexes of OIr. *-cht* and *-cc* merge into [xg]. Comparison of *bolachdan* with PGm. *balikōn- and Lat. *fulica* 'coot' suggests PC *bo/ula/okko-, yielding OIr. *-cc*. This reconstruction differs from the Germanic and Latin only in that the stem-final velar is geminated and in having PC *a/o against *i elsewhere. Germanic and Latin are moreover formally incongruent with each other in that the former has qPIE *a/o in the first syllable, and a stem-final qPIE *g, while the latter has *u and *k (EDPG 50, EDLI 248).

3.11 Cup/Head

qPIE *kapu-k- > PC *kaɸuko- > OIr. *cuäch*, W *cawg* 'cup, dish'
qPIE *kapu-t- > PC *kaɸuto- > OIr. *cuäd* 'cup, mug'

qPIE *kaput- > Lat. *caput* 'head'
qPIE *ka(u)pe/ut- > PGm. *ha(u)beda- ~ *ha(u)buda- > Go. *haubiþ*, ON *hǫfuð*, OE *hæfud*, *hafud*, *hēafod*, OHG *houbit* 'head'
qPIE *kapV(t)l-[20] > PGm. *hafe/a/ulan- > OE *hafola*, *-ala*, *-ela*, OFri. *heila*, *holla* 'head'
?qPIE *kapo/ā/ēl- > Skt. *kapā́la-* 'cup, jar, dish'

The intra-Celtic alternation between *k and *t in *cuäch* and *cuäd* is difficult to account for in IE terms (Boutkan 1998: 111). They cannot be analysed as inherited roots with the past participle suffix *-to- and the appurtenance suffix *-ko-: this would imply a PIE root *kH/apu-, but roots of this shape are impossible. A PC reconstruction *kaɸut-ko- with loss of the dental is similarly unlikely, because such a cluster would be expected to yield a geminate, cf. PC *ad-kid̯ā > OIr. *aicce* 'nearness', W *ach* 'lineage'. The relationship between the two is nevertheless unmistakable in view of a shared element *kapu- and the identical meaning 'cup'. These

19 It is formally possible to extract the abstract suffix *-achd* from *bolachd-*, but there is no semantically relevant base to attach this to.
20 A qPIE *t may have been lost regularly in this position: the parallel of PIE *deh_2u-t/d^hlo- > PGm. *tōla- 'tool' shows that the cluster *-tl- would regularly assimilate to *-ll-, which would in turn degeminate to *-l- in an overlong syllable (EDPG 520). A parallel to this formation could be Lat. *capitulum* 'little cup'.

two words may be further compared to European words for 'head' containing intervocalic **p*; the correspondence between a hiatus between vowels in Irish and *p* outside of Celtic suggests that this word was borrowed into Celtic twice before the Proto-Celtic loss of PIE **p*. For the semantic shift 'cup' (Celtic, Sanskrit) to 'head' (Latin, Germanic), compare Dutch *kop* 'cup; head', from Late Lat. *cuppa*, and Fr. *tête* 'head', from Lat. *testa* 'pot'.

Beekes (1996: 218–220) compares Lat. *caput* and PGm. **ha(u)buda-* only, and notes that on face value a root vowel **a* is found. Beekes moreover considers *a~au* alternation a feature of non-IE borrowing, and adduces Lat. *caupō* 'merchant' and Gr. κάπηλος 'retail-dealer, huckster' as a parallel. EDPG (215) accounts for the *a~au*-alternation by positing an old proterodynamic paradigm nom. **hafuþ* < **kh₂p-ut*, gen. **habweþaz* < **kh₂p-uet-(o)s*, with a secondary root **haub-* from metathesis of oblique cases in **habw-*. Such a paradigm would allow for the word to be connect to a European root **kap-* or **kh₂ep-* 'to hold'. However, Boutkan (1995: 2) remarks that there are no cases of ablaut within suffixes ending in a stop.

Within Germanic, OE *hafola*, *-ela*, OFri. *heila*, *holla* 'head' < **kapV(t)l-on-* also appears related. If so, Germanic provides further evidence that the first element is to be segmented **kapu-t/k/l* rather than **kap-ut-* (Schrijver 1997: 295). This speaks against derivation from a European **kap-* or **kh₂ep-* 'to hold'. If Skt. *kapā́la-* 'cup, jar, dish' is adduced (*contra* Boutkan 1998: 111), we may be dealing with a wide-ranging or very early *Wanderwort*; at any rate the relationship between *kapā́la-* and the other words cannot be one of shared inheritance from PIE, except under the assumption of a PIE phoneme **a*, because *kapā́la-* is incompatible with **kh₂p-*, **kep-* or **kop-*.

3.12 Drone

qPIE **tron-* > PC **s(t)a-tron-* > OBret. *satron*, Bret. *sardon*, OCorn. *sudron*[21]

qPIE **dʰren-* > PGm. **drenan-* > OHG *treno*, OS *dreno* 'drone',
qPIE **dʰron-* > PGm. **drana(n)-* > OE *dran*, *drane*, *dræn* gl. *fucus*, OS *drano*, G *Tran* 'drone'
qPIE **tron-* > Lith. *trãnas*, Latv. *trans* 'drone'

21 The element *sa-* in OBret. *satron* remains unexplained. Perhaps it continues PIE **steh₂-* 'to stand' in the meaning 'to be lazy, idle', a fitting description of the inactive drone. The regular outcome of OBret. *-tr-* in *satron* would be MBret. *-er-*. This means that the metathesis *-dr-* > *-rd-* observed in Breton must antedate this vocalization of the dental.

qPIE **tront-* > Slk. *trút*, SerbCS *trutъ* 'wasp', SCr. *trût*, Po. *truteń*, Ru. *trúten'* 'drone'[22]
qPIE **dʰrēn-* > Gr. θρήνη 'drone'
qPIE **dʰrōn-* > Gr. θρῶναξ 'drone'

The alternation between **t* and **dʰ* precludes a PIE origin for this word. Within Germanic, the variation in *e*-vocalism and *o*-vocalism can be accounted for by assuming an ablauting *n*-stem (Kroonen 2011: 152–153, EDPG 101), but the same alternation in vowel quality is found within Greek and an alternation in vowel quantity is found between Greek and Germanic. In Greek many variant forms are found (e.g. ἀνθρήνη, ἀνθηδών, ἀνθρηδών, τενθρήνη, τενθρηδών), but these can all be explained as secondary from θρήνη (van Sluis 2022).

3.13 Dung

qPC **ai(t)l-iko-* > OIr. *aílech*, MoIr. *aoilech*, ScG *aolach* 'dung, manure'; PC **aitl-ito/u-* > OIr. *aíledu* (acc. pl.) gl. *stercora* 'dung'
qPC **atl-* > W *hadl* 'ruined, decayed, rotten, weak'

qPIE **atl-* > PGm. **adl-* > OE *adela*, MDu. *adel*, G *Adel* 'filth, liquid manure'

Schrijver (1997: 305) suggests OIr. *aílech*, W *hadl* and OE *adela* are related through a shared substrate, and that they show a substratal alternation between **ai* in Goidelic and **a* in Brittonic, although the **t* is conjectural for Irish, and W *h-* must be secondary.

3.14 Elm

qPC **leu̯mo-* > PBrit. **lēmo-* > W *llwyf* 'elm'
qPC **le/imo-* > OIr. *lem*, MoIr. *leamhán*, ScG *leamhan* 'elm'

qPIE **a/olm-* > PGm. **alma-* > ON *almr* 'elm'
qPIE **elm-* > PGm. **elma-* > OE *elm*, OHG *elm-boum* 'elm'

22 Words ending in *-enь* can be analogous to Po. *szerszeń*, Ru. *šéršen'* 'hornet' or Ru. *slepén'* 'horsefly' (van Sluis 2022).

qPIE **(o)lm-* > Lat. *ulmus* 'elm'
qPIE **(i)l(i)m-* > PSl. **jьlьmъ* > Ru. *íl'm*, *ílem*, *íl'ma*, Po. *ilm*, *ilem*, LSrb. *lom*, Polab. *jelm* 'elm'

The irregularity within Celtic between Irish *lem* < PC **lemo-* or **limo-* and Welsh *llwyf* PC **leu̯mo-* or later **lēmo-* suggests that the word was borrowed into Celtic twice after Proto-Celtic.[23] Irish *lem* can be connected with some difficulty to the extra-Celtic comparanda by assuming both root ablaut and Schwebeablaut: **h₁lemo-* > OIr. *lem; *h₁elmo-* > OE *elm; *h₁lmo-* > Ru. *íl'm*, **h₁olmo-* > Lat. *ulmus*, ON *almr*. However W *llwyf* cannot be connected to these comparanda within an IE framework even under these assumptions.[24] Schrijver (1997: 311) instead compares variation in vowel location with the *a*-prefix, although the vocalism in the prefix is not restricted to *a* in this instance. The alternation between Celtic CV and VC elsewhere can also be analyzed as a metathesis; if so, this metathesis is a feature of the donor form rather than a borrowing-induced feature because it occurs in the seemingly independently borrowed Irish and Welsh forms.

3.15 Frequent/Many

qPIE **menekk-* > PC **menekki-* > OIr. *meinicc*, MoIr. *minic*, W *mynych*, MCorn. *men(o)ugh* 'frequent'

qPIE **mo/an-o/agʰ-* > PGm. **man-ag-* > Go. *manags*, ON *mangr*, OS *manug*, OHG *manug*, ODu. *manug* 'many', PWGm. **man-īg-* > OE *manig*, *monig*, *menig*, OHG *manig*, *menig*, MHG *menic*, *manic*, *-ec*, MLG *mennich man(n)ich*, OFri. *monich*, *manich*, MDu. *men(n)ich*, *man(n)ich* 'many'
qPIE **munogʰ-* > PSl. **mъnogъ* > OCS *mъnogъ*, Ru. *mnógij*, SCr. *mnȍgī* 'much, many, numerous'

A substratal **mVnVK-* may be reconstructed. The final velar **kk* in Celtic is voiceless, which differs from the voiced aspirated **gʰ* found in Balto-Slavic and Germanic. Celtic is also unique in that this velar is geminated here, although there is

23 Koch (2020: 74) furthermore adduces the Hispano-Celtic names *Lemaui*, Λεμαυων, *Lemava*, *Lemaus*, but their original meaning cannot be verified, as they are personal names.
24 W *llwyf* could be to a PIE **h₁lmo-* by assuming a regular development to PC **limo-* followed by the creation of a neo-full-grade **leimo-* through *Ablautentgleisung*. This scenario requires that ablaut remained productive well after Proto-Celtic, even affecting a thematic formation, but there is little evidence for productive nominal ablaut by this period.

no geminate-simplex contrast in this position in Germanic and Slavic.[25] LEIA (M-37) suggests a root **men-* also found in *méit* 'multitude', but this would leave the formation unexplained, and would be irreconcilable with the *u*-vocalism found in Slavic. Outside IE, Fi. *moni* 'many' can be adduced (EDPG 352), but in view of the probable existence of cognates in Permic it may be an accidental lookalike; its lack of a stem-final velar speaks against a shared etymon.[26]

3.16 Garlic

qPIE **kremu-* > PC **kremu-* > OIr. *crem*, *crim* 'wild garlic'
qPIE **kramV-* > PC **kramV-* > W *craf* 'wild garlic'

qPIE **kro/amus-* > PGm. **hramusan-*, **hramusjōn-* > OE *hromsa*, OS *hramusia*, MLG *ramese*, *remese* 'wild garlic'
qPIE **krom(m)us-* ~ **kremus-* > Gr. κρόμμυον, κρόμυον, Hsch. κρέμυον 'onion'
qPIE **kermus-* > Lith. *kermušẽ*, dial. *kermušà* 'tip of a drill, top of a flail, wild garlic'
qPIE **ḱ/kerm(o)us-* > Ru. *čeremšá*, SCr. *crȉjemuša*, *srȉjemuša* 'wild garlic'

The intra-Celtic variation **e*~**a* is not reconcilable to IE ablaut: a zero grade would yield W ***cryf*, cf. **kʷrmi-* > W *pryf* 'insect, larva'; a reconstruction with a laryngeal **ḱrHm-* would likely yield W ***craw(f)* and is anyway incompatible with other IE languages. The fact that two Proto-Celtic forms are reconstructible may mean that the ancestors of Irish and Welsh borrowed this word independently. It appears that a substratal **krVmus* was borrowed into Celtic as a *u*-stem, while in other IE languages this **krVmus* was taken to be the stem upon which a variety of native endings was added. Greek provides another irregularity in that the **-m-* was adopted as both a simplex and a geminate consonant (Beekes 2000: 29).

25 The West Germanic forms in **-ig* show regular suffix substitution (Boutkan 1998: 124–125; 2003: 23).

26 Boutkan (1998: 124–5) adduces OFri. *meni, mene* 'multitude' as a comparandum lacking a stem-final velar, separating it from OFri. *menie*, for which he reconstructs **manigī-*. This is decidedly uneconomical, particularly in view of the absence of comparanda without a velar elsewhere in Germanic.

3.17 Goat

qPIE **kad^{h}(V)l-* > PC **kad(V)lot-* > OIr. pl. *cadlaid*, Ir. *cadhla* 'goat'

qPIE **katVl-* > Lat. *catulus* 'young of an animal, puppy', Umb. *katel* 'certain sacrificial animal'
qPIE **kat/d^{h}Vl-*> PGm. **hadVlan-* > MHG *hatele* 'goat'

A substratal **kaTVl-* is attested with a voiced dental in Irish that is voiceless in Latin; the original voicing quality of the dental is ambiguous in Germanic (LEIA C-5, Blöndal 1989: 297, EDLI 98, EDPG 163).

3.18 Goat/Sheep

qPIE **g(^{h})ab/pr-* > PC **gabro-* > OIr. *gabor*, W *gafr*, OCorn. *gauar*, gl. *capra vel capella*, Bret. *gavr* 'goat', OIr. *gabor* (feminine) 'horse'
qPIE **kaper-* > PC **kaɸero-* > OIr. *cauru* 'sheep', W *caer-iwrch* 'roebuck'

qPIE **kapr-* > Lat. *caper*, Umb. *kaprum* 'he-goat, buck'
qPIE **kapr-* > Gr. κάπρος 'wild boar'
qPIE **kapr-* > PGm. **hafra-* > ON *hafr*, OE *hæfer* 'he-goat'

The Celtic onset **g-* against **k-* elsewhere suggests a borrowing scenario (EDLI 88, EDPG 198). Two forms are reconstructible for Proto-Celtic, and both forms are found in both Goidelic and Brittonic. This fact, as well as loss of **p* found in reflexes of PC **kaɸero-*, suggest that the word was borrowed twice already by Proto-Celtic, both of which survived independently in different meanings. The original meaning may have been '(male) goat' with a subsequent semantic broadening towards the male of similar animals in W *caeriwrch*; OIr. *cauru* appears to be a feminine derivative of this secondary meaning, with a semantic narrowing to 'sheep' (cf. Stifter 2020).

3.19 Heather

qPIE **u̯roiḱ-* > PC **u̯roiko/ā-* > OIr. *fróich*, *fráech*, MW *gwrug*, W *grug*, Gaul. **uroicā* (>> Bret. *brug*, Prov. *bruga*, Fr. *bruyère*) 'heather'

qPIE **u̯erḱ-* > PSl. **versъ* > Cz. *vřes*, Ru. *véresk*, *véres*, Cr. *vrijês* 'heather'
qPIE **u̯irǵʰ* > Lith. *vìržis*, Latv. *vîrzis* 'heather'
qPIE **(u̯)ereiḱ-* > Gr. ἐρείκη 'heather'

The variation in vocalism between the various reflexes cannot be accounted for in IE terms; the vocalism of Gr. ἐρείκη moreover implies a disyllabic root with two full grades (EDG 452). Lith. *vìržis* and Latv. *vîrzis* moreover alternate in voicing with the other languages, although LED (1680) provides a number of language-internal explanations for this.

3.20 Holly

qPIE **kolis-n-* > PC **kolinno-* > OIr. *cuilenn*, W *celyn*, OCorn. *kelin* gl. *ulcia*, MBret. *quelenn-enn* 'holly'
qPIE **kela(s)-str-* > PC **kelastr-* > Bret. *kelastrenn* 'holly branch'

qPIE **kulis-* > PGm. **hulisa-* > MDu. *huls*, OHG *hulis*, *huls* 'holly'
qPIE **kule/in-* > PGm. **hule/ina-* > OE *holegn*, *holen*, MoE *hollen*, *hollin*, *holm* 'holly'
qPIE **kul(V)p/bʰ(r/Vs)-* > PGm. **hulbra-* or **hulbe/aza-* > ON *hulfr* (>> ME *hulfere*, E dial. *hulver* 'holly')
qPIE **kē/ālastr-* > Gr. κήλαστρος 'holly'
qPIE **go/ust(i/u)li-* > Arm. *kostł* 'bird-lime', *kostłi* 'holly'
qPRom. **kulis-* > Poitevin *coux* 'holly'
qPRom. **kol(V)str-* > Pic. *keûstria* 'mistletoe'
qPRom. **golost(r)i-* > Srd. *golóstḍu*, *golóstri* 'holly'

An inherited etymology involving PIE **kel-* 'sharp' is to be rejected because it cannot regularly yield the majority of attested forms (EDPC 213). It rather appears that a substratal **kVlVs-* was borrowed into various European languages, sometimes with the suffixes *-n-* and *-str-* designating plant-names.

The geminate *n* in PC **kolinno-* is confirmed by MBret. *quelennenn;* PC **-nn-* is a regular outcome of **-sn-* and reconstruction of such a cluster most closely reconciles this form with a substratal **kVlVs-* found in various other languages of Europe (van Sluis, Jørgensen, and Kroonen 2023). This sound change postdates the

split between Hispano-Celtic and the remaining Celtic languages, allowing for the date of borrowing of **kolis-no-* to be placed before this split (Stifter 2023a). On the other hand, the absence of a regularly matching comparandum in any other branch of IE makes it unattractive to date borrowing to a period in which Celtic was still part of an undifferentiated Indo-European or Italo-Celtic proto-language. The borrowing of Bret. *kelastrenn* can date to any time after Joseph's rule (**eRa* > **aRa*) on account of its cluster *-ela-*. This rule is found across the Celtic languages, but not outside Celtic, and it must postdate the decomposition of syllabic nasals (EDPC 8). The decomposition of syllabic nasals itself postdates the Italo-Celtic split. Barring a scenario in which both Celtic holly-words were borrowed in widely differing periods, the only timeframe within which they could both be borrowed appears to be around Proto-Celtic.

Arm. *kostłi* 'holly' appears to be the product of a metathesis from **go/ul(i/u)-stiV-* to **go/ust(i/u)liV-*, while Arm. *kostł* 'bird-lime' seems back-formed from *kostłi* after *-i* was analyzed as a tree-suffix and *kostł* as its produce, cf. *kałin* 'acorn, hazelnut' > *kałni* 'oak, hazel tree'. Outside Indo-European, the word is attested in Bsq. *gorosti*.[27] Pre-Basque did not have word-initial voiceless stops, so the adoption in Basque as *gorosti* suggests that it entered Basque in a period when this restriction was still in place. This period includes Latin and early Romance loanwords, cf. *bake* 'peace' < Lat. *pacem*. The word has a western European and Mediterranean distribution, which parallels the distribution of the plant (*Ilex aquifolium*) itself.

Some of the forms in Germanic and Romance may be intra-IE borrowings, and therefore need not indicate direct borrowing from a substrate. OE *holen* may be a borrowing from PC **kolinno-* after Celtic **sn* > **nn* and before the Germanic consonant shift with sound substitution of PC **o* in the initial syllable with PGm. **u*. Fr. dial. *coux* 'holly' may be a borrowing from an early Germanic reflex of PGm. **hulisa-*, cf. PGm. **hamiþja-* > VLat. *camisia* 'shirt' for the substitution of PGm. **h-* with VLat. *c-*.

3.21 Hook

qPIE **bakk-* > PC **bakko-* > OIr. *bacc* 'bill-hook, angle, bend', W *bach*, Bret. *bac'h*, MCorn. *bagh* 'hook, peg'

qPIE **bak-(e)(t)l-* > Lat. *baculum* 'stick, staff', Lat. *bac(c)illum* 'small stick'

27 The *-r-* in *gorosti* need not be old: **VlV* regularly becomes *VrV*, cf. *(h)aizkora* 'axe' < Lat. *asciola*. Therefore this word cannot be identified as an instance of liquid alternation (Section 4.10).

qPIE **bak-tr-* > Gr. **βακτήρ* > Gr. βάκτρον ‘stick, cudgel’, βακτηρία ‘staff, cane’
qPIE **bak-* > PGm. **pagjō-* > MDu., ME *pegge* ‘pin, cone, peg’

The reconstructed phonemes **b* and **a* were marginal in PIE. This fact, as well as the alternation between Celtic **kk* and **k* elsewhere make the word unlikely to be inherited from PIE (LEIA B-2, EDLI 67, EDPC 52, EDPG 395). The meaning ‘hook, hooked instrument’ is the primary meaning found in Celtic, while the meaning ‘stick’ predominates elsewhere; its original meaning was probably ‘shepherd’s crook’.

The suffix of Lat. *baculum* can either be instrumental suffix **-tlo-*, *or* the agentive/diminutive suffix **-(e)lo-*. An instrumental suffix is also found in Gr. βάκτρον, but the shape of the suffix, **-tro-*, differs from **-tlo-* possibly found in Latin. Since the formations differ, only a shared stem **bak* can be reconstructed between them. PGm. **pagjō-* is also attested in the diminutive **pagila-* > OE *pægel* ‘wine-vessel’, E *pail*, MDu. *pegel*, *peil* ‘measuring-knob in vessel for liquid goods’ (EDPG 395).

3.22 Lark

qPC **alaudā* > Gaul. **alauda-* > Gallo-Lat. *alauda* ‘lark’

qPGm. **laiwiz-akōn-* > OE *lāwrice*, WFri. *ljurk*, OHG *lērahha* ‘lark’

The correspondence **d* ~ **z* can be analyzed as a stop-fricative alternation. If the word was borrowed into Celtic when a PC **d* was already pronounced [ð] intervocalically, and into Germanic before dental fricatives appeared, a single donor form with [ð] may be reconstructed. The *a*-prefix in Celtic is accompanied by syncope in the following syllable when compared to Germanic, and Celtic **a* alternates with Germanic **ai* (Schrijver 1997: 309–310, EDPG 324).

3.23 Lead

qPIE **ple/ou̯dʰi̯-* > PC **ɸloudi̯o-* > MIr. *lúaide* ‘lead’

qPIE **plo/undʰu̯-* > Lat. *plumbum* ‘lead’
?qPIE **moliu̯d- (vel sim.)*> Myc. *mo-ri-wo-do* /moliwdos/, Gr. μόλυβδος, μόλυβος, μόλιβος, βόλυβδος, βόλιμος, βόλιβος ‘lead’
?qPIE **mlīu̯-* > PGm. **blīwa-* > ON *blý*, OS *blī*, OHG *blīo* ‘lead’

The Celtic and Latin words for 'lead' have a similar shape, but neither form can be a direct borrowing from the other.[28] They rather appear borrowed from a form such as **pl(o)u(n)dʰ(u)-*. Huld (2012: 336) proposes a shared Italo-Celtic formation **plou-dʰ(H)om*, from PIE **pleu-* 'to flow, float', but the added nasal in Latin would be irregular. If a prehistoric borrowing is assumed, one may furthermore tentatively adduce a non-IE **m(V)liu̯(d)-* underlying PGm. **blīwa-* and Gr. μόλυβδος (Thorsø et al. 2023: 116).

Outside Indo-European, a similar-looking word is found in Berber and Basque. The Berber material, Ahaggar Tuareg *ăhâllun* 'tin, lead', Iwellemmeden Tuareg *aldom* 'tin', Ghat Tuareg *ahellum* 'lead', Kabyle *aldun* 'lead', Sous Berber *aldun* 'lead', Mzab *buldun* 'lead', etc., yields several Proto-Berber reconstructions: **βaldūn ~ *βāldūn ~ *būldūn ~ *βaldūm* (Boutkan and Kossmann 1999: 92–93). This variation suggests borrowing into Berber after Proto-Berber. Bsq. *berun* 'lead' potentially reflects a form **bl(e)un(P)*, and as such it may be an independent early adoption from Italo-Celtic or its source, but a later Romance borrowing from e.g. Gasc. *ploum* cannot be excluded.

3.24 Leek/Garlic

qPIE **kasn-ē/īn-* > PC **kannīnā* > OIr. *cainnenn* 'garlic, leek', W *cennin* 'leek, daffodil', MBret. *quinghenn*, *quinhenn* 'onion', Bret. *kignen* 'garlic', OCorn. *kenin* gl. *allium*

qPIE **kesn-* > PSl. **česnъ*; **česno*; **česnъkъ* > Ru. *česnók*, Cr. *čèšnjak* 'garlic'

Schrijver (1995; 2000a) suggests a substrate etymon because the Celtic *a*-vocalism is irreconcilable with Slavic *e*-vocalism. The Celtic expansion in *-n-* may moreover be reconciled with a substratal *n*-suffix. PSl. **česn-* has alternatively been derived from PSl. **česàti* 'to scratch, comb', but this is difficult to sustain on semantic grounds. Falileyev and Isaac (2003) reject the substratum etymology on the grounds that a PIE **kes-no-*, to the root of the aforementioned verb, could regularly yield the attested Slavic forms in the *e*-grade. For Celtic, they propose that a zero-grade **ksno-* would regularly yield **kasno-* with regular anaptyxis in initial consonant clusters. However their parallels for this anaptyxis always involve two stops, and cannot be relied upon.

28 Proto-Celtic **ϕloudio-* is later borrowed as Proto-(West-)Germanic **lauda-* > OE *lēad*, OFri. *lād*, Du. *lood*, G *Lot* (van Sluis, Jørgensen, and Kroonen 2023).

3.25 Leg

qPIE **g(ʰ)(a)rs-* > PC **garri/o/ā-* > OIr. *gairr* 'calf of the leg', W *gar*, Bret. *garr* 'leg', Gaul. **garrā* >> OFr. *gare* 'leg', Prov. *garra* 'hollow of knee'

qPIE **a-k(a)r-* > Gr. ἄκαρα 'legs' (Cretan, Hesychius)

Schrijver (1997) argues for a substrate etymon with voicing alternation and an *a-* prefix. A competing etymology by EDPC (152) requires metathesis of PIE **ǵʰesr-* 'hand' to **ǵʰers-* of which a thematicized zero grade **grsV-* would yield PC **garrV-*, while shifting in meaning from 'hand' to 'leg'. Prov. *garra* 'hollow of knee' and related forms within Romance can be borrowed from a Gaulish cognate of W *gar*, Bret. *garr*, as can Bsq. *garro* 'tentacle', either directly or through Romance (FEW 4: 65ff).

3.26 Lightning/Sulphur

qPIE **strab(ʰ)-* > PC **strabi-* > OIr. *sraib*, *straif*, *straiph* 'sulphur', *sraiftine* 'lightning'

qPIE **a-str(a)p-* > Gr. ἀστραπή 'lightning'
qPIE **(a-)st(e)rop(ʰ)-* > Gr. στεροπή, ἀστεροπή, στροφή 'lightning'

A substrate origin for Celto-Greek isogloss may be assumed on account of the intra-Greek variation in voicing and *a*-prefixation (Schrijver 1997: 310, EDG 156, Beekes 2014: 47–48). Within Celtic OIr. *sraib* is likely the most archaic form: *-f* in OIr. *straif*, *straiph can be* backformed from *sraiftine* 'lightning', a compound with *téine* 'fire'. The original meaning was most likely 'spark' or 'lightning'. The semantic shift to 'sulphur' can be understood from the usefulness of sulphur in producing sparks to start a fire. The presence of the original meaning in compounds with *teine* 'fire' suggests a shift of markedness. Stokes (1888: 242, *apud* LEIA S-181) suggests that *sraib* is borrowed from Fr. *soufre* 'sulphur' with metathesis of *r*, but this poorly accounts for the *a*-vocalism.

3.27 Man/Hero

qPIE **karut-* > PC **karut-* > OIr. *caur*, gen. *curad* 'warrior, hero'

qPIE **kale/ut-* > PGm. **haleþ-* ~ **haluþ-* 'man, hero' > ON *hǫldr*, poet. *halr*, OE *hæleþ*, *heleþ*, OHG *helid* 'man, hero'

The correspondence Celtic **r*, Germanic **l* cannot be accounted for in PIE terms. PGm. **hale/uþ-* has also been related to ToB *kālśke*, *kālyśke* 'youth, young brahmin' < **kHl-sko-*, but this requires it to be segmented **kal-Vt-* (EDPG 204). The equation with Celtic requires no such assumption and is closer semantically. The intra-Germanic alternation of unstressed **-e* ~ **-u* is discussed in Section 5.2.

OIr. *caur* has alternatively been connected to W *cawr* 'giant; hero' and Gaulish PNs in *cauaro-* (LEIA C-50), but this proposal is formally problematic as it does not account for its dental flexion. The Germanic tribal name *Harudes*, Χαροῦδες can be reconciled to a pre-PGm. **karut-*, and as such it may be a direct borrowing from or into Celtic before the Germanic consonant shifts (Thurneysen 1917: 71). However, a link cannot be substantiated because the original semantics of onomastic material cannot be established.

3.28 Nit

qPIE **snid-* > PC **snidā* > OIr. *sned* gl. *lens*, W *nedd*, MBret. *nez*, LCorn. *nedhan* 'nit(s)'

qPIE **ḱonid-* > Gr. κονίς 'nit; flea; bug', Alb. *thërí*, *thëní* 'nit'
qPIE **h_2/h_3/a/o-(s)nid-* > Arm. *anic* 'nit'
qPIE **ḱnid-* > PGm. **hnit-* > ON *gnit*, OE *hnitu*, OHG *hniz*, *niz* 'nit'
qPIE **g(^{h})nid-* > PBSl. **gni?da?* > Latv. *gnīda*, Lith. *glìnda*, Ru. *gnída*, SCr. *gnjȉda*, Sln. *gnída* 'nit'

Various PIE reconstructions, such as **ḱonid-*, and **dḱnid-*, along with a number of contaminations or tabooistic sound changes have been proposed (EDLI 334, EDSIL 169), but irregularities found in the various reflexes are paralleled by other words of plausible non-IE origin, and the word is isolated to Europe and Armenia. Kroonen (2012a: 247) suggests that the variation between endings in **-ind-* and **-id-* allows for the identification of a substrate suffix **-īT* also found in Gr. -ινθος and possibly in RUSHES, but this would leave only a consonantal onset as the stem in some forms. The alternation between initial **s-*, **ḱ-*, **g-*

likely represents a palatal affricate in the source language. The onset of Arm. *anic* allows for identification of a substratal *a*-prefix. Alternatively Arm. *a*- can descend from an **o* (< **h₃*), cf. Gr. κονίς, Alb. *thërí*, but without the initial consonant.

EDPC (349) speculates that PC **snidā* may be regular from an inherited **sknid-*, but adduces no parallels for loss of **k* in this position, and admits that word-initial **kn-* is regularly preserved.[29] Lat. *lēns*, *-endis* 'nit' is often adduced with a reconstructed onset **gl-* on the basis that such an onset matches that of Lith. *glìnda*. However, the *-l-* for *-n-* in is likely secondary within Lithuanian, cf. Žemaitian *gnýda*. Consequently Lat. *lēns* is formally too dissimilar to be adduced.

3.29 Nut

qPIE **knu(H)-* > PC **knuu̯-* > OIr. *cnú*, W *cnau*, Bret. *kraoñ* 'nut(s)'
?qPIE **knut-* > PC **knutV-* > W *cnwd* 'produce', MBret. *cnod* 'offspring'

qPIE **knud-* > PGm. **hnut-* > ON *hnot*, OE *hnutu*, OHG *nuz* 'nut'
qPIE **knuk-* > Lat. *nux* 'nut'

The Celtic words for 'nut' can be reconciled under a single PC **knuu̯-*, which itself straightforwardly reconstructs to qPIE **knuH-* with a stem-final laryngeal. In Brittonic, **-uH-* in prevocalic position would yield PC **-uu̯-* > **-ou̯-*, correctly yielding the attested forms (SBCHP: 329–330). Under an alternative reconstruction without a laryngeal, a qPIE full-grade **-eu-* or **-ou-* would equally well yield the Brittonic forms. OIr. *cnú* can be reconciled with a laryngeal-less reconstruction by assuming a qPIE zero-grade **-u-* that was lengthened to *ú* in the nom. sg. as a result of vowel lengthening in absolute auslaut of stressed syllables. The alternation between what in IE terms appears to be **H* or zero, **d*, and **k* (Section 4.3) is difficult to reconcile in IE terms, and suggests borrowing from a non-IE source (EDLI 418, 420, EDPC 212, EDPG 237). A dental may also be found in Celtic **knutV-* > W *cnwd*, although its more general meaning 'produce' makes its appurtenance doubtful.

29 Similar clusters involving other resonants also speak against such simplification, cf. the preservation of **k* in PC **se-skrid-* > OIr. *scaird* 'strips' (Schumacher 2004: 582).

3.30 Oats

qPIE **korki-* > PC **korkio-* > OIr. *corcae*, MIr. *corca*, *coirce*, W *ceirch*, Bret. *kerc'h* 'oats', OCorn. *bara keirch* gl. *panis avena*

qPIE **kokr(i)-* > PGm. **hagrja-* > Da. *hejre* 'brome grass'; PGm. **hagran-* > OSw. *hagri*, Nw. dial. *hagre* 'oats'

A substratal **korki- ~ *kokri-* with metathesis may be reconstructed (EDPC 216; EDPG 199; van Sluis, Jørgensen and Kroonen 2023). There is moreover Ir. *coirce*, *cuirce* 'crest, tuft, headdress', which can be a semantic development of *coirce* 'oats' on account of the long awns hanging from the oat plants. Nw. dial. *hagr*, *harg* 'coarse hair from a horse's mane or tail' < **kokro ~ *korko-* shows a similar semantic shift. In both cases, the direction of the semantic shift is difficult to establish, so the original meaning may have been 'tuft'. The meaning 'oats' is more widespread in both Celtic and Germanic, however. Outside IE, Fi. *kattara* 'brome' may be related. If so, the interchange between *t* in Finnish and qPIE **k* elsewhere suggests an original glottal stop.

3.31 Pine

qPIE **gisust-* > PC **gisusto-* > OIr. *giús*, ScG *giuthas*, MoIr. *giumhas*, *giúis* 'fir-tree, pine'

qPIE **gisn-* > PGm. **kizna-* > OE *cēn* 'pine tree, spruce', MLG *kēn* 'pine cone, pinewood', OHG *kien* 'pine tree, pinewood torch'

A substrate element **gis(u)-* may be reconstructed with an *n*-suffix in Germanic, while the *st*-suffix in Celtic may be the *st(r)*-suffix (van Sluis, Jørgensen and Kroonen 2023).

3.32 Poplar

qPIE **a-pt(V)l-* > PC **axt(V)l/nV-* > MBret. *ezlen*, W *aethnen* 'poplar', OCorn. *aidlen* gl. *abies* 'fir'

qPIE **to/apol-* > PSl. **topolь* > OCS *topolь*, Ru. *tópol'*, Cz. *topol* 'poplar'
qPIE **tpel-* > Myc. *pte-re-wa*, Gr. πτελέα 'elm tree'
qPIE **a-pel-* > Gr. ἀπελλόν (Hesych.) 'black poplar'

?qPIE **tele/i̯-* > Lat. *tilia* 'linden tree'
?qPIE **teli-* > Arm. *tʿeł* 'elm'
?qPIE **pōpVl-* > Lat. *pōpulus* 'poplar'

W *aethnen*, MBret. *ezlen* 'poplar', OCorn. *aidlen* gl. *abies* 'fir' together allow for a reconstruction to PC **axt(V)l/nV-* < **aCt(V)l/nV-* followed by a feminine singulative suffix. The consonantism of MoBret. *evl*, 'poplars', sgtv. *evlenn* is most likely influenced by Lat. *ebulum* 'danewort (berry)' (Deshayes 2003: 223). OIr. *aidlen* is a ghost-word (LEIA A-27).

An element **tpel-* may be gathered from the Greek and Balto-Slavic comparanda. Proto-Greek **ptel-* > Myc. *pte-re-wa*, Gr. πτελέα 'elm tree' is probably borrowed into PIt. **teli̯ā* > Lat. *tilia* 'linden tree', and possibly also into Arm. *tʿeł* 'elm', with simplification of the onset πτ-. An *a*-prefix is likely found in **a-pel-* > Gr. ἀπελλόν (Hesych.) 'black poplar'. Lat. *pōpulus* 'poplar' can be adduced under the assumption of reduplication of the first syllable, as may the German toponym *Vielbaum* (if from PGm. **fī-fel-*); alternatively the first **t* was assimilated to **p* in anticipation of the next syllable (Matasović, this volume). Together they suggest that a substrate element **tpel-* was borrowed into various IE dialects with the *a*-prefix in Gr. ἀπελλόν and with various occasions where a consonant cluster **tp-* was simplified.

If the Celtic evidence belongs to this substratal **tpel-*, an *a*-prefixed donor form **a-ptlV-* or perhaps **a-tplV-*[30] would regularly yield PC **axtlV-*, which in turn regularly yields MBret. *ezlen*, suggesting that W *aethnen* with *-n-* is an innovation. There is a remote possibility that MoBret. *evlenn* continues a variant form with early simplification of **pt* to **p*, because **a-plV-* regularly yields PC **ablV-* > MoBret. *evl* if followed by an element causing *i*-affection.

3.33 Pubic Hair

qPC **katirā-* > OIr. *caither*, *caithir*, *cathair* 'body hair; adult'
qPC **keto/urā-* > MW *cedor* 'pubic hair'
qPC **ka(x)te/airV-* > LPBrit. **kaiθ/dē̯/ēr* > OW *caitoir* gl. *pube*
qPC **ka(xt/d)o/urV-* > LPBrit. **kaiθ/ðo/ur* > MBret. *quaezour*, Bret. *kêzour*, *kaezour* 'pubic hair; puberty'

30 PIE **-pt-* regularly yields the required PC **-xt-*, but it is uncertain what a cluster **-tp-* in world-medial position would yield in Celtic. The metathesis found in **-tk-* > **-(x)t-* in PIE **h₂rtḱo-* > PC **ar(x)to-* 'bear' suggests that PC **-xt-* may be the regular outcome outcome of **-tp-* through an intermediate stage **-pt-*.

The Celtic words for 'pubic hair' cannot be reconciled into a single PC form. OW *caitoir* and Bret. *kaezour* are most similar: If OW *t* stands for /θ/, they share a PBrit. **kaiθ-* < PC **kaxt-*, but the vocalism of the second syllable differs. OW *oi* may stand for W *wy* (< PC **ei* or **u* before secondary yod) or for W *oe* (< PC **ai* or **o* before secondary yod) in this position. However, each of these possibilities is incompatible with Bret. *kaezour*: the expected Breton reflex for W *wy* would be Bret. ***o(e)* (cf. Bret. *baradoz*, *baradoez* 'paradise', borrowed from British Latin **paradēsus*), and for W *oe* it would be MBret. ***oa*, yielding Bret. ***o* in final syllables. It has been suggested that OIr. *caither* was borrowed into Brittonic (VKG II: 50, Bauer 2018: 23), but it is not clear how *caither* could yield the Brittonic vocalism in either form.

W *cedor* does not appear to be a direct descendant of OW *caitoir* in view of the vocalism; if OW *t* stands for /d/, then it agrees with W *cedor* in consonantism, but then the consonantism of OW *caitoir* must be separated from Bret. *kaezour*. The further etymology is unknown. Borrowing from Lat. *charactēr* 'mark' (VKG II: 50) requires several *ad hoc* assumptions to work formally, and is semantically weak. It is conceivable that at least some of the intra-Celtic variation is due to taboo deformation in view of the semantic field. Still, the alternation in stem vocalism between OIr. *caither* on the one hand and OW *caitoir*, Bret. *kaezour* on the other can be analyzed as *a~ai* alternation (Section 4.6), although in this case it is Late British **ai* (< PC **ax/g/d*) alternating with **a*, rather than PC **ai*. This suggests borrowing into Brittonic after PC **ai* > Brit. **ɛ̄*, which can be dated to the first century CE (LHEB 324–330).

3.34 Raspberry

qPC **mab-* > W *mafon* 'raspberries'

qPRom. **a-mb-* > MFr. *ambre*, Wall. *õpr*, Prov. *ambra* 'raspberry'
qPRom. **a-mp-* > MFr. *ampe*, Rht.-Rom. *ámp(w)a*, Wall. *ãpun*, Prov. *ampoun*, Lad. *ampón*, Lomb. *ampóma*, Rht.-Rom. *ámpia* 'raspberry'
qPRom. **a-m-* > Wall. *amõn*, Fr.-Prov. *ẽmęr* 'raspberry'

A substrate lexeme **a-mP-* ~ **mab-* may be proposed, with an *a*-prefix and no vowel between the **m* and the labial in Romance, and without the *a*-prefix but with a medial vowel in Welsh. W *mafon* contains the plural suffix *-on* frequently found in names for berries, cf. *aeron* 'berries'. The following suffixes are found in Romance: collective **-(a)ra*, augmentative *-óne*, diminutive *-ul*, and the obscure **-wa* (FEW 21:93). W *afan* 'raspberries' may also be compared, either as a derivative of *mafon* if a lenited **fafan* lost its first *f* in dissimilation, or as a borrowing from (the source of) **a-m-* > Wall. *amõn*.

3.35 Rock

qPIE **k(ar)ra/ek-* > PC **kra/ekī* > W *craig* 'rock'
qPIE **karre/ik-* > PC **karre/ikā* > W *carreg*, Bret. *karreg* 'stone'
qPIE ?**karn-* > PC **karno-* > OIr., W, MCorn. *carn* 'heap of stones, tumulus', Gaul. *karnitu* 'has erected [a tombstone]'

qPIE **kark-* > PGm. **hargu-*[31] ~ **harha(n)-* > ON *hǫrgr* 'pile of rocks, sanctuary', OE *hearg* 'pagan temple, idol', OHG *harug* 'grove', Elfd. *ar* 'bedrock', Nw. *har(e)* 'cliff, rocky bottom', Du. dial. *hare* 'hillock'

It is possible to resolve the intra-Celtic variation between W *craig* and W *carreg* by assuming an original **karrek-* whose initial vowel was syncopated before a resonant, cf. W *crydd* < **cerydd* 'shoemaker'; a parallel case whereby syncope occurs before an original geminate **rr* is W *crach*, LCorn. *cragh* 'scabs', a likely borrowing from Ir. *carrach* 'scabby'. MIr. *crec*, *crac*, ScG *creag* 'crag, rock' appear borrowed from British Celtic, as does MIr. *carraic* 'rock' (VKG I: 23). This leaves an alternation between Celtic **karrek-* and pre-proto-Germanic **kark-*; if an onset **kar-* can be isolated, it is possible to furthermore connect PC **karno-* (OIr., W, MCorn. *carn* 'heap of stones, tumulus') and a Mediterranean **karrā* allegedly found in Bsq. *harri* 'rock' and Béarnais *carroc* 'rock' (FEW 2: 408–412), however the segmentation into **kar-* and **-k-*, **-n-*, respectively, appears arbitrary; identification of the IE adjectivizing suffixes **-ko-* and **-no-* is precluded by the formal irregularity between PC **karre/ikā* and PGm. **hargu-*, whose irregularity suggests that a form with a velar was already present in the substrate language.

3.36 Rowan

qPC **karV-(dīno-)* > W *ceri* 'service-tree', Old British *caerdin*, W *cerddin*, LCorn. *kerden*, Bret. *kerzhin* 'rowan'
qPC **kaḭrā-(tīno-)* > OIr. *cáer* 'globular mass', *cáerthann* 'rowan'

The relationship between *cáerthann* and its Brittonic counterparts cannot be one of shared inheritance on account of the alternation *a~ai*. Brittonic to Irish borrowing is suggested (VKG I 23, VKG II 659), but this is based on Welsh *cair*, a ghost-word (LEIA C-8). Loth (1920: 144) proposes an ancient alternation between

31 Pre-PGm. **karuko-* for *karku-* (LEIA C-40–C-42) is outdated (EDPG 211).

Irish *áe* < **ai* and Brittonic *e* < **a*.[32] SBCHP (251) cautions that Brittonic **a* may go back to an **ai* shortened in pretonic position, but evidence against pretonic shortening comes from Welsh *ceri* 'service-tree', which may reflect a form without the **dīno*-suffix.[33] Here, such pretonic shortening would also have to have happened, but without the suffix there is no closed syllable which seems to be a conditioning factor for at least some types of pretonic shortening.

3.37 Rushes

qPC **sem-* > OIr. *sim(a)* 'stalk, stem', *simin(n)*, *seimen(n/d)* 'rushes, reed'
qPC **seb-* > OIr. *sibin(n)*, *sifin(n)* 'rushes, reed'

qPIE **sem-(id-)* > PGm. **sem-(it-)* > OS *semith*, OHG *semida* 'rushes, reed', G *Simse* '(bul)rush'
qPIE **seb^{h_1}-* > PGm. **seb-* > ON *sif*, MHG *sebede* 'rushes, reed'

The Irish alternation between medial *b* and *m* can be explained by assuming an original cluster **bn* with assimilation of the nasal quality (Stifter 2015: 101), but it is also possible to assume a dissimilation of *m* to *b* before *n* (Ó Maolalaigh 2003: 127–128). However the fact that this same alternation is found within Germanic suggests a deeper origin (van Sluis, Jørgensen, and Kroonen 2023). LEIA (S-110) suggests that *siminn* is derived from inherited *sim* 'chain, loop', but the suggested semantic pathway from an instrument to a raw material is rather unexpected, and a shift in the reverse direction with backformation from *simin(n)* would be more likely. This *-in(n)* is synchronically analyzable as a diminutive suffix, but may also be compared to Gr. -ινθος found in substrate words. OS *semith*, OHG *semida* contain the **-eþ-* suffix denoting groups of trees and plants, but G *Simse* suggests PGm. **semĭt-*, and similarly invites comparison with Gr. -ινθος, cf. Gr. ἐρέβινθος 'chickpea' ~ PGm. **arwīt-* 'pea'.

32 The alternation also appears within Ogham inscriptions, cf. *maqi-cairatini* ~ *maqi-caratinn*, but this may be merely orthographical as in Ogham the *i* in *i*-diphthongs is not written consistently (David Stifter, p.c.). The alternative form *cárthann* may be the result of a later sporadic change *áe* > *á* before *r*, *l* and a consonant.
33 W *ceri* is mechanically reconstructible to an *s*-stem plural **karesā*, but this plural ending may well be analogous with other tree-names such as W *dâr*, pl. *deri* 'oak'.

3.38 Sedge/Iris

qPIE **a-lis(k)-str-* > PC **alistro-* > OIr. *ailestar* 'flag iris', W *elestr*, pl. *elystr* 'iris', OCorn. *elestren* gl. *carex*, LCorn. *elestren*, *elester*, OBret. *elestr* gl. hibiscum, MBret. *elestren* 'iris'

qPRom. **liska* > MLat. *lisca* gl. *carecto* (Reichenau glosses), OFr. *lesche*, Fr. *laîche* 'sedge'
qPRom. **līska*> Ligurian *lisca* 'acute sedge'
qPRom. **loska* > Clairvaux *lauche*, *lôche* 'sedge'
qPRom. **lọska*, **lōska*, or **lūska* > Pierrecourt, Broye-les-Pesmes, Minot *louche* 'sedge'
qPRom. **laska* > Alpine Dauphinois *lātša* 'sedge'
qPGm. **lisk-* > OHG *lesc(a)*, *lisca* 'sedge', MDu. *lessche*, *lissche* 'iris', OE ?*lisc* (in place-names)

qPGm. **leuskō* > MHG *liesche* 'sedge', MLG *lēsch*, MDu. *lies(che)*, OS *lius* 'alga'
qPGm. **lūsk(j)ō* > MLG *lusch* 'sedge', Du. obs. *luusch*, Rhinelandic *lusch*, MLG *lüsch*, Rhinelandic *lüsch*, WFri. *ljiske* 'sedge'

A number of secondary forms exist: Ir. *soilestar* 'marsh flag, flag iris' appears folk-etymologically influenced by Lat. *salicastrum* 'kind of vine', or *solus* 'bright'. The ending of MoIr. *feileastram* 'wild iris, flag' is due to influence of *alastram* 'alexander'. Initial *f-* and *g-* in Ir. *feileastram* and W *gelestr* are secondary through reanalysis of vowel-initial forms as lenited. Pedersen (VKG I: 192) suggests that *ailestar* is borrowed from Lat. *salicastrum* 'a kind of vine', but this is semantically unconvincing and requires irregular loss of *s-*. O'Rahilly (1942: 172–173) reconstructs PC **elestro-*, to a PIE element **pel-* (now **pelH-*) also found in Lat. *palūs* 'marsh', G *Felber* 'white willow', with the *str*-suffix for flora. Stifter (2005: 170–172) reconstructs PC **alistro-* because **elestro-* is incompatible with OIr. *ailestar*, and adduces Gallo-Lat. *(h)alus*, *hal* 'comfrey' and W *alaw* 'lily' and ancient Celtic river-names with *alauno-*.

The potential presence of the suffix **-str* in **alistro-* makes the word of likely substrate origin. A velar before **str-* disappears regularly in Celtic, cf. W *rhwystr* 'obstruction' < **reig-s-tro-*, to PIE **reig-* 'bind, tie', OIr. *lestar*, W *llestr* 'vessel, wickerwork' < PC **ɸlestro-* < PIE **pleḱ-s-tro-*, to **pleḱ-* 'to plait' (Stifter 2023b). A substrate element **alisk-* ~ **lisk-* may therefore be proposed between Celtic, Romance, and Germanic; in Celtic this element is found with the *a*-prefix and the *str*-suffix.

The same word without this suffix may be found in an element **lVsk-* found in Germanic and Romance, with a variation in vocalism in **i*, **a*, **u* and diphthongs

thereof that cannot be explained in IE terms. Hubschmidt (1954) accounts for the Germanic and Romance in PIE terms by assuming that PIE **leuH-* 'to cut' underlies a more primitive root **el-* that was variously inflected as an *i*-stem, a *u*-stem, and a laryngeal stem, only for each of these inflections to be independently suffixed with **sk* and to take the same meaning 'sedge'. Aside from the formal issues with this reconstruction, it is rather unlikely that such an archaic process would yield such a host of by-forms within Germanic and Romance whose meaning is identical, so it is to be rejected. De Vries (1971: s.v. *lis*) and FEW (5: 372–374) suggest that the vacillation in vocalism as well as the limited distribution may point to a substrate word.[34]

In Celtic, **alistro-* usually means 'iris', but the meaning 'sedge' is found in Old Cornish and Modern Irish; **lVsk-* in Germanic and Romance conversely means 'sedge' in most cases, but the meaning 'iris' is found in Du. *lis* and Rhinelandic *Lüsch*. The semantic shift between these meanings therefore appears surmountable.

3.39 Silver

qPC **silabur-* > Celtib. *silabur* 'silver, money'

qPIE **silubʰ/pr-* > PGm. **silubra-* 'silver' > Go. *silubr*, ON *silfr*, *sylfr*, OE *seolfor*, OHG *silabar* 'silver'
qPIE **sirebr-* > OCS *sьrebro* 'silver'
qPIE **sido/abr-* > Lith. *sidãbras* 'silver'

A widespread word found in various languages of western and northern Europe. Celtic outside the Iberian Peninsula preserves the inherited PIE word for 'silver', suggesting that this word was borrowed into Celtiberian only after it split off from Proto-Celtic (van Sluis, Jørgensen and Kroonen 2023). Similarly in Balto-Slavic, the different reconstructions required for Baltic and Slavic suggest the word was borrowed into these languages after the Balto-Slavic split. The word refers to an easily tradeable good; this, as well as the word's apparent late borrowing and wide distribution suggests an ancient *Wanderwort*.

Beyond Indo-European, it is found in Bsq. *zilhar*, whose *-h-* can be from **-p-*, and Proto-Berber **ẓrĭp-/ẓrŭp-* 'silver' (DveV 45, Thorsø et al. 2023: 108), as well as Proto-Semitic **ṣarp-* > Akk. *ṣarpu*, Arab. poet. *ṣarīf* 'silver'. Boutkan and Kossmann (1999) reject the appurtenance of Proto-Semitic **ṣarp-* because the meaning

34 De Vries also takes MLat. *lisca* as a borrowing from Germanic, but Germanic borrowing cannot account for all the vocalic alternations found within Romance.

'silver' is only marginally attested and may be secondary from 'to burn, purify, refine'. However, this language-internal etymology can also be adduced to establish Semitic as the ultimate source of this *Wanderwort* (Meid 1993: 113).

OCS *sьrebro* shares its **r* in the onset of the second syllable with Semitic against **l* in Germanic and Celtic, suggesting that (Balto-)Slavic was borrowed along a different pathway. However, it is also conceivable that the alternation **r* ~ **l* is the result of either dissimilation **r. . .r* > **l. . .r* or assimilation **l. . .r* > **r. . .r* in the target language(s). The widely differing vocalism outside the initial syllable suggests that this word's non-initial vowels in the donor language poorly mapped onto the vowel systems of the receiving IE dialects. Perhaps these were highly reduced vowels with a central position in the vowel chart such as [ə].

3.40 Son/Servant

qPC **mak^{w}k^{w}o-* > OIr. *macc* 'son'
qPC **mak^{w}o-* > W, Bret. MCorn. *mab* 'son'
qPC **magu-* > OIr. *mug*, MCorn. *maw*, Bret. *mav* 'servant', W *meu-dwy* 'hermit' (lit. 'servant of God'), W pl. *maon* 'subjects'

qPIE **mag^{h}u-* > PGm. **magu-* > Go. *magus*, ON *mǫgr*, OE *magu* 'boy, son, servant'
qPIE **mag^{wh}-* > PGm. **mawī-* > Go. *mawi*, ON *mær* 'girl'
qPIE **mag^{h}-* > PGm. **magaþ-* > OE *mægeþ*, OHG *magad*, OFri. *megith*, *megeth*, *magad*

The geminate-simplex alternation in Celtic can be explained by invoking hypocoristic gemination in kinship terms (LEIA M-1), see WOMAN. This gemination is not limited to words of substrate origin, cf. W *geneth* 'girl' < PC **genettā* < **ǵenh$_1$-et--eh$_2$*. Boutkan (1998: 123) analyzes the dental expansion in PGm. **magaþ-* as a substrate feature, although the parallel with OIr. *maccdacht* 'childhood' must be rejected (van Sluis, Jørgensen and Kroonen 2023).

3.41 Stalk

qPIE **slatt-* > PC **slattā* > OIr. *slatt* 'stalk, stem, branch', W *llath* 'rod, staff', Bret. *lazh* 'pole, rod'

qPIE **lat-* > PGm. **laþ(þ)a/ōn-*, **latta(n)-* > OE *lætt*, ME *laþþe*, MoE *lath*, *lat*, MDu. *latte* 'lath', OHG *lad(d)a/o*, *lat(t)a/o* 'lath, shoot'

Within Germanic, the variation in consonants can be explained by assuming an original *n*-stem, with nom. **laþō* < **lat-ōn-*, gen. **lattaz* < **lat-n-os* (Kroonen 2011: 214; van Sluis, Jørgensen and Kroonen 2023). PC **slattā* cannot be borrowed from Germanic (*contra* LEIA S-127), because its initial **s-* is not found in Germanic.[35] Because borrowing from Germanic to Celtic is excluded, and because the geminate-simplex alternation within Germanic can be reconstructed to an original simplex, a substrate word with geminate-simplex alternation as its irregularity may be reconstructed.

3.42 Turnip

qPIE **a-rb$^{(h)}$-* > PC **arbīno-* > W *erfin*, Bret. *irvin* 'turnips'

qPIE **rāp-* > Lat. *rāpum* 'turnip'
qPIE **rab^{h}-* > Gr. ῥάπυς,[36] ῥάφυς 'turnip', ?ῥάφανος, ῥέφανος 'cabbage', ῥαφανίς, 'radish'
qPIE **rāp-* > Lith. *rópė* 'turnip'
qPIE **rē/aip-* > OCS *rěpa*, Ru. *répa*, SCr. *rȅpa* 'turnip'
qPIE **rāp-* > PGm. **rōb(j)ōn-* > MDu. *roeve*, OHG *ruoba*, *ruoppa* 'turnip'

The Celtic word for 'turnip' can be reconciled with European **raP-* ~ **rāP-* by assuming an *a*-prefix with syncope (DveV 64; Iversen and Kroonen 2017: 121). Irregularities in vowel length and voicing also speak against inheritance (EDLI 514). The suffix **-īno-* in PC **arbīno-* can be the inherited adjectivizing suffix **-iHno-*. Bsq. *arbi* 'turnip' appears borrowed from Celtic, because the *-i* can be etymologized as from PC **-īno-* (*pace* DveV 64).

3.43 Vessel

qPIE **kelpurn-* > PC **kelɸurno-* > OIr. *cilorn* 'pitcher, vessel', W *celwrn* 'pail, pitcher, vessel', Bret. *kelorn* 'tub'

qPIE **kalpar-* > Lat. *calpar* 'vessel (for wine)'
qPIE **kalp-* > Gr. κάλπις, -ιδος 'pitcher'

35 E *slat* is borrowed from OFr. *esclat* and therefore unrelated.
36 The Greek variant with -π- is not reliably attested and does not provide evidence for a reconstruction with **p* (Anthony Jakob, p.c.).

A substratal **kVlp* may be reconstructed whose stem vowel alternates between Celtic **e* and **a* elsewhere. The ending **-urno-* may indicate an Etruscan proximate source, as such clusters are common in Etruscan-Latin onomastic material and in words loaned into Latin through Etruscan.[37] Lat. *calpar* is only attested in glossaries, and EM (88) suggests borrowing from Greek through either Oscan or Etruscan in view of the suffix *-ar*. The Celtic comparandum speaks in favour of Etruscan. Borrowing of Etruscan personal names and morphology is amply attested in Cisalpine Celtic (McCone 2005: 396), although evidence for Etruscan influence in Celtic beyond this area is disputed for the historical period. If this identification is correct, it seems to be a Mediterranean *Wanderwort* that was borrowed from Etruscan into pre-Proto-Celtic (in view of loss of **p*) (EDPC 198), and with its origins further east, particularly if Ass. *karpu*, *karpatu* 'vase, pot' is related (EDG 627).

3.44 Woman

qPC **u̯raggā* > OIr. *frac*, *frag*, ScG *fràg* 'woman', W *gwrach*, Bret. *gwrac'h* 'hag'
qPC **u̯rakī* > W *gwraig*, Bret. *gwreg* 'woman'

A parallel for the correspondence of Ir. *g* ~ W *ch* in PC **u̯raggā* is provided by PC **biggo-* > Ir. *beag*, W *bych* 'little'. The gemination can be described as hypocoristic gemination occasionally seen in kinship terms, cf. W *geneth* 'girl' < PC **genettā*, OIr. *geined* 'offspring' < PC **geneto-* (to PIE **ǵenh₁-* 'to beget'), W *mab* 'son' < PC **makʷo-*, OIr. *macc* 'son' < PC **makʷkʷo-* (cf. SON/SERVANT). However there also appears to be a voicing alternation between **u̯raggā* and **u̯rakī*. This, too, is arguably paralleled by PC **magu-* 'servant' and PC **makʷ(kʷ)o-* 'son' (cf. SON/SERVANT). OIr. *frac* and W *gwraig* may be mutual borrowings if it is assumed that either the Irish *ā*-stem flexion or the Brittonic **ī*-stem flexion are innovative, but OIr. *frac* is an archaic and poetic word, while **ī*-stem flexion is pan-Brittonic. It lacks convincing comparanda outside Celtic.[38]

37 Cf. Lat. *cisterna* 'cistern', borrowed through Etruscan from Gr. κίστη 'vessel' and *lanterna* 'lantern', from Gr. λαμπτήρ (Niedermann 1916: 152)

38 Pedersen (VKG I: 159) sought to reconcile these forms with Lat. *virga, viragō* through a shared compound **uir-gʷen-* with *ad hoc* loss of **i*. Ranko Matasović (p.c.) suggests borrowing from a reflex of PGm. **wrakjan-* 'persecutor, champion, exile', from PGm. **wrekan-* 'to pursue' (EDPG 594). It is borrowed from OLFra. **wrakkjo* into Fr. *gars, garçon* 'servant, boy, son', but also *garse* 'female servant, girl, daughter'. For Celtic one may propose similar borrowing events (whereby

3.45 Wood/Copse

qPIE *k^wres-n- > PC *k^wrenno- > Gaul. *prenne*, OIr. *crann*, W, Bret. *prenn*, OMCorn. *pren* 'tree, wood'
qPIE *k^wres-ti- > PC *k^wresti- > W *prys(g)* 'copse, grove' (>> ScG *preas* 'bush, shrub, thicket')
?qPIE *k^werstV- > PC *k^werstV- > OIr. *ceirt* 'apple tree; ogham letter Q', W *perth* 'hedge, bush, thicket'

qPIE *k^wrs-ti- > PGm. *h(w)ursti- > OE *hyrst*, OS *hurst*, OHG *hurst* 'crest, copse'
?qPIE *ksu̯ors-(s)ti- > PSl. *xvórstъ > OCS *xvrastije*, Ru. *xvórost* 'brushwood, bush', SCr. *hrâst* 'oak, tree'

An element *k^wres- appears borrowed into Celtic with both an *n*-suffix and a *st(r)*-suffix; *k^wres-n- was borrowed before PC *-*sn*- became *-*nn*- (cf. HOLLY). The appurtenance of OIr. *ceirt*, W *perth* is uncertain: it is more commonly considered inherited from PIE *$perk^wu$- 'oak', but its meaning, particularly that of W *perth* 'hedge', speaks for its appurtenance to *k^wres-. The Slavic onset *xv*- (quasi < *ksu̯-) alternates with Germano-Celtic *k^w-; this irregularity may represent a different approximation of whatever sound found in the donor language, e.g. [x]. PSl. *xvórstъ may alternatively be borrowed from Germanic, but this requires an unattested *o*-grade PGm. *hwarsti-.

3.46 Wooden frame

qPIE *karb- > PC *karbanto- > OIr. *carpat*, Gallo-Lat. *carpentum* 'chariot', W *carfan*, Bret. *karvan* 'weaver's beam, frame; jawbone'
?qPIE *korb- PC *korbo- > OIr. *corb* '?chariot'

qPIE *ka/orb- > PGm. *harpōn- > OE *hearpe*, MDu. *harpe*, OHG *harfa*, ON *harpa* 'harp'
qPIE *korb- > Lat. *corbis* 'basket'
?qPIE *ko/arb^h- > Lith. *kařbas* 'basket', Ru. *kórob* 'box, basket'

---

the Irish is borrowed through Brittonic on account of its final /g/) and semantic shifts (now with complete loss of the older meaning 'boy'), but a suitable time frame is lacking.

A substratal **kVrb-* with **a* in Celtic and **o* in Italic may be reconstructed. This irregularity could in principle be resolved by assuming Germanic-to-Celtic borrowing after PIE **o* > PGm. **a*, but then a PIE root **korb(ʰ)-*, containing a rare **b*, would have to be reconstructed. PC **karbanto-* has alternatively been related to **(s)kerb-* 'to bend, shrink' > ON *skreppa* 'to slip, pull back', OHG *rimpfan* 'to shrink', Lith. *skrèbti* 'to shrink, dry' (LEIA C-40, IEW 948–949), but this root cannot account for Celtic *a*-vocalism: a zero grade would yield ***(s)krib-*.[39]

W *carfan*, Bret. *karvan* for expected **carfant*, **karvant* present an irregularity in that they show no trace of stem-final **t*, but the distribution of **-ant-* among the genetically and geographically distant Goidelic and Gaulish suggests that presence of stem-final **t* is archaic and its absence is innovative.[40] OIr. *corb* is a glossary word glossed by *carpat*, and it may be a creation by Old Irish etymologists rather than a word with any pedigree. W *cerbyd* 'car, carriage' is a borrowing from OIr. *carpat*. Irish *garman* 'weaver's beam' appears borrowed from Brittonic; its unlenited *-m-* may be secondary in a nasal environment (Ó Maolalaigh 2003).

Lith. *kařbas* and Ru. *kórob* may be independently borrowed from a non-IE source (Koch 2020: 136). If so, circumflex accentuation in Lith. *kařbas* 'basket' suggests a qPIE **ko/arbʰ-* that cannot be reconciled to PGm. **harpōn-* and that violates PIE root constraints. However, it is difficult to exclude that Lith. *kařbas* is borrowed from Latin *corbis* through Germanic (cf. MHG *korp*) and/or Proto-Slavic **korbъ*, **korba* (cf. Ru. *kórob* 'box, basket') (EDSIL 234).

3.47 Young animal

qPIE **me/ind-* > PC **me/indo-* 'young animal, young goat' > OIr. *mindu* (acc. pl.), MIr. *menn*, W *myn*, MBret. *menn*, OCorn. *min*.
qPIE **m(a)nn-* > LLat. *mannus* 'small Gallic horse', Gaulish PNs in *mandu-*, Old British PN *Cat-man*

39 The origin of the suffix **-ant-* is unclear. It is homophonous to the present participle suffix sometimes nominalized in e.g. PC **karant-* 'friend', but here the verbal base required for such a formation is lacking.

40 Two scenarios are conceivable to account for the Brittonic absence of *-t*. The first scenario is that *-ant* was substituted with the diminutive suffix *-an* < PC **-agno-* (GPC s.v. *carfan*). However, this suffix is not otherwise found in Breton. The second scenario is borrowing from Welsh to Breton, as in Welsh *-nt* in unstressed syllables regularly became *-n* by the end of the MW period (cf. MW *aryant* 'silver' > MoW *arian*, MW *ugeint* '20' > MoW *ugain*, but Bret. *arc'hant, ugent*).

qPIE **mand-* > PRom. **mandiu* '(male) calf; colt' > Wall. *monse* 'sterile cow', It. *manzo*, Rht.-Rom. *manzi* 'colt', Friul. *manǧ* 'ox', Lad. *manz* 'bull', Rom. *mânz* 'foal'.

For PC, **me/indo-* may be reconstructed; the reconstruction **menno-* given by EDPC (265) must be rejected because a stem-final *-d* can likely be adduced from OIr. *mindu* (Barrett and Stifter 2019: 69). This precludes a root connection with PIE **meiH-* 'become small' (EDPC 265). LLat. *mannus* 'small Gallic horse' is described by Consentius as a Gaulish word (LEIA M-38), and appears independently borrowed, differing from the other forms in having *-nn-*, not *-nd-*.

Romance **mandiu* is often described as borrowed into Romance, but a known source is lacking; ML (384) describes it as an old Alpine word. Alb. *mëz* 'foal; two-three year old colt' can be borrowed from **mandiu*. Outside Indo-European, Bsq. *mando* 'mule; sterile', which is unlikely to be inherited within Basque in view of its initial *m-*, may be adduced (Trask 1997: 128).[41] A competing etymology by which Lat. **mandiu* is inherited adduces the root **mend-* 'to suck' (IEW 729; FEW 6/1: 155). This root is found as a verb in Alb. *mënd* < PAlb. **mand-* (Demiraj 1997: 265), but in view of its restriction as a verb to Albanian the PIE status of this root is doubtful, and it cannot be excluded that the verb is denominal from **mandiu*.

The ultimate origin of this **mVnd/n(i)u-* is likely in the western Mediterranean, as this area contains the largest variation in forms. Here we find variation between **nn* and **nd* as well as between forms with and without a yod. Forms found further north and east can be explained from a single Roman-era spread of **mandiu-*.

4 Recurring alternations

In this section recurring alternations are reviewed and their use in identifying and dating borrowed vocabulary in Celtic is established.

41 If *mando* is borrowed from Romance, its ending *-o* is unexpected: yod in Lat. *-iu* is often not borrowed into Basque, cf. *diru* 'money', from Lat. *denarius*, and *-u* is often borrowed as *-o*, cf. *bil(h)o* 'hair', from *pilus*. However, the more common outcome of Latin loanwords in *-iu* is *-i*, as in *marti* 'March' < *mārtius*, suggesting borrowing from a yodless **mandu-* instead.

4.1 The *a*-prefix

Table 1: The *a*-prefix in Celtic.

Lexeme	With *a-*	Without *a-*
?BEE	**a-pi-* > Lat. *apis*	**bʰi-* > OE *bēo*
BLACKBIRD	**a-msl-* > OHG *amsala*	**mesal-* > W *mwyalch*
LARK	**a-laud-* > Gallo-Lat. *alauda*	**laiwiz-* > E *lark*
LEG	**a-kar-* > ἄκαρα	**g(ʰ)ar-* > W *garr*
LIGHTNING/SULPHUR	**a-str(a)p-* > Gr. ἀστραπή	**strab-* > OIr. *sraib*
?NIT	**a-snid-* > Arm. *anic*	**snid-* > OIr. *sned*
POPLAR	**a-ptl-* > MBret. *ezlen*	**ptel-* > Gr. πτελέα
RASPBERRY	**a-mb-* > MFr. *ambre*	**mab-* > W *mafon*
SEDGE/IRIS	**a-lis(k)-str-* > OIr. *ailestar*	**lisk-* > MLat. *lisca*
TURNIP	**a-rb-* > W *erfin*	**rāp-* > Lat. *rāpum*

The *a*-prefix (Table 1) can be identified when an etymon is found with initial **a-* in some cases while lacking this vowel in other cases. Forms with **a* usually show syncope compared to their unprefixed alternants. This alternation is not found with both alternants occurring within Celtic, suggesting that words with this feature were as a rule borrowed into IE after PIE, but before any discernible secondary nodes in the family tree such as Proto-Celtic and Proto-Germanic. The exception is Greek, within which both prefixed and unprefixed forms are found for the same lexeme. This implies that the language of the *a*-prefix remained in contact with Greek for longer than elsewhere.

The *a*-prefix appears to be a pan-European phenomenon: it is found in comparanda as far as Greek in LEG, LIGHTNING/SULPHUR, NIT and TURNIP, but also in words restricted to Western or Northern Europe, i.e. SEDGE/IRIS, RASPBERRY, LARK, BLACKBIRD. It thus seems that the language of the *a*-prefix was spoken all over Europe, and words were borrowed into Celtic in multiple locations.

4.2 The *n-* and *st(r)*-suffixes for flora

Table 2: The *n-* and *str*-suffixes in Celtic.

Lexeme	With *-st(r)-*	With *-n-*	No suffix
BREAD		**bʰarege/i-n-* > OIr. *bairgen*	**bʰarege/i-* > W *bara*
CLOVER	**(s)mel-sti-* > W *meillion*		**smēr-* > Icel. *smári*
HOLLY	**kela(s)-str-* > Bret. *kelastrenn*	**kolis-n-* > OIr. *cuilenn*	**kulis-* > OHG *hulis*

Table 2 (continued)

Lexeme	With *-st(r)-*	With *-n-*	No suffix
LEEK/GARLIC		**kasnē-n-* > OIr. *cainnenn*	**kesn-* > Ru. *česnók*
PINE	**gis-ust-* > OIr. *giús*	**gis-n-* > OE *cēn*	
SEDGE/IRIS	**a-lisk-str-* > W *elestr*		**lisk-* > MLat. *lisca*
WOOD/COPSE	**kʷ(u)rs-sti-* > OE *hyrst*	**kʷres-n-* > W *pren*	

Words with the *n*-suffix (Table 2) have been argued to belong to a layer of prehistoric loanwords associated with Neolithic farmers, cf. the European word for 'bean': PGm. **baunō-* ~ Lat. *faba* ~ PSl. **bobъ* (Kuiper 1995: 80). Celtic evidence also attests to somewhat early borrowing, in Celtic it occurs after **s* in HOLLY and WOOD/COPSE where in both cases the resulting cluster undergoes the common Celtic sound change **-sn-* > **-nn-*.

The *st(r)*-suffix is found in much of western Europe (Bertoldi 1930), and co-occurs with the *n*-suffix in the same stem in HOLLY, PINE, and WOOD/COPSE. This co-occurrence suggests that one source language underlies both suffixes.

Both the *st(r)*-suffix and *n*-suffix appear in secondary formations seemingly postdating Proto-Celtic: in BREAD Goidelic forms in *-n* differ from Brittonic forms without it, and in CLOVER Brittonic forms in *-st* differ from Goidelic forms without it. Both examples imply independent borrowing into Brittonic and Goidelic.

4.3 Dental-velar-zero alternation

Table 3: Dental-velar-zero alternation in Celtic.

Lexeme	Dental	Velar	Neither
BEE	**bʰit-* > Lith. *bitė*	**bʰek-* > OIr. *bech*	**bʰi-on-* > OE *beó*
CUP/HEAD	**kaput-* > Lat. *caput*	**kapuk-* > W *cawg*	?**kapu-lo-* > OE *hafola*
NUT	**knud-* > ON *hnot*	**knuk-* > Lat. *nux*	**knu(H)-* > OIr. *cnú*

Alternation following a vowel between a dental stop, a velar stop, and no consonant is found in three reliable cases (Table 3). For NUT, it has been suggested that this alternation may indicate an original glottal stop (Kroonen 2012a: 248). This alternation appears old: the correspondence between a **p* and the hiatus found in Old Irish reflexes of CUP/HEAD suggests that a form with **p* was borrowed into Celtic before the weakening and subsequent loss of PIE **p*. Comparanda are always found in Ger-

manic and Boutkan (1998) provides a number of putative substrate words found only in Germanic with this alternation.[42]

4.4 A tendency for front vowels

The substrate lexicon of Celtic contains seven lexemes where a front vowel (*e*, *i*) found in one branch alternates with a back vowel (*o*, *u*) found in another branch. Alternation between a front vowel and a back vowel is found in seven words: DRONE, ELM, FREQUENT/MANY, GARLIC, HOLLY, SEDGE/IRIS, WOOD/COPSE. It may be observed that here Celtic continues the form with a front vowel, and in all words except DRONE and HOLLY, Celtic has at least one of the front vowels *e*, *i*, but none of the back vowels *o*, *u*.[43] Conversely, whenever the vowel *u* alternates with another vowel between IE branches, the form with *u* is never found in Celtic.[44]

4.5 Alternation between *e* and *a*

Alternation between *e* and *a* is found in ELM, LEEK/GARLIC, VESSEL, GARLIC, ?YOUNG ANIMAL, and BULL. Both *e* and *a* are found in Celtic, and no obvious date can be assigned to the alternation: BULL appears borrowed into Celtic much earlier than ELM, for example. This alternation can easily reflect a cross-linguistically common phone not found in early IE, perhaps [æ]. It is therefore not possible to connect it to a specific stratum or donor language.

42 Some of these have Celtic comparanda, i.e. CUP/HEAD, FREQUENT/MANY and SON/SERVANT.

43 The two exceptions stand out in that the full variation in vocalism is also found within another branch of IE. In DRONE, the front vowel *e* is only found in languages that also have back vowel *o* (Greek and Germanic), and in HOLLY, front vowel *e* and back vowel *o* are found side by side within Celtic. Another possible exception is COOT where $^{*}i$ is found in Italic and Germanic, and Celtic has either $^{*}a$ or $^{*}o$.

44 MAN/HERO has Germanic forms with *e* corresponding with Celtic *u*, but all the vocalic variation is intra-Germanic.

4.6 Alternation between *a* and *ai*

Schrijver (1997) identifies this alternation as a particularly northwest European phenomenon, and here too it is found that secure cases are restricted to words only found in Celtic or shared with Germanic (ROWAN, LARK, DUNG, PUBIC HAIR).[45] PUBIC HAIR is the one case where the variant with **ai* occurs in Brittonic; here borrowing may be dated positively to an even later date, because **ai* is borrowed as the secondary Late British **ai* rather than Proto-Celtic **ai*, suggesting this word was borrowed into British after the first-century CE development whereby PC **ai* became PBrit. **ē*. This alternation appears restricted to northwestern Europe, and often even postdates Proto-Insular-Celtic. It appears early, however, in that it co-occurs with the otherwise early *a*-prefix in LARK and possibly TURNIP. Schrijver (1997) mentions more cases from Germanic without Celtic comparanda, so even if there are few cases in Celtic, its presence is a reliable indicator that a word is of a substratum origin.

4.7 Vowel length

When a reconstructed long vowel is found in the Celtic substrate lexicon, it is always found alternating with a short vowel. This suggests that vowel length was not distinctive in the source language, or at least that any such distinction in the substrate language could not be reproduced consistently in IE and early Celtic. The long vowels found in Germanic reflexes of SEDGE/IRIS and BEE appear secondary in view of the variation within this branch.

Variation in vowel length is also found between OIr. *lem* and W *llwyf* in ELM. The existence of both variants within Celtic suggests that borrowing of this word occurred twice, at a relatively late date following the break-up of Celtic into Goidelic and Brittonic, and close in distance. This makes it quite likely that a single donor form yielded both *lem* and *llwyf*, and that a single vowel in the substrate language was variously interpreted as either long or short. Other lexemes with long vowels are CLOVER, DRONE, HOLLY, TURNIP, and here in each case the long vowel stands in an open syllable. These findings suggest that vowels standing in open syllables in the substrate language(s) could be interpreted by early Indo-European speakers as long. There is consequently no evidence that any of the substrate languages had phonemic vowel length.

45 Some instances proposed cannot be upheld, i.e. W *creÿr,* Early HVann. *querhair* 'heron', Bret. *frao,* and W *baedd* (van Sluis, Jørgensen, and Kroonen 2023).

4.8 Voicing alternation

Table 4: Voicing alternation in Celtic.

Lexeme	Voiced	Voiceless
?BADGER	**tazg-* > MIr. *Tadg*	**task-* > OW *Teuhant*
?BEE	**b(ʰ)ek-* > OIr. *bech*	**api-* > Lat. *apis*
COOT	**bʰolig-* > OS *beliko*	**bʰulik-* > Lat. *fulica*
DRONE	**dʰron-* > OE *dran*	**s(t)a-tron-* > OBret. *satron*
FREQUENT/MANY	**mo/anagʰ-* > OE *manig*	**menekk-* > OIr. *meinicc*
GOAT	**kad(V)l-* > Ir. *cadhla*	**katVl-* > Lat. *catulus*
GOAT/SHEEP	**gabr-* > OIr. *gabor*	**kapr-* > Lat. *caper*
HEATHER	**u̯irǵʰ-* > Lith. *viržis*	**u̯roik-* > MW *gwrug*
HOLLY	**go/ust(V)li-* > Arm. *kostł*	**kelastr-* > Bret. *kelastrenn*
LEG	**g(ʰ)ar-* > W *garr*	**akar-* > Gr. ἄκαρα
LIGHTNING/SULPHUR	**strab(ʰ)-* > OIr. *sraib*	**astr(a)p-* > Gr. ἀστραπή
NIT	**g(ʰ)nid-* > Latv. *gnīda*	**ḱonid-* > Gr. κονίς
?NUT	**knud-* > OE *hnutu-*	**knut-* > W *cnwd*
RASPBERRY	**mab-* > W *mafon*	**amp-* > MFr. *ampe*
SON/SERVANT	**magʰu-* > Go. *magus*	**makʷ-* > W *mab*
TURNIP	**arb(ʰ)-* > W *erfin*	**rāp-* > Lat. *rāpum*
WOMAN	**u̯ragg-* > OIr. *frac*	**u̯rak-* > W *gwraig*

Many substrate words resist reconstruction into a single PIE lexeme because a voiceless stop corresponds to a voiced stop or to a voiced aspirate (Table 4).[46] In most cases, the alternation is between IE branches, i.e. NIT, COOT, DRONE, ?BEE, HOLLY, GOAT, LEG, LIGHTNING/SULPHUR, FREQUENT/MANY, and TURNIP. Cases where the alternation is found within Celtic are less frequent, i.e. WOMAN, GOAT/SHEEP, and BADGER. RASPBERRY and HEATHER show voicing alternations within Romance or Balto-Slavic. Most cases of voicing alternation are found in early borrowings, the alternation is found more frequently between IE branches than within Celtic, and voicing quality is consistent within a branch. Even in the intra-Celtic voicing alternation in GOAT/SHEEP, both variants appear to be old: other branches attest to a pre-form **kapro-*, while Celtic has **gabro-* < **gapro-* and **kaɸero-* < **kapero-*, but both forms show PIE treatment of **p* in their respective positions. This suggests both Celtic by-forms of GOAT/SHEEP entered Celtic before the Proto-Celtic loss of PIE **p*. Stifter (this volume) offers an alternative analysis for some specific cases of this alternation, namely cases where Celtic **b* corresponds to **p* elsewhere. Such words may be bor-

46 Putative alternations between plain voiced and aspirated voiced stops are irrecoverable in Celtic, because these series merged in Proto-Celtic.

rowings postdating the Proto-Celtic loss of /p/ from its phoneme inventory with sound substitution of a foreign [p] as **b*. Words that are potentially dateable to Proto-Celtic by this criterion are LIGHTNING/SULPHUR, RASPBERRY and TURNIP.

Voicing alternation by itself is a poor criterion to designate a word as substrate in origin. Stops are rather common in IE languages, so that words that are in reality chance resemblances may be identified as substrate words, and even inherited words are sporadically voiced or devoiced, e.g. in the Celtic word for 'tongue', PIE **dnǵʰ-ueh₂* > PC **tangʷāt-* > OIr. *tengae*, W *tafod* (cf. Stifter 2017: 1190).

4.9 Geminate-simplex stop alternation

Table 5: Geminate-simplex stop alternation in Celtic.

Lexeme	Simplex	Geminate
COOT	**bʰulik-* > Lat. *fulica*	**bʰo/ulVkk-* > ScG *bolachdan*
HOOK	**bak-* > ME *pegge*	**bakk-* > OIr. *bacc*
SON/SERVANT	**makʷ-* > W *mab*	**makʷkʷ-* > OIr. *macc*
STALK	**lat-* > MoE *lath*	**slatt-* > OIr. *slatt*
WOMAN	**u̯rak-* > W *gwraig*	**u̯ragg-* > W *gwrach*

Geminate-simplex stop alternation mostly appears late and restricted to Western and Northern Europe (Table 5). It is found with only Celtic comparanda in WOMAN, and it is found only within Celtic in the Celto-Germanic SON/SERVANT, while STALK, and COOT are otherwise restricted to Germanic and/or Italic. HOOK has a wider distribution, but it lacks diagnostic sound laws that allow for its borrowing into Indo-European to be dated; its geminate is at any rate restricted to Celtic.

The Indo-European origins of geminated consonants in Celtic are mostly unclear, except for *r*, *l*, *n* (McCone 2005: 407; Stifter 2023a), and alternation between simplex and geminate consonants has already been identified as a substrate feature in Germanic (Kuiper 1995). This makes many Celtic etyma with geminated stops already suspect of being borrowed, and those with an alternation between gemination and lack thereof all the more so.

4.10 Liquid alternation

Table 6: Liquid alternation in Celtic.

Lexeme	*l*	*r*
CLOVER	**smel-* > Bret. *melchon*	**smēr-* > Icel. *smári*
MAN/HERO	**kale/ut-* > OE *hæleþ*	**karut-* > OIr. *caur*
?SILVER	**silabur-* > Celtiberian *silabur*	**sirebr-* > OCS *sьrebro*

Alternation between the liquids **r* and **l* is rare in Celtic, and is found in the geographically restricted CLOVER and MAN/HERO and possibly the late *Wanderwort* SILVER (Table 6). The alternation does not seem to be a feature of borrowing from a specific substrate language, and may instead be introduced by the borrowing process, or by language-internal liquid assimilations and dissimilations.

Borrowings with comparanda in Semitic show Indo-European **l* against Semitic **r*. In VESSEL, Ass. *karpu*, *karpatu* and in SILVER Proto-Semitic **ṣarp-* have have **r* where in European comparanda *l* predominates. The two liquids were distinct phonemes in both PIE and Proto-Semitic. This may imply that these words were not borrowed directly from Semitic to Indo-European, but through an intermediary language where /r/ did not contrast with /l/.[47]

4.11 Velar-s alternation

Table 7: Velar-*s* alternation in Celtic.

Lexeme	Stop	Fricative
BREAD	**bare/ig(ʰ)-* > OIr. *bairgen*	**bare/is-* > Go. *bariz-*
NIT	**g(ʰ)nid-* > Latv. *gnīda*	**snid-* > OIr. *sned*

BREAD and NIT show a pattern whereby qPIE (palato-)velar stops alternate with **s*, constituting an alternation in both place and manner of articulation (Table 7). This alternation may represent a source-language palatal affricate, perhaps [ʧ].

47 Cf. Kroonen (2012b) for a similar scenario involving Semitic in the word for 'garlic'.

Words showing this alternation consistently appear borrowed early before most IE languages developed such palatal sounds, and have a wide distribution.[48]

4.12 Metathesis

Among early borrowings, metathesis is found in HOLLY, where Arm. *kostłi* < **gVst(V)l-* contrasts with a European **KVlVst-* and in BULL, where Celtic **tarṷ-* contrasts with **teu̯r-* elsewhere. In both cases only one IE branch shows metathesis of an otherwise consistent sequence suggesting that it was not a grammatical feature of the source language(s), but that the borrowing process or developments internal to Celtic and Armenian are responsible for these metatheses.

In later and geographically restricted borrowings, metathesis can be shown to be a feature of the source language(s). Celto-Germanic OATS and BADGER share a structure CVCC where the stem-final consonant cluster has descending sonority in Celtic (**kork-*, **tasK-)* and ascending sonority in Germanic (**kokr-*, **taKs-*) in the final consonant cluster. In ELM, Celtic continues **lVm-* against **Vlm-* elsewhere, and the variation in vowel quality found in words with the shape **Vlm-* suggests this shape was borrowed from a substrate language several times over. The variation in vowel quantity between OIr. *lem* and W *llwyf,* however, suggests that a substratal **lĕm-* was borrowed into Celtic twice, and that the shape **lVm-* was also found in the substrate itself. This word therefore suggests that a metathesis had already occurred within the language(s) from which ELM was borrowed, rather than metathesis being the result of borrowing or language-internal processes in Celtic. ELM and CLOVER share an alternation (C)CVC ~ (C)VCC with metathesis of the medial vowel-consonant group. In ELM, (C)CVC is found in Celtic (**lĕm-)* against (C)VCC elsewhere (**Vlm-),* and in CLOVER the alternation runs through Celtic, with Gaulish and Irish sharing a structure **sVm(a)r-*, and Germanic and Brittonic sharing a structure **smĕr/l-*. It seems that syllables with heavy onsets were generally preferred over heavy codas in the substrate donor language(s) to Celtic.

Some instances of the *a*-prefix have an **a* in the non-*a*-prefixed non-syncopated form, so that the alternation can be analyzed as a metathesis: RASPBERRY, TURNIP (cf. GPC s.v. *erfin*), LIGHTNING/SULPHUR. Metathesis and the *a*-prefix may ultimately reflect the same phenomenon in the substrate language in ELM, where Schrijver (1997: 311) observes an alternation between an unsyncopated form without a prefix and a pre-

48 Another possible substrate word with this alternation, but without Celtic comparanda, is Lat. *avēna* 'oats' < **au̯e(C)s-n-*, Fi. *vehnä* 'wheat' < Western Uralic **wešnä,* Lith. *avižà* 'oats' < **au̯iǵ(ʰ)-*, Ru. *ovës* 'oats' < **au̯iḱ-*. See Kroonen et al. (2022) for a discussion of how this word cannot be inherited, but rather appears borrowed.

fixed form with syncope, even if the vowel prefix is not always *a*. The early date and pan-European distribution of the *a*-prefix contrasts with a late date and more restricted distribution of ELM, so it is conceivable that the variation seen in ELM reflects a later development of the *a*-prefix within the substrate language.

4.13 Stratification

At least two layers of substrate words may be identified with little overlap in features: an early layer predating most sound changes between PIE and Proto-Celtic, often with comparanda in other IE branches across Europe, and a late layer postdating Proto-Celtic sound changes with alternations within Celtic and with more locally distributed comparanda in other languages. The early layer consists of words borrowed before most of the sound changes separating PIE from Proto-Celtic, most visibly loss of PIE **p* in most environments. This early layer is characterized by the *a*-prefix, voicing alternation, dental-velar-zero alternation, velar-*s*-alternation, and the *n*-suffix. The late layer contains words borrowed after Celtic developed geminates, as well as other sound changes leading up to Proto-Celtic. This layer is characterized by geminate-simplex alternation, metathesis and *a~ai*-alternation. Liquid alternation and the *st(r)*-suffix cannot be associated with a specific timeframe.

Dateable substrate alternations may be used to date otherwise undateable words (Section 2.3). RASPBERRY and POPLAR are two lexemes without compelling evidence as to their date of borrowing. The words cannot be demonstrated to have undergone any sound changes predating Proto-Celtic, and neither word is found in Goidelic so that it is unknown whether a Goidelic cognate would show intra-Celtic regularity or irregularity. However, both words contain the *a*-prefix, and having established the *a*-prefix as early, both words appear early borrowings into Celtic. Conversely, the geminate-simplex stop alternation is a feature of later borrowings, so its presence in HOOK suggests that this word was borrowed rather late.

An obvious explanation for the existence of the distinct layers in the alternations is that speakers of early IE and Celtic were in contact with speakers of different substrate languages in different stages of prehistory. Such differences could be the result of prehistoric population movements, or of developments within the substrate languages. However, changes in the substrate(s) need not be the only cause of changes in alternations over time. Differences between the early layer and the late layer may also reflect the same source phenomenon being borrowed differently in different stages of the target language as its phonological system evolved.

One such case where phonological change within Celtic drove the formation of distinct layers may be the contrast between voicing alternation and geminate-simplex alternation. Voicing alternation is primarily found in the early layer, while geminate-simplex alternation is found chiefly in the late layer. However, this contrast may well be accounted for with reference to the historical phonology of Celtic itself. PIE did not have geminates, but did have opposition in voicing and aspiration, so any substratal opposition in length, aspiration or voicing would be borrowed as a difference in voicing or aspiration in early IE. In early Celtic, distinction in aspiration was lost with the merger of voiced stops and voiced aspirates. Among simplex consonants, lenition of intervocalic voiced stops led to their realization as fricatives, leaving only voiceless stops to be pronounced as stops. The loss of distinctive aspiration and voicing in this part of the Celtic stop system was balanced by the development of an opposition between simplex stops and geminates. Thus from early Celtic onwards, any hypothetical substratal distinction in length, aspiration or voicing of stops would likely be borrowed as a simplex-geminate opposition. It may well be that the stop inventory of the substrate language(s) was stable, and that differences between early and late borrowings reflect phonological developments in the target language. No closer inference can be made as to the phonology of the substrate on the basis of these two contrasting alternations, or even whether the same language or multiple languages underlie them.

5 Substrate affiliation

Vasconic, Finnic, and Afro-Asiatic are the known language families with multiple comparanda in this lexicon. Having established several layers of borrowings into Celtic without prejudiced comparison with a particular source language, it becomes possible to evaluate the explanatory power of assuming these language families as substrates (McCone 2005: 406). It also becomes possible to compare the lexicon with substrate languages that are not directly attested, but inferred from earlier substrate studies. One attempt to identify multiple languages in substrate vocabulary comes from Kuiper (1995), who distinguished three layers in the Germanic lexicon of which he suspected a substrate origin based on the prevalence of qPIE **a*, and which he named A1, A2, and A3.

5.1 Avidic

The oldest of Kuiper's substrate languages, A1, is primarily identified by the reconstruction of qPIE **a* and a wide geographic distribution, with the added observation that voiced aspirates are common. These criteria are rather unspecific, but these features, the early date, and the pan-European distribution are consistent with the *a*-prefix, which has itself been characterized as a language of Neolithic farmers (Iversen and Kroonen 2017). Schrijver (1997) identified this prefix in many western IE bird names, leading Stifter (2010) and now the present author to adopt the term "Avidic" for this language.

A particularly early borrowing from Avidic into Celtic is found in BLACKBIRD. Here, a single Italo-Celtic **mesl-* can be reconstructed with later independent vocalization of a syllabic **l* in both Celtic and Italic, suggesting that the word was borrowed into a dialect of IE in which Celtic and Italic were undifferentiated. Other early borrowings can be LEG, SULPHUR/LIGHTNING on account of their wide distribution. Evidence for later Avidic borrowings is rarer. LARK appears to be a later borrowing, as shown by the co-occurence of the otherwise early *a*-prefix and otherwise late *a~ai*-alternation.[49] The metathesis in ELM can be analyzed as a substrate-internal development of the *a*-prefix (Section 4.1). Because ELM appears borrowed after Proto-Celtic, this word, too, suggests that Avidic survived after Proto-Celtic.

The wide temporal and geographical spread of Avidic means that it has little diagnostic value in dating and locating Proto-Celtic. However it may well give insight into Neolithic Europe. The exceedingly early borrowing of BLACKBIRD as well as the later geographically restricted borrowing of LARK suggests a somewhat homogenous late linguistic landscape of late Neolithic Europe. The apparent late survival of this language suggests that the Indo-Europeanization of northern Europe was a protracted process.[50]

Schrijver (2011, 2018) connects the *a*-prefix to a Hattic prefix *ha-*. Hattic words may take this prefix, and when they do, they exhibit syncope-like changes in vocalism. Thus arguably a language related to Hattic, or at least with similar morphology in this regard, appears to have been spoken from Anatolia all the way to

49 This co-occurrence has several possible explanations. One possibility is that the language of *a~ai* alternation borrowed the word from the language of the *a*-prefix before entering Germanic. Another possibility is that both features are found in the same substrate language, but the *a~ai* alternation stems from later developments within this language. Both scenarios suggest that Avidic was present in western Europe.

50 It may be similar in this regard to southern Europe, where the earlier advent of writing confirms the existence of many non-IE languages.

Western Europe before its replacement by Indo-European. It is likely that this language survived in the vicinity of Greek comparatively late, because alternation between forms with and without the prefix is found within Greek.

5.2 Pre-Germanic

Kuiper's second oldest layer, A2, is somewhat less widespread. This layer is characterized by initial clusters *kn-*, *kl-*, *a/i/u*-vocalism in derivational suffixes, prenasalization, geminate-simplex stop alternation and voicing alternation.[51] Boutkan (1998) furthermore adduces disyllabic stems of the shape C(R)VCVC- and various semantically neutral extensions in labial, dental and velar stops and **l* as a feature of this substrate layer, and perhaps this phenomenon may be equated with dental-velar-zero alternation (Section 4.3). Some words in the Celtic substrate lexicon contain one or more of these features. Table 8 contains words with initial velar plus resonant (**KR-)*, vowel variation in non-initial syllables (**V ~ V)*, and words with disyllabic stems of the shape C(R)VCVC-. Substrate alternations otherwise found in these words are the *n-* and *st(r)-*suffixes for flora, liquid alternation and various types of vowel alternation, but the *a*-prefix is not found.

All of the words given in Table 8 except PUBIC HAIR have comparanda in Germanic. Most words are otherwise confined to Balto-Slavic and/or Italic, but GARLIC, HOLLY and SILVER have wider distributions. One may conclude that this pre-Germanic layer primarily influenced Germanic, and its influence on Celtic and other languages neighbouring Germanic was less intense. The Celtic substrate lexicon confirms Kuiper's and Boutkan's findings that velar-resonant clusters and disyllabic stems belong to a Germanic substrate; if they had collected words sharing these shapes through mere chance resemblance, then Celtic substratum words with the same shape would not be expected to have Germanic comparanda to the same extent.

Among the words given in Table 8, CUP/HEAD was borrowed into Celtic at an early date, as is shown by its PIE treatment of **p*. Other words appear borrowed into Celtic or other IE branches at a later date, however. GARLIC and PUBIC HAIR both require multiple reconstructions within Celtic, while SILVER requires multiple reconstructions within Balto-Slavic. The geminate-simplex stop alternation found

51 Alternation in voicing and gemination is ubiquitous in the Celtic substrate, but its value in identifying specific substrate sources is doubtful (Section 4.13). Prenasalization (i.e. the alternation of nasal and stop with a stop alone) does not reoccur in Celtic. It is found in the *Wanderwort* LEAD (Lat. *plumbum*), but restriction of prenasalization to Latin suggests a Latin-internal development here. These features are therefore not treated as diagnostic of a pre-Germanic origin here.

Table 8: Lexemes containing features identified as indicative of pre-Germanic by Kuiper (1995) and Boutkan (1998).

Lexeme	Reconstruction	*KR-	*V ~ V	*C(R)VCVC-
BREAD	**bareC-*			+
COOT	**bʰVla/iK-*		+	+
CUP/HEAD	**ka(u)pe/uC-*		Within Gm.	+
FREQUENT/MANY	**mVne/aK-*		+	+
GARLIC	**krVmus-*	+		
GOAT	**katVl-*			+
HOLLY	**kVli/a/os-*		+	
MAN/HERO	**kale/ut-*		Within Gm.	+
NUT	**knuC-*	+		
PUBIC HAIR	**kVti/u/ar-*		+	?
SILVER	**sila/u/ePr-*		+	
WOOD/COPSE	**kʷres-*	?		

in FREQUENT/MANY also suggests borrowing at a later date. All in all, both strata identified in Section 4.13 appear represented in the Celtic substrate lexicon of pre-Germanic origin, but later borrowings predominate.

Pre-Germanic likely also accounts for the one Finnic comparandum in the Celtic substrate lexicon, i.e. OATS. It is not likely that pre-Germanic was itself Finnic (*pace* Hyllested 2010: 123–124): its characteristic **kn/l-* clusters are incompatible with Finnic phonotactics, where such initial clusters are not allowed. Hyllested (2016) argues that there was a Finnic substrate in Celtic not shared with Germanic. However, the one word also found in Finnic also has comparanda in Germanic, so the existence of a Finnic substratum in Celtic but not Germanic cannot be confirmed.

Pre-Germanic was most likely spoken in prehistoric northern Europe. This location is suggested by the fact that comparanda are found in northern European languages, plus Italic. The presence of this otherwise northern European substrate in the Italic lexicon suggests that the IE language that would later develop into Italic had some presence north of the Alps, or alternatively that this substrate language reached south of the Alps.

5.3 Vasconic

Kuiper's A3 refers to the language of Old European hydronymy sometimes equated with Vasconic. The existence of such a substrate layer in western Europe is championed by Vennemann (2003), but Schrijver (1997) already rejected association of the

Celtic substratum with this layer. Support for a Vasconic substrate layer in Celtic could otherwise come from substrate words with comparanda in Basque.

The corpus contains six Basque comparanda, but most of these can be shown not to be inherited in Basque. Bsq. *gorosti* (HOLLY) and *berun* (LEAD) show a substitution of an initial voiceless stop with a voiced stop, as is common in Roman-era borrowings (Trask 1997: 129). Bsq. *zilhar* (SILVER) and *arbi* (TURNIP) hardly look inherited because they can be etymologized within Semitic and Celtic, respectively. Bsq. *mando* (YOUNG ANIMAL) appears borrowed into Basque in view of its initial *m-*. Only BADGER can be an early borrowing into IE from a pre-form of Bsq. *aizkon*. One may conclude from these findings that Vasconic and Celtic were most likely in contact with a common substrate, and that the substrate itself was not Vasconic.

5.4 Afro-Asiatic

Seven words have comparanda in Afro-Asiatic. These can be further divided in words otherwise found across Europe, i.e. BULL, BEE, and SILVER, and words otherwise restricted to the Mediterranean, i.e. VESSEL, BERRY, LEAD. They can be stratified into distinct temporal layers. BULL has the widest distribution reaching Greek. Its wide distribution as well as its regular reconstructability to Proto-Celtic and other secondary branches of PIE suggest exceedingly early borrowing. BEE also appears borrowed early on account of its fairly wide distribution, but separate reconstructions for Baltic and Slavic suggests its borrowing into IE postdates Balto-Slavic unity. It is restricted to northern Europe, and contains a dental-velar-zero alternation characteristic of the pre-Germanic substrate. This word therefore provides evidence that vocabulary shared with Afro-Asiatic reached prehistoric IE dialects of northern Europe, but also that any putative contact between Afro-Asiatic and Indo-European was mediated by pre-Germanic.[52] SILVER has a northern and western European distribution, but absence of this word from Insular Celtic and its irregularities within Balto-Slavic suggest a *Wanderwort* that spread across Europe along the Atlantic after Proto-Celtic. VESSEL, BERRY, LEAD have a Mediterranean distribution plus Celtic, and their presence in Celtic is perhaps best viewed as the final leg of Mediterranean trading networks.

52 If Lat. *apis* 'bee' is adduced as an *a*-prefixed form of BEE, then Avidic also served as a conduit for *Wanderwörter* originating in Afro-Asiatic.

None of these words suggest that Afro-Asiatic itself rather than an intermediary language was present along the Atlantic coast of prehistoric Europe (*contra* Vennemann 2003). Substrate words shared between Proto-Celtic and Afro-Asiatic are alternatively shared with northern European languages or with Mediterranean languages. Earlier work on contact between Afro-Asiatic and Celtic focused on morphosyntactic parallels such as conjugated prepositions or verb-subject-object word order (Hewitt 2007). However the aforementioned lexical evidence shows that Afro-Asiatic contact was often intermediated by local substrates, and not particularly intense with Celtic compared to neighbouring IE dialects. Thus, if these morphosyntactic parallels were the result of Afro-Asiatic influence, one should expect to find them in these neighbouring dialects as well as Celtic. As long as these parallels are not found in those branches of IE, it is preferable to assume a language-internal origin for these putative contact phenomena (*pace* Eska 2010).

6 A timeline of Celtic prehistory

The various stages of contact between Indo-European and non-Indo-European described above allow for the reconstruction of some prehistoric developments in the Celtic languages and their speakers. The earliest contacts between IE and its European substrata can be inferred from geographically widespread substratum words that cannot be shown to postdate any sound changes postdating the dissolution of PIE. Most words falling into this category refer to the agricultural produce and the European natural environment: BEE, BREAD, BULL, DRONE, GOAT/SHEEP, LEG, LIGHTNING/SULPHUR, NIT, POPLAR, and TURNIP. To the extent that the source language can be determined, Avidic is the most common. These borrowings may reflect changes in subsistence strategies as early IE-speaking pastoral nomads adopted some of the subsistence strategies associated with the more sedentary farmers of Neolithic Europe (Iversen and Kroonen 2017).

Other words similarly predate most known sound changes between PIE and its daughter languages but have a more limited geographical distribution variously restricted to northern Europe, western Europe and the Mediterranean. Such words may be borrowed as early as the first category, but there are fewer arguments to support such an early date. Words falling in this category are BEER, BERRY, BLACKBIRD, BREAD, COOT, CUP/HEAD, GOAT, HOOK, LEAD, NUT, SEDGE/IRIS, RASPBERRY, VESSEL and WOODEN FRAME. These words may refer to metals, wooden tools and pots as well as agricultural produce and the natural environment. This stage at-

tests to borrowings from both Avidic and pre-Germanic, but some craftsmanship terms appear to have a southern origin.

Two closely related words meaning 'holly' ultimately yielding Bret. *kelennen* and Bret. *kelastrenn* appear borrowed from a non-IE substrate language into Celtic. As discussed under the lemma HOLLY, borrowing of these words can be dated to roughly the Proto-Celtic stage on the basis of the sound laws that must predate and postdate these borrowings. These findings suggest that Proto-Celtic speakers had likely moved into western Europe at least as far westwards as where the holly (*Ilex aquifolium*) is found; its distribution is shown in Figure 1.

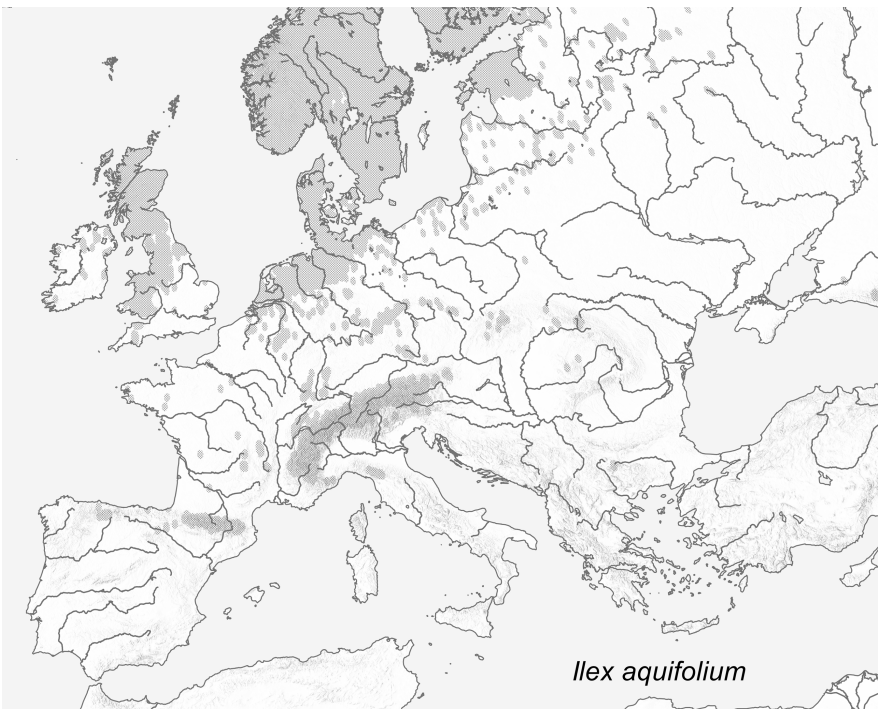


Figure 1: Distribution of the common holly (*Ilex aquifolium*). Data from GBIF.org. Map by author.

It is likely also around this stage that many words exclusively shared with Germanic and other northern European languages were borrowed: FREQUENT/MANY, LARK, MAN/HERO, OATS, PINE, ROCK, RUSHES, SON/SERVANT, STALK, and WOOD/COPSE. The absence of these words in Italic suggests that borrowing of these words postdates a

putative Italo-Celtic subnode.[53] The shape and distribution of most of these words suggest a pre-Germanic origin, but LARK shows continued contact with Avidic in this period.

Words positively dateable to after the dissolution of Proto-Celtic on the account of their intra-Celtic substrate alternations are BADGER, CLOVER, DUNG, BAT, PUBIC HAIR, ROWAN, YOUNG ANIMAL and WOMAN.[54] GARLIC and ELM similarly require multiple reconstructions to Proto-Celtic, but are more geographically widespread. Possibly these are early borrowings with multiple forms surviving, but they may also be later *Wanderwörter*. The *Wanderwort* SILVER is borrowed into Continental Celtic exclusively, implying that borrowing postdates separation from Insular Celtic. Substrate words shared with Basque mainly date to this period. No borrowings from Avidic can be shown to date to this period. The phonological structure of some words in this layer can be equated to pre-Germanic, but this may be accidental.

Words relating to society (MAN/HERO, PUBIC HAIR, SON/SERVANT, WOMAN) are limited to Germanic and Celtic or Celtic alone, and appear borrowed late. These borrowings may reflect changes in Celtic society from the latter half of the Bronze Age onwards. Atlantic Europe at the turn of the first millennium BCE saw a re-emergence of enclosed settlements and sword warfare, and luxury goods became more common in burials from the Late Bronze Age onwards (Mallory 2016).[55] One may speculate that the increased societal complexity and political clout of the people who built these settlements and graves led to the assimilation of non-IE speaking people into Celtic (and Germanic) society. Support for such a scenario comes from Patterson et al. (2022), who show that the Middle to Late Bronze Age saw large-scale migration into Britain from the continent, most probably from France, leading to an increase in the size of the mating pool; similar trends are also observed elsewhere in Europe. It is likely that this enlarged mating pool entailed intensified contact between IE speakers and non-IE speakers

53 The large amount of Italo-Celto-Germanic substrate words with pre-Germanic features suggests that an undifferentiated Italo-Celtic was in contact with pre-Germanic, while the pre-Germanic words shared with Germanic alone suggests that this contact between Celtic and pre-Germanic endured after this split.

54 The relative scarcity of this later stratum may be the result of the chosen methodology whereby only irregular correspondences are considered. David Stifter (this volume) shows that a methodology collecting words with necessarily non-inherited phoneme sequences or morphological elements alone yields more borrowed vocabulary postdating Proto-Celtic.

55 The period in which words relating to society were borrowed into Celtic is comparable to the exchange of vocabulary in this semantic domain between Celtic and Germanic (van Sluis, Jørgensen, and Kroonen 2023). The archaeological vectors accounting for these Celto-Germanicisms may therefore also account for this layer of the Celtic substrate lexicon.

throughout Europe. The adoption of non-IE institutional vocabulary may well be the result of the integration of previously isolated substrate populations into IE-speaking institutions.

These findings can feed into the debate on where the Celtic homeland was located. The two competing hypotheses are the traditional hypothesis, which places Proto-Celtic in the eastern Alps of the Late Bronze Age, and the Celtic from the West hypothesis which sees Proto-Celtic as the result of dialect levelling among Indo-European speakers facing the Atlantic over the course of the Bronze Age (Sims-Williams 2020, with references). As discussed under its lemma, borrowing of HOLLY is estimated to date to around Proto-Celtic and the absence of its referent from eastern Europe suggests that Proto-Celtic developed no further east than the Alps. Influence from the pre-Germanic substrate appears strongest before Proto-Celtic and appears located in northern Europe. These insights are difficult to reconcile with a Celtic homeland in the eastern Alps. An early Celtic homeland along the Atlantic, however, would suggest early contact with Vasconic, but what is found instead is that such mutual borrowings postdate Proto-Celtic. A third option offered by Sims-Williams (2020), with the Proto-Celtic homeland in second-millennium BCE Gaul, is the most compatible with the Celtic substrate vocabulary.

6.1 Prehistoric gene flow and the Insular Celtic substrate

The arrival of Celtic languages in the British Isles has two possible genetic vectors. The first is the near-complete replacement of the Neolithic farmer ancestry by a Corded Ware-derived ancestry profile originating in north-central Europe with the arrival of the Bell Beaker phenomenon in the Early Bronze Age (Olalde et al. 2018). This ancestry profile itself contained a large component originating from the Pontic-Caspian Steppe currently thought to be where Proto-Indo-European was spoken. This scenario suggests that the British Isles were first Indo-Europeanized by an Indo-European dialect that would evolve into Celtic *in situ*, and as such is most compatible with the Celtic from the West hypothesis. The second possible genetic vector is a more limited population replacement occurring over the Middle and Late Bronze Age originating in France (Patterson et al. 2022). This vector is more in line with current estimations of Proto-Celtic in west-central Europe around the Middle to Late Bronze Age (Sims-Williams 2020).

The first scenario is difficult to square with words such as HOLLY, which suggest that Celtic did not arrive as far west as the Atlantic seaboard in the form of undifferentiated Indo-European. It is also difficult to reconcile this scenario with the existence of substrate words limited to the British Isles and postdating the split of the Insular Celtic languages. The Insular Celtic substrate would have to

survive the millennia separating the arrival of the Bell Beaker complex from the formation of distinct branches of Goidelic and Brittonic, only to disappear right before the historical period. Such a scenario could be sidestepped by equating the second of the above-mentioned genetic vectors with the Insular Celtic substrate, but then it would be awkward to identify precisely a continental genetic signature with vocabulary only found on the British Isles.

Identification of the Middle to Late Bronze Age genetic vector with the initial Celticization of the British Isles makes it easier to account for the existence of an Insular Celtic substrate, and the northern European genetic signature of the population of the British Isles preceding this influx also accounts for the otherwise northern European distribution of substrate words postdating Proto-Celtic. Under this scenario, however, the Insular Celtic substrate is associated with an Indo-European-associated genetic signature with its origins in the Steppe, and as such it is unexpected to find no evidence of an Indo-European Insular Celtic substrate.

A third possibility is to identify the Insular Celtic substrate with the few people descending from Neolithic farmers holding out in northern Britain and in Orkney (Dulias et al. 2022), but these holdouts seem rather peripheral to the more southerly location where Brittonic and Goidelic must have formed.

Whatever the case may be, the high genetic diversity and multitude of migrations reconstructible for prehistoric Atlantic Europe suggests the presence of many now-extinct languages, both Indo-European and non-Indo-European, and much work still remains to be done for linguists in uncovering these languages in the lexicon of surviving languages.

References

Allentoft, Morten E., Martin Sikora, Karl-Göran Sjögren, Simon Rasmussen, Morten Rasmussen, Jesper Stenderup, Peter B. Damgaard et al. 2015. Population Genomics of Bronze Age Eurasia. *Nature* 522(7555). 167–72.

Anthony, David W. 2010.*The horse*, *the wheel*, *and language: How Bronze-Age riders from the Eurasian steppes shaped the modern world*. Princeton, NJ: Princeton University Press.

Barrett, Siobhán, David Stifter. 2019. Blathmac's stanzas 260–303 on judgement day. *Celtica* 31. 19–89.

Bauer, Bernhard. 2018. Irish loanwords in the Southwest British Celtic languages. In Michael Hornsby & Karolina Rosiak (eds.), 23–33. *Eastern European Perspectives on Celtic Studies*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Beekes, Robert S. P. 1996. Ancient European loanwords. *Historische Sprachforschung / Historical Linguistics* 109. 215–36.

Beekes, Robert S. P. 2000. European substratum words in Greek. In Michaela Ofitsch & Christian Zinko (eds.), *125 Jahre Indogermanistik in Graz*, 21–32. Graz: Leykam.

Beekes, Robert S. P. 2014. *Pre-Greek: phonology*, *morphology*, *lexicon*. (Brill Introductions to Indo-European Languages). Leiden/Boston: Brill.

Bertoldi, Vittorio. 1928. Antichi filoni nella toponomastica mediterranea incrociantisi nella Sardegna [Ancient layers in the Mediterranean toponomastics meeting in Sardinia]. *Revue de linguistique romane* 4. 222–250.

Bertoldi, Vittorio. 1930. Gallico **cabrostos* «ligustro» (da **cabros*, il corrispondente gallico del greco κάπρος, ecc.) [Gaulish **cabrostos* 'privet' (from **cabros*, the Gaulish cognate of Greek κάπρος, etc.)]. *Revue Celtique* 47. 184–196.

Blöndal Magnússon, Ásgeir. 1989. *Íslensk orðsifjabók* [Icelandic etymological dictionary]. Reykjavik: Orðabók Háskólans.

Boutkan, Dirk F. H. 1995. Altgermanisch 'Held' und die Entwicklung von unbetontem **e* im altnord- und westgermanischen. *Amsterdamer Beiträge zur älteren Germanistik* 41. 1–7.

Boutkan, Dirk F. H. 1998. On the form of north European substratum words in Germanic. *Historische Sprachforschung* 111. 102–33.

Boutkan, Dirk F. H. & Maarten G. Kossmann. 1999. Some Berber Parallels of European Substratum Words. *Journal of Indo-European Studies* 27. 87–100.

Cassidy, Lara M., Rui Martiniano, Eileen M. Murphy, Matthew D. Teasdale, James Mallory, Barrie Hartwell & Daniel G. Bradley. 2016. Neolithic and Bronze Age migration to Ireland and establishment of the Insular Atlantic genome. *Proceedings of the National Academy of Sciences* 113(2). 368–73.

de Vries, Jan. 1971. *Nederlands etymologisch woordenboek* [Dutch etymological dictionary]. Leiden: Brill.

Demiraj, Bardhyl. 1997. *Albanische Etymologien*. Amsterdam/Atlanta, GA: Rodopi.

Deshayes, Albert. 2003. *Dictionnaire étymologique du breton*. Douarnenez: Chasse-Marée.

Dulias, Katharina, M. George B. Foody, Pierre Justeau, Marina Silva, Rui Martiniano, Gonzalo Oteo-García, Alessandro Fichera, Simão Rodrigues, Francesca Gandini, Alison Meynert, Kevin Donnelly, Timothy J. Aitman et al. 2022. Ancient DNA at the edge of the world: Continental immigration and the persistence of Neolithic male lineages in Bronze Age Orkney. *Proceedings of the National Academy of Sciences* 119(8). e2108001119.

DveV= Oštir, Karel. 1930. *Drei vorslavisch-etruskische Vogelnamen*. Ljubljana: Znanstveno društvo.

EDG = Beekes, Robert S. P. & Lucien van Beek. 2010. *Etymological dictionary of Greek*. 2 vols. Leiden/Boston: Brill.

EDLI = de Vaan, Michiel. 2008. *Etymological dictionary of Latin and the other Italic Languages*. Leiden/Boston: Brill.

EDPC = Matasović, Ranko. 2009. *Etymological dictionary of Proto-Celtic*. Leiden/Boston: Brill.

EDPG = Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic*. Leiden/Boston: Brill.

EDSIL = Derksen, Rick. 2008. *Etymological dictionary of the Slavic inherited lexicon*. Leiden/Boston: Brill.

EM = Ernout, Alfred,Antoine Meillet & Jacques André. 2001. *Dictionnaire étymologique de la langue latine: Histoire des mots*. Paris: Klincksieck.

Eska, Joseph F. 2010. Contact and the Celtic languages. In Raymond Hickey (ed.), *The Handbook of Language Contact*, 538–549. Blackwell Publishing.

Falileyev, Alexander. & Graham. R. Isaac, 2003. Leeks and garlic: the Germanic ethnonym *Cannenefates*, Celtic **kasn-* and Slavic **kesn-*. *North-Western European Language Evolution* 42. 3–12.

FEW = Wartburg, Walter von, Margarete Hoffert, Carl Theodor Gossen, Jean-Pierre Chambon & Jean-Paul Chauveau. 1922. *Französisches etymologisches Wörterbuch: eine Darstellung des galloromanischen Sprachschatzes*. 29 vols. Bonn: Klopp Verlag.

Fleuriot, Léon. 1964. *Dictionnaire des gloses en vieux breton*. (Collection linguistique 62). Paris: Klincksieck.

Furnée, Edzard J. 1972. *Die wichtigsten konsonantischen Erscheinungen des Vorgriechischen*. The Hague/Paris: Mouton.

GPC = Bevan, Gareth A., Patrick J. Donovan, Andrew Hawke, G. Angharad Fychan, Manon W. Roberts, Gwenno Sven-Myer, R. Geraint Gruffydd. 2014. Geiriadur Prifysgol Cymru: A dictionary of the Welsh language. http://www.geiriadur.ac.uk (6 May 2022).

Haak, Wolfgang, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick, Bastien Llamas, Guido Brandt, et al. 2015. Massive Migration from the Steppe Was a Source for Indo-European Languages in Europe. *Nature* 522(7555). 207–11.

Hamp, Eric P. 1990. The pre-Indo-European language of northern (central) Europe. In T. L. Markey & J. A. C. Greppin, (eds.), *When Worlds Collide: The Indo-Europeans and the Pre-Indo-Europeans*, 291–305. Ann Arbor: Karoma.

Hewitt, Steve. 2007. Remarks on the Insular Celtic/Hamito-Semitic question. In Raimund Karl & David Stifter (eds.), *The Celtic World*, Vol. 4: *Celtic Linguistics*, 230–268. London: Routledge.

Hubschmidt, Johannes. 1954. Fr. *laîche*, dt. *lische*: eine gallische und germanische Wortfamilie. *Zeitschrift für celtische Philologie* 24. 81–90.

Huld, Martin E. 2012. Some observations on the development of Indo-European metallurgy. In Martin E. Huld, Karlene Jones-Bley & Dean A. Miller (eds.), *Archaeology and Language: Indo-European Studies Presented to James P. Mallory*, 281–356. Washington D. C.: Institute for the Study of Man.

Hyllested, Adam. 2010. The Precursors of Celtic and Germanic. In Stephanie W. Jamison, H. Craig Melchert, Brent H. Vine, Angelo Mercado (eds.), *Proceedings of the 21st Annual UCLA Indo-European Conference: Los Angeles*, *October 30th and 31st*, *2009*. 107–28. Bremen: Hempen.

Hyllested, Adam. 2016. Again on pigs in ancient Europe: The Fennic connection. In Bjarne Simmelkjær Sandgaard Hansen, Benedicte Nielsen Whitehead, Thomas Olander & Birgit Anette Olsen (eds.), *Etymology and the European Lexicon: Proceedings of the 14th Fachtagung Der Indogermanischen Gesellschaft*, *17–22 September 2012*, *Copenhagen*, 183–96. Wiesbaden: Reichert.

IEW = Pokorny, Julius. 1959–1969. *Indogermanisches etymologisches Wörterbuch*. 2 vols. Bern, München: A. Francke.

Iversen, Rune & Guus Kroonen. 2017. Talking Neolithic: linguistic and archaeological perspectives on how Indo-European was implemented in southern Scandinavia. *American Journal of Archaeology* 121. 511–525.

Koch, John T. 1992. Gallo-Brittonic *Tasc(i)ouanos* 'badger-slayer' and the reflex of Indo-European *g^{wh}*. *Journal of Celtic Linguistics* 1. 101–118.

Koch, John T. 2020. *Celto-Germanic: Later prehistory and post-Proto-Indo-European vocabulary in the north and west*. Aberystwyth: Canolfan Uwchefrydiau Cymreig a Cheltaidd Prifysgol Cymru.

Kroonen, Guus. 2011. *The Proto-Germanic* n-*stems. A study in diachronic morphophonology*. (Leiden Studies in Indo-European 18). Amsterdam: Rodopi.

Kroonen, Guus. 2012a. Non-Indo-European root nouns in Germanic: evidence in support of the Agricultural Substrate Hypothesis. In Riho Grünthal & Petri Kallio (eds.), *A Linguistic Map of Prehistoric Northern Europe*, 239–260. Helsinki: Société Finno-Ougrienne.

Kroonen, Guus. 2012b. On the etymology of Greek *ἄγλις* and *γέγλις* 'garlic': an Akkadian loanword in Pre-Greek. *Journal of Indo-European Studies* 40. 289–299.

Kroonen, Guus, Anthony Jakob, Axel I. Palmér, Paulus van Sluis, and Andrew Wigman. 2022. Indo-European Cereal Terminology Suggests a Northwest Pontic Homeland for the Core Indo-European Languages. *PLOS ONE* 17 (10). e0275744.

Kuiper, F. B. J. 1995. Gothic *bagms* and Old Icelandic *ylgr*. *North-Western European Language Evolution* 25. 63–88.
LED = Smoczynski, Wojciech. 2018. *Lithuanian etymological dictionary*. Berlin: Peter Lang.
LEIA = Vendryes, Joseph,Édouard Bachallery & Pierre-Yves Lambert. 1959–1996. *Lexique étymologique de l'irlandais ancien*. Dublin: Dublin Institute for Advanced Studies.
LHEB = Jackson, Kenneth H. 1953. *Language and history in early Britain: a chronological survey of the Brittonic languages, first to twelfth century A.D*. Edinburgh: Edinburgh University Press.
Loth, Joseph. 1920. Alternances *ai* : *a* ; *ou* : *o* en goidélique et en brittonique. *Revue Celtique* 38. 144–51.
Mac Eoin, Gearóid. 2007. What language was spoken in Ireland before Irish? In Hildegard L. C. Tristram, (ed.), *Celtic Languages in Contact*, 113–125. Potsdam: Potsdam University Press.
Mallory, James P. 2016. Archaeology and language shift in Atlantic Europe. In John T. Koch & Barry W. Cunliffe (eds.), *Celtic from the West 3: Atlantic Europe in the Metal Ages: Questions of Shared Language*, 387–406. Oxford: Oxbow Books.
Marstrander, Carl. 1908. Über irisches *loscann* und einige andere indogermanische Namen der Kröte. In Magnus Bernhard Olsen (ed.), *Sproglige og historiske afhandlinger viede Sophus Bugges minde*, 240–246. Kristiania: H. Aschehoug & Co.
Matasović, Ranko. 2012. The Substratum in Insular Celtic. *Journal of Language Relationship* 8. 153–68.
McCone, Kim. 2005. Mögliche nicht-indogermanische Elemente in den keltischen Sprachen. In Gerhard Meiser & Olaf Hackstein, (eds.), *Sprachkontakt und Sprachwandel*, 395–436. Wiesbaden: Dr. Ludwig Reichert Verlag.
McNeill, Murdoch. 1910. *Colonsay: its plants, climate, geology, etc*. Edinburgh: David Douglas.
Meid, Wolfgang. 1993. *Die erste Botorrita-Inschrift: Interpretation eines keltiberischen Sprachdenkmals*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
Michelena, Luis. 1990. *Fonética histórica vasca*, San Sebastián: Diputación Foral de Guipúzcoa.
Mikhailova, T. A. 2020. Dokel'tskiye yazuki Britanskikh ostrovov [Pre-Celtic languages of the British Isles]. In Jurij Borisovič Korjakov, (ed.), *Paleoevropejskie jazyki*, 323–336. Sankt-Peterburg: Nestor-Istorija.
ML = Meyer-Lübke, Wilhelm. 1911. *Romanisches etymologisches Wörterbuch*. Heidelberg: Carl Winter.
Niedermann, Max. 1916. Zur indogermanischen Wortforschung. *Indogermanische Forschungen* 37. 145–155.
OED = Oxford English Dictionary. www.oed.com (25 January 2022).
Olalde, Iñigo, Selina Brace, Morten E. Allentoft, Ian Armit, Kristian Kristiansen, Thomas Booth, Nadin Rohland et al. 2018. The beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555(7695). 190–96.
Olalde, Iñigo, Swapan Mallick, Nick Patterson, Nadin Rohland, Vanessa Villalba-Mouco, Marina Silva, Katharina Dulias et al. 2019. The genomic history of the Iberian peninsula over the past 8000 years. *Science* 363(6432). 1230–34.
Ó Maolalaigh, R., 2003. Processes in nasalization and related issues. *Ériu* 53. 109–132.
O'Rahilly, T. F. 1942. Notes, mainly etymological. *Ériu* 13. 144–219.
Patterson, Nick, Michael Isakov, Thomas Booth, Lindsey Büster, Claire-Elise Fischer, Iñigo Olalde, Harald Ringbauer, et al. 2022. Large-scale migration into Britain during the Middle to Late Bronze Age. *Nature* 601. 588–594.
SBCHP = Schrijver, Peter C. H. 1995. *Studies in British Celtic Historical Phonology*. (Leiden Studies in Indo-European 5). Amsterdam/Atlanta, GA: Rodopi.
Schrijver, Peter C. H. 1995. De etymologie van de naam van de Cannenefaten [The etymology of the name of the Cannenefates]. *Amsterdamer Beiträge zur älteren Germanistik* 41. 13–22.

Schrijver, Peter C. H. 1997. Animal, vegetable and mineral: Some Western European substratum words. In Alexander Lubotsky, (ed.), *Sound Law and Analogy*, 293–314. Amsterdam: Rodopi.

Schrijver, Peter C. H., 2000a. Keltisch of niet: twee namen en een verdacht accent [Celtic or not: two names and a suspicious accent], in: Rijcklof Hofman, Bernadette Smelik & Lauran Toorians, (eds.), *Kelten in Nederland*, 69–87. Utrecht: de Keltische Draak.

Schrijver, Peter C. H., 2000b. Varia V. Non-Indo-European surviving in the first milennium AD. *Ériu* 51. 195–199.

Schrijver, Peter C. H., 2005. Varia I. More on Non-Indo-European surviving in Ireland in the first milennium AD. *Ériu* 55. 137–144.

Schrijver, Peter C. H., 2011. La langue hattique et sa pertinence possible pour les contacts linguistiques préhistoriques en Europe occidentale. In Coline Ruiz Darasse & Eugenio R. Luján, (eds.), *Contacts linguistiques dans l'Occident méditerranéen antique*, 241–255. Madrid: Casa de Velázquez.

Schrijver, Peter C. H., 2018. Talking Neolithic: the case for Hatto-Minoan and its relationship to Sumerian. In Guus Kroonen, James P. Mallory & Bernard Comrie, (eds.), *Talking Neolithic: Proceedings of the Workshop on Indo-European Origins Held at the Max Planck Institute for Evolutionary Anthropology*, *Leipzig*, *December 2–3*, *2013*, 336–374. Washington D. C.: Institute for the Study of Man.

Schumacher, Stefan. 2004. *Die keltischen Primärverben: ein vergleichendes*, *etymologisches und morphologisches Lexikon*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Sims-Williams, Patrick. 2020. An alternative to 'Celtic from the East' and 'Celtic from the West.' *Cambridge Archaeological Journal* 30. 1–19.

Sjögren, Karl-Goran, Iñigo Olalde, Sophie Carver, Morten E. Allentoft, Tim Knowles, Guus Kroonen, Alistair W. G. Pike, Peter Schröter, Keri A. Brown, Kate Robson Brown, Richard J. Harrison, Francois Bertemes, David Reich, Kristian Kristiansen, Volker Heyd. 2020. Kinship and social organization in Copper Age Europe. A cross-disciplinary analysis of archaeology, DNA, isotopes, and anthropology from two Bell Beaker cemeteries. *PLoS ONE* 15, e0241278.

Stifter, David. 2005. Zur Bedeutung und Etymologie von altirisch *sirem*. *Die Sprache: Zeitschrift für Sprachwissenschaft* 45. 160–189.

Stifter, David. 2010. The invisible third. The Basque and Celtic words for 'swallow.' *Ériu* 60. 145–57.

Stifter, David. 2015. The language of the poems of Blathmac. In Pádraig Ó Riain (ed.), *The Poems of Blathmac Son of Cú Brettan: Reassessments*, 47–103. London: Irish Texts Society.

Stifter, David. 2017. The phonology of Celtic. In Jared S. Klein, Brian D. Joseph & Matthias Fritz, (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, 1188–1202. Berlin/Boston: De Gruyter Mouton.

Stifter, David. 2020. Old Irish etymology through the ages. *Language & History* 63(1). 24–46.

Stifter, David. 2023a. The rise of gemination in Celtic [version 1; peer review: awaiting peer review]. *Open Research Europe* 3(24). https://doi.org/10.12688/openreseurope.15400.1.

Stifter, David. 2023b. With the back to the ocean. The Celtic maritime vocabulary. In Kristian Kristiansen, Guus Kroonen & Eske Willerslev (eds.), *The Indo-European Puzzle Revisited: Integrating Archaeology*, *Genetics*, *and Linguistics*, 172–192. Cambridge: Cambridge University Press.

Stika, Hans-Peter & Andreas G. Heiss. 2013. Plant cultivation in the Bronze Age. In Harry Fokkens & Anthony Harding (eds.). *The Oxford Handbook of the European Bronze Age*. 348–369. Oxford: Oxford University Press.

Stokes, Whitley. 1888. H. 2. 17. *Revue Celtique* 9. 241–244.
Thorsø, Rasmus, Andrew Wigman, Anthony Jakob, Axel I. Palmér, Paulus van Sluis & Guus Kroonen. 2023. Word mining: metal names and the Indo-European dispersal. In Kristian Kristiansen, Guus Kroonen & Eske Willerslev (eds.), *The Indo-European Puzzle Revisited: Integrating Archaeology*, *Genetics*, *and Linguistics*. Cambridge: Cambridge University Press.
Thurneysen, Rudolf. 1917. Irisches. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 48. 48–75.
Trask, R. Larry. 1997. *The history of Basque*. London/New York: Routledge.
van Bezooijen, Renée & Marinel Gerritsen. 1994. De uitspraak van uitheemse woorden in het Standaard-Nederlands: een verkennende studie The pronunciation of foreign words in Standard Dutch: an exploratory study. *De Nieuwe Taalgids* 87. 145–160.
van Sluis, Paulus S. 2022. Beekeeping in Celtic and Indo-European. *Studia Celtica* 56. 1–28.
van Sluis, Paulus S., Anders R. Jørgensen & Guus Kroonen. 2023. European prehistory between Celtic and Germanic: the Celto-Germanic isoglosses revisited. In Kristian Kristiansen, Guus Kroonen & Eske Willerslev (eds.), *The Indo-European Puzzle Revisited: Integrating Archaeology*, *Genetics*, *and Linguistics*, 193–243. Cambridge: Cambridge University Press.
Vennemann, Theo. 2003. Syntax und Sprachkontakt: mit besonderer Berücksichtigung der indogermanischen Sprachen des Nordwestens. In Alfred Bammesberger & Theo Vennemann, (eds.), *Languages in Prehistoric Europe*, 333–364. Heidelberg: Carl Winter.
VKG I = Pedersen, Holger. 1909. *Vergleichende Grammatik der keltischen Sprachen. Vol. I*. Göttingen: Vandenhoeck und Ruprecht.
VKG II = Pedersen, Holger. 1913. *Vergleichende Grammatik der keltischen Sprachen. Vol. II.* Göttingen. Vandenhoeck und Ruprecht.
Weiss, Michael. 2010. Observations on the *littera* rule. https://www.academia.edu/1709398/Observations_on_the_Littera_Rule (6 May 2022)
Witczak, Krzysztof Tomasz. 2005. Iranian **paina-* 'honey' and **hangu-* '(queen) bee.' *Studia Etymologica Cracoviensa* 10. 205–9.

Anders Richardt Jørgensen

5 A bird name suffix *-*anno*- in Celtic and Gallo-Romance

1 Introduction

Interest in possible substrate vocabulary of the Celtic languages has seen a resurgence since the publication of Schrijver (1997), upon which many of the more recent studies build. Schrijver focuses on establishing systematic irregularities in words with roughly the same meaning across various European Indo-European branches, words which look alike, but which cannot be related by the established sound laws. This is done by making a tentative alternative set of possible correspondences valid for a larger group of proposed "cognates", reflecting either independent transfers from the same substrate into varieties of already differentiated Indo-European or regular alternations in the substrate language whose distribution is no longer recoverable.

The approach adopted in the present study[1] is somewhat different. The suffix *-*anno*-[2] discussed here is only described as substratal due to the absence of an etymology taking it back to Proto-Indo-European, whereas the "cognates" identified through systematic irregularities are, if accepted, by definition not inherited, but either substratal or borrowed. Hence, the signal identified in the present study might be stronger, but the origin of the signal (substratal, loanword or inherited) is more difficult to establish.

Trying to identify substratal bird names comes with particular challenges. These were recently discussed by Matasović (2020). Among other things Matasović points out (2020: 331–334) that birds are frequently named after the sounds they

1 The research in the present study was carried out as part of the research projects "A new perspective on French Historical Phonology – What loan words in Breton can tell us" funded by the Swedish *Vetenskapsrådet* (ref. 421-2014-1119), "Connecting the Dots: Reconfiguring the Indo-European Family Tree" funded by the Independent Research Fund Denmark (ref. 9037-00086B) and "Languages and Myths of Prehistory (LAMP)" funded by *Riksbankens Jubileumsfond* (ref. M19-0625:1). I furthermore wish to thank Herve Bihan as well as Paulus van Sluis, Guus Kroonen, Peter Schrijver, David Stifter and the rest of the participants at the *Sub-Indo-European Europe* workshop in Leiden, August 30–31, 2021, for useful discussion and input.

2 As discussed below (§ 4), it is not entirely clear whether we should reconstruct this suffix as *-*anno*- or *-*asno*-, since the contrast between the resulting *-*ænno*- and *-*anno*- may not have been preserved outside the initial syllable in later Celtic. In the following it will simply be reconstructed as *-*anno*-, but this should be understood as possibly covering both possibilities.

https://doi.org/10.1515/9783111337920-005

make, hence unrelated onomatopoeic creations will often look alike across different languages. Likewise, there typically appears to be a large turn-over of bird names in speech communities, making the semantic field inherently unstable. These difficulties are less pertinent to the present study, however, since the object identified is a "suffix" occurring in several different words in the same semantic field. It is therefore unlikely to be onomatopoeic or due to random noise. Indeed, the fact that the element **-anno-* apparently is restricted to bird names only strengthens the case that it really is a meaningful string of phonemes (whether originally a suffix, a noun class suffix or a compositional element). This stands in contrast to e.g. the situation in Pre-Greek (as presented by Beekes 2014: 28–42), where many of the suffixes identified mainly appear to be composed of recurring strings of phonemes towards the end of words typically without a clearly identifiable meaning.

2 The material

In the following section the bird names argued to contain the element **-anno-* will be treated, along with a discussion of the Proto-Celtic reconstruction and possible extra-Celtic cognates. Evidence from Breton will prove particularly relevant for the reconstruction due to the preservation in this branch of phonemic contrasts lost early on in both Goidelic and in the other Brittonic languages.

2.1 Proto-Celtic **kau̯anno-* 'owl'

LLat. *cauannus** (← Gaulish; Delamarre 2003: 111; Lambert 2003: 195).
MW *cuan* (GPC 626).
OBret. *couann* (DVB 120, Bauer 2008: 137), MBret. *couhen* (Cms.), *couhenn* (Ca.), Late MBret. *caoüen* (Nom.), pl. *caouennet* (VEach 16), Early MoBret. *caouen*, *caouenn*, pl. *-ennet* (SCger.), MoBret. *kaouenn* fem.; Early HVann. *cohann*, *cohàn.*

The first bird name to be treated is PCelt. **kau̯anno-*[3] 'owl'. It is attested in Late Latin as *cauannus**, presumably borrowed from Gaulish. The borrowing is contin-

3 Strictly speaking, the reconstructions provided here and in the following are not necessarily of Proto-Celtic age. On the basis of the attestations alone we cannot be certain that the words actually date back this far. However, the Proto-Celtic stage will be used since it provides a useful point of phonological reconstruction.

ued in Gallo-Romance in a multitude of forms, e.g. OFr. *chouan*, OOcc. *cauana*, MoFr. *chouan*, *chat-huant* reflecting either a Gaul. **kau̯anno-* or its feminine counterpart **kau̯annā* (FEW 2: 548–550).[4] The word is not attested in Goidelic and only marginally in Welsh, namely twice (with the second attestation immediately following the first) in the early Middle Welsh tale *Culhwch ac Olwen* (lines 871–872 in the edition by Bromwich and Evans 1992). There it is given as the name of a legendary animal, the *Cuan Cwm Kawlwyt*, 'the *Cuan* of *Cwm Kawlwyt*'. In later folklore, this creature is named *Tylluan Cwm Cawlwyd*, 'the owl of *Cwm Cawlwyd*', allowing us to infer that *cuan* in all likelihood had the meaning 'owl' (cf. Bromwich and Evans 1992: 143). Such an interpretation, based on the text of *Culhwch ac Olwen*, may also lie behind the much later attestation provided by Lhuyd (1707: 234a). Therefore this and subsequent lexicographical attestations do not necessarily indicate that the word survived in spoken Welsh, a possibility which warrants some caution in taking the spelling at face value. There does not appear to be any reflexes in Cornish, but the word is amply attested in Breton, appearing already in Old Breton as *couann* (DVB 120; Bauer 2008: 135). In Middle Breton we find *couhen* in the Cms. (1464) and *couhenn* /kou̯(h)enn/ in the Ca. (1499) with a surprising *-en(n)* for the expected **-ann* (on this see below § 3.2).[5] The Early Haut-Vannetais forms *cohann*, *cohàn* on the other hand are as expected, directly reflecting a form with PBrit. **-ann* < PCelt. **-anno-/-ā* (cf. § 3.1).

The most straightforward reconstruction appears to be PCelt. **kau̯anno-*, a form which effortlessly explains the Late Latin and Gallo-Romance forms. In Breton we may see a change **a'u̯V* > *ou̯V* (further to *o(h)V* in Haut-Vannetais), as in MBret. *louazr* /lou̯aðr/ 'trough', HVann. *lo(h)er* < **lau̯adr* (Gaul. *lautro*), MBret. *louen* /lou̯en/ 'joyous' < **lau̯en* (MW *llawen*, Gaul. pers. name *Lauenus*), MBret. *moues* /mou̯es/ 'woman', HVann. *moéz* < **mau̯es* < **magu-* + fem. **-issā* (MCorn. *mav* 'lad; servant', Gaul. *magu-*). However, there are counterexamples to this sound change, e.g. MBret. *auel* 'wind' < **au̯el* (W *awel*) and MBret. *cauell* 'cradle' < **kau̯ell* (W *cawell*) and reconstructions with initial **kou̯-*, **kuu̯-* (Schrijver 1995: 335; Zair 2012) or even **kāu̯-* may also be possible. Be that is it may, the vocalism

4 The relevance of the apparent reflexes of a Gaul. **kău̯ā* in OFr. *choue*, OPic. *chaue* 'owl' and its diminutive derivative *chouette* is not entirely clear (not to be confused with the homophonous OFr. *choe* 'jackdaw', MFr. *chouette* 'crow' borrowed from WGm. **kawō*). As interesting as they may seem, it cannot be excluded that they were formed within Gallo-Romance by substitution of the perceived suffix of **kau̯anno-*.

5 The Middle Breton adjective *couen*, *coen* /kou̯en/ 'frightful, terrible' occurring in the poetic corpus must be etymologically unrelated. It is consistently spelled with single final <-n> and rhymes with words with single /-n/ (J 99/2044, Pm. 273/6026, M 1175, 1609, 1741, 2027, 2256, 2290, 2413, 2567, 2607, 2675, 3070, 3582).

of the first half of the word should not affect the argument presented here, which only concerns **-anno-*. As to further etymological connections, *cauannus* has been compared to OHG *hūwo* 'owl', *hūh* 'id.' and PSl. **sova* 'owl'.[6] However, it is very likely that all these words for 'owl' are onomatopoeic in origin. This naturally does not preclude an etymological connection, but it makes it very difficult to demonstrate. Whether or not we accept the proposed extra-Celtic connections, PCelt. **-anno-* remains unaccounted for.

2.2 Proto-Celtic **gulbanno-* 'sparrow'

MIr. *gelbann*, *gealbhonn* masc. *o*-stem 'sparrow' (DIL G-60).
MoW *golfan* 'house-sparrow' (GPC 1451).
OCorn. *goluan* gl. *passer* (Graves 1962: 221; Campanile 1974: 50)
MBret. *goluan* (Cms., Donoet), *goluann* /golvann/ (Ca.), fem. *goluannes* (Donoet), pl. *guelvenn* /gelvenn/[7] (Dag. 195), Late MBret. *goluen*, dim. *goluennicq* (Nom.), Early MoBret. *golven*, pl. *guelvin*, *gueluen* (SCger.); Early HVann. *golvàn.*

Inherited reflexes of PCelt. **gulbanno-* 'sparrow' are only found in OCorn. *goluan* and in Breton, from MBret. *goluan* (Cms.), *goluann* (Ca.) onwards. Therefore its status as Proto-Celtic or even Proto-Brittonic is not certain. We do admittedly also find a reflex of the word in Middle and Modern Irish, e.g. MIr. pl. *min-gelbuind* 'little sparrows', MoIr. *gealbhan* 'sparrow'. However, the vocalism /e/ and lenition[8] in the cluster **-lb-* strongly hints at this being a borrowing from some form of Brittonic, either from an unattested OMW sg. **gəlvan(n)* < PBrit. **gulβann* or from a Brittonic plural with final and internal *i*-affection (thus Pedersen 1909–1913, 1: 184; Campanile 1974: 50), as seen in MBret. pl. *guelvenn* /gelvenn/ (Dag. 195). Welsh *golfan*, first attested in the 17th c., is likewise a borrowing, but a purely lexicographical one, the word being a modernization of the OCorn. *goluan*.

6 As e.g. by IEW (536), Schrijver (1995: 335), Matasović (2009: 196, 2020: 333 fn. 1), Lloyd and Lühr (2009: 1199–1200, 1302), Derksen (2008: 461–462).

7 From PBret. pl. **gölβe̯nn* (with final and internal *i*-affection) < **gulbann-ī* < **-oi̯.* Le Bihan (2013: 109), following a suggestion by Ernault (in Largillière 1928–1929: 672), takes *guelvenn* to be an error for *quelyen* 'flies' (W *cylion*). However, the surrounding lines clearly demand a rhyme in /-enn/, which is fully compatible with *guelvenn* /gelvenn/, but difficult to square with a plural/collective *quelyen* ending in *-yen* /-i̯en/ < PBret. **-i̯on.* Since both 'sparrows' and 'flies' fit the context, I see no reason not to take the manuscript at face value.

8 A reconstruction such as PCelt. **ge/olobanno-* vel sim. would presumably account for both the Brittonic and the Goidelic forms. However, this would require positing ablaut *e/o* and it would furthermore sever the attractive connection to the Celtic words for 'beak'.

The single final <-n> in OCorn. *goluan* cannot be taken as an indication of a PBrit. single *-*n* since the *Vocabularium Cornicum* consistently spells PBrit. *-*nn* with a single <-n>.[9] The word is well-attested in Breton, from Middle Breton onwards. In the Ca. it is written *goluann*, clearly indicating /-ann/ from PBrit. *-*ann*, a reconstruction supported by Early HVann. *golvàn* (cf. § 3.1). The reconstruction with *-*ann* finds additional indirect support in the later LTK Breton forms ending in *-en(n)* (cf. § 3.2).

The word is reconstructed here as **gulbanno*- rather than the equally possible **golbanno*-. This is solely due to etymological considerations: it is very likely that it belongs to a well attested cluster of words in Celtic based on an apparent Proto-Celtic "root" **gulb(V)*- 'beak'. This is seen in OIr. *gulban*, MW *gyluin*, OCorn. *geluin* 'beak' < PCelt. **gulbīno*-, MW *gwlf* 'one of the ends of a bow having a groove or a notch cut into it to provide a firm seat for the end of the bowstring' < PCelt. **gulbo-/u-/i*-, MW *gylf* 'sharp pointed instrument; beak' < **gulbi̯o*-, Fr. *gouge* 'gouge' (a kind of chisel) < LLat. *gulbia* ← Gaul. **gulb(i)i̯ā* (Delamarre 2003: 184), MoBret. (Vann.) *gilvig*[10] < dim. **gulbīko*-. A PCelt. **gulb-anno*- 'sparrow' < 'beak-bird', would make sense semantically, since sparrows have strong and thick beaks. A semantic parallel is provided by French *gros-bec*, literally 'thick-beak', the name of various grain-eating passerine birds. In Serquiais Norman French [gro bɛ:k] specifically translates Fr. *moineau* 'sparrow' (ALF map 866, point 398).

The further origin of Celtic **gulb(V)*- 'beak' is unclear. It has been related to Baltic and Slavic words for 'swan', Lith. dial. *gulbìs*, PSl. **kъlpь* (cf. Derksen 2008: 261, 2015: 192–193), and both have been taken to come from a Pre-Indo-European substrate (thus e.g. Matasović 2009: 169, 2011: 9). Irrespective of whether this is correct or not, *-*anno*- behaves like a derivational suffix added to **gulb(V)*- 'beak' internally in Celtic.

9 As shown by e.g. *pen* 'head', *guan* 'weak', *guyn* 'white', *pren* 'tree', *glan* '(river)bank', *len* 'cloth' and the numerous singulatives in *-en*, such as *steren* 'star'.

10 The name of various marine birds with characteristic beaks, originally presumably 'little beaked one'. The word appears to be restricted to Île-de-Groix. The following attestations are given in the *Ichthyonymie bretonne*: [gilvic bek ˈpla:t] 'razorbill; *Alca torda*' (Berr 1986, 2: 473), [gilvic bek ˈpik] 'the common murre; *Uria aalge*' (Berr 1986, 2: 474), [gilvic beg ˈry] 'Atlantic puffin; *Fratercula arctica*' (Berr 1986, 2: 475).

2.3 Proto-Celtic **u̯aḭlanno-* 'seagull'

OIr. *foílenn* /foḭl'ənn/ masc. *o*-stem/fem. *ā*-stem (DIL F-22), MoIr. *faoileann* fem. 'fair maiden', *faoileán* masc. 'seagull' (with the dim. suffix *-án*, possibly foreshadowed by the peculiar Old Irish hapax <foilan>).
MW *gwylan*, pl. *gwylein* 'seagull; fair maiden' (GPC 1760)
OCorn. *guilan* gl. *alcedo* (Graves 1962: 218–219; Campanile 1974: 56), LCorn. *gullan* (George 2009: 236).
OBret. pl. *guilannou* (DVB 191; Bauer 2008: 72), MBret. *goelann* (Ca.), Early MoBret. *goulen*, pl. *gouleny* 'goilan, oiseau de mer' (GReg.); Early HVann. *gouilann*.

As argued by Schrijver (1995: 115–116), the attested forms are best explained departing from a PCelt. **u̯aḭl-*. This reconstruction is possible if we accept that W *gwylan* is a borrowing from SWBrit., where PCelt. **u̯aḭl-* > **u̯ēl-* > PBrit. **u̯oɪl-* would regularly give**(g)u̯ɯl-* due to the change of pretonic **(g)u̯o-* to **(g)u̯u-*. OIr. *foílenn* admittedly shows the wrong diphthong, /oḭ/ instead of the expected /aḭ/. Due to this irregularity, *foílenn* has been taken to be a borrowing from Brittonic, from a form such as OW **(g)u̯oɨlan(n)* or **(g)u̯uɨlan(n)* (Pedersen 1909–1913, 1: 23, 184). However, it may also be inherited, showing an early case of confusion between the two diphthongs (cf. McCone 1996: 139), possibly due to the preceding labial consonant.[11]

As for complete reconstructions, we have Stokes and Bezzenberger's PCelt. **u̯oḭlenno-* (1894: 285), which is completely incompatible with the Brittonic data and should be abandoned. Matasović (2011: 42) reconstructs a masc. **u̯aḭlino-* (presumably to account for the palatalization of *-l-* in OIr. *foílenn*) and a fem. **u̯aḭlanā* (to account for the vocalism in the final syllable of W *gwylan*). However, he does not explain the origin of this unparalleled suffixal allomorphy, masc. **-ino-*/fem. **-anā*. Elsewhere, Matasović (2012: 157) argues that the word most likely stems from an unknown substrate language. If this is the case, Goid. **-ino-*, Brit. **-anā* might have an explanation in the donor language (although this is not a pattern identified in other potential substrate words). Even if such a scenario cannot be excluded, we should ideally attempt to reconstruct only one form that could account for both the Goidelic and the Brittonic attestations. There is a single reconstruction which will account for the palatalization in Irish and the vocalism in Brittonic, namely PCelt. **u̯aḭlanno-/ā.* This would regularly give Goid. **u̯aḭlenno-* (cf. McCone 1996: 56–57,

11 Cf. the reflex of PCelt. **maḭlo-* 'bald' (MW *moel*, MBret. *moal*) in Ogam Ir. *MOLE-GOMRID* (McManus 1991: 121) and OIr. *móel caich* in the Stowe Missal (Stokes and Strachan 1901–1903, 2: 251) and the reverse confusion in OIr. *maídem* (Wb., *prima manus*) for the usual *moídem* 'boasting' and Ogam Ir. Gsg. *VRAICCI*, next to OIr. Gsg. *froích* 'of heather' < **u̯roḭkī* (McManus 1991: 121).

106) with subsequent palatalization of the *-l-* between two front vowels. PCelt. **u̯aḭlanno-/ā* would also regularly produce SWBrit. **(g)u̯ulann*, from which we get OBret. *guilannou*, MBret. *goelann* and, with loss of the length contrast after unstressed vowel, MW *gwylan* and OCorn. *guilan*. For the Modern Breton forms with *-en(n)*, see § 3.2 below.

As to the further etymology of PCelt. **u̯aḭlanno-/ā*, an internal Celtic connection with the word **u̯aḭlo-* 'wolf' has been suggested (Schrijver 1995: 116; cf. MIr. *fael* 'wolf'). This makes sense semantically, at least if **u̯aḭlo-* itself is a derivative of PCelt. **u̯aḭ* 'woe' (IEW 1111). The original meaning of **u̯aḭlo-* would then be 'wailer' and the formation **u̯aḭlanno/ā* 'wailer-bird', an apt description of the seagull. If this analysis is correct, **-anno/ā* behaves very much like a derivational suffix, possibly to an inherited stem.

2.4 Proto-Celtic **giguranno-* 'barnacle goose'

OIr. *gigrann*, *giugran* masc. *o*-stem 'a wild goose, barnacle goose' (DIL G-81).
MoW *gwyran* masc./fem., pl. *gwyrain* 'barnacle-goose', pl. also 'barnacles' (GPC 1781).
OBret. *goirann* (DVB 178; Bauer 2008: 102).

The last proper bird name under consideration is 'barnacle goose'. It is attested in Irish, Welsh and Old Breton, but unknown in Cornish and later Breton. The first part has been treated in detail by Schrijver (1995: 358), who convincingly reconstructs PCelt. **gigur-*. This effortlessly accounts for the Irish form and it is most likely compatible with PBrit. **gu̯ɪr-*, assuming that **gigur-* was syncopated to **gɪɣ'r-* to give **gēr-* and finally, with regular diphthongization, PBrit. **gu̯r-*. The full reconstruction **giguranno-* given here, instead of Schrijver's **gigurano-*, is solely based on the Old Breton form *goirann* (DVB 178). Neither Welsh, with its early loss of the contrast between long and short resonants after unstressed vowels, nor Irish, with MacNeil's Law, provide any information on the matter. Accordingly, it is simplest to take OBret. *goirann* at face value as pointing to **-anno-*.[12]

12 The collection of glosses in which OBret. *goirann* occurs (in BN ms lat. 10290; cf. Lambert 1982 and Bauer 2008) is otherwise quite consistent in the use of postvocalic single <-n> and double <-nn> for /-n/ and /-nn/ respectively.

With etymological /-n/ we have plurals/collectives in /-i̯on/: *doguormaheticion*, *meltion*, *moruion*, *unsillabochion*; diminutives in /-an/: *Antunan*, *belann*, *bodaran*, *catoinan*, *cauellan*, *ceneuan*, *coloinan*, *genouan*, *guinan*, *lohan*, *Patrican*, *poplan*, *uinan*; various derivatives in *-in* /-in/: *bledin*, *brientin*, *euin*, *gilbin*, *orin*; various other words ending in /-Vn/: *calonn*, *cicguan*, *cochin*, *din*, *enuen*, *erbin*, *genn*, *glan*, *guan*, *ou hun*, *milin*, *morbran*, *tan*, *taran*, *un*.

The further etymology of **giguranno-* is unclear. It is generally assumed to be of onomatopoeic origin (GPC 1781), which seems reasonable. Stifter (2019) relates it to the generic Celtic word for 'goose' (OIr. *géd*, MoW *gŵydd*, MBret. *goaz*). He argues that the only regular reconstruction for this word is **giɣðo-*, with a unique cluster **-ɣð-*. This cluster cannot be inherited from Proto-Celtic (PCelt. **-gd-* would probably result in an unlenited Ir. /d/ rather than the attested /ð/) and accordingly, the word must have entered at a later stage, possibly Insular Celtic. This **giɣðo-* is then compared to our **gigur(anno)-* (phonetically probably [giɣur-]). This is possible if we assume that the words were borrowed from the same substratum language with internal dialectal differences, at different times or with different adaptations of substratal phonology. If this is correct, **giguranno-* appears to contain a segmentable part **-anno-* comparable to the **-anno-* in the preceding words.

2.5 Proto-Celtic **barannīko-* 'limpet, barnacle'

OIr. *bairnech* masc. *o*-stem 'limpet; barnacle' (DIL B-19).
MoW coll. *brennig*, sglt. *brennigen* fem. 'limpet' (GPC 320).
LCorn. coll. *brennik*, sglt. *brenigan*, *bernîgan* (George 2009: 106).
MBret. coll. *brinicq* (Am. 538), MoBret. *brennig*, *berinig* 'limpets' (→ OFr. *bernicle* 'limpet' and 'barnacle goose').

The final word, here tentatively reconstructed as **barannīko-* or **barannīkā*, means 'barnacle; limpet', i.e. it is not actually a bird name but either the name of a crustacean (barnacle) or snail (limpet) with a hard shell which attaches itself to rocks and other objects in the sea. It is well attested in both Goidelic and Brittonic. In the latter branch the resulting **bęŗennig* underwent a somewhat irregular syncope of the first vowel to **brẹnnig*.[13] Dialectally in Breton a form without syncope

---

With etymological /-nn/ we have singulatives in /-enn/: *aballen*, *cennenn*, *notenn*, *tonnenn*, *torocenn*; various other words with /-Vnn/: *bonn*, *genn*, *penn*, *tonn*, *unpenn*.

The only irregularities are *belann* /belan/(?) '(little) bee' (derivationally unclear but presumably a diminutive in *-an*; alternatively = MIr. *bechlann* 'beehive'), *genn* /gen/ 'jaw' (possibly influenced by the gloss *genn* /genn/ 'wedge' occurring on the same page), *aballen* /aβallenn/ 'apple tree' and possibly *calonn* /kalon/(?) 'heart' (Middle and Modern Breton point to /kalon/, but the MW pl. *calonneu* points to /-nn/).

13 We have similar instances of irregular syncope between a consonant and a liquid in e.g. MW *cryd* 'shoemaker' (< **kęrɪð* < PCelt. **kariio-*, cf. MBret. *quere*), MW *yspryt* 'spirit' (< **spɪrɪd* < Lat. *spīritus*, cf. MBret. *speret*), MBret. *coulm* 'dove' (< **kolom* < Lat. *columba*, cf. MW *colomen*, OCorn. *colom*). In **bęŗennig* 'limpets, barnacles' the first vowel will have been continuously unstressed

survives, K *berinig*,[14] and this form was borrowed by Old French as *bernicle* (with several secondary forms developing from this, cf. FEW 2: 2–3). The connection to birds is evident in Fr. *bernicle*, which means both 'barnacle' and 'barnacle goose'.[15] The reason for this double meaning is the widespread popular belief that barnacle geese, whose eggs and young birds were never encountered, actually grew from barnacles (cf. FEW 20: 2; White 1945; Buckeridge 2011; we now know that they simply migrate to the Arctic and breed there). In light of this belief, it is at least possible that **barannīko-* is a derivative formed with the very common adjectival or diminutive suffix **-īko-* from a bird name in **-anno-*.

As to the further etymology, OIr. *bairnech* is traditionally derived from *bairenn* fem. *ā* 'large stone; rocky district' (Stokes and Bezzenberger 1894: 162; LEIA B-9; DIL B-19). While this makes sense semantically, the palatalization in *bairenn* is incompatible with a Proto-Celtic reconstruction **baranno-* > Goid. **barenno-* (McCone 1996: 116). To address this problem, the following possibilities are available:

a. OIr. *bairenn* and *bairnech*, *brennig*, etc. are etymologically unrelated.
b. OIr. *bairenn* and *bairnech*, *brennig* reflect independent derivatives of the same putative Proto-Celtic noun **bar(V)-* 'rock(?)',[16] namely PCelt. **bar-īnā* 'rocky (place/object)'[17] and **bar-anno-* 'rock-bird'[18] (i.e. 'barnacle goose; barnacle') respectively. These derivatives would be parallel to **gulb-īno-* 'beak-like thing; beak' and **gulb-anno-* 'beak-bird' (i.e. 'sparrow') discussed in § 2.2.

---

from Old British onwards, meaning that the time of its loss cannot be dated in relation to the accent retraction.

14 See Berr (1986, 2: 218–221) for an overview of the dialectal forms. MBret. *brinicq* and Modern Breton K forms show a single *-n-*, probably taken from the singulative *bren(n)igenn* where the *-nn-* followed an unstressed vowel.

15 Cf. MoW *gwyrain* pl. 'barnacle geese' and 'barnacles' and MoBret. *garreli* 'brant goose' (Berr 1986, 2: 453) and 'barnacle' (Berr 1986, 2: 265–266; probably a borrowing from Gallo-Romance, cf. OPic. *garlon* 'pigs' feet', FEW 4: 68, and Vann. *treid-moh* 'barnacles', lit. 'pigs' feet' for the semantics). The double meaning is also encountered in MMoFr. *cravan* 'barnacle goose' and 'barnacle' (FEW 2: 1266), cf. § 6.

16 Possibly from PIE **g^{w}r̥H(-i)-* (Matasović 2009: 57; cf. Skt. *girí-* 'mountain', Alb. *gur* 'rock'). Alternatively from the root **b^{h}erH-* 'to fashion with a sharp tool' (IEW 134; Bernardo Stempel 1999: 459).

17 As reconstructed by Bernardo Stempel (1999: 459), alternatively **barinā* (Matasović 2009: 57). Neither form can be the basis of W *brennig*, Bret. *brinnig* due to the PBrit. **-nn-*.

18 One may note the vague similarity to Germanic names for the barnacle goose and similar aquatic birds, e.g. Du. *brandgans* 'barnacle goose; *Branta leucopsis*', E *brant* or *brent (goose)* '*Branta bernicla*', Sw. *brandgås* 'the common shelduck; *Tadorna tadorna*'. Unless the first part of this is simply Germanic **branda-* 'a fire, burning', it is at least possible that the hypothesized PCelt. **baranno-* 'barnacle goose' and PGm. **branda-* may reflect independent borrowings of the same word from an unknown language.

c. Palatalization has been introduced from cases with syncope, i.e. Gsg. and NADpl.
d. The reconstruction **baranno-* is wrong and *bairenn*, *bairnech*, *brennig*, etc. have nothing to do with the bird names in **-anno-* (a distinct possibility). We should instead reconstruct **barinno-* as the basis, to account for the palatalization in Irish *bairenn* and the long *-nn-* in Brittonic. However, it is possible that a **barinnīko-* should lead to MW ***berynnig*.

In light of the evidence for a bird name suffix **-anno-* in **kaṷanno-*, **gulbanno-*, **ṷaiḷlanno-* and **giguranno-* given above, it is at least possible that a **baranno-* 'barnacle; barnacle goose' once existed, forming the derivational basis for **barann-īko-* '(one) pertaining to a barnacle goose; a little barnacle goose'.[19] The proposals under a. and b. above are compatible with this assumption. The proposal under c. is more complicated, since *bairenn* would then indicate that the meaning of **barannā* was 'large rock, rocky district', rather than 'barnacle (goose)', as would be expected if **baranno/ā* really contained a bird name suffix. There may be ways of accounting for this discrepancy, such as interpreting **barannā* as an old neuter collective 'mass of barnacles' → 'place of barnacles' → 'rocky district', but it would be preferable not having to rely on such additional assumptions.

2.6 Possible additional material

Apart from the bird names (and related) discussed in § 2.1–5, there are a number of other words in Celtic with uncertain etymology seemingly displaying a "suffix" **-anno-*. However, these are nowhere nearly as safely reconstructible. OMIr. *eidenn* masc. *o*-stem 'ivy' is most straightforwardly derived from a PCelt. **edanno-* or **edenno-* in order to explain the palatalization of **d* and lack of raising of the initial **e-*. However, MW *eidew*, *eido* and MBret. *ilyeauenn*, *ilyo* 'ivy' (with irregular *-l-*) are incompatible with this reconstruction and point to a longer proto-form. OMIr. *léibenn* 'level surface' and MoW *llwyfan* 'stage, platform' may be mechanically reconstructed back to PCelt. **leiḃanno-*. However, a derivation from the Proto-Celtic root **sleiḃ-* 'smooth surface' (cf. OIr. *slíab* 'mountain' < **sleiḃ-es-* and MW *llyfn* 'smooth, level', OIr. *slemon* < **slib-no-*) seems very probable, meaning that the word can only be inherited in Brittonic (due to **sl-* > **l-* here). The Irish

19 One could consider including Gasc. *barenne* 'scallop' (Arcachon; FEW 21: 267) in the discussion as a reflex of a Gaulish **barennā* < **barœnnā* < **barannā*. However, *barenne* is poorly attested, geographically very restricted and presents an imperfect semantic match. Therefore it is most likely not relevant to the question at hand.

word must then instead be a borrowing from Brittonic. It is also possible that MIr. *loscann*, *loiscenn* masc. *o*-stem 'toad' and MW *llyffant* masc. 'id.' reflect a preform with *-*anno*-. However, the medial consonant cluster is difficult to reconstruct and the final *-t* in Brittonic will have to be secondary. Explaining these irregularities as being due to the words having been independently borrowed from a third language is possible, but one should note that we find no such irregularities in the bird names. Furthermore, a reconstruction with PCelt. *-*ando*- rather than *-*anno*- may also be possible in this instance due to the lack of sufficiently early attestations in Irish.

3 Reconstruction of the "suffix": *-*n*- or *-*nn*-?

In the Proto-Celtic reconstructions given in the above, I have reconstructed the "suffix" as *-*anno*- (or its feminine counterpart *-*annā*) in all instances. However, several other Proto-Celtic reconstructions are encountered in the treatments of the individual words, e.g. **golbano*- (implied by Campanile 1974: 50), **u̯oi̯lenno*- (Stokes and Bezzenberger 1894: 285), **u̯ai̯lino-*/**u̯ai̯lanā* (Matasović 2011: 42), **gigurano*- (Schrijver 1995: 358; Stifter 2019: 315), **barennīkā* (Stokes and Bezzenberger 1894: 162), **barinīko*- (implied by Matasović 2009: 57). In light of this it may be relevant to review the material, this time focussing specifically on the question of whether we should reconstruct *-*n*- or *-*nn*-.

In Old Irish the operation of MacNeill's Law (cf. Stüber 1998: 39–44) obscures the difference between old *-*Vnno*- and *-*Vno*- in both *foílenn* and *gigrann*. However, as mentioned above (§ 2.3) it is easier to explain the palatalization in *foílenn* in Irish by positing PCelt. **u̯ai̯lanno*-, a form compatible with the Brittonic reflexes. This would regularly give **u̯ai̯lœnno*- > **u̯ai̯lenno*- in Goidelic (cf. § 4) and produce the observed palatalization in OIr. *foílenn* (cf. Schrijver 1993: 34, 1995: 455–456; McCone 1996: 56–57). Welsh is not of much help due to the early merger of long and short resonants in word-final position after unstressed vowels. And although it is possible that the distinction was still present in the Cornish of the *Vocabularium Cornicum*, both word-final /-nn/ and /-n/ are consistently spelled <-n>, even in monosyllabic words. Hence OCorn. *goluan* and *guilan* may equally well reflect /-ann/ and /-an/.

3.1 The Breton evidence for Proto-Celtic *-*nn*-

This leaves us with the Breton material. Old Breton shows <-ann> in *couann*, *guilannou* and *goirann*, most straightforwardly explained as representing /-ann/. In

the Classical Middle Breton of the Catholicon (Cms., Ca.), the mystery and miracle plays and the older religious poetry the distinction between post-tonic long *-nn* and short *-n* is usually preserved, both in spelling (though not quite reliably and depending on the specific text) and in rhyme (much more reliably, but apart from the plural *guelvenn* /gelvenn/ 'sparrows' with final and internal *i*-affection in Dag. 195, none of our bird names are attested in poetry). In the first printed edition of the Catholicon (Ca., 1499), which tends to be very reliable when it comes to spelling, we find *couhenn*, *goluann*, *goelann*, all pointing to *-nn.* In Late Middle Breton and Modern LTK Breton there is a shortening of post-tonic long *-nn* (just as in Welsh and Cornish) leading to a complete merger between the phonemes **-nn* and **-n* in this position. However, Vannetais Breton, the most divergent dialect which split off early from LTK Breton, preserves the distinction between **-Vn* and **-Vnn* in polysyllabic words due to its final stress (although this distinction is not always consistently reflected in the orthography of the texts). We shall therefore have a brief look at the spelling of our bird names and other words with final **-ann* and **-an* in three Early Haut-Vannetais sources, namely the *Dictionnaire François-breton ou françois-celtique du dialecte de Vannes* (L'Arm., 1744), the vocabulary part of the first edition of the *Vocabulaire nouveau, ou dialogues français & bretons* (pp. 1–56; here abbreviated VN, cf. Le Goaziou 1950: 24–25) and the entire text of the *Magasin Spiriuel* (MS, 1790). The latter two texts are ascribed to Abbé Marion and are dated to appr. 1790 (cf. Le Goff in Belz 1986: 223–224).

Reflexes of 'owl', 'sparrow' and 'seagull':
PBret.[20] **kou̯(h)ann* > *cohann* (L'Arm. 184, VN 12), *cohàn* (MS 344, 345)
PBret. **golβann* > *golvàn* (VN 13), *golvan*[21] (L'Arm. 243)
PBret. **gu̯ulann* > *gouilann* (L'Arm. 174)

Reflexes of words with etymological **-ann*:
PBret. **rou̯hann*(?) > *rohann* 'a span' (a measure) (L'Arm. 129, VN 20, 21)
PBret. **trɪdeð-rann*[22] > *terderàn* 'a third' (MS 22), *derderann* (L'Arm. 381)
PBret. **pedu̯areð-rann* > *pœrann* 'a quarter' (VN 21, 54, L'Arm. 240, 316)
PBret. **öiθɥed-rann* > *évédrann* 'an eighth' (L'Arm. 316, VN 21)

20 PBret. = Proto-Breton, the stage immediately preceding any dialectal division within Breton.
21 The apparent reflex of **-an* in this attestation might be explained by attraction to the diminutive suffix **-an.*
22 A compound of the ordinal **trɪdeð* 'third' and **rann* 'part'. Likewise, **pedu̯areð-rann* and **öiθɥed-rann* were originally transparent compounds with the meaning 'fourth-part' and 'eighth-part'.

Reflexes of words with etymological *-*an*:
PBret. *lu̯an > *louan* 'rope' (L'Arm. 19), *luan* (L'Arm. 81)
PBret. *lɪdan > *lédan* 'wide' (L'Arm. 214)
PBret. *u̯nan > *unan* 'one' (L'Arm. 404, VN 22, 23, MS 10, 22)
PBret. *bu̯an > *buan* 'quick' (L'Arm. 403), *béan* (VN 52, MS 113), *bean* (MS 106)
PBret. **bihan* > *bihan* 'little' (L'Arm. 281, VN 33, MS 8, 17)

As can be seen, these texts spell the outcome of PBrit. *-*ann* as <-ann> or <-àn> and the outcome of PBrit. *-*an* as <-an>.[23] Hence Early Haut-Vannetais confirms what we already saw in Old and Classical Middle Breton, namely that the bird names reflect older *-*ann*.

3.2 A Middle Breton sound change: Unstressed *-ann* > *-enn*

As noted above, Late Middle and Modern Breton show an unexplained *-en(n)* in *kaouenn*, *golven*, *goulen* (for *kaouenn* already in Classical Middle Breton) instead of the expected *-an(n)*. At first sight, one could assume that the change was due to attraction to the singulative suffix *-enn*. Alternatively, influence from plural forms with final *i*-affection might be suggested. However, when we look at how other words ending in unstressed *-*ann* behave in Late Middle Breton and Modern Breton, it rather looks like the workings of a regular sound change. Below an extensive but not exhaustive list is given of the attestations of PBrit. **kɵu̯-rann* 'part' (W *cyfran*, OIr. *comrann*) in Middle Breton (MoBret. *kevrenn*). As can be seen, there is a shift from *queffrann* to *queffren* at some point during the Middle Breton period (roughly coinciding with the transition from Classical to Late Middle Breton).[24]

/-ann/	>	/-en(n)/
queffrann B 29/173, 34/202, 264/1608 (rhyme /ann/), J 52/1048 (rhyme /ann/), Ca. 110a, 168a, 172b; *quefurann* J 228/4654 (rhyme /ann/); *quefran* M 2114 (rhyme		*queffren* Catechism 15^{v}, 16^{v}, H 49, Cnf. 15^{r}, 20^{r}, 22^{r}, etc., Bel. 6^{v}, 9^{v}, 10^{r}, etc., Nom. 295, [343], [345], Be. 56, 147, CAntiquou 96, 100, 108, etc.; *qveffren* CAnti-

23 As expected, the same distribution is found in monosyllabic words, i.e. <-ann> from *-*ann* in e.g. *er hann* 'the full moon' (VN 3, L'Arm. 223) < **kann*, *goann* 'weak' (VN 33, MS 86, 235, L'Arm. 160) < **gu̯ann*, *splann* 'clear, bright' (MS 49, 105, 162 L'Arm. 60) < **splann* as opposed to <-an> from *-*an* in e.g. *tan* 'the fire' (VN 3, 38, L'Arm. 155) < **tan*, *bran* (L'Arm. 76), *ur Vran* 'a raven' (VN 12) < **bran*, *glouan* 'wool' (VN 19, L'Arm. 212) < **gu̯lan*.
24 Deshayes (2003: 390, 613) instead takes *queffrann* and *queffren* to be two etymologically distinct words. This fails to explain the clear chronological distribution.

/ann/); *keffran* Jer. 368; *queffran* Be. (sic!) 148; Deriv. *queffrannec* J 14/212 (rhyme /ann/), Ca. 168b.

quou 18, 24; *queffrenn* Bel. 121^{r}; *quefren* Cathechism 4; H 56; Nom. 22 (2x), 141, 210 (2x), etc.; *queffr'en* Be. 148; *quen-/fren* Be. 160; Plur. *queffrennou* Cnf. 69^{r}; Bel. 8^{r}, 9^{r}, 9^{v}, etc.; *quefrennou* Catechism 16^{v}; Nom. 76, 295, [337] (2x), etc.

Although we do not have nearly as many attestations for other words with etymological, post-tonic **-ann*, we find the same pattern.[25]

Reflexes of etymological **-ann*:
PBret. **trɪdeð-rann* > MBret. *trederann* (B 264/1607, J 13/209; rhyme /ann/) > *trederenn*[26] (Dag. 95; rhyme in /ann/!) > MoBret. *trederenn* 'a third' (Vann. *terderann*)
PBret. **pedu̯areð-rann* > OBret. *petg[u]arerann* (DVB 284) > MBret. **peuarann* > MoBret. *pevarenn*[27] 'a quarter' (Vann. *pœrann*)
PBret. **gu̯ur-ɣlann* (cpd. with **glann* 'shore', lit. '(that which is) on the shore'[28]) > MBret. *gourlann* (Cms.) > *gourlen* (Ca.), *gourleũ* (Cb., for *gourleñ*, i.e. *gourlenn*) 'seaweed left by the tide'[29]
PBret. **gu̯ur-ɣrann* (cpd. with **grann* 'bristles', lit. 'over-bristles') > MBret. **gourrann* > MoBret. *gourren(n)* 'eyebrow'
PBret. **rou̯hann*(?) > MBret. *rouhenn* (Ca.), MoBret. *rahouenn* 'span' (Vann. *rohann*, MW *rhychwant*)

25 MBret. *qualan mae* (Pm. 272/6019, rhyming with *splann*, hence /kalann/) from PBret. **kalann* 'first day of the month' ← Lat. *kalendae* (cf. MW *kalan*, OMIr. *calland*) does not show the change to *-enn* in later LTK Breton. However, this word does not develop regularly in Breton. The Modern Breton form *kala* was abstracted from the phrase *qualan mae* which underwent nasal assimilation of /-annm-/ to /-amm-/ (thus Jackson, 1967: 795–796). Other attestations show reanalysis to *kal-an-* as if containing the definite article of the following noun, e.g. EMoBret. *Kal ar goan* 'All Saints' Day' (SCger. 119). These factors explain the absence of Modern Breton attestations in *-en(n)* and the word does not constitute a counterexample to the proposed sound change.
26 Hemon (1984: 269–270) assumes that the development of *trederann* to *trederenn* is due to attraction to the singulative suffix *-enn*.
27 MBret. *renn* 'a quarter' (a measure) must have been back-formed from **peuarenn* (cf. Ernault 1893–1895: 462, 570).
28 For the compositional type, compare MoBret. *empenn* 'brain', lit. '(that which is) in the head'.
29 Ernault (1893–1895: 286) takes this word to be etymologically identical to MoBret. *gourlanv* 'the stand of the (high) tide', consisting of *gour-* 'over' and *lanv* 'flood'. However, in light of the difference in form and meaning, it seems more likely that they have different origins.

PBret. **glas-tann* > MBret. sglt. *glastannenn* (Ca.) > sglt. *glastennenn* (Cb.) 'a holm oak' *(*OCorn. *glastannen)*
PBret. **ahanann*[30] > MBret. adv. *ahanann*[31] (J 52/1049), *ahanan*[32] (Jer. 101) /aˈhanann/ > *ahanenn* /aˈhanenn/[33] > MoBret. *ac'hann*[34] [(a)ˈxãn], [(ə)ˈhãn] 'from here' (ALBB map 6)

This sound change may also allow us to derive the Modern Breton personal name *Youen(n)* from the Old Breton personal name *Iohan*, *Iohann* /i̯o(h)ann/ (Credon),[35] ultimately from Lat. *Iohannes.* However, apart from the inherent difficulties associated with the sound changes in personal names, it is not always possible to separate *Youen(n)* from the reflexes of etymologically different, but superficially similar personal names (several of which are identified with French *Yves*, cf. ALBB map 182), in particular the modern reflexes of OBret. *Eudon* /eu̯ðøn/, MBret. *Euzen* /eu̯ðen/.

It is important to note that the proposed sound change does not occur in Modern Breton words with etymological **-an* as demonstrated by the following collection of examples.

30 Probably from a petrified 1pl. prep. **a-han-ann* 'from us' of the preposition *a* 'from' (*pace* Ernault 1893–1895: 20–21 and Deshayes 2003: 49). An exact cognate is found in MCorn. *ahanan*, which functions both as the 1pl. prep. 'from us' (e.g. OM 612, 1101, BM 609) and as an adverb 'from here' (BM 2300). The regular Middle Breton 1pl. form is attested as *à hanomp* (Nl. 90) with *-m(p)* imported from the verbal system as is usual. Parallel to *ahanann*, *ahanenn* 'from here' we find the Middle Breton adverb *ahane*, *ahano* 'from there', which appears to come from a PBret. 3pl. prep. **a-han-ꝡβ* 'from them' (cf. MW 3pl. *ohonu*), formed to a shorter (and probably older) stem than the usual 3pl. prep. *aneze/anezo* 'from them' < PBret. **a-han-Vð-ꝡβ* (with unstressed 3pl. **-ꝡβ* > MoBret. *-e*, *-o* as seen in OBret. *tut lub* /tꝺd-lꝺβ/ > MoBret. *tule*, *tulo* 'navelwort', cf. Schrijver 1995: 146). Other adverbs formed from petrified forms of conjugated prepositions are found in Welsh and Cornish, e.g. the original 2sg. forms MW *uchot*, OCorn. *huchot* adv. 'above' (MW *uch*, *uwch* prep. 'above, on top of') and MW *issot*, OCorn. *isot* adv. 'below' (MW *is* prep. 'under, beneath').

31 The final /ann/ is secured by rhyme with *splann* /splann/, *queffrann* /kevrann/ and *rann* /rann/.

32 The final /ann/ is secured by (a misplaced) rhyme with *splan* /splann/ and *enebran* /enebrann/.

33 The final /enn/ is secured by rhyme in e.g. B 243/1476, 258/1572, J 42/835, 116/2395, etc.

34 With haplological reduction in LTK Breton resulting in final instead of the usual penultimate stress, cf. MBret. *amanenn* /aˈmanenn/ (Ca.) > MoBret. *a'mann* 'butter' and MBret. *balaznenn* /baˈlaðnenn/ (Ca.) 'a broom plant' > **baˈlaenenn* > MoBret. *baˈlaenn* 'a broom'. The alternative Modern Breton form *ac'halen* 'from here' probably shows dissimilation of the first nasal instead (whence *ac'halemañ* 'from here', *ac'halese* 'from there' formed with the deictic suffixes *-mañ*, *-se*).

35 Modern Breton *Yoan(n)* would then be the Vannetais variant of this name.

Reflexes of etymological *-*an*:
PBret. **garan* 'crane' > MMoBret. *garan* 'instrument used for making grooves in barrels; the groove itself' (cf. Gaul. *TRIGARANVS* '(with) three cranes')
PBret. **luɯan* > MBret. *louffan*, MoBret. *louan* 'rope' (cf. OIr. *loman*)
PBret. **lɪdan* > MMoBret. *ledan* 'wide' (cf. OIr. *lethan*)
PBret. **bʉan* > MMoBret. *buan* 'quick'
PBret. **ʉnan* > MMoBret. *unan* 'one'
Diminutives in **-an*: MBret. *amprefan*, MoBret. *amprevan* 'bug, insect', *korrigan* 'goblin', etc. (cf. OIr. *-án*), MMoBret. *bihan* 'little' (cf. OIr. *becán*)
Verbal nouns in **-ɯan*:[36] MBret. *goeluan* 'grieving, weeping' (Ca., M, J; rhyme in /-an/), MoBret. *gouelvan*; MBret. *queinuan* 'moaning' (Ca.), *queinoan* (M; rhyme in /-an/), *queynuoan* (Nl.; rhyme /-an/), EMoBret. *queinvan* (SCger. 92b)

Based on this material we can posit a regular sound change for the LTK Breton dialects that formed the basis of Middle Breton and the later literary language, stating that word-final unstressed *-ann* gives *-enn*. This sound change probably took place in the first half of the 15th century (given that already Cms. has *couhen* < **kou̯(h)ann*). It provides extra support for the reconstruction with **-ann* in our bird names, since the three names discussed here with reflexes in Middle and Modern Breton show the development to *-en(n)*.[37] Welsh, Cornish and Vannetais (with its historical final stress) remain untouched by this sound change. It also appears to be the case that some peripheral varieties of LTK Breton were unaffected.[38]

36 This verbal noun suffix, corresponding to W *-fan*, must be reconstructed with **-an* rather than **-ann* (*pace* Schumacher 2000: 123–125). The spelling and rhyming patterns in Classical Middle Breton consistently point to /-an/ rather than /-ann/.

37 Establishing this sound change also allows us to exclude bird names in *-an* attested solely in Late Middle Breton and Modern LTK Breton, even though post-tonic *-n* and *-nn* had merged by this time. Thus LMBret. *pochan* (Nom.), MoBret. *poc'han* [poːhɑ̃n], dim. [pohɑ̃ːnik], [bohɑ̃ːnik] 'Atlantic puffin; *Fratercula arctica*' (Berr 1986, 2: 475–476; whence probably E *puffin*) must be a diminutive derivative in *-an* from *boc'h* 'cheek', i.e. 'little cheek-one' rather than a bird name in **-anno-* (cf. the later addition of *-ig*, yet another diminutive suffix).

38 E.g. Molènes [ɛrˈhoˑlvɑ̃n] (Madeg 2021: 62), [gɔlvɑ̃n] 'the ruddy turnstone; *Arenaria interpres*', pl. [gɔlvɑ̃ːnɛt] (Berr 1986, 2: 460), Southeastern Cornouaille [golvɑ̃nik] dim. (Bouzec, Goapper and Souffez 2017: 295). Alternatively, these could show early attraction to the diminutive suffix *-an*. Influence from plurals in *-ed* and diminutives in *-ig*, where the *-ann-* would be stressed, is also possible. On the other hand, LTK attestations with /-an/ such as Go. <golañn> (Chapalain 2010: 90) and those given in Berr (1986, 2: 466–471) are difficult to evaluate, since they could represent more recent re-borrowings of Fr. *goéland* (itself borrowed from Breton).

4 Proto-Celtic *-*anno*- or *-*asno*-?

In § 3 we saw that there is ample evidence from Breton for a reconstruction with *-*nn*-. Before we can focus on the etymology of the suffix, we will need to address the question of the origin of *-*nn*-, i.e. whether we should reconstruct PCelt. *-*anno*- or *-*asno*-.[39] The Brittonic and Gaulish reflexes are compatible with either reconstruction. This leaves us with the Irish evidence. So far in this study it has been assumed that PCelt. *-*ann*- would be needed to account for the palatalization seen in OIr. *foílenn* and *bairnech*, by assuming a development via *-*ænn*- to *-*enn*- in Goidelic. However, this is not necessarily the only valid reconstruction. While there is a quite securely established distinction between PCelt. *-*asn*- > -*ann*- (in e.g. *-*u̯l̥sno*- > *-*u̯lasno*- > *-*u̯lanno*- > OIr. *flann* 'blood-red') and PCelt. *-*ann*- > -*ænn*- (in e.g. *-*gʰn̥-n-d*- > *-*gann*- > *-*gænn*- > OIr. *ro-geinn* 'is contained') in initial syllables (cf. Schrijver 1993: 34, 1995: 455–456 and McCone 1996: 56–57),[40] we appear to lack solid examples of the contrast outside the initial syllable. The available etymologies simply do not allow us to ascertain if PCelt. *-*asn*- in this position would yield Goidelic *-*enn*- or *-*ann*-. This means that it is not entirely clear if the palatalization observable in OIr. *foílenn* and *bairnech* points to PCelt. *-*ann*- or *-*asn*-.[41]

5 The origin of the suffix

The fact that we appear to find a second part *-*anno*- (or *-*asno*-) in these bird names – and not in other Celtic words – may be taken as an indication that it was a meaningful element in the language where it originated, whether it is inherited or substratal. But this still leaves us with the question of the origin of the suffix. As we saw above in § 4, it is not entirely clear whether the suffix should be reconstructed as PCelt. *-*anno*- or if PCelt. *-*asno*- will suffice. If the latter is the case, it is conceivable that a convincing Proto-Indo-European etymology may be found. One could consider deriving *-*asno*- from the Proto-Indo-European root *$pet(h_2)$- 'to fly', as found in e.g. PCelt. *-*(ɸ)etno*- > OIr. *én*, W *edn*, MBret. *ezn* 'bird', PCelt. *-*(ɸ)atanV*- > MBret. *hadan* 'nightingale', PCelt. *-*(ɸ)atanī* > MW *adein* 'wing'. This

39 Alternatively PCelt. *-*ænno*- opposed to *-*anno*-, if the assimilation of *-*sN*- > *-*NN*- is assumed to have taken place already in Pre-Proto-Celtic.

40 The difference between the two is mainly visible in Goidelic, where *-*ann*- gives OIr. -*enn*- (and -*inn*- by raising) and *-*asn*- gives OIr. -*ann*-. It is also possible that PCelt. *-*ann*- occasionally yields what is written -*enn*- in Gaulish (whether actually /enn/ or /ænn/), rather than the usual -*ann*-.

41 PCelt. *-*ansno*- would probably also produce the required outcomes.

would require a formation like *-*(ɸ)atsno*-, which might find a partial parallel in OLat. *pesna* 'feather' < **petsnā*. If only a reconstruction **-anno*- is valid, it seems much more likely that the element is ultimately of non-Indo-European origin, since a PCelt. **-anno*- (≠ **-asno*-) is quite difficult to generate from Proto-Indo-European to Proto-Celtic with the commonly accepted sound laws. The most straightforward projections back to Proto-Indo-European would be *-*n̥d(ʰ)no*-, *-*n̥no*- or *-*n̥nd(ʰ)o*-[42] but these do not appear to match anything meaningful and otherwise reconstructible (and the latter two could probably only be generated in nasal presents[43]). A PIE *-*n̥t-no*- might work better morphologically – we do have diminutive animal names in **-nt*- – but the phonology would be without parallel and a **-no*-stem directly suffixed to an **-nt*-stem does not inspire much confidence. Furthermore, apart from nasal infix presents, words containing **-ann*- do not appear to have solid Proto-Indo-European etymologies (though they in several cases have apparent "cognates" in Germanic and nowhere else):

PCelt. **bannā* 'peak; horn' > OIr. *benn* 'mountain, peak; horn', MW *bann* 'peak', Occ. *bano* 'horn' (borrowed from Gaul. **bannā*). Traditionally compared to PGm. **pinta*- 'penis; tip' (IEW 96–97; LEIA B-36) under the assumption of a root **bend*-, but this is no more than a possibility. A Proto-Indo-European reconstruction seems unlikely, not least because of the initial **b*- indicated by Germanic.
PCelt. **glanno*- > Gaul. *glanno*- in the place name *Glanno-venta*, OIr. *glenn*, Gsg. *glinne* 'valley' (neuter *s*-stem secondarily?), MW *glann* 'shore, riverbank', MBret. *glann* 'id.' (MW *glynn* 'valley', MBret. *glenn* 'land; earth' borrowed from Goidelic?). Most likely connected to LG *klint*, ON *klettr* < PGm. **klinta*- (cf. Delamarre 2003: 179–180).
PCelt. **granno*- 'bristles, beard, moustache' > MIr. *grend*, *grenn* 'bristles, beard', MW *grann* 'id.', OFr. *grenon* 'moustache' (dimin., borrowed from Gaul. **grenno*-). PGm. **granō*- 'moustache' is probably related, but exactly how is not clear (cf. Delamarre 2003: 183 for references).

It is quite likely that these nouns containing **-ann*- do not go all the way back to forms with PIE *-*n̥d(ʰ)no*-, *-*n̥no*- or *-*n̥nd(ʰ)o*-, but were rather borrowed from an unknown language at a later point. The same may apply to PCelt. **gobann*- 'smith' with apparent PCelt. **-ann*- (or **-asn*-?) outside the first syllable (Gaul. *gobann*-,

42 Theoretically *-*h₂end(ʰ)no,* *-*h₂enno*- and *-*h₂ennd(ʰ)o*- would also be possible.
43 As in **gann*- 'to be contained, to fit in' < **gan-n-d*- (KPV 330–331); PCelt. **glann*- 'to choose, to collect' < **glan-n-d*- (KPV 334–337); **skann*- 'springs, darts, flies off' < **skan-n-d*- (KPV 574–575); **sṵann*- 'to play (an instrument)' < **sṵan-n*- (KPV 607–608).

OIr. *gobae*, *-ann*, MW *gof*, pl. *goueyn*, MBret. *goff*, pl. *gueuing* /geviŋ/) which has yet to receive a convincing Indo-European etymology.

In light of the difficulties in finding a suitable etymology for the element *-*anno*- found in bird names, it seems probable that it comes from a Pre-Celtic substrate language, as a derivational suffix or a noun class suffix carried over into Celtic with borrowed words. One could try to make a connection with the apparent bird name suffixes in *-*Vnd-/-Vmb-/-Vng*- discussed by Anthony Jakob in the present volume, specifically the variant *-*Vnd*-. This could point to a Pre-Proto-Celtic reconstruction *-*anndo*- or *-*n̥ndo*- (cf. PCelt. **bannā* ~ PGm. **pinta*- and PCelt. **glanno*- ~ PGm. **klinta*- discussed above). However, chance resemblance can hardly be ruled out without more supporting evidence. What we can say is that the suffix does not look like it derives from Proto-Indo-European by the known sound laws.

6 Gallo-Latin *-*annu*-?

In the preceding part I have argued for the existence of a Celtic bird name "suffix" *-*anno*-. This is not an entirely new idea, however. Georg Cohn, in his study on suffix substitutions in Vulgar Latin and pre-literary French, noted an unexpected development in certain bird names (Cohn 1891: 139–144), exemplified by OFr. *faisan(t)*, MoFr. *faisan* 'pheasant'. This word derives from Lat. *phasiānus* and is therefore expected to give OFr. **faisien*. Yet the attested form *faisan(t)* appears to reflect an unattested Gallo-Latin **phasiannus* with *-nn-*. A reconstruction **phasiannus* would explain both the vocalism and the fact that we find a frequent alternative Old French form in *-nt* (also the source of English *pheasant* and Middle Breton *faessant*).[44] The potential relevance for Celtic lies in Cohn's proposed solution. He suggested that our hypothetical Gallo-Latin form **phasiannus* came about by influence from Gaulish, more specifically the "suffix" seen in Gallo-Latin *cauannus* 'owl' (cf. § 2.1 above). He furthermore suggested that several other French bird names unexpectedly pointing to Latin *-annus* are to be explained as due to Gaulish influence.[45]

44 The *-nt* developed regularly from *-*nn*-, when this collided with the inflectional *-s* in the nominative singular and the oblique plural, i.e. Lat. *-*Vnnus* > OFr. *-nz* /-nts/ (cf. Lat. *annus* > OFr. *anz* 'year'). A similar pattern with secondary *-t* is seen in OMFr. *tirant* (next to *tiran*) from Lat. *tyrannus*. Unsurprisingly, E *tyrant* and MBret. *tirant* also show *-nt*.

45 Another possible instance of Celtic *-*anno*-/*-*asno*- in Gallo-Romance is MMoFr. *cravan* 'barnacle; barnacle goose' which could derive from a hypothetical Gaul. **krăCanno*-. This *cravan* is already taken by Gamillscheg (1928: 273–274, followed by FEW 2: 1266) to be a borrowing from

While there does appear to be an unexpectedly high number of French bird names which display a "suffix" behaving as if deriving from a Lat. *-*annu*-, it should be admitted that many of the words adduced by Cohn, such as *faisan(t)* and *milan* 'kite'[46] may be explained as borrowings from Occitan (thus FEW 8: 375). Others cannot be accounted for in this way, however. OFr. *cormoran(t)* is derived from a Lat. *corvus marīnus*, lit. 'sea raven' (cf. Occ. *cormarin*, FEW 2: 1239–1240), but such a proto-form is expected to give OFr. **cormarin*. The most common form *cormoran(t)* (whence E *cormorant*) behaves as if from *corvus *morannus*. Here Occitan offers no help, but influence from Celtic seems like a distinct possibility, supported by the change in vocalism from *mar*- to *mor*- (cf. PCelt. **mori*-, OIr. *muir*, MoW *môr*, MoBret. *mor* 'sea').

Cohn's suggestion of a Gaulish origin of the "suffix" in these words has not been generally accepted, in part because he did not present much evidence for a Gaulish suffix **-anno*- apart from *cauannus* 'owl', in part because he was overinclusive in his selection of material (as pointed out by FEW 8: 375). However, as I hope to have shown above, one can at least make a case for the existence of a Celtic suffix **-anno*- specifically used in bird names. Whether or not this then has anything to do with Fr. *-an(t)* in e.g. *faisan(t)* and *cormoran(t)* is at present unclear.

7 Conclusion

As argued above, we can identify a formerly unknown element **-anno*- in a number of Celtic bird names, most clearly in **kaṷanno*- 'owl', **gulbanno*- 'sparrow', **ṷaiilanno*- 'seagull' and **giguranno*- 'barnacle goose'. This apparent "suffix" is mostly attached to roots of unknown or onomatopoeic origin. The reconstruction of

Gaulish and reconstructed as **kraganno*-. However, **krăbanno*-, **krăpanno*- or **krăṷanno*- would fit the attested forms better (and **krăbanno*- would even allow a tentative connection with MoW *crafu*, MoBret. *kravañ* 'to scrape'). In the absence of older forms and actual evidence from Celtic (MoBret. *kravan* 'barnacle' is most likely borrowed from French) it remains too uncertain to be included as an example of the bird name suffix **-anno*-. It should also be noted that including material not actually attested in Celtic languages would open the door to other Western Romance words, such as a hypothetical "Gaulish" **kăpannā* 'hut' (OOcc. *cabana*, It. *capanna*, Sp. *cabaña*; suspected to be of Gaulish origin in FEW 2: 244–246) and **karkanno*- 'neck ring' (OPic. *carcant*, *charcan*, *carchan*, OFr. *charchant*, MoFr. *carcan*; supected of being of Germanic origin in FEW 2: 361–362, but a suitable Germanic source remains to be identified). If these two words were to be included, we would have to review the status of the suffix as being restricted to bird names.

46 From **mīl(u)ānus* a derivative of Lat. *mīluus* 'kite'.

*-*anno*- is based on evidence mainly from Goidelic and Brittonic. The evidence from Breton is crucial due to the late preservation of the contrast between Proto-Brittonic post-tonic *-*n* and *-*nn* (from PCelt. *-*VnV*- versus *-*VnnV*-) in this branch. Additionally, a new sound change of post-tonic *-ann* > *-enn* in the course of Middle Breton was identified, further strengthening the reconstruction of the suffix as PCelt. *-*anno*-. As to the age of the suffix, the indirect Gaulish attestation of *cauannus* and the fact that relatively uniform proto-forms can be reconstructed allow us to date it at least as far back as the common predecessor of Goidelic, Brittonic and Gaulish (barring undetected borrowing between these branches) and quite possibly significantly older.

Since we know that bird names belong to a semantic field where substrate vocabulary (or just plain loanwords) would not be surprising, it remains a distinct possibility that these words in *-*anno*- come from an unknown substrate language, at least until a plausible Proto-Indo-European source has been identified. The fact that we are dealing with an apparently meaningful element *-*anno*-, found in four or five different words, rather than a single word of an unusual shape and without a known etymology, may be taken as an indication that this is indeed *something*. It is another matter altogether exactly what *kind* of something it is. Whether it is truly substratal (or simply borrowed[47]) or it is merely the case that the Proto-Indo-European source of the element has not yet been identified remains to be seen.

References

Abbreviations of primary sources follow DEVRI for Breton and George (2009) for Cornish.

ALBB = Pierre Le Roux. 1924–1963. *Atlas linguistique de la Basse-Bretagne*. Rennes/Paris: Plihon et Hommay.

ALF = Jules Gilliéron & Edmond Edmont. 1902–1910. *Atlas linguistique de la France*. Paris: Honoré Champion.

Bauer, Bernhard. 2008. *Studien zu den Altbretonischen Glossen*. Vienna: University of Vienna MA thesis.

Beekes, Robert S.P. 2014. *Pre-Greek. Phonology, Morphology, Lexicon.* Leiden/Boston: Brill.

Belz, Jorj (ed.). 1986. *Œuvres de Pierre Le Goff (1860–1941)*. Rennes: Institut Cultural de Bretagne – Skol-Uhel ar Vro.

47 Cf. Matasović (2020: 331–332). In light of the time depth, it may be impossible to determine the mechanism by which a given word entered Celtic, whether as part of a language shift from a Pre-Indo-European language to Celtic (substratal) or as a loanword from a language that has since become extinct.

Bernardo Stempel, Patrizia de. 1999. *Nominale Wortbildung des älteren Irischen*. Tübingen: Max Niemeyer.
Berr, Alan-Gwenog. 1986. *Ichthyonymie bretonne*. Brest: Brud Nevez.
Bouzec, Mona, Jos Goapper & Yannick Souffez. 2017. *Le breton des rives de l'Aven et du Bélon*. Lannuon: An Alarc'h.
Bromwich, Rachel & D. Simon Evans. 1992. *Culhwch and Olwen*. Cardiff: University of Wales Press.
Buckeridge, John. 2011. Of trees, geese and cirripedes: Man's quest for understanding. *Integrative Zoology* 2011(6). 3–12
Campanile, Enrico. 1974. *Profilo etimologico del Cornico antico*. Pisa: Pacini.
Chapalain, Bernard. 2010. *Geriadur ar brezhoneg komzet (Goueloù, Treger ha Kernev-Uhel)*. Lannuon: An Alarc'h.
Cohn, Georg. 1891. *Die Suffixwandlungen im Vulgärlatein und im vorlitterarischen Französisch*. Halle an der Saale: Max Niemeyer.
Delamarre, Xavier. 2003. *Dictionnaire de la langue* gauloise. Arles: Errance.
Derksen, Rick. 2008. *Etymological Dictionary of the Slavic Inherited Lexicon*. Leiden/Boston: Brill.
Derksen, Rick. 2015. *Etymological Dictionary of the Baltic Inherited Lexicon*. Leiden/Boston: Brill.
Deshayes, Albert. 2003. *Dictionnaire étymologique du breton*. Douarnenez: Le Chasse-Marée.
DEVRI = Martial Menard & Herve Le Bihan. *Devri: Le dictionnaire diachronique du breton*. http://devri.bzh (accessed 29 June 2022).
DIL = Ernest Gordon Quin et al. 1983. *Dictionary of the Irish Language – Compact Edition*. Dublin: Royal Irish Academy.
DVB = Fleuriot, Leon. 1964. *Dictionnaire des gloses en vieux-breton*. Paris: Klincksieck.
Ernault, Émile. 1893–1895. *Glossaire moyen-breton*. Paris: E. Bouillon.
FEW = Walther von Wartburg & Hans-Erich Keller. 1922–1967. *Französisches etymologisches Wörterbuch: Eine Darstellung des galloromanischen Sprachschatzes*. Bonn: Klopp, Heidelberg: Carl Winter, Leipzig/Berlin: Teubner, Basel: R. G. Zbinden.
Gamillscheg, Ernst. 1928. *Etymologisches Wörterbuch der französichen* Sprache. Heidelberg: Carl Winter.
George, Ken. 2009. *An Gerlyver Meur*. n.l.: The Cornish Language Board.
GPC = Richard J. Thomas, Gareth A. Bevan & Patrick J. Donovan. 1967–2002: *Geiriadur Prifysgol Cymru*. Caerdydd: Gwasg Prifysgol Cymru.
Graves, Eugene Van Tassel. 1962. *The Old Cornish Vocabulary*. Columbia University dissertation. Michigan: University Microfilms.
Hemon, Roparz. 1984. *A Historical Morphology and Syntax of Breton*. Dublin: The Dublin Institute for Advanced Studies.
IEW = Julius Pokorny. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern/München: Francke.
Jackson, Kenneth Hurlstone. 1967. *A Historical Phonology of Breton*. Dublin: The Dublin Institute of Advanced Studies.
KPV = Stefan Schumacher. 2004. *Die keltischen Primärverben*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
Lambert, Pierre-Yves. 1982. Les gloses du manuscrit BN Lat. 10290. *Études celtiques* 19. 173–213.
Lambert, Pierre-Yves. 2003. *La Langue gauloise*. Arles: Errance.
Largillière, René. 1928–1929. Le dialogue entre Arthur et Guinclaff. *Annales de Bretagne* 38. 627–674.
Le Bihan, Herve. 2013. *An Dialog etre Arzur Roe d'an Bretounet ha Guynglaff*. Rennes: Tir.
Le Goaziou, Ad. 1950. *La longue vie de deux "Colloques François et Breton" (1626–1915)*. Quimper: Librairie Le Goaziou.

LEIA = Joseph Vendryes, et al. 1959–. *Lexique étymologique de l'irlandais ancien*. Dublin: Dublin Institute for Advanced Studies, Paris: Centre National de Recherche Scientifique.
Lhuyd, Edward. 1707. *Archæologia Britannica*. Oxford: Printed at the Theater for the author.
Lloyd, Albert L. & Rosemarie Lühr. 2009. *Etymologisches Wörterbuch des Althochdeutschen*, vol. IV (*gâba – hylare*). Göttingen: Vandenhoeck & Ruprecht.
Madeg, Mikael. 2021. *Glossaire molénais / Brezhoneg molenez*. Lannuon: An Alarc'h.
Matasović, Ranko. 2009. *Etymological Dictionary of Proto-Celtic*. Leiden/Boston: Brill.
Matasović, Ranko. 2011. *Addenda et corrigenda to Ranko Matasović's Etymological Dictionary of Proto-Celtic (Brill*, *Leiden 2009)*. https://mudrac.ffzg.hr/~rmatasov/EDPC-Addenda%20et%20corrigenda.pdf (accessed 22 May 2022).
Matasović, Ranko. 2012. The substratum in Insular Celtic. *Journal of Language Relationship / Вопросы языкового родства* 8. 153–168.
Matasović, Ranko. 2020. Language of the bird names and the pre-Indo-European substratum. In Romain Garnier (ed.), *Loanwords and Substrata. Proceeding of the Colloquium held in Limoges (5th–7th June*, *2018)*, 331–344. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
McCone, Kim. 1996. *Towards a Relative Chronology of Ancient and Medieval Celtic Sound Change*. Maynooth: Department of Old and Middle Irish, Saint Patrick's College.
McManus, Damian. 1991. *A Guide to Ogam*. Maynooth: An Sagart.
Pedersen, Holger. 1909–1913. *Vergleichende Grammatik der keltischen Sprachen*. Göttingen: Vandenhoeck & Ruprecht.
Schrijver, Peter. 1993. On the Development of Vowels before Tautosyllabic Nasals in Primitive Irish. *Ériu* 44. 33–52.
Schrijver, Peter. 1995. *Studies in British Celtic Historical Phonology*. Amsterdam/Atlanta: Rodopi.
Schrijver, Peter. 1997. Animal, vegetable and mineral: some Western European substrate words. In Alexander Lubotsky (ed.), *Sound Law and Analogy. Papers in honor of Robert S.P. Beekes on the occasion of his 60th birthday*, 293–316. Amsterdam/Atlanta: Rodopi.
Schumacher, Stefan. 2000. *The Historical Morphology of the Welsh Verbal Noun*. Maynooth: The Department of Old Irish, National Universtity of Ireland.
Stifter, David. 2019. Further to Avidic. *Historische Sprachforschung / Historical Linguistics* 132. 312–316.
Stokes, Whitley & Adalbert Bezzenberger. 1894. *Wortschatz der keltischen Spracheinheit*. Göttingen: Vandenhoeck & Ruprecht.
Stokes, Whitley & John Strachan. 1901–1903. *Thesaurus Palaeohibernicus*, vol. 1–2. Cambridge: At the University Press.
Stüber, Karin. 1998. *The Historical Morphology of* n-*Stems in Celtic*. Maynooth: The Department of Old Irish, National University of Ireland.
VN = Anonymous. n.d. *Vocabulaire nouveau, ou dialogues français & bretons*. Vannes: J. M. Galles.
White, Beatrice. 1945. Whale-Hunting, the Barnacle Goose, and the Date of the "Ancrene Riwle." Three Notes on Old and Middle English. *The Modern Language Review* 40(3). 205–207.
Zair, Nicholas. 2012. Schrijver's rules for British and Proto-Celtic *-*ou̯*- and *-*uu̯*- before a vowel. In Philomen Probert & Andreas Willi (eds.), *Laws and Rules in Indo-European*, 147–160. Oxford: Oxford University Press.

David Stifter

6 Prehistoric layers of loanwords in Old Irish

1 Introduction

This chapter is concerned with the complex question of layers of loanwords in the attested Celtic languages from unknown, prehistoric languages. The focus is specifically on the Goidelic branch of Celtic, represented by Old Irish and its younger descendants, the modern Gaelic languages Irish, Manx and Scottish Gaelic. A perspective from the British (or Brythonic) languages Welsh, Cornish and Breton would entail slightly different points of emphasis, but *mutatis mutandis* most of what will be said in section 2 can easily be transferred to all other Celtic languages. This study approaches the matter deductively from a survey angle and through a discussion of the involved methodological issues. In doing so, it complements the inductive approach adopted in Van Sluis' study of substratal loanwords in this volume. In the preliminary methodological sections 1.1. and 1.2., the sources that can be quarried for data are introduced and the practical challenges that are faced in this process are briefly discussed. The central part of the study is section 2. We can reckon with four separate chronological layers at which traces of prehistoric substratal languages in the form of loanwords can be expected to have entered the Celtic, and specifically the Gaelic, languages. Candidates for such loans are scrutinised in section 2, as well as the criteria which can be used to detect them. Where possible, concrete proposals will be made as to which prehistoric languages correspond to the layers. Genetically identifiable populations will serve as a proxy for the languages when necessary. This survey is restricted to lexical evidence for language contact and borrowing. The question of structural influence, which is so elementary, for example, for the Hamito-Semitic substratum theory,[1] will not be addressed.

1 For an overview of the history of the Hamito-Semitic or Afro-Asiatic substratum hypothesis in Celtic and the arguments used, see Hewitt (2007) and Isaac (2007).

Acknowledgements: I thank Anthony Jakob, Guus Kroonen, Anders Jørgensen, and especially Paulus van Sluis for their helpful advice, for many valuable suggestions, and for the opportunity to see preliminary versions of their own contributions to this volume. All disclaimers apply.

https://doi.org/10.1515/9783111337920-006

1.1 Sources

The generic lexicon of Gaelic, as well as that of the other Celtic languages, typically, and naturally so, receives most attention in the search for traces of substratal borrowing. However, this is not the only source that can be scoured for evidence. The following types of sources must be considered for a full investigation into the matter:

1.1.1 Lexicographic collections

The bulk of the evidence comes from dictionaries. eDIL, the *electronic Dictionary of the Irish Language* (eDIL 2019), takes a special place among them since it is the dictionary for the early stages of the Irish language, chiefly covering Old and Middle Irish up to the 12th, but also including Modern Irish material up to the 16th century. eDIL derives its material exclusively from texts contained in manuscripts from the 8th to approximately the 17th century. For the modern Goidelic languages, the primary reference points are Dinneen's dictionary of Modern Irish (Dinneen 1927) and Dwelly's dictionary of Scottish Gaelic (Dwelly 1911).[2] They draw on older dictionaries and dialect collections of the spoken languages of the 19th and early 20th centuries, but some of their material derives also from earlier manuscripts.

Data for the British languages come from *Geiriadur Prifysgol Cymru* (GPC) for Welsh, from Favereau's *Geriadur Bras* (2016–) and *Devri – Le dictionnaire diachronique du breton* (Menard 2021) for Breton, and from *An Gerlyver Meur* (George 2020) for Cornish. Their sources are manuscripts and the spoken languages and dialects. Unlike the Goidelic dictionaries, they cover the entire life of their languages, although the emphasis is on the modern stages.

For Gaulish, the various collections of Delamarre provide useful resources, namely the *Dictionnaire de la langue gauloise* (Delamarre 2018), *Noms de personnes celtiques dans l'épigraphie classique* (Delamarre 2007), *Noms de lieux celtiques de l'Europe ancienne* (Delamarre 2012), and *Dictionnaire des thèmes nominaux du gaulois* (vol. I : Delamarre 2019). For Celtiberian, Jordán Cólera (2019) is the most comprehensive and up-to-date collection.

2 Work on more comprehensive online lexicons of Modern Irish and Scottish Gaelic is currently underway both in Ireland and Scotland.

1.1.2 Placenames

Toponomastic material offers a type of evidence that by definition contains geographical information and that, if the data is sufficiently large, allows to plot distribution maps. Although the majority of placenames in Ireland and in the Gaelic parts of Scotland are transparently Irish and probably very recent formations, occasionally names are encountered that may continue from pre-Gaelic population groups. Starting points for Irish placename studies are the monumental collection *Onomasticon Goidelicum* by Hogan in Ó Corráin's almost 3000 page-long revision of 2017 ('eHogan'), and the *Historical Dictionary of Gaelic Placenames* (Ó Riain et al. 2003–), which has progressed to letter D so far. Broderick 2013 offers a history of scholarship on the pre-Celtic linguistic situation in Ireland and Britain, albeit strongly biased towards theories of Hamito-Semitic substrates.

1.1.3 Personal names

If non-Goidelic population groups persisted in Ireland up to the early medieval period, they might be expected to have left traces in the earliest strata of historical sources, especially the genealogies. Because of the huge amount of data there is no fully comprehensive collection of Irish anthroponomastics. A convenient starting point are the indexes to *Corpus Genealogiarum Hiberniae* (O'Brien 1962) and *Corpus Genealogiarum Sanctorum Hiberniae* (Ó Riain 1985), as well as indexes to historical sources such as those for the *Annals of Tigernach* by Ó Murchadha (1997). Medieval Irish personal names have been studied overwhelmingly in the light of Indo-European naming traditions. To my knowledge, no systematic search for unetymologisable name elements has been carried out yet. O'Brien's typological observations about types of names in medieval Ireland (published by Baumgarten 1973: 226–230) remain an important attempt at identifying structural substratal influence in anthroponomastics.

1.1.4 Ogam

The earliest written documents in Ireland, the corpus of inscriptions in the ogam script, constitute a unique type of source. The corpus consists overwhelmingly of anthroponomastic data, which has not been studied under the aspect of substratal vestiges yet. Since the earliest inscriptions date to the earlier 5th century or even before that, there is in theory the potential for them to record substratal onomastic material, especially if the hypothesis is correct that a pre-Celtic language

was still spoken in Ireland at the time. A case in point is the name RONANN, Old Irish *Rónán*, which is discussed in Appendix 1. However, because of the many practical obstacles that ogam inscriptions pose, such as physical damage to the objects, fluctuating spelling conventions, and a strong geographical bias to the south and south-west of Ireland, ogam names are often difficult to interpret. The data of the Irish and south-British ogam inscriptions is collected in Ziegler (1994). The database *Ogham in 3D* (White 2013) will achieve a full collection of the entire material in the near future.

1.1.5 Syntax and typology

Finally, a word has to be said about non-lexical evidence for substratums in the form of structural, i.e. syntactical features. This pertains foremost to the notorious question of typological influence that precursor languages in Britain and Ireland, usually suspected to be related to the Hamito-Semitic/Afro-Asiatic language family, are supposed to have exerted on Insular Celtic. According to this hypothesis, the pre-Celtic inhabitants of the Western Archipelago spoke languages that were related with or similar to the languages of North Africa and the Near East. The popularity of this hypothesis has come and gone in waves for over two centuries. The history of scholarship, together with the main facts and arguments, is conveniently summarised in Hewitt (2007) and Isaac (2007). Despite the ink spilled over this question, linguists have not reached a consensus as to how to interpret the structural similarities between Insular Celtic and Afro-Asiatic.

Recent progress in the palaeogenetics of Britain and Ireland has effectively put an end to traditional approaches to this particular substratum hypothesis. Prehistoric insular genetic links are firmly first with European hunter-gatherer societies and later with Neolithic agriculturists from Anatolia, but not with populations in North Africa. This does not invalidate syntactical-typological research into substrates as such, but it can no longer be done simplistically with the Afro-Asiatic model.

1.2 Methodological remarks

In another chapter in this volume, Paulus van Sluis attempts to identify possible substratal loans in Celtic on the basis of non-canonical correspondences with other Indo-European branches. By necessity, his approach can say nothing about items that are restricted to a single language or branch of Indo-European, and it has a ‘blind spot’ for words that were borrowed relatively recently in history

after the evolution of the individual European languages known today. In contrast to Van Sluis' approach, this study goes a step further and also takes isolated items into account that lack parallels in other languages. This more speculative approach is done in the full awareness of its potential pitfalls. In the absence of non-canonical correspondences in other languages, aspects to take into account in identifying suspect items are the lack of Indo-European root etymologies and of more distantly related formations; the presence of unusual sounds or sound-combinations; and reference to items in the real world that were probably unknown to speakers of Proto-Indo-European. None of these criteria are conclusive in themselves, but cumulatively they contribute to a plausible case for substratal candidates.

The expression 'unusual sounds' refers to segmental phonology, to violations of distributional restrictions of sounds, and to supra-segmental features, such as unusual phonotactics. An example of the former are the presence of initial *p* and medial *f* in Irish words. While *p-* is downright excluded in native Irish words, medial *-f-* is expected to be very rare (see the discussion in sections 2.4.3. and 2.4.4.). The emphasis on unusual sounds means, of course, that this methodology works not so well in the deep historical perspective: the further back we go in time, the more likely it is that sounds, which were foreign then, have been adapted to the native phonological system and have become unrecognisable, or unremarkable, by the time they appear in our extant sources.

Uncommon combinations of adjacent sounds and unusual distributions of non-contiguous sounds in a word are other manifestations of unusual phonotactics, especially in the case of sound combinations that should have been eradicated by regular sound change acting on the inherited lexicon, or clusters that should not or could not have arisen in the first place because of phonological and morphological constraints. The historical sound changes of Celtic and Goidelic tightly limit the range of sounds that are possible in specific parts of words. For example, the occurrence of *é* is subject to many restrictions in Old Irish. It either continues PC **eį*, in which case it can appear as *é* only before palatalised consonants; otherwise it is 'broken' to *ía*. Or it is the result of compensatory lengthening of **e* or **a*, but this is in turn restricted to specific contexts that involve the loss of a consonantal segment before a non-lenited stop or before a lenited liquid or nasal. In late Primitive Irish and the early stages of Old Irish, the two variants of long *é* must have been phonetically distinct, but ultimately they can be analysed as synchronic allophones of each other in Old Irish. The upshot of all this is that the sequence of an *é* followed by a non-palatalised fricative (i.e., the lenited outcome of an earlier obstruent) cannot be generated from inherited structures by regular sound change. Words of such a structure are therefore suspect of being loans. Two examples of such an 'impermissible' sequence are *géd* 'goose'

and *bréch* ‘wolf’. In Stifter (2019b: 313–314), I concluded that the Insular Celtic word for ‘goose’ underlying OIr. *géd*, W *gŵydd*, Bret. *gwaz*, has to be reconstructed as **giγδ-*. Since the cluster **γδ* cannot have arisen in inherited Celtic words,[3] the word must be borrowed from an unknown source. Equally in the case of *bréch*, there is no possible Proto-Celtic reconstruction that would generate the sequence *-éch*.

The putative preform of MoIr. *gúgán* ‘whelk’ is **gūggo-* (see section 2.4.2). There is no way conceivable how a voiced geminate could have arisen after a long **ū* in the root syllable of a Primitive Irish word. A long *ú* followed by a voiced stop thus stands out from among the permissable sound combinations of inherited Celtic words. On the other hand, there is no synchronic restriction in Modern Irish any longer that would disallow any long vowel to be followed by any voiced obstruent. This gave the speakers of the language the opportunity to create neologisms with a highly expressive or sound-symbolic potential – if this word is an *Urschöpfung*. More details about phonological criteria for identifying substratal borrowings in Celtic can be found in Paulus van Sluis’ discussion in this volume.

The intuitively most obvious non-linguistic criterion for substratal borrowings in a language are words for unfamiliar items, i.e., words for real-world items in a habitat that had been foreign to the speakers of a language before their migration into the new environment, or words for technological innovations. When the speakers of such a language find themselves in a situation where they have to refer to objects and concepts for which their language lacks the terminology, one way to resolve it is the transfer of existing words to a new referent or the expansion of the reference of existing words through metonymy. Another solution to fill the lexical gap is to borrow items from a local language. Both strategies are amply attested for Irish with regard to the Atlantic fauna (Stifter 2023: 181–186). Several examples for borrowing will be provided in the following chapters. An example for lexical transfer is *bled*, itself probably a loanword, which must originally have referred to the ‘wolf’ in Celtic, but which in Old Irish means ‘whale’. Another expression, also for the ‘whale’, is *míl mór* ‘big animal’, in which the

3 I cannot think of any certain examples of inherited words that would allow us to see how word-internal **gd* would have behaved in a regular way in Insular Celtic. From the treatment of other clusters of two different obstruents, assimilation to the second component might be expected word-internally. Blažek (2015: 125) wants to derive OIr. *séd* ‘deer’ from **segdo-*, but too many philological problems beriddle his argument to make this proposal convincing. Word-initially the cluster **gd* is simplified to **d* in the Insular Celtic languages, if we can trust the example PIE **dʰĝʰomi̯o-* ‘earthling’ > PC **gdoni̯o-* > Gaul. *χtonikon* (?), OIr. *duine*, W *dyn*, Corn., Bret. *den* ‘human’.

most generic zoological term for 'animal' is used, *míl*. A similar situation obtains in Welsh, which has *morfil* 'sea animal' or *morfarch* 'seahorse' for large maritime mammals. The lack of a common Celtic expression for the whale is clearly connected with the land-locked habitat of the Proto-Celtic speech community on the European Continent, where they had no exposure to this kind of animal, prior to their migration to the Western Archipelago. In the case of OIr. *mucc mara* 'pig of the sea = porpoise', an expression for a domesticated animal underwent semantic expansion to a sea mammal, not the whole-sale transfer seen in *bled*.

1.3 Practical challenges

Hunting for substratal items means walking a very thin line. For the methodology of dealing with words that are attested across several Indo-European branches, but that exhibit non-conventional correspondences, I refer to other contributions in this volume. As for words that are found only in a single language, there are many practical obstacles to identifying them as substrate borrowings. When an inherited Indo-European word has been lost in all other branches of the family, its isolated presence in one language can lead to its false identification as a loanword. Words that are exclusive to a single branch, but whose phonological and morphological structure can be explained within an Indo-European framework, can therefore not be positively identified as borrowings, unless additional factors make such an analysis unavoidable.

There is always the danger of falling victim to linguistic pareidolia. *Pareidolia* is the tendency of the human mind, usually of its visual faculty, to perceive meaningful patterns in random inanimate objects, where there are none. An equivalent linguistic pareidolia haunts language comparison, namely the tendency to recognise correspondences and phonological and morphological rules in random items.[4] It is, for example, a manifestation of Indo-Europeanist pareidolia to interpret vowel variations as reflexes of *ablaut*. Ir. *fáecha* vs. W *gwichen* 'periwinkle' and Ir. *blóesc* vs. W *blisg* 'shell, husk' are illustrations of this (see section 2.4.1. and Appendix 4). They display a contrast between **u̯oi̯k-*/**u̯īk-* and **bloi̯sk-*/**blīsk-* respectively that, on the surface, is reminiscent of *ablaut* variation between an *o*- and a zero-grade. Arguments against such an analysis are the lack of an Indo-European root etymology for **bloi̯sk-*/**blīsk-* and the fact that *ablaut* in the root is almost completely absent from nominal inflection and derivation in Celtic other-

4 Teeter's Law, the tendency to consider the language that one researches oneself as most archaic in a language family, is a variant of this tendency.

wise, except for very archaic formations. These words would therefore have to belong to a remarkably archaic layer of the lexicon, which, given their meaning, is flatly implausible.

Substratal pareidolia, on the other hand, is the readiness to ascribe anything to substrates that cannot be readily explained by a traditional etymology. This is illustrated by MoIr. *gúgán* 'whelk' for which various scenarios are sketched in Appendix 4, namely either a substratal loan straight away, or a 'British' treatment of a word corresponding to Ir. *fáecha*. A more sober assessment would be to not award *gúgán* a chronologically deep etymology at all. In view of its simple phonological structure, namely two open, reduplicating syllables to which the very productive Irish suffix *-án* has been added, an original creation in the language, i.e. a comparatively recent neologism, cannot be excluded.

The patchiness of the available documentation is another stumbling block in recognising substratal loans. Specialised vocabulary for fauna, flora, topography, but also for other aspects of local life, going back to substratal languages, may have never gained a wide currency outside local speech communities or regional dialects. Such material may either not have been adopted into the common language at large or into the written standard, or it may not have been collected before the dialect became extinct, or it is only known from modern sources. For this reason, our knowledge of the extent of loan relationships can only ever be patchy.

Patchy and late data also mean that it can be very difficult to establish protoforms. Terms for maritime microflora illustrate the point. Because of the genre bias of medieval Irish literature, they are very rarely mentioned in Old Irish sources. The earliest available evidence is typically Middle or even Modern Irish. On the other hand, the dictionaries of the modern Gaelic languages, including specialised dictionaries for flora and fauna, contain an abundance of words, but their spelling and phonology can differ substantially between the sources. Words for some plants and animals are only first recorded in dialect collections of the second half of the 20th century. Even if their word formation was regular and transparent in the distant past, in their currently accessible form they can defy etymological analysis. Alternations such as MoIr. *coirleach* ~ *coirleagannach* ~ *coilleaganach* ~ *cál leannógach*, all for 'strap wrack, oarweed (*Laminaria digitata*)', illustrate the dilemma. Without old attestations, the direction of change is difficult to determine. Which is the more archaic term, *cál leannógach* or *coirleach*? *Cál leannógach* is readily understandable as 'cloaked cauliflower'. This could be an instance of metonymic, folk-taxonomic nomenclature. Alternatively, it could have arisen through folk etymology in an effort to rationalise an obscure word such as *coirleach* or one of the other variants, which allow no surface analysis. This consideration finds support in the fact that the adjective *leannógach*, which seems to contain *leann* 'cloak', does not otherwise exist in the

language. This does not, however, clarify the question as to which of the other three forms is the most archaic, or whether any of them is archaic at all. All could be 'distortions' of yet another variant that happens not to be recorded. Someone with an intimate knowledge of Modern Irish dialects and their tendencies of productive word formation might be able to see clearer.

2 Layers

The main part of the study is concerned with two questions: at what points in the prehistory of Irish can we expect potential loanwords to have entered the lexicon, and are there specific phonological, morphological and semantic features that are characteristic for those layers? Cut-off points for several separate sections of loanwords offer themselves, although the precise details are not always clear-cut in every single instance.

2.1 Pre-Celtic contacts

If we leave aside the Indo-European protolanguage for practical reasons,[5] the first stage when interaction with non-Indo-European languages may have left traces in the Celtic lexicon is the period when the precursor speech communities of Celtic, coming from the Pontic-Caspian *Urheimat*, entered Central or Eastern Central Europe. It is *a priori* plausible that some of the local languages that the incoming speakers of Indo-European encountered were those of the Neolithic farming communities identified by recent aDNA studies (Lazaridis et al. 2013; Allentoft et al. 2015; Haak et al. 2015). It cannot be excluded, however, that at that time, perhaps in marginal areas, communities of the earlier hunter-gatherer societies could still be found in Europe as well, with whom interaction and linguistic exchange would also have been possible. Although hunter-gatherers gave way to Neolithic farmers very quickly in Central Europe, in some pockets traces of the

5 The absence of parallels in other branches and/or the violation of regular sound laws by apparent cognates are among the most important criteria for identifying substratal loans in the individual branches of Indo-European. For obvious reasons, these criteria cannot be applied to the protolanguage itself since there are no external branches to which it can be compared. Unusual phonology or phonotactics can still serve as hints at potential suspects for loans into Proto-Indo-European, but this is a different methodology from the quest for substratal loans in the individual branches.

earlier life-style can be discovered well into the Neolithic, for example the Vlaardingen Culture.

There is a small body of substratal loans in Celtic that must have entered the lexicon during that period in Central Europe, mostly words for items that refer to the natural world. Research in the past decades has led to the identification of a considerable number of lexical items in the documented Indo-European languages that are suspect of originating in the language of the agriculturalist populations of Neolithic Europe. Given the wide geographical extent and the long period over which such loans may have entered Indo-European languages, it is likely that we cannot simply assume a single, monolithic language as the donor, but that we have to reckon with interactions with several dialects or related languages or even entirely distinct languages. This would explain some of the divergences between similar-looking words across the attested languages. On the other hand, it cannot aprioristically be assumed that the linguistic situation of prehistoric Europe north of the Alps was as diverse and complicated as the historically documented Appenine and Iberian models of the 1st millennium B.C. might suggest. Even though it is tempting to regard the linguistic diversity of pre-Roman Italy or of the Iberian Peninsula as typical of the rest of prehistoric Europe, as Sims-Williams cautiously suggests (2018: 7),[6] this is by no means necessarily the case. It is conceivable that the very different topographic character of Central and Eastern Europe favoured languages, or at least language families, with a wider, more homogenous extent than the small-scale topography of, for instance, the Apennine Peninsula.

Peter Schrijver (1997) has drawn attention to a particular layer of loanwords that appears to be connected to the pre-Indo-European agriculturalists of Europe. Because of the prominence of birds in his initial collection of examples, Schrijver (1997; 2001: 419) referred to it as the 'language of the bird names', for which I proposed the shorter moniker 'Avidic' (Stifter 2010: 155). An alternative, more technical designation is '*a*-prefix language'. A non-trivial morphological characteristic of Avidic loans is that the morphophonology of many words is conspicuous of a sort of ablauting behaviour: when the optional prefix *a*- of unclear function is added to a nominal stem, a vowel in the interior of the stem tends to be syncopated (more examples in Iversen & Kroonen 2017: 518; cf. also Campanile 1976: 134). Another morphological alternation is that between the absence or presence of a 'suffixal' nasal before a voiced consonant (Jakob, this volume; see also Kuiper 1956 for prenasalisation in Prehellenic loans into Greek). Plant names often con-

6 I cite Sims-Williams only as an example for this notion, which is found expressed in publications by other scholars as well.

tain the suffix *-tr-* or *-str-*. Kroonen (2012) highlighted root-noun inflection of non-inherited nouns as a diagnostic feature for such loans.[7] Jørgensen (this volume) draws attention to *-ann* as a possible suffix in a number of names for birds in Celtic and Romance.

Schrijver's phonological criteria for Avidic loans are the presence of a vowel or diphthong that is reflected either as *a* or *ai*, and which he therefore sets up as **aə* (Schrijver 1997: 313); the presence of the sounds **x* and **δ*; and the presence of initial consonant clusters. Stifter (2019b) added **γ* to the sound inventory of the donor language. However, most equations and reconstructions made in the past years by other scholars lack (or avoid?) fricative sounds and rather display words with a simple obstruent consonant system. Other occasional features of loans are irregular sound correspondences between internal *r* and *l*.[8] Appendix 2 contains further, more speculative thoughts about formal characteristics of loans from this language. Schrijver (2011; 2015: 202–203; 2018) has developed the Avidic hypothesis further in that he now connects this language with ancient Hatto-Minoan.

Loans from Avidic have been postulated mainly for Western European languages, namely Celtic, Germanic, Italic, but newer proposals have been increasingly encompassing more easterly branches as well, namely Greek, Albanian, and Balto-Slavic (e.g. Matasović 2020; Jakob, this volume). I made a proposal that would have involved Basque as a recipient (Basque *ain(h)ara* 'swallow'; Stifter 2010: 152–153), but this has been rebutted by Bascologists (Ariztimuño and Egurtzegi 2013), who suggest a Romance origin for the word.[9] Schrijver (1997) identified words from the semantic fields for birds, animals, plants, and minerals in his

7 *Nota bene*: The treatment of foreign words as root nouns in Germanic may also apply to borrowings from other sources.

8 This is only found in a few items, namely in 'fern' (Da. *bregne* < **bʰreg-n-*, Lat. *felix*, *filix* < **bʰelVk-*, Gr. βλῆχνον < **blēgʰ-n-*, and Fr. dial. *breuze* < **brelik-*; Cid Swanenvleugel, pers. comm.) and perhaps in 'holly' (Gr. κήλαστρος < **kelastro-*, against Arm. *kostł* < **gostVl-*). Probably not too much should be read into this, but it could either mean that the phonetic realisation of those sounds in the donor language was different from that in ancient Indo-European languages, or that the donor language only had a single liquid, the realisations of which could be perceived as either *r* or *l* by speakers of languages with a phonetic opposition between these sounds. In this context it is noteworthy that Minoan, the pre-Greek language of the Linear A inscriptions, appears to know only a single liquid sound.

9 It would, in fact, be of great interest if experts in Basque could look into the question if Avidic loans can be identified in that language. To my knowledge, such evidence is lacking so far. If it could be established that Basque had not adopted loans from Avidic, this would of course have interesting implications for models of how early Neolithic farmers and their language or languages spread across Europe.

seminal article. Since then, more birds have joined the flock (Stifter 2010, 2019b; Matasović 2020; Jakob this volume; Jørgensen, this volume), but agricultural plants have emerged as another focal point (Kroonen 2012; Iversen & Kroonen 2017).

Since loans from Avidic are found across many branches of Indo-European, they must have occurred at a very early date, perhaps before the separate branches crystallised with their recognisable features. The window of opportunity for Avidic words to have entered Indo-European languages ranges therefore from the immigration of Indo-Europeans into Central and Western Europe in the early 3rd millennium B.C. until, very broadly speaking, probably the end of the 2nd millennium B.C. Groups with Neolithic farmer ancestry have been shown to have persisted in remote pockets until the middle of the 2nd millennium (Furtwängler et al. 2020).

Criteria for assigning words in Celtic to loans at this early, somewhat amorphous Western Indo-European stage, before we can speak of Proto-Celtic in the strict sense, are the presence of similar lexemes in other branches across the Continent, and the presence of sounds that underwent typical Proto-Celtic sound changes, such as **p* > **φ* > Ø. A possible example for the latter is **kap(e)r-* 'goat', if it is not an inherited word. OIr. *cauru*, gen. *cáerach* 'sheep' < **kaperoh₃kᵘ-* 'looking like a goat' (Stifter 2020: 31–34) shows the regular Celtic segmental loss of **p*, while Gaul. *gabro-*, OIr. *gabor*, W *gafr*, OCorn. *gauar*, LCorn. *gavar*, Bret. *gavr* 'goat' < **gabro-* < **kapro-* shows its regular voicing to **b* before **r* (the voicing of intial **k* > **g* in this word remains unclear). Other examples, where ostensibly substratal **p* is treated like Indo-European **p* in Celtic, include OIr. *elada* 'craft, skill', W *elydn*, *elydr* 'brass, bronze, pewter' < **pelotn/r-* (cf. Sp. *peltre*, OFr. *peautre* 'pewter'); OIr. *lúaide* 'lead' < **ploud̂ʰi̯o-* (cf. Lat. *plumbum*); and OIr. *cilorn*, W *celwrn* 'vessel' < **kelpurno-* (Lat. *calpar*, Gr. κάλπις) (see Van Sluis, this volume). MBret. *ezlen* 'poplar' < **aptl-* (cf. Gr. πτελέα, Lat. *tilia*) is less significant because of the extensive phonological neutralisation that occurred in the consonant cluster.

The change **sn* > **nn* is perhaps not of Proto-Celtic age, since Celtiberian may still retain the unassimilated cluster (cf. Stifter 2023: 3.1.1.). However, it probably occurred very early in the history of all other Celtic languages, since the change has been carried out exceptionlessly in Gaulish, Irish and British Celtic, and it is visible in Celtic loans into Germanic.[10] One item that has undergone this change is **kᵘresno-* 'tree' > Gaul. *prenno-*, OIr. *crann*, W, OCorn. *pren*, Bret. *prenn*. The **s* of the 'root' is visible in the related formations **kᵘr(e)sto-* > W *prys* 'copse, grove', Corn. *Preeze*, ScG *preas*, and OHG *hurst* 'thicket' < **kᵘr̥st-*, Ru. *xvórost* 'brushwood'

10 In the case of PIE **bʰrus-en-* > PC **brusn-* 'breast' → **brunn̥i̯o-* 'breast, chest' (OIr. *bruinne* 'id.', W *bryn*, Corn. *brenn* 'hill'), the Proto-Germanic loan **brunjōn-* reflects the variant with assimilation.

< *k^{h}*u̯orst-*.[11] Another suspect for this is hypothetical *k^{u}*esno-* > Celtic *k^{u}*enno-* 'head', cf. Gaul. *penno-*, OIr. *cenn*, W, OCorn. *pen*, Bret. *penn*, although no external evidence is known to me for the original presence of **s* in this lexeme.

A practical limitation of tracing early loans via the operation, or not, of Celtic sound changes, concerns the vowels. Avidic loanwords are preponderantly reflected with short vowels in European languages, while in the transition from pre- to Proto-Celtic, it was only the long vowels that were subject to major phonological changes.

In the foregoing, the emphasis has been on loanwords that can be connected with the hypothetical Avidic language or languages, but this is not the only conceivable source of borrowing. Proposals have been made for substrata, traces of which are particularly noticeable in the languages of Northern Europe, for a long time. Hamp (1990) defines a 'Northern Group' of loans, affecting Germanic, Balto-Slavic and Albanian, to which Celtic is peripheral, but for which he also adduces the "prehellenic element" in Greek (1990: 294). Huld (1990) distinguishes two groups of substratum words in European languages: One, which he calls 'Alpine', left loans preponderantly in Celtic and Italic, but also in Germanic; the other one, 'North Balkan', encompasses Baltic, Slavic and Greek. He makes the unsupported claim that the language responsible for the Alpine borrowings was spoken in Central Europe until the 2^{nd}–4^{th} centuries A.D. Huld specifies medial geminates as a feature of his Alpine layer, but of his six proposed items, only two are reconstructed with a geminate, one of which, **g/katto-* 'cat', is more likely to be a *Wanderwort*.[12] Other characteristics are disyllabic bases and "ambiguity of voice" in the borrowing languages. North Balkan loans, on the other hand, typically have "open syllables except for diphthongs and clusters of nasal and stop. The former suggests palatalization [. . .]. The latter point hints at either nasalized vowels or prenasalized stops" (Huld 1990: 394). Salmons (1992) stresses the loss of distinction between **a* and **o*, the presence of **b*,[13] and the violation of the constraint on the

11 The fact that the sound that is reflected in Celtic and Germanic as *k^{u} behaves as two segments in Slavic either means that this was perceptible as two sounds in the donor language, or that labiovelars had already lost their labiality in Slavic at the time of borrowing, so that a foreign labiovelar had to be reinterpreted as two segments.

12 This is not the place to go into a detailed discussion of Huld's article, but there are several aspects that deserve critique. For example, in the article he speaks of seven items borrowed from 'Alpine', but he only lists six in the Appendix. His assessment of the ablauting behaviour of some of the words, e.g. 'apple' and 'dwelling' (Huld 1990: 412, 417), is partly wrong, and Welsh *matog* 'hoe' cannot continue his preform **mattuka*, but must be a late loan from Latin, Romance or Germanic.

13 Salmons (1992: 269) quotes Hamp's substratal reconstruction **ɔblu* for the 'apple' (Hamp 1990: 296: **əblu-*). This is an opportunity to point out that the Celtic facts are usually misrepresented in the discussion of this item. As shown in Stifter (2019a), the Celtic data is perfectly explicable on

co-occurrence of two voiced non-aspirated stops in Indo-European roots as diagnostic criteria for substratal loans (cf. also Hamp 1990: 298).

Kuiper (1995: 65–76) postulates three layers of loanwords, in all of which the vowel **a* features prominently: A1 'European' is further characterised by the absence of plain voiced stops, only aspirated voiced stops can be reconstructed. A2 belongs chiefly to loans into Proto-Germanic. This layer is characterised by a simple tripartite vowel system *a i u*, initial clusters *kn-* and *kl-*, as well as complex alternations of voice, gemination and prenasalisation of root-final consonants. A3 'Old European' is represented by hydronymy across the Continent. Kuiper's ideas about substratum layer A2 are taken up by Schrijver (2003: 220–224) who compares the variation in root-final consonants in Germanic with the phonological behaviour of the Saami languages and thinks of a common North-West European substratum underlying both language families. Furthermore, Schrijver (2003: 195–220) regards the reduced four-vowel system of North-West Germanic as an imprint of this presumed substrate (cf. also Hamp 1990: 296–298).

The significance of these diagnostic features for substratum loans into Celtic plays out on several levels. For any criterion where voiced stops and voiced aspirated stops are involved, like in the violation of the Indo-European root structure constraint, Celtic cannot add anything on its own, since the distinction between the two classes had been neutralised except in the labiovelars. It is only in comparison with other Indo-European branches that Celtic provides supporting evidence. Although there is little for the Celtic languages on the level of concrete etymologies among Kuiper's layer of A2 loans, it is noteworthy that many words suspect of substratal borrowing into Celtic must be reconstructed with root-final geminate consonants (Stifter 2023b: 13.1.). The four-vowel system of Schrijver's and Hamp's North-West European substratum has a weak structural parallel in the reduced long-vowel system of Proto-Celtic, which consists only of the three vowels **ā*, **ī*, **ū*, while, on the other hand, the number of five inherited short vowels remains as such in Proto-Celtic and is even increased somewhat in Insular Celtic. One notable morphological feature of Hamp's Northern Group are *u*-stems (1990: 298), of which there are many in the Celtic languages.

One pre-Germanic item that is also shared by Celtic is OIr. *cuäch*, W *cawg* 'cup' < **kapuko-*, OIr. *cuäd* 'drinking-vessel' < **kaputo-*. The loss of **p* proves the deep chronological horizon of this borrowing. Cognates are not only restricted to Germanic, e.g., Go. *haubiþ*, OE *hēafod*, OHG *houbit* 'head', but are also found in

---

the basis of a quasi-Indo-European vocalism and morphology, namely PC **abūl* < **$h_2$ebōl*. The notorious question of the **b* is a separate matter.

Lat. *caput* 'head' and maybe even Indo-Iranian, if Skt. *kapā́la-* 'cup, jar, dish' belongs here (Van Sluis, this volume).[14]

The question whether Celtic shares similarities, in particular innovations, with Italic to the exclusion of all other Indo-European branches is a controversial issue of Western Indo-European. During the pre-Celtic stage, speakers of that language are likely to have been in vicinity of those of pre-Italic, perhaps in an area north of the Alps (Mallory 2013: 24). Whether this amounts to a genuine genetic node of Italo-Celtic or is rather to be conceived of as a period of geographic proximity with tendencies of a *Sprachbund* does not make as much practical difference for the lexicon of the involved languages as it does for their phonological and morphological profile.[15] If we choose to operate with the genetic concept of Italo-Celtic, it makes even more sense to speak of a South-Western Indo-European or 'Lusitano-Veneto-Italo-Celtic' node. In view of the geographic proximity of the fragmentary branches Lusitanian and Venetic to, and in view of their apparent similarities with, the better attested branches Italic and Celtic,[16] it is conceivable that, if there was a genetic node, it encompassed those languages as well. How-

14 A remarkable feature of a small number of loanwords is the variation of word-final *-d/t*, *-k*, *-H* across the languages, which Kroonen (2012) suggests as perhaps reflecting a glottal stop in the donor language. Words with that feature usually have a 'northern' distribution, i.e. such words are found in Germanic, Celtic, Italic and Balto-Slavic, that is, in branches that are suspected to have had an intermediary prehistoric homeland north of the Alps or in Eastern Europe. Apart from 'head, cup', two items have been assigned to this group so far: 'nut' (**knuH-* > OIr. *cnú*, W *cnau*; **knud-* > OE *hnutu*; **knuk-* > Lat. *nux*) and 'bee' (**bʰi-* > OE *bēo*, **bʰit-* > Lith. *bìtė*, **bʰik-* > OCS *bьčela*, **bʰek-* > OIr. *bech*, W *begegyr*, **a-pi-* > Lat. *apis*). If we cast the net wider, we can perhaps catch more candidates. Another possible item is **selik-*/**selit-* (ON *síld*, Lith. *šil̃kė*, Fi. *silli*, *silakka* 'herring'; pers. comm. Anthony Jakob). To push the speculation a step further, we can ask the question if **salik-* 'willow-tree', with a similar-looking suffix and a Europe-centred distribution, belongs in this group, too (Lat. *salix*, OIr. *sail*, gen. *sailech*, W *helyg*, PGm. **salihōn-* > OHG *salaha*, OE *sealh*; Gr. ἐλίκη, Myc. *e-ri-ka* shows a divergent vocalism). Perhaps even the spread of the *k*-suffix to inherited words such as **n̥h₁trih₂-* 'serpent' (cf. W *neidr* < **natrī* vs. Lat. *natrix*, OIr. *naithir*, gen. *nathrach* < **natrik-*) was modelled on the borrowed variation of **-H/k*?

15 See the recent discussions of Italo-Celtic in Eska (2009: 23), Weiss (2012; 2022), Schrijver (2016). Mallory (2013: 21) is somewhat sceptical of the concept, whereas Zair (2018) takes a critical position.

16 The relationships of the well-attested branches Italic and Celtic to the fragmentary branches Lusitanian and Venetic have received diverse assessments in the past years, e.g. Witczak (2005: 398–410), Wodtko (2017: 7–8), Mallory (2013: 21–27). Prósper (2002: 429–433; 2008: 63) observes that Lusitanian shows more similarities with the Italic languages than with Celtic. She proposes, with some caution, that Lusitanian may be genetically closely related to Italic. Schrijver (2016: 499–500) postulates a Proto-Italo-Celtic language, from which Celtic was the first to branch off, then followed by Venetic, which he thus regards closely related to Italic.

ever, because of the very small amount of surviving Lusitanian and Venetic material, they are not relevant for the present lexicographic study.

Shared borrowings from substrata have so far not figured as an argument in the Italo-Celtic controversy, probably because this question is usually tackled from an Indo-European angle, that is to say, studies of it concentrate on innovations within the inherited linguistic system to the neglect of shared external innovations, including loans. Taking Matasović's – by no means exhaustive – list of possible non-Indo-European elements in the Celtic lexicon (2009: 441–443) as an example, only one item in it forms an exclusive equation with Italic, namely PC **bodi̯o-* 'yellow' with Lat. *badius* 'bay'. The other seven cognates with Italic in Matasović's list have parallels also in other branches of Indo-European, usually in Germanic, and they were therefore probably borrowed earlier than the suspected Italo-Celtic sub-node.[17] Additional examples adduced by Van Sluis (this volume) are PC **bak-* 'berry' and Lat. *bāca*, It. *bacca* 'berry', which appears to be an exclusive Italo-Celtic equation; and, very tentatively, PC **kurmi-* 'beer' and Lat. *cremor* 'thick broth' (see the discussion below). Another item of an odd phonological shape, namely **gʰirūd-/gʰiruzd* or **ɣirūd-/*ɣiruzd-* 'slimy, aquatic animal', will be presented in Appendix 4. In summary, however, it appears that there is no sizeable Italo-Celtic layer of loanwords that sets it noticeably apart from the more general Western European layer. Those cases where Celtic and Italic exclusively share words could be due to the chance loss of related words elsewhere. However, the absence of an, as it were, Italo-Celtic layer of substratal loanwords does not constitute evidence against an Italo-Celtic node. The date of this node can have been late, e.g. in the 2nd millennium B.C., and in a part of the Continent where speakers of non-Indo-European languages were no longer left.

2.2 Proto-Celtic

The next layer of substratal loans is that of words that were adopted after the Celtic proto-language had crystallised out of the amorphous Western Indo-European. In practical terms it is difficult to separate items borrowed at this stage from those from the preceding pre-Celtic or from the subsequent Insular Celtic period. Strict

17 They are: 1. PC **baski-* / Lat. *fascis* 'bundle' + perhaps Alb. *bashkë* 'fleece'; 2. PC **kage/o-* 'to get', **kag̑io-* 'fence' / Lat. *incohāre* 'to begin', Osc. *kahad* 'to take' + PGm. **haga-* 'hedge, fence'; 3. PC **knū-* / Lat. *nux* 'nut' + PGm. **hnut-*; 4. PC **mazdi̯o-* 'stick' / Lat. *mālus* 'mast' + PGm. **masta-*; 5. PC **mesal(s)kā-* / Lat. *merula* 'blackbird' + PGm. **amslōn-*; 6. PC **sǫraṷo-* 'crow' / Lat. *parra* 'a bird' + PGm. **sparwa/en-*, OPr. *spurglis* 'sparrow'; 7. PC **i̯oi̯ni-* / Lat. *iūncus* 'rushes, reed' + PGm. **jainja-* 'juniper'.

criteria for assigning items to borrowings specifically at the Proto-Celtic stage are their exclusive presence in several sub-branches of Celtic, but not outside Celtic, or, if manifest cognates exist in other branches, phonological behaviour that requires Proto-Celtic sound changes to already have taken place and/or unusual morphology and sound combinations that are substantially different from those of the cognates.

The strict application of these criteria is usually not possible in actual practice. Celtiberian branched off first from the rest of Celtic and did not participate in several developments that affected the rest of the Celtic languages, hereafter referred to as 'Core-Celtic' (Gaulish, British, Goidelic). Strictly speaking, therefore, only such items can be projected to Proto-Celtic that are attested in Celtiberian as well as in one or more of the other branches of Celtic. However, due to the constraints imposed by the very fragmentary documentation of Gaulish and especially of Celtiberian, some flexibility has to be applied and Insular Celtic reconstructions often have to serve as a proxy for Proto-Celtic. Such a procedure is warranted in cases where a word has close cognates in other Indo-European languages since, by necessity, such words must have been part of the Proto-Celtic lexicon. It is impossible to achieve the same level of confidence for substratal loans. For practical purposes, Insular Celtic or Core-Celtic protoforms will be treated as equivalent to Proto-Celtic ones, unless there are strong reasons not to do so.

The demarcation of Proto-Celtic substratal loans against those adopted in the pre-Celtic phase only works if it can be shown that specifically Celtic sound changes had already operated at the time of borrowing. Positive proof for this is rarely available. To judge from the evidence as reflected in the European languages, most suspected loans from Avidic or other prehistoric languages are reconstructed with a comparatively simple consonant system, containing sounds that were present both at the pre-Celtic stage of the language and in Proto-Celtic.[18] The only Indo-European consonant sound that was famously lost from the phonological system of Proto-Celtic is **p*. It can therefore serve as a signal for a Proto-Celtic as against a pre-Celtic borrowing if a word with *p* elsewhere shows a substitute sound in Celtic instead. Among the suggested Avidic equations, three items qualify for such an interpretation:

1. W *erfin*, Bret. *irvin* 'turnips' < **arb*$^{(h)}$*-īno-* vs. Lat. *rāpa*, OHG *ruoba* < **rāp-*, Gr. ῥάφανος < **rap*^{h}*-ano-* or **rab*^{h}*-ano-*, OCS *rěpa* < **rē/aip-*;
2. OIr. *sraif, straif* 'sulphur' < **strab*$^{(h)}$*i-* vs. Gr. στροπά, ἀστραπή, ἀστεροπή 'lightning' < **strop-*, **a-str̥p-*, **a-sterop-*;
3. W *mafon* 'raspberries' < **mab*$^{(h)}$*-* vs. MFr. *ampe* < **amp-*.

18 Huld (1990: 394–395) raises the important *caveat* that the sounds of substratum words as reflected in our known European languages only say something about the phonetics of the recipient languages, but little about the donor language or languages.

The divergent behaviour of these Celtic words from the other languages is commonly ascribed to variation in voice, either in the donor language or operating during the borrowing process, but in these three instances it can be argued that Proto-Celtic **b* substitutes **p* of the donor language.

A more speculative case is OIr., W, (M)Corn., Bret. *bran* 'raven' and perhaps Gaul. *bran(n)o-*, which presuppose PC **brano-* 'raven'. This has been compared with OCS *vrana*, Lith. *várna*, Latv. *vãrna*, OPr. *warne* 'raven', under the assumption of an ill-defined anlaut alternation **b*~**v* (or **b*~**u̯*) in the donor language. If, however, the substrate word was, for instance, **βrano-* or **vrano-* with a voiced fricative, Celtic and Slavic could have substituted the initial sound in different ways. In Slavic **u̯* (> **v*) may have been the best approximation. It is likely that the allophonic lenition of voiced stops in intervocalic position was already a feature of Proto-Celtic. In that case, speakers of the language could have naturally interpreted **β* as an allophon of **b* and substituted the stop at the beginning of the word. This variation is roughly reminiscent of the double representation of substrate labial sounds as **u̯* and **b* in Greek (Kuiper 1956: 217).

An item that is exclusively found in Celtic, but that shows phonological features not incompatible with Avidic, and that fits the semantic profile, is a set of words for 'goose', reconstructable as **giγδo-* and **giγuranno-* (Stifter 2019b). The presence of Avidic words that were borrowed after the emergence of the typical phonological features of Celtic would have the corollary that Avidic speech communities still persisted in Europe or on the off-shore islands at the Proto-Celtic period (whenever that was). However, Jørgensen (this volume) makes the point that the suffix *-anno-*, albeit possibly of substratal origin, may have been productively added to words in Celtic, so that the pre-Celtic existence of the complex formation **giγur-anno-* can by no means be taken for granted.

The non-Indo-European layer of Proto-Celtic is not just limited to names of birds and other animals. For instance, a candidate for a Proto-Celtic loanword is **kurmi-* 'beer' (McCone 2005: 404–405), which is attested in three, potentially even four of the four branches of Celtic: Gaul. *curmi*, *korma*, OIr. *cuirm*, W *cwrw(f)*, OCorn. *coref*, *coruf*, LCorn. *cor*, Corn. *korev* and perhaps the word underlies the gentilic name *kurmilokum* in Celtiberian. Contrary to Matasović (2009: 217) and De Vaan (2008: 142), the Proto-Celtic etymon of all these words has to be set up as **kurmi*, not as **kormi*, since there is no regular way how PC **o* could become **u* in Gaulish, British and Goidelic. Within the Indo-Europeanist framework, the word has been compared with Lat. *cremor* 'a thick juice made by boiling grain' and *cremō* 'to burn' (De Vaan 2008: 142, from the putative root **k$^{(u)}$remH-* 'to burn'), and with OCS *krъma* 'food', Ru. *korm* 'fodder' (Matasović 2009: 217; Matasović derives the latter from a putative root *?*kerm-*, but Derksen 2008: 262 connects it with PIE **k̂erh$_3$-* 'to satiate, feed'). Although both suggestions are se-

mantically conceivable, the formal difficulty alluded to above casts serious doubt on the suggested connections and makes **kurmi-* rather suspect of external borrowing. In addition, the often-noted comparison with Lat. *cervēsia*, *cervisia* 'beer' (borrowed from Gaulish according to Pliny) is formally even more problematic. If there is a connection with Lat. *cremor* and *cremō*, they may have been borrowed independently from the same or a related substratal source, although the integration of a substratal loan into the verbal system is rather unusual. Morphologically, the inflection of PC **kurmi-* as a neuter *i*-stem is noteworthy, since substratal loans are more commonly assigned to *o*- or *ā*-stems or to the nasal inflection.

In the *Etymological Dictionary of Proto-Celtic* (Matasović 2009), less than 6% of the vocabulary do not have Indo-European etymologies (85 out of 1490 items; Matasović 2012: 157). This contrasts drastically with the assessment of Campanile (1976: 138) who finds that 28% of his corpus of 745 Old Cornish uncompounded words are of unknown origin. Of these, all but 8 have cognates in other Celtic languages. The two figures are not really commensurate. Campanile's figure is skewed since his data consists of nouns, which are much more amenable to borrowing than verbs, which constitute a considerable portion of Matasović's data. Campanile himself grants that some of those words may yet be identified as loans from historically known sources. Furthermore, Campanile talks about the lexicon of Old Cornish, a historically attested language stage, while Matasović covers a reconstructed language, Proto-Celtic. He includes inherited Indo-European words that are only reflected in a single language, but only allows for substrate words to be projected back to Proto-Celtic if they are found in at least two sub-branches. For any actual Celtic language, the proportion of words with non-Indo-European background is probably somewhere in between those two figures.

For words that entered the lexicon at a time when one can speak of a genuinely distinct Celtic variety of Indo-European, the question about the geographical region where the loan scenario played out is of importance, but the Celtic '*Urheimat*' has proven notoriously hard to pin down. For most of the 20th century, Proto-Celtic had been associated with the archaeologically defined Late Hallstatt and Early La Tène cultures, but this notion has been more questioned than endorsed in the past decades. It is upheld in Hamp (1990: 295); the most recent scholar to defend the traditional view is Zimmer (2020: 55–57). Although Mallory (2013: 19–26) avoids any link with archaeological cultures, he considers a Central European homeland north of the Alps most likely. McCone (2005: 405) tentatively hints at the possibility of structural influence from a language similar to Basque, which would place Proto-Celtic rather in the south-west of Europe. For related reasons, Schrijver (2015: 200) opts for south-west France or northern Italy. In 2016 his focus had shifted to Italy or the Alpine region north of Italy. Sims-Williams

(2020) provides a comprehensive survey of the divergent proposals across time and concludes with a new proposal of his own that Proto-Celtic may have been situated in Western Central Europe, namely "somewhere in Gaul [. . .] perhaps including part of Cisalpine Gaul" (Sims-Williams 2020: 13). The recent contributions by Van Sluis et al. (2023) and Stifter (2023) are compatible with this. Independently, and for different reasons, they put forward linguistic arguments that favour an inland location of Proto-Celtic probably in Western Central Europe.

In terms of the chronological horizon for the break-up of Proto-Celtic, the latter part of the 2nd millennium B.C. appears to be the most plausible upper limit for Proto-Celtic from a linguistic perspective. This date can be triangulated by considering the degrees of similarities and differences of the oldest documented stages of the Celtic languages. Lepontic, a Celtic language that is already notably differentiated from Proto-Celtic, is first documented as early as the 7th century B.C.; at the same time, all Celtic languages known from or reconstructable for the end of the 1st millennium B.C. appear to be very similar to each other in phonology and noun morphology. This implies that the breakup of their unity cannot have been very far back in time.[19] The Insular Celtic languages, and to a lesser degree Gaulish, are only affected by major differentiating phonological changes in the centuries after the beginning of the Christian era.

An entirely different model, hard to square with the linguistic facts, is favoured by archaeologists in the Insular academic world. According to this scenario, the absence of a notable horizon of large-scale immigration in the archaeological record of the Iron Age implies that an ancestral form of the Insular Celtic languages must have been present in Britain and Ireland at least since the arrival of the Corded Ware/Bell Beaker culture in the 3rd millennium B.C. Celtic arose in the Early Bronze Age as a trade or vehicular language along the Atlantic seaboard that stretches from the Iberian Peninsula to Britain and Ireland and further north. The glottogenesis would have happened as early as the 3rd or 4th millennium B.C. in the early Bronze Age (see, for example, Cunliffe 2018: 54–58; Koch and Cunliffe 2016). This hypothesis has numerous implications for the Celtic languages and for the linguistic geography of ancient Western Europe, implications that are flatly contradicted by the observable facts. Historically documented vehicular languages from around the world fulfil very different sociolinguistic functions from those assumed for the prehistoric 'Atlantic' Celtic. These so-called pidgin and creole languages typically exhibit a simple grammar and a strongly restricted lexicon, often drawn from multiple lexifiers, and the former do not serve as community languages. The Celtic

19 It is, however, very difficult to put a precise figure on how long these developments would have needed to unfold.

languages on the other hand preserve complex grammatical features of the ancestral Indo-European language unchanged and their rich and diverse lexicon derives in large parts directly from Proto-Indo-European. From the assumed time-depth of this hypothetical 'Atlantic' Proto-Celtic one would expect the Celtic languages in the region to be strongly dialectally diversified. However, when the Celtic languages come into the light of history, they all look comparatively similar and homogenous, as if they had only undergone separation and expansion relatively recently, with the sole exception of Celtiberian. At the same time, a number of languages, e.g., Lusitanian, Proto-Basque, perhaps the shadowy North-West-block in the Low Countries, are documented along the Atlantic seaboard at the beginning of the historic period, languages whose presence in the region would be difficult to explain if Celtic had been the *lingua franca* for millennia in that vast space.[20]

In contrast to these linguistically unsubstantiated ideas, it is evident that Proto-Celtic, or the earliest stages of Celtic, had considerable linguistic contact with pre- or Proto-Germanic (cf. Van Sluis et al. 2023; Koch 2020). This supports traditional hypotheses about the geographic position of Proto-Celtic. Taking into account the staging area for Germanic probably in northern Central Europe and the comparatively shallow time-depth of Proto-Germanic, the contact zone for the two branches is best assumed to be first-millennium B.C. Continental Central Europe, not the Atlantic seaboard.

2.3 The insular experience – Britain and Ireland

Moving forward in time from Proto-Celtic to Goidelic, the question of where the developments played out geographically does not become clearer. The beliefs, often following sharp disciplinary lines, about how Irish came to be spoken on the island of Ireland, could not be more antithetical to each other. Due to the absence of physical evidence for any major wave of immigration during the Iron Age, it is a commonly held view among Irish archaeologists that the precursor of the Irish language had been spoken on the island for a very long time before its earliest documentation. However, while this view may find expression in oral or personal communication, it is almost impossible to find it expressed or elaborated in writing, apart from the fact that it feeds into the 'Celtic from the West' hypothesis that was discussed in the previous section.

20 For more detailed criticism of these theories, see Sims-Williams (2020: 6–8) and Isaac (2004: 49–52).

The matter looks very different from a linguistic point of view. From what can be seen in the earliest accessible sources of Irish, that is, the ogam inscriptions from the 4th century A.D. onward, the grammatical make-up of Primitive Irish differs very little from that of other ancient Celtic languages (Mallory 2013: 19–20; cf. Schrijver 2014: 72–87). This comes with the *caveat* that ogam allows us only glimpses at the phonology and at a small section of noun morphology of the language, while verbal morphology and syntax remain opaque. It is those latter sections of grammar where the well-documented medieval Irish and British languages diverge most noticeably from the mainstream of Indo-European languages. The available material gives no indication as to how early or how late the syntactic peculiarities of the Insular Celtic languages started to emerge. With phonology and nominal morphology as guides, it seems that the Irish branch of Celtic separated from the rest of Core Celtic not very long before the historical period. In a naïve interpretation of the facts, it could be concluded that Irish came to Ireland only comparatively recently, perhaps in the last centuries B.C. Schrijver (2014: 72–86; 2015: 206–207) even thinks that Irish is a very recent offshoot from British and arrived in Ireland as late as the 1st century A.D., only shortly before the geographer Claudius Ptolemaeus recorded the first placenames from Ireland. Be that as it may, most linguists in the past have placed the advent of Irish sometime between the two extreme poles of the Early Bronze Age and the Middle Iron Age.

New pertinent evidence of a very different nature has come to light in palaeogenetic studies. Patterson et al. (2022) report a population influx into southern Britain from southwest France at the end of the Bronze Age between 1200–800 B.C. This could be a window of opportunity for an early form of Celtic to have entered Britain, but further critical studies of all the implications of this hypothesis are needed. In particular, more data is needed to see if Ireland was affected by this migration at all. The currently available data seems to speak against an immigration during that period (Cassidy et al. 2016).

Two routes have been suggested, one being more plausible than the other, along which Celtic speakers could have arrived in Ireland. It is a popular and surprisingly pertinacious idea that speakers of Celtic immigrated to Ireland from the Iberian Peninsula. This notion has its roots in the romantic belief that the medieval literary myth of the five invasions of Ireland related in the *Lebor Gabála Érenn*, the 'Book of the Taking of Ireland', has a historical kernel. The last of those invasions, that of the so-called *Milesians*, the ancestors of the Gaels, is said to have started in Spain. This myth is fictional pseudo-history and is ultimately based on the view, reported by classical authors, that Ireland lies across the sea 'opposite Spain' (Caesar, *bell. Gall.* 5,13; Tacitus, *Agr*. 24). The implication, namely that Ireland can be reached from Spain via a straight cruise across the sea, is true for modern navigation, but not for ancient seafaring, which proceded along the

coasts. This evidently wrong idea that Ireland was settled directly from Spain would not deserve a discussion outside literary studies if the Spanish sea-route would not still be occasionally cited as a serious option. For a short while in the middle of the 20th century, after the discovery of Celtiberian as a separate branch of Celtic, this notion was buoyed by the observation that both Celtiberian and Irish are *q*-Celtic languages, i.e. that they retained the Proto-Celtic sound *k^u, whereas the *p*-Celtic languages, Gaulish and the British Celtic languages, had changed it to *p (cf. Mallory 2013: 20). With the insight that the traditional distinction between *q*- and *p*-languages does not have a major significance for the internal classification of Celtic (cf. McCone 1996: 67–68), the apparent connection between Irish and Celtiberian and, as a consequence, the necessity for a geographical link between the two, has fallen by the wayside. In summary, the belief in an Iberian connection of Irish must be abandoned.

It is more natural to assume that Ireland was Celticised via Britain, possibly across the North Channel of the Irish Sea where the distance between the two islands is shortest, just about 35 km, or even less from some remote points in Scotland (Mac Eoin 2007: 117). The corollary of this scenario is that the precursor of Irish must have been spoken for a while on the island of Britain. This creates the opportunity for language contact with local prehistoric languages in Britain, and thus the potential for a shared layer of loanwords from Britain in Irish and British Celtic, in addition to the specific loans that Proto-Goidelic picked up after its transfer to Ireland. Conversely, British Celtic will have had potentially longer exposure to the local substratum in Britain than Irish did. A number of controversial proposals for insular loans in Goidelic and British are discussed in Appendix 3.

A small group of words without Indo-European explanation, characterised by long vowels as well as geminates, seem to be exclusive to the Insular Celtic languages and are therefore suspect of being local loans. Long vowels generally seem to be rare among substratal loans in the European languages. Candidates for this layer are *$ētt^s$*(V)lūm(b)*- ‘bat’ (OIr. *íatlu*, W *ystlum*; Van Sluis, this volume), **nūsso*- ‘beestings, milk of a newly calfed cow’ (OIr. *nús*, perhaps borrowed into W *nus*, Bret. *luzenn*, unless they continue **n/lou̯sso*- with a different ablaut grade). PC **ātti*- ‘place’ (OIr. *áitt*) may also be included in this group, despite numerous unconvincing attempts at an Indo-European etymology (see Stifter 2023: 13.1. (1)).

There is no dearth of hypotheses about which historically known languages the substrates encountered on the Western Archipelago may have been related to. Until very recently, this discourse was dominated by syntactic considerations, not by loanwords. Because of their striking typological similarities, Hamito-Semitic or Afro-Asiatic languages have for a long time enjoyed a particular popularity, almost a monopoly as potential substrates in the Westerns Isles (see section 1.1.5), e.g., in the very influential ideas of Julius Pokorny or Heinrich Wagner. De-

spite the strong belief by some scholars in structural influence, which is outside the focus of the present study, the number of proposed lexical equations with Afro-Asiatic is tiny. Four words in Van Sluis' corpus of substratum words in Celtic (this volume) have comparanda in Afro-Asiatic, namely 'bull', 'bee', 'cup/head' and 'silver'. However, none of these links is exclusive to Celtic, but cognates are found in other European languages. These alleged cognates do not require a direct contact situation in the Western Archipelago, but are reconcilable with early Mediterranean or European loanword layers. Some of the words could also be early *Wanderwörter*.

Pokorny's and Wagner's ideas, which were so influential in the mid-20th century, still underpin the hypotheses of Adams (1980) and Broderick (2013). Adams considered languages related to Basque and Berber as the most likely candidates for the prehistoric insular world, but like the works of his predecessors, his arguments are structural and typological, not lexical. Broderick (2013) is a broad and convenient survey of various substratum theories (which have to be glossed over here). While for the most part he, too, repeats structural arguments, Broderick does also quote several lexical proposals made by previous scholars, specifically with reference to placenames. For example, Coates (1988: 21–23) drew attention to the similarity between the names of the islands *Ibiza* (Lat. *Ebusus*) and *Uist* (ScG *Uibhist*), for which he suggested a Semitic etymology, as did Vennemann for a number of other insular placenames (see Broderick 2013: 298–300). The nature of placename evidence is, of course, different from that of borrowings into the generic lexicon, and many methodological question marks beset these proposals.

The evidence for identifiable loanwords in Celtic from an ancient precursor of Basque, such as Aquitanian-Vasconic or an earlier stage of it, is slim. The best candidate is Aquitanian and Bsq. *andere* 'lady, woman'. This word has striking parallels in OIr. *ainder* 'young woman', W *anner*, Bret. *annoar* 'heifer', W *enderig* 'bullock, bull-calf' (with unusually retained *d*, perhaps under influence from the verbal root *dar-* 'to bull'?; pers. comm. Paul Russell), and, with greater uncertainty, perhaps in Gaul. *anderon*, if that word means 'of women', and Fr. *landier* 'firedog'. Although numerous attempts have been made to explain the word natively within Celtic and to treat Bsq. *andere* as a loan from Celtic (see the literature referenced in Schrijver 2002: 205; differently De Bernardo Stempel 2014), Gorrochategui (1995: 42; 1996: 12–14) has made a strong case that it is in fact a compound formed productively within Aquitanian-Vasconic from the native elements **and-* (= **hand-* 'big'?) + **er(h)e-* 'woman'. Consequently, if there is a connection with the Celtic languages, the latter must be the ones that borrowed. However, the time-depth remains a puzzle. As Schrijver (2002: 214–216) has argued, OIr. *ainder* 'young woman' cannot be an old loanword, but its phonology rather points to a loan not earlier than the middle of the 1st millennium A.D. No

case can therefore be made for the borrowing of this word already at the Proto-Celtic stage. The contact between the two languages families may rather have taken place in the early historic period, although the details remain bafflingly obscure.

Finally, it needs to be stressed that recent studies of ancient DNA do not support the idea of close genetic links with populations mentioned in older hypotheses. The prehistoric linguistic precursor on the Western Archipelago may have been a lost branch of Indo-European ('Bell-Beakerish'), but the old 'favourites' Hamito-Semitic, Basque or Berber are definitly out of the race.

2.4 Diagnostic features of substratum words

This overview has so far demonstrated that it is plausible that a diversity of source languages contributed to the lexicon of Irish, and of Celtic in general, depending on the geographical regions in which their ancestral stages were spoken (Continental Europe *grosso modo*; Western Central Europe; Britain; Ireland). Furthermore we have to reckon with the possibility that at each stage several different languages, not just a single one, contributed loans to the mix (cf. also similar remarks about Celtic by Van Sluis, and about Greek by Meester in this volume).

In this section I want to discuss four formal features of suspected substratal borrowings into Celtic or Irish. They are in addition to those mentioned in the preceding sections and to those discussed by Van Sluis (this volume). They consist of one morphological and three phonological criteria that are conspicuous in Irish and Celtic words that lack Indo-European etymologies. Two of the criteria appear to be exclusive to substratal loans into Primitive Irish (sections 2.4.3.–4.) and may therefore be specific to the local pre-Celtic language of Ireland. The criteria in sections 2.4.1.–2. have a wider application and are reflected in languages other than Irish. It is evident that these four criteria do not just reflect a single source, but probably belong to several layers of borrowing and indeed to several substratal languages.

2.4.1 *sk*?

It seems as if *-sk-* can be identified as a specifically substratal suffix, not only in Irish or Celtic, but also in other Western Indo-European languages. The complex suffix **-sko-*/*-sk̂-* is well-known in many Indo-European languages, and **-sk-* is fairly pro-

ductive in inherited formations of Celtic.[21] However, it also appears in a number of words, mostly for animals, that have no obvious Indo-European pedigree. In the semantic field of marine molluscs (see also Appendix 4), one word for a type of 'shell', attested twice in genuinely Old Irish sources, is *áesc*. In post-Old Irish sources, the same word is found with an inorganic *f-* as *faísce* as an epithet of nuts, and as MoIr. *faoisce* 'shell, shellfish' and *faoisceán*, *féascán* 'crab, shellfish, mussel'.[22] The small Old Irish evidence is sufficient to project it back to pre-Irish **aiskā-*. Although there are many formally suitable roots of the required structure **aiC-*/**h₂eiC-*, no semantically satisfactory Indo-European or Celtic analysis suggests itself for this word. It is therefore a candidate for a substratal loan. Another obscure word for 'shell, husk' with **-sk-*, but without any further root etymology, is *blóesc* or *bláesc*. No sufficiently early attestations exist to decide between *áe* and *óe*, but the word is evidently cognate with W *blisg* 'husk, shell'. The relationship of the vowels is reminiscent of that between OIr. *fáecha*, **fóecha* and W *gwichan* (discussed in Appendix 4.).

In other words for animals, *-sk-* is typically added to a simpler basis, which is itself suspect of being a substratal loan. This is illustrated by *blesc* 'harlot', which must go back to **bled-skā-* 'female wolf', in turn derived from the etymologically obscure word **bledo-* 'wolf'.[23] In OIr. *bled* 'whale', the word has undergone a semantic shift, but the original meaning is preserved in MW *bleid*, OCorn., OBret. *bleit*, MCorn. *blyth*, LCorn. *blaidh*, Corn. *bleydh*, Bret. *bleiz* 'wolf' < **bledi̯o-*, Gaul. *Bledinus*, OBret. *bledin* 'wolf cub', and perhaps in various placenames on the Continent (Delamarre 2003: 79). The suffix *-sk-* may also be present in **taski̯o-* in the name of the Old British king *Tasciouantis*, OW *Teuhuant*, if this can be analysed as suffixally extended **tazg-sk-i̯o-* of simpler **tazgo-* (cf. the OIr. personal name *Tadg*). Alternatively the variation **tazg-* ~ **task-* may just be an instance of voice alternation in loanwords. The meaning 'badger' is inferred from the comparison with the name for that animal in PGm. **þahsa-* and Bsq. *azkon*.

One possible preform of W *mwyalch*, OCorn. *moelh*, Corn. *molgh-dhu*, MoBret. *moualc'h* 'blackbird', Corn. *molgh* 'thrush' is **mesalska-* (Schrijver 1997: 307), in which case *-sk-* would appear in one of the staple suspects for a loan from Avidic.[24]

---

21 For instance, in Old Irish it is found as a suffix that forms nouns (e.g. *uisce* 'water' < **udn̥skii̯o-* << **u̯edor/n-*; *géscae* 'branch' < **gank-skii̯o-* << **ǩn̥k-*), adjectives (e.g. *mesc* 'drunk' < **med-sko-*; *lesc* 'lazy' < **legʰ-sko-*), and verbs (e.g. *nascaid* 'to bind' < **nHd-sḱe/o-*; *mescaid* 'to mix' < **miḱ-sk-ā-*).

22 Without the Old Irish evidence, we could be tempted to compare *fáesc* with W *gwich*.

23 Stefan Höfler (pers. comm.) proposes an alternative explanation as **mḷd-sḱ-eh₂-* 'soft one, dear one' from the PIE root **meld-* 'soft'.

24 The reconstruction **mesal-ka-* is also possible. The substratum origin of the word for 'blackbird' is disputed; explanations within a traditional Indo-European framework have been suggested (cf. IEW 35–36; Kroonen 2013: 25–26) on the basis of a root **h₂ems-* 'black' (cf. Gr. ἄσις 'mud', Skt. *ásita-* 'dark, black').

Similarly, W *alarch*, OCorn. *elerhc*, Bret. *alarc'h* 'swan' can continue **alar-sko-* or **alar-ko-*. Matasović (2020: 334) hints at the possibility that the name for this bird belongs to the Avidic layer. The suffixal extension **-sk-* is also found in bird names outside Celtic, namely in Germanic. OE *þræsce* 'thrush' continues PGm. **þra(st)-skōn-*, while the Old Norse cognate *þrǫstr* < **þrastu-* does not exhibit the suffix (Kroonen 2013: 545). Van Sluis (this volume) thinks that W *tresglen* and Bret. *draskl* are borrowed from the Old English word with different sound-substitutions for *þ* and *æ*. The unsuffixed stem is also found in OIr. *truit* 'starling', if this is from **trozdi-* (IEW 1096), with cognates in Lat. *turdus* < **tor(z)do-* or **tur(z)do-*, Arm. *artoyt* < **droṷd-*, Lith. *strãzdas* < **strozdo-*, and OPr. *tresde*. The matter is complicated by the fact that OIr. *truit* has manifest cognates in OBret. *trot*, MoBret. *tred*, W *trydw*, OCorn. *troet*, LCorn. *trojen* 'starling'. The root-final consonant in all these names represents Proto-British **d*. This, however, cannot go back to **zd*, which would have resulted in **θ*. Either the name for the bird in the Insular Celtic languages continues **troddi-*, which is unsatisfactory with regard to the extra-Celtic cognates, or the British forms are loans from Irish.[25] The cladistic distribution of the word **(s)trosdo-* (*vel sim.*) across many Indo-European branches in Europe is compatible with an Avidic loanword. If so, **trozd-* and **trozd-sk-* could reflect a variation within Avidic.

Reference can also be made to the occasional occurrence of suffixal *-sk-* in plant names. In a genuinely productive formation in Celtic, it is found in OIr. *duilesc*, W *delysg*, *dylysg* 'edible seaweed' < **dolisko-*, derived from the basis **doli-* < PIE **d^holH-i-* 'leaf' (W *dail*, Corn., MBret. *del* 'leaf/leaves', and OIr. *duilne* 'leaf, foliage' with a singulative suffix; also underlying Lat. *folium*). OIr. *rúsc*, W *rhisgl* 'bark' (OCorn. *rusc* and Bret. *rusk* 'id.' must be borrowings from some variety of Latin) could be set up as **ruH-sk-o-*, from the PIE root **reu̯H-* 'to tear open, rip', but Campanile (1976: 135–136) notes the presence of related words in Sardinian, Sicilian and other dialects of Italian. He regards this geographical distribution as indicative of a substratal loan. Other examples are uncertain. At a first glance, PGm. **aska-* 'ash tree', Alb. *ah* 'beech', Arm. *hacc'i* 'ash tree', Gr. ὀξύα 'beech', which continue **osk-* 'ash tree', may look pertinent here, too. But not only is it undecided if the word is inherited or substratal (see Matasović, this volume), but **os-* (< **Hh_3os-*?) is a more basic root of this word (cf. PC **osno-* in W, LCorn. *on*, Bret., Corn. *onn*, **osVno-* in Lat. *ornus*), so that the suffix could equally be **-k-*. The root of the – probably Avidic – loanword for 'sedge, iris' (e.g. OIr. *ailestar*, W *elestr* < **alisk-stro-*) appears to

25 In Stifter (2021: 174), I was still leaning towards PC **troddi-*. Other Irish words with *-sc* are recent loans. MoIr. *troisc* 'thrush' is a borrowing from English, and *trosc* 'cod' is a loan from ON *þorskr*. But note the *-sk-* in the Germanic name of that fish, **þurska-*.

be **lisk-*, cf. MLat. *lisca* (Van Sluis, this volume). With our current knowledge, it cannot be broken down into more nuclear constituent parts, so *-sk-* could be part of the root, but cf. also Lat. *lilium* and Gr. λείριον.

The combined evidence from several European languages suggests that *-sk-* may have been a suffix associated with loanwords for animals and perhaps for plants. The divergence between presence or absence of the suffix not only across branches, but sometimes even within branches of Indo-European could be due to separate loans, but it could also mean that the suffix remained mildly productive within the recipient languages. Its similarity with a well-known inherited suffix of Indo-European could mean that the donor language formed part of *Indogermania Submersa*,[26] unless suffixal *-sk-* is just a chance formal similarity between genetically unrelated languages. The presence of *-sk-* is only an indicator for a possible substratal loan, it is not by itself a proof. This was seen in OIr. *duilesc* above, but it also applies to OIr. *loscann* 'frog', for which the Indo-European etymology **plu-sko-* > **φlusko-* from the root **pleu̯-* 'to swim, hover' has been proposed. The Germanic word **fruska-* 'frog' < **prusko-* (e.g., G *Frosch*) may also be compared, which could be a deverbative formation from the root **preu̯-* 'to jump' (Marstrander 1908). The suffix *-ann* is reminiscent of the avian suffix proposed by Jørgensen (this volume).

2.4.2 geminates

The next criterion for the identification of substratal loans has to do with the observation that many words in Irish and other Celtic languages that lack Indo-European etymologies contain sounds that seem to continue geminate consonants, especially geminate stops. In Irish, the reflexes of these geminates typically show up as word-internal unlenited sounds. At the same time it is noteworthy that geminates seem to be rather rare in the corpus of substratal loans of other Indo-European languages.[27] It is hardly a coincidence that geminates surface

26 For example, it could be a lost 'Bell-Beakerish' branch of Western Indo-European that arrived with the immigration of Bell-Beaker people with Indo-European steppe ancestry in Britain and Ireland from around 2450 B.C. (Cassidy et al. 2016: 369–370; Olalde et al. 2018: 193).

27 The only examples I could find are: 1. Gr. κάμμαρος 'lobster' < **kammar-*, which stands beside ON *humarr* < **kuma/or-* with a single **m*. W *cimwch* and MoIr. *gliomach*, which appear to belong here, also presuppose a geminate **mm*. 2. Lat. *bacca* 'berry', which, however, stands beside more common *bāca*. The geminate in this word could therefore be an internal development of Latin; W *bagad* 'cluster, bunch' reflects a preform with a simple **k*. Germanic is the only major exception to the rarity of geminates in substratal loans. A specific layer of unetymologised lexical items in the Germanic languages, labelled 'A2' by Kuiper (1995: 68–72; cf. also Schrijver 2003: 220–224), is characterised by geminates.

chiefly in the one group of European languages in which an opposition between 'weak' (i.e. lenited) und 'strong' (i.e. unlenited) sounds plays an important phonological role. This could mean that, at least in a subset of cases, Celtic geminates do not necessarily reflect a genuinely double or long sound in the donor language, but rather a fortis sound that was cognitively equated with non-lenition in Celtic – the phonological manifestation of non-lenition in word-internal position being gemination. Like in most Indo-European languages, geminate stops are very rare in root-final position in inherited Celtic words, but they occur more frequently in words without evident Indo-European explanation. The phenomenon of gemination in Celtic is treated in great detail in Stifter (2023b).

Appendix 4 presents three words for molluscs and shells that will serve as an illustration of this phenomenon, namely OIr. *giritán*, *fáecha* and MoIr. *gúgán*, as well as their cognates from other Celtic languages. Since the real-life items are specific to the Atlantic littoral, there is an inherent plausibility that terms for them may go back to precursor languages. However, things are not so simple. Sometimes geminates may only be apparent, or they are artifacts of the way how we reconstruct prehistoric forms.

2.4.3 The *f*-substrate

Gearóid Mac Eoin (2007: 120–122) mentioned a criterion for identifying local Irish loans. He observed that a number of etymologically obscure words contain word-internal *f*. He quotes *bréife* 'ring', *cufar* 'leg', *cuifre*/*cuipre* 'indulgence', *fafall*/*fabhal* 'filth', *lufe* 'female', *slife* 'broadening', *strophais* 'a covering of straw'.[28] Most of these words are very rare or are restricted to glossaries, both of which indicates that they may have had a limited geographical or demographic distribution. Medial *-f-* also occurs in placenames, namely *Aífe, Bréifne, Crufait, Dún Gaifi, Faffand, Grafand/Rafann/Raphe, Grafrenn, Life, Máfat*. These placenames, some of which refer to more than one place, come from all over the island, but they appear to be gravitating particularly towards the center-east of Ireland.

Word-internal *f* is not excluded in inherited words of Irish as such, but since it can basically only continue the Proto-Celtic cluster **-su̯-*, which is not a common combination,[29] Mac Eoin's argument is valid in principle and his list of 'suspicious' generic nouns and placenames with *f* deserves attention. However, it is necessary

28 Meyer's (1891: 462–432) analysis of *strophais* as a compound of Old Norse *strá* 'straw' + unattested **peiss* < VLat. *pexa* 'tunic, shirt' strikes me as contrived.

29 For example, OIr. *mifir* 'faint-hearted, despondent' < **mis-u̯ir-i-* 'bad-manly'. See on this question also Stifter (2019a: 207–212). Jasanoff (2017: 334) identified **-βɬh-*, which arose through an

to exercise caution. One item with a Celtic etymology, and therefore rightly not included in Mac Eoin's list, is *grafann* 'horse-race', which can be explained as PC **gregi-sṷannā-* (*vel sim.*), a compound of **gregi-*,[30] a collective noun for 'horses, stud, herd', + the verbal root **sṷann-* 'to chase'. As the simplification of PC **-gVsṷ-* > **-γ†hṷ-* > *-f-* demonstrates, Old Irish *-f-* can be the outcome of rather complex clusters and it is conceivable that not all possibilities have yet been exhausted in reconstructing the items in Mac Eoin's list. A possible candidate for such an approach is the placename *Bréifne*. Perhaps there is a link between it and *bréch*, an obscure glossary word for 'wolf'. For the use of the latter in placenames, one can compare *Bréchmag* 'wolf plain', which Ó Corráin (2017: 468–469) lists as the name of ten different places across Ireland, among them modern *Breaffy*. As for the placename *Grafand/Rafann/Raphe*, Mac Eoin's argument against equating it with the word for 'horse-race', which closely resembles one of the three attested variants, is that the name refers to an uphill location that is unsuited for horse-racing.

Mac Eoin includes *Aífe* among the placenames, but it is better known as a female name in medieval literature. In the form *Aoife*, it is one of the most popular Irish girls' names today. Formally, the name can be easily explained as Celtic **aḭsṷiiā-*, the female version of the name of the Gaulish god *Aesus*/*Esus* < **aḭsu-*, which has a satisfying Indo-European etymology in the root **h₂eḭs-* 'to venerate'.[31] Under this analysis, the placename *Aífe* may reflect a secondary usage of that divine and personal name. The alternative is of course also conceivable, namely that *Aífe* started out as a substratal placename from which the name of a mythological figure was then abstracted. Eventually it was adopted as a personal name for ordinary humans. There are parallels for such a sequence of events.

2.4.4 The *p*-substrate

The chronologically youngest criterion, relevant only for Irish, is easiest to define. Peter Schrijver (2000; 2005; 2014: 82–83) has drawn attention to a small number of words in Irish that start with *p-*. Since Primitive Irish famously lacked this sound, words of the *p*-stratum can only have entered the language after the adop-

irregular, 'precocious' syncope before the well-known syncope of Primitive Irish, as another possible source for OIr. *-f-*. If anything, this is structurally an even rarer context than **-sṷ-*.

30 The vowel *a* in the first syllable of *grafann* presupposes the specifically Irish change of **e* > **a* before a palatalised **g* (McCone 1996: 111). The reconstruction **grego-sṷendnā* of IEW 382 is therefore impossible. The non-palatalised *-f-* of *grafann* may perhaps be due to the depalatalising effect of **h*, cf. Proto-Goidelic **æmbi-sūs* > **imb†hu* > OIr. *impu* 'around them'.

31 Cf. Stifter (2019a: 209 fn. 48); a similar proposal had been made before by O'Rahilly (1946: 5).

tion of *p-* as a foreign sound into the phonemic inventory of Irish. This occurred approximately in the late 5th or early 6th century. Despite the methodological criticism of Isaac (2003), no convincing etymologies can be found in the known contemporary neighbouring languages for several of those words in *p-*. Therefore the conclusion is inevitable that they were borrowed from an unknown local language that was still spoken in Ireland at that time. Words that belong to this layer are *partán/portán/partlach* 'crab', *petta* 'pet', *pluc* 'round mass', *prapp* 'rapid', *pait* 'bottle made of skin'. In the case of some of these words, such as *prapp* or *pluc*, one could also think of sound-symbolic *Urschöpfungen* (original creations or neologisms) within Irish.[32] *Pell* 'horse' is doubtful, since it could be metonymically from *pell* 'animal skin', a loan from Lat. *pellis* (cf. OW pl. *guapeli*, W *gobell* 'saddle' < **u̯o-pell-*), or a shortened form of *capall* 'horse' < **kappilo-* (cf. W *ceffyl*, OBret. *chefel* 'horse', Bret. *kefeleg* 'woodcock'; Stifter 2023b: 13.1. (18)). Even after its adoption into the phonemic inventory, the sound *p* still retained a high markedness within the grammatical system of Irish. This would have made it particularly apt for utterances with a high symbolic load.

Pattu 'hare' and *pít, fít* 'portion of food' were originally proposed to belong to this group as well, but Schrijver retracted them because of formal ambiguities after Isaac's criticism.[33] There are a few more possible items which are not included in Schrijver's original list, although they meet the phonological criterion. *Parn* is once attested in a glossary (*Sanas Cormaic* 1049), where it is explained by the equally obscure *bloach mara* '*bloach* of the sea', which in turn is elsewhere glossed as 'whale' (Stifter 2023: 183). *Lupait*, a word for a 'young pig', has an internal voiceless *-p-*, cf. MoIr. *lupaid*. *-ait* < **-antī-* < **-n̥tih₂-* looks like an inherited suffix, but the presuffixal part of the word has no obvious etymon. Perhaps the original loan was **lupp*, which was then influenced by inherited *birit* 'sow' < **bʰer-n̥tih₂*. OIr. *poll* 'hole, cavity' is regarded as a loan from W *pwll*, MCorn. *pol(l)*, Bret. *poull* 'hole, pool'. The British words themselves are believed to be borrowed from OE *pōl* 'pool'. This is difficult chronologically, semantically (the meaning 'hole' seems to be primary in Celtic), and, not least, phonologically: the British Celtic word has a short

32 Paulus van Sluis (pers. comm.) speculates if *prapp* goes back to a sort of reduplication **(ra)p-rap(ide)* of Lat. *rapidus*, colloquially curtailed in situations where it was used as a term for urging on people.

33 In personal communication, Peter Schrijver has mentioned the idea that *pattu*, if its original meaning was 'the one with the notable paws < *feet', could represent a treatment of PIE **pod-* 'foot' in a lost local Indo-European; the hypothetical Bell-Beakerish language mentioned in footnote 26 would be a suitable candidate. Schrijver likewise noted that **kappilo-* 'horse' could have a similar connection with PIE **ĥoHpo-/ĥopHo-* 'hoof'.

*u, not a long *ō, and a geminate, not a simple *l*. Perhaps this word belongs to the same layer of loanwords as those of the *p*-group in Irish.

Schrijver very tentatively hints at the possibility that a remote region in the west of Ireland in counties Mayo and Galway, home to the *Partraige* 'the crab-people', could have been the last refuge of the speakers of the language from which these words with *p* were possibly borrowed (2005: 137).

3 Conclusion

In this chapter, I have tried to sketch a panorama of the potential layers of substratal loanwords that the Celtic languages, and especially its Goidelic branch, absorbed in prehistory. The time-frame encompasses the post-Proto-Indo-European stage in the 3rd millennium B.C. before the crystallisation of Celtic as a recognisable linguistic entity, through Proto-Celtic approximately in the late 2nd millennium B.C., to Insular Celtic and Goidelic in the millennia on either side of the turn of the era. Geographically, I have moved along a route from Central Europe through Western Europe to Britain and then Ireland. The stages of this route correspond approximately to the chronological stages. At each of the stages it is methodologically possible to detect the influx of substratum words according to criteria specific to that stratum. I do not imply that the influx must equally have been from four different sources. It could be more or fewer sources that contributed to the substratal mix of the Celtic languages.

Many of the early loans are clearly to be identified with Avidic, the language of the Neolithic Early European Farmers of Anatolian descent who left traces in most languages of Europe. It is conceivable that the adoption of words from that source continued from the *voreinzelsprachlich* to the *einzelsprachlich* period of Celtic. No criteria exist to identify with certainty items that could go back to the lexicon of the Mesolithic hunter-gatherer populations that inhabited Western Europe before the spread of the Neolithic farmers.

Separate lexical layers can be detected in the languages of the Western Archipelago. Given what is known about the genetic make-up of its Bronze-Age population, it stands to reason that this stratum of loanwords into Celtic can be identified with 'Bell-Beakerish', i.e. the presumably Indo-European language that had been brought to these islands around 2450 B.C. (Patterson 2022). This need not have been a single, uniform language at the time of interaction with early Celtic, but it is more likely to have become a series of diversified, but related dialects or languages across the islands and within Ireland. There is a small amount of evidence to indi-

cate that the last remnants of pre-Celtic tongues were spoken in Ireland until the early middle ages.

This panorama does not aim at proposing a solution to the prehistory of Celtic, many stages of which are still obscure, but it is meant as a contribution that can indicate the direction that research can take in the future. A lot of discoveries still remain to be made.

Appendices

1. One etymologically difficult item, for which ogam provides crucial information for assessing its possible status as a loanword, is OIr. *rón* 'seal'. A loan from OE *hran, hron* 'small kind of whale, mussel', perhaps via British, has been proposed (cf. LEIA R-42; Pedersen 1909: 21), despite the difference in vowel length. However, such a loan is practically excluded for chronological reasons. The name ROṆ[A]ṆN = *Rónán* 'little seal' appears on the ogam stone from Arraglen (I-KER-012 = CIIC 145). On linguistic grounds this stone is dated approximately to the second half of the 6th century. Several very early Irish saints of that name, by necessity rather obscure, are also alleged to have lived in the 6th century, namely St Ronan of Locronan, an Irish pilgrim who became a hermit and saint in Brittany, and St Rónán of Ulster, whose death is said to have occurred on 11 January 535. The earliest occurrence of the name in a manuscript source is in an entry in the Annals of Ulster for the year 624, which reports the death of Rónán mac Colmáin. Although we have no other information about this man, it is fair to assume that he must have received his name in the late 6th century at the latest, i.e. approximately at the time of the ogam stone from Arraglen. It is hardly credible that an Old English loanword could have reached Irish and could have entered the onomastic lexicon as early as the 6th or even the late 5th century. Even though it does not clarify the origin of *rón*, the ogam evidence at least helps to exclude one source that has been postulated in the past.

Rón has cognates in MW *moelrawn*, MCorn. *ruen*, Corn. *reun*, Bret. *reunig* 'seal'. The only way how the Irish and British vowels can match regularly is by reconstructing Proto-Celtic (= PC) **rau̯no-* with the diphthong **au̯*.[34] This reconstruction is homonymous with PC **rau̯no-* 'coarse animal hair', attested in W

34 Lexical equations with PC **au̯* are very rare, e.g., PC **au̯-bero-* 'vain, inane' > OIr. *óbar*, *úabar*, W *ofer* (with pretonic reduction of the first vowel), Bret. *euver*, MCorn. *ufer*. In the majority of cases OIr. *ó* corresponds to British **ïŭ* (written < u > in the British Celtic languages) < PC **ou̯*, or, if the length of OIr. *ó* is caused by compensatory lengthening of short **o*, to a British sequence of *o*

rhawn, Bret. *reun*, and in MoIr. *rón* and in the singulative OIr. *rúainne*, *róinne* 'a single hair' < **rau̯n-in-i̯o-*.[35] Despite their formal identity, it is unclear if 'seal' and 'coarse hair' are etymologically the same word. Seals do have very thick and dense outer hair, so that their Insular Celtic name could be interpreted as a metaphoric use of 'animal hair', but morpho-semantic and pragmatic doubts remain. It would be surprising if the word for 'seal' derived from 'hair' without an overt derivative suffix. One rather expects the animal to have been called 'the one having coarse hair' with one of the common adjectival suffixes (e.g., PC **-(i)i̯o-* or **-āko-*), not just simply 'coarse hair'; I am not aware of a parallel for such an appellative strategy.

The situation is further complicated by the Baltic words for the 'seal', Lith. *rúonis* (beside *rùinis*) and Latv. *ruônis* 'seal' (Ariste 1971–2: 10).[36] That such similar-

+ a consonantal segment (which in turn may be vocalised); while words with W *aw*, MCorn. *ue*, Bret. *eu* (the latter two phonetically [ø]) preponderantly have cognates with OIr. *á* < PC **ā*. Only when OIr. *ó* corresponds to W *aw* (or shortened *o*), and SW-Brit. **ø* do we have to operate with a common preform **au̯*.

35 PC **rau̯no-* 'coarse ainmal hair' has manifest cognates for 'hair' or 'wool' across the Indo-European language family, but the devil is in the details. The only plausible reconstruction for Celtic **rau̯no-* within the framework of traditional Indo-European phonology is PIE **reh₂uno-*. The various Indo-European nouns for 'hair' or 'wool' are probably derived from a verbal root for 'tearing (open)', which LIV (p. 510) sets up as **reu̯H-*. The semantic motivation for the nominal formations lies in the practice of plucking animal hair to gather wool. If the suggested Proto-Celtic reconstruction is correct, the laryngeal could be determined as **h₂*. LIV remarks that in Vedic this root can show up secondarily as an *aniṭ*-root, e.g., in the verbal adjective *rutá-*. Related nominal formations, with different suffixes, are ON *rǫggr*, *rǫgg* 'shaggy hair or wool' < **rawwa/ō-* < **rou̯H-ó-* (Jasanoff 1978: 89–90, no. 11; Rasmussen 1990: 440; **HrouH-ó-* in Kroonen 2013: 407), which seems to confirm the presence of a laryngeal (in personal communication, Guus Kroonen informs me that he believes that Germanic *Verschärfung* need not be triggered by laryngeals, and that he considers a reconstruction **(H)rou̯-ó-* equally viable). Skt. *lóman-*/*róman-* 'hair' < **reu̯-mon-* seems to continue the *aniṭ*-root. The same is true for Slavic **rūnò* 'fleece' (e.g., Ru. *runó*, Cz. *rouno*), which belongs to accentual paradigm b and therefore presupposes a laryngeal-less preform. Derksen (2008: 440) reconstructs **(H)rou-(m)no-* for Slavic and compares not only Skt. *róman-* '(body-)hair', but also OIr. *rúamnae* 'blanket'. However, the latter is almost certainly a ghostword (see Greene 1982: 163); the alleged sole attestation in Sg. 69a9 belongs rather to the unrelated word *rúamnae* 'red colour, dye'. The true Celtic cognates of Slavic **rūnò* are the above-mentioned OIr. *rón*, *rúainne*, W *rhawn* etc., which, however, cannot continue **(H)rou-(m)no-*, but, as demonstrated above, require the diphthong **au*, whether it come from PIE **eh₂u* or from an original **au̯*. The only way out of this dilemma is to assume that Slavic and Indic are built on a laryngeal-less variant of the root (perhaps the laryngeal was lost in the heavy cluster **(H)re/ou̯(H)mn-*), while Germanic and Celtic retained the laryngeal, but with metathesis of **(H)reu̯h₂no-* > **(H)reh₂uno-* in Celtic. The laryngeal is also evidenced by Proto-Germanic **rūjan-* 'to tear' < **HruH-i̯e-* (Kroonen 2013: 416).

36 I thank Anthony Jakob for suggesting this item and for discussing the phonological problems with me.

looking words are found for the animal in so genetically and geographically distant branches like Baltic and Celtic is unlikely to be due to independent, parallel developments or to borrowing between the two branches. According to the traditional view, the sources of Baltic *uo* are Indo-European **ō* and **ōu̯* < **oHu*.[37] If the word were inherited from the common proto-language, the Celtic and Baltic words could be reconciled under the assumption of an ablaut and suffixal alternation between Celtic **reh₂uno-* ~ Baltic **roh₂uni-*. However, since Proto-Indo-European or an early form of post-Proto-Indo-European is not known to have been spoken in a region where seals were native, common inheritance is very unlikely. Lacking any compelling alternative, the best solution to this impasse that remains for the moment is to regard **rōnV-* as a borrowing from a prehistoric language in northern Europe into both Insular Celtic and Baltic. Paulus van Sluis (pers. comm.) reminds me that a late, Roman-period foreign **ō* can occasionally result in the same distribution of vowels as inherited **au̯*, cf. Lat. *hōra* 'hour' > W *awr*, Bret. *eur*, OIr. *ór*, *úar*; or Lat. *nōna* 'ninth (hour)' > W *nawn*, OIr. *nóin*. We could therefore be confronted with a relatively recent borrowing during the first half of the 1st millennium A.D.

2. In this paragraph, I want to follow down the rabbit-hole of a very speculative observation: Schrijver (2018: 363–370) has made the very tentative comparison between **aule*, which he postulates to be the Minoan word for 'pig', and Lith. *kiaũlė* 'pig' < **keu̯l-*, *kuilỹs* 'boar', and the first element of the Middle Welsh legendary figure *Culhwch* < **keu̯lo-sukko-* (W *hwch* = 'pig'). The relationship between the two etymons can either be described as including an initial velar (?) sound that did not have a graphic representation in Linear A, or as involving the presence or absence of a velar prefix *k-*. Furthermore the vowels *a~e* alternate. Speaking about various possibly substratal words for 'goat', Kroonen (2012: 245–247) reconstructs **ai̯di̯-/*aȷ-* as the ancestor of Gr. αἴξ, Arm. *ayc*, Skt. *ajá-*, Lith. *ožỹs* 'goat' (cf. here, by the way, OIr. *ag* 'bovine animal'). These stand beside **gʰai̯d-* (Lat. *haedus* 'kid', PGm. **gait-* 'goat'), PSem. **gadi̯-* 'goat', and another protoform **kăĝ⁽ʰ⁾-* (OCS *koza*, Alb. *kedh* 'goat'). To this group we may perhaps also add OIr. *cadla* 'goat' < **kadVlāt-* and PGm. **hadVlan-* (MHG *hatele*) and **hadVnōn-* (ON *haðna*) 'goat'. In a maximally reductive process this series of words can be simplified to the primitive structure **(K)a(i̯)D-*, where again presence or absence of a velar initial can be observed. With an eye on Schrijver (1997: 303–307, 312) we may even consider reconstructing **(K)aəD-*. There is, in fact, a third pair with 'mobile *k*': Lat. *aper*, PGm. **ebura-* < pre-Gm. **epr-*, and OCS *veprъ* 'wild boar' stand beside Gr. κάπρος 'wild boar', as well as the family

37 But see the critical discussion by Villanueva Svensson (2015), who argues that the sound law *uo* < **ōu̯* is based on dubious evidence.

of Lat. *caper*, PGm. **hafra-*, OIr. *gabur* '(male) goat'. In addition to mobile *k*, this item also shows vowel alternation between *a~e*, which, incidentally, is also observable in words for 'bull', i.e. **(s)taur̥o-/steur̥o-*. The initial mobile velar sound could either be a sound that was foreign to the recipient languages and that therefore is reflected by different Indo-European substitutes, e.g. some sort of glottal sound, or, given that it occurs in a number of different items, it could have been a class or gender prefix in the donor language.

3. Matasović (2012: 157) is sceptical of the scenario that British and Goidelic shared a common phase in insular prehistory. Instead he maintains that "the large majority of substratum words [. . .] in Goidelic and Brythonic [. . .] is not shared by these two languages, which probably means that the sources were different substrates of, respectively, Ireland and Britain". He quotes words for six animals and one plant in support of this claim:

	Irish	**Welsh**
1. 'fox'	*sinnach*	*cadno*
2. 'rowan-tree'	*luis*	*cerddinen*
3. 'duck'	*lacha*	*hwyad*
4. 'blackbird'	*lon*	*mwyalch*
5. 'beetle, chafer'	*dega*	*chwilen*, *gordd*
6. 'weasel'	*ness*	-
7. 'crab'	*partán*	*cranc*

However, under a critical scrutiny this list needs to be revised profoundly, with the consequence that little remains from this handful of words to back up his critical attitude towards a shared insular prehistory of Goidelic and British. For methodological reasons, inherited words, productively created words, and words that were borrowed before or at the Proto-Celtic stage have to be removed from this list, if Matasović's claim of external influence from two different substrates on the two islands is to be proven. Not much remains after such a critical sifting of the evidence:

1. OIr. *sinnach* 'fox' is not a substratal loan, but finds a perfectly language-internal explanation in PC **senŭnāko-* 'old one', an adjectival formation in **-āko-* from the *on*-stem **senū*, **senon-* 'old one', with generalised full grade of the suffix as a morphological marker of high animacy. This is a *noa* word that replaced the inherited word for the 'fox', probably PC **loφerno-*, for taboo reasons. Likewise, W *cadno* is a transparent compound of *cad* 'battle' + the common male name element *-no-* < **-gnāur̥o-* (*vel sim.*; see Schrijver 1995: 300). This is identical with the personal name *Cadno*, which was transferred

onto the animal, like in French the personal name *Reynard* > *renard* replaced *goupil*, the inherited word for the fox. Other Welsh personal names used for the 'fox' are *Madog* and *Madyn*.

2. W *cerddinen* 'rowan-tree' has an Irish cognate in *cáerthann*. Even though their formal relationship is notoriously difficult, it is obvious that they are related, so *cerddinen* cannot be called an exclusive loan of British (cf. Schrijver 1997: 305–306). On the other hand, the Irish word *luis*, which Matasović cites for 'rowan-tree', cannot even be proven to belong here. It is the name for a letter in the ogam alphabet. Its identification with the 'rowan-tree' is probably a late construct by native grammarians, as can be demonstrated for a series of other letter names (McManus 1988: 150; Hayden and Stifter 2022). The original meaning of the word, if it had any, may have been 'light, shine' < **luk-s-i-* or 'plant, vegetation' < **h₁lud*ʰ*-ti-* (cf. OIr. *lus* 'plant' < **h₁lud*ʰ*-tu-*).
3. For OIr. *lacha* 'duck', Matasović himself (2020: 341) has proposed an Avidic etymology. If the word had entered Goidelic only in Ireland, this would imply a distinct community of Avidic speakers on that island at a relatively late date. One would rather assume that the word was borrowed together with the bulk of Avidic borrowings, i.e. on the Continent in the Bronze Age. W *hwyad* is possibly a formation that ultimately contains the Indo-European word for 'bird' **h₂eu̯i-* (see Schrijver 1995: 102).
4. W *mwyalch* 'blackbird' has been one of the chief suspects for an Avidic loan (Schrijver 1997: 307–308). It therefore belongs to the pre-insular stratum of loans. Moreover, it has even been defended as an inherited Indo-European word (see fn. 24).
5. OIr. *dega* 'beetle' has been compared with E *tick*, G *Zecke*, perhaps from **digāt-* (LEIA D-38), which points to a loan at the pre- or Proto-Celtic stage. W *chwilen* may go back to the Indo-European root **su̯ei̯H-* (IEW 1041), which also forms the basis for a number of other formations in Welsh. GPC suggests that *gordd*/*ordd* 'small beetle in a fulling mill' is the same word as 'hammer', but it provides no explanation for the alleged semantic connection. **Ordo-* is also found in OIr. *ord* 'hammer' and in the name of the ancient British *Ordouices* 'hammer fighters'. The *g-* of W *gordd* can be due to a reanalysis of the initial in lenited contexts. Although its further etymology is unknown, this word is at least common to British and Irish. If *gordd*/*ordd* is indeed the same item etymologically for 'beetle' and 'hammer', no word for 'beetle, chafer' can be demonstrated to be of insular origin.
6. W *cranc* is perhaps a loan from Latin *cancer*, in which case it cannot be adduced as evidence for a local substratum.

> What remains are *lon*, *ness*, and *partán*, all words exclusively found in Irish (for *partán*, evidently a very late borrowing, see section 2.4.4). What Matasović's small selection therefore demonstrates is that the Goidelic branch was influenced by a substratum specific to the island of Ireland, perhaps without parallels outside of the island. A supporting criterion for a local borrowing is sometimes the connection with the Irish habitat. Such a layer of local substratal loan items, after the ties of Goidelic (or Irish) with the other Celtic languages had been severed, is hardly a surprise. The presence of this layer does not rule out a common intermediate homeland of Irish and British on the island of Britain. Irish can have been exposed to influence in both places. The presence of 'suspicious' items in both branches does not, however, constitute unambiguous proof that the word was borrowed at a common stage before the separation of the two branches. Since it is known that Goidelic and British borrowed from each other during the historical period, and probably already before that, it is conceivable that some words that are found in both branches may in fact be the result of borrowing from one into the other, and that only one of them originally adopted a substrate lexeme from a precursor language.

4. In the following three sub-sections, proposals that I made in a paper in Gothenburg 2018 and that are published in Stifter (2023: 185–186) are further developed. The suggestions below supersede the discussion in Stifter (2023).

4.1. giritán

OIr. *giritán* is once attested in a glossary, explained as *faochain mara* 'periwinkle (*Littorina littorea*)', a type of sea-snail. In Modern Irish, *gioradán* means 'edible periwinkle'. A mechanical language-internal reconstruction leads to the Primitive Irish preforms **giruddo-* or **geruddo-*. A connection with Indo-European **ĝʰerH-* 'intestine, gut' is semantically not attractive and would leave many morphological questions. The dedication to *Matribus Gerudatiabus* 'the *Gerudatian* Mothers' in Southern Gaul (Aix-en-Provence; CIL 12, 505) does not help as a parallel either. In so far as the name is amenable to analysis, a segmentation into *geru-* and *dati-* seems plausible (Scheungraber 2020: 298–300), which is excluded for the Irish word.

In Stifter (2023: 186), I left it at this and concluded that, because of its structure, *giritán* is probably an isolated substrate borrowing. In the meantime, I have realised that Latin *hirūdō* 'leech' has a very similar structure. This word has no established etymology either (cf. De Vaan 2008: 286), but its mechanical reconstruction **gʰiruHdon-* or **ĝʰiruHdon-* is deceptively similar to that of *giritán*. Both can be united under the non-Indo-European looking reconstruction **gʰirūd-* (in which case Irish would have to show an extra gemination of unclear nature) or

g^hiruzd-,[38] or even **γirūd-*/**γiruzd-*, if we are prepared to leave the path of traditional Indo-European phonology. Leeches and periwinkles are very different species, even though both inhabit the water, but folk taxonomy may have stressed the commonalities between them. This etymon refering to a 'slimy aquatic creature' can therefore be added to the possible substratal borrowings that are exclusive to Celtic and Italic.

4.2. fáecha

Another word for 'sea-snail' or 'periwinkle' is Early Irish *fáecha*, *fáechan*, *fáechóc*, often cited as a chief suspect for a substratal loan. The word is only found in late, i.e. post-Old Irish manuscripts. Therefore the spelling with *-áe-* may just as well represent old *-óe-*, since the two diphthongs had fallen together by the Middle Irish period. The surviving medieval attestations of the word, nom. sg. *fáecha*, nom. pl. *faechain*, prep. pl. *fáechnaib* (eDIL dil.ie/21002), are consistent with an original *n*-stem inflection. On this basis, the word can be mechanically projected back to Primitive Irish **u̯aiki̯Vi̯/u̯on-* or **u̯oiki̯Vi̯/u̯on-*, where *V* is an undefined back vowel. The *n*-stem inflection may be an indication that this formation is of some age within the language. In view of the spiral shell of the animal, it is possible to speculate about an Indo-European root etymology for postulated OIr. **fóechu*. Semantically and formally fitting roots are PIE **u̯ieh$_1$-* 'to wrap, wind' or **u̯iek^u-* 'to wrap'. In both cases one has to operate with *Schwebeablaut*. In the case of the former root, laryngeal loss and a guttural suffix also need to be invoked. Either root would yield an acceptable semantic explanation for the snail, either as 'winder' or as 'wrapping animal'.

The British Celtic languages on the island of Britain have a word that is doubtlessly related, even though the details of the relationship are hard to specify, namely W *gwichiad*, *gwichen*, *gwichyn*, and LCorn. *gwihan*, Corn. *gwigh* 'periwinkle'. A mechanical reconstruction, ignoring the various suffixes that are due to productive word-formation within the British languages, leads to the possible pre-forms **u̯īkko-* (with a geminate) or **u̯īkso-* or perhaps **u̯ī(k)ski̯o-* (with an avatar of the suffix **-sk-*).[39] Ir. *fáechan* has been suggested to be a borrowing from British *gwichan* (e.g., GPC s.v. *gwichiad*; Matasović 2012: 157), but it is difficult to see how this should work. If it were an early loan, OIr. ***fích-* or something similar would be expected; if it were a late loan, ***guich-* or a similar approximation of the Welsh sounds should result. If the loan went in the opposite direction, namely from Irish into British, it is hard to see why the diphthong *áe* or *óe* would

38 For the unclear treatment of **zd* in unstressed syllables in Irish, see Stifter (2009).

39 OIr. *blóesc/bláesc* and W *blisg* 'husk, shell' appear to display the same kind of vowel correspondence.

be reflected by British *i*. It would have to be a late loan when the Irish diphthong had already become the long front monophthong [eː] or [ɪː] in Middle Irish. At the same time, we would have to assume that initial *f-* was etymologically correctly substituted by Welsh *gw-*. A loan relationship between the two branches of Insular Celtic therefore evidently requires a series of additional assumptions that reduce its overall plausibility.

An alternative way of looking at the material is to speculate that British **u̯īk-* stands in a sort of *ablaut* relationship with Goidelic **u̯aiḭk-* or **u̯oiḭk-*. If, for example, the root **u̯ieh₁-* 'to wrap, wind' is involved, the putative Irish form **fóechu* could continue **u̯oiḭh₁kVḭon-* (with loss of the laryngeal), whereas the British word could go back to **u̯ih₁kk(i)o-*, **u̯ih₁kso-* or **u̯ih₁k-skḭo-*. However, this explanation requires that root *ablaut* in nominal derivation was still a productive process in Insular Celtic, which is hard to believe. In the end, no simple conclusion can be offered. Although a measure of good will and imagination allows us to arrive at an Indo-European etymology, the amount of special pleading required at each step results in a solution that looks contrived and implausible, and the word is perhaps better regarded as a candidate for a substratal loan. Given the differences in the details between Irish and British, the borrowing must have occurred independently in the two branches.

4.3. gúgán

The word *gúgán* for another mollusc, 'whelk (*Buccinum undatum*)', is only attested in recent Modern Irish. Biologically, this is a very different animal from periwinkles, belonging to a different order within the subclass of Caenogastropoda. However, a glance into Modern Gaelic dictionaries shows that this and the aforementioned terms were and are being used indiscriminately for all kinds of maritime molluscs. If we take the dictionary form *gúgán* seriously, it can be projected back to pre-Irish **gūggo-*, for which no Indo-European connection suggests itself. The apparent geminate could be regarded as a marker of foreign origin. I say 'apparent', because if the word was borrowed into late prehistoric Irish after lenition had acquired phonemic status, the Irish single voiced stop may simply reflect a single voiced stop in the donor language. This is borne out by loans from the Old British Celtic languages, where a single voiced stop is the regular product of lenition of a single voiceless Proto-Celtic stop. British or rather Welsh loans into Irish that show this treatment are OIr. *brocóit* 'bragget' < OW *bracaut*, W *bragawd* < **mrakāto-* (Bauer 2015: 14–15), OIr. *carrac* 'rock' < W *carreg* < **karrikā-* (Bauer 2015: 17–18), or OIr. *Combrec* 'the Welsh language' < W *Cymraeg* < **kombrogikā-* (Bauer 2015: 25–26). Since these loans presuppose the British lenition of intervocalic voiceless stops of around the late 5th century, the borrowing must have occurred comparatively late in Irish.

Keeping this possibility in mind and treating, for argument's sake, **gūggo-* as a British loan, the internal **gg* could reflect an Old British **g* < **k*. The initial **g-* could represent British **gw-* < Proto-Celtic **u̯-*, unrounded before the following **ū*. The long **ū* itself could be a Gaelic substitution for British **ū̃* < earlier **ū* < Celtic **oi̯*. *Voilà*, what we arrive at is exactly the same stem **u̯oi̯k-* that is one of the potential preforms of genuine Irish *fáecha*! *Gúgán could* reflect an otherwise lost British cognate of Ir. *fáecha*. But all of this could just be coincidence. The earliest evidence for *gúgán* known to me is from 1978. Beside it, we find similar-sounding words for 'bivalve shell-fish' such as *gruán* or *grúgam*, which may be local variants of *gúgán*. The nature of the evidence does not allow a decision about which of these forms is the oldest or 'most original'. In the end, their comparatively basic phonology and simple syllable structure renders them suspect of being sound-symbolic *Urschöpfungen* or neologisms. Instead of being the simplified outcome of an earlier geminate, the *-g-* could be original, since there are no longer any positional restrictions to its occurrence in Modern Irish.

References

Adams, George B. 1980. Place-names from pre-Celtic languages in Ireland and Britain. *Nomina* 4. 46–63.

Allentoft, Morten E. et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522. 167–172.

Ariste, Paul 1971–1972. The earliest substrata in the Baltic-Finnic languages. *Néprajz és Nyelvtudomány* 15. 5–11.

Ariztimuño, Borja & Ander Egurtzegi 2013. Remarks on the etymology of the Basque word for 'swallow' and its potential relation to Celtic. *Ériu* 63. 79–90.

Bauer, Bernhard 2015. *Intra-Celtic loanwords.* PhD-thesis. University of Vienna.

Baumgarten, Rolf (ed.) 1973. Old Irish personal names: M[ichael] A. O'Brien's 'Rhŷs Lecture'-Notes, 1957. *Celtica* 10. 211–236.

Blažek, Václav 2015. Celtic 'deer'. *Études Celtiques* 41. 121–127.

Broderick, George 2013. Indo-European and non Indo-European aspects to the languages and place-names in Britain and Ireland: an overview. In Cathinka Hambro & Lars Ivar Widerøe (eds.), *Lochlann. Festskrift til Jan Erik Rekdal på 60-årsdagen. Aistí in ómós do Jan Erik Rekdal ar a 60ú lá breithe*, 282–314. Oslo: Hermes.

Campanile, Enrico 1976. Indo-European and non-Indo-European elements in the Celtic dialects. *Journal of Indo-European Studies* 4. 131–138.

Cassidy, Lara M. et al. 2016. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proceedings of the National Academy of Sciences of the United States* 113(2). 368–373.

Coates, Richard 1988. *Toponymic topics. Essays on the early toponymy of the British Isles*. Brighton: Younsmere Press.

Cunliffe, Barry 2018. *The Ancient Celts.* 2nd edition. Oxford: Oxford University Press.

Cunliffe, Barry and John T. Koch (eds.) 2016. *Celtic from the West 3. Atlantic Europe in the Metal Ages: Questions of Shared Language.* Oxford: Oxbow Books.

De Bernardo Stempel, Patrizia 2014. Indogermanisch **h₂endʰe(s)-ro-s/-rā/-ro-m*, **neigʷ-sk-ā* und ihre Fortsetzungen im Keltischen und den Nachbarsprachen. *Wékʷos* 1. 47–73.

Delamarre, Xavier 2003. *Dictionnaire de la langue gauloise. Une approche linguistique du vieux-celtique continental*. Paris: Éditions Errance.

Delamarre, Xavier 2007. *Noms de personnes celtiques dans l'épigraphie classique. Nomina Celtica Antiqua Selecta Inscriptionum*. Paris: Éditions Errance.

Delamarre, Xavier 2012. *Noms de lieux celtiques de l'Europe ancienne (-500 / +500). Dictionnaire*. Paris: Éditions Errance.

Delamarre, Xavier 2018. *Dictionnaire de la langue gauloise. Une approche linguistique du vieux-celtique continental*. Paris: éditions errance.

Delamarre, Xavier 2019. *Dictionnaire des thèmes nominaux du gaulois. I. Ab- /Iχs(o)-*. Paris: Les Cent Chemins.

Derksen, Rick 2008. *Etymological dictionary of the Slavic inherited lexicon*. Leiden/Boston: Brill.

De Vaan, Michiel 2008. *Etymological Dictionary of Latin and the other Italic Languages*. Leiden/Boston: Brill.

Dinneen, Pádraig S. 1927. *Foclóir Gaedhilge agus Béarla. An Irish-English Dictionary, being a thesaurus of the words, phrases and idioms of the Modern Irish language. New edition, revised and greatly enlarged*. Dublin: Irish Texts Society.

Dwelly, Edward 1911. *Faclair Gaidhlig gu Beurla le Dealbhan. The Illustrated Gaelic-English Dictionary*. Glasgow: Gairm Publications.

eDIL 2019 = *An Electronic Dictionary of the Irish Language, based on the Contributions to a Dictionary of the Irish Language (Dublin*: *Royal Irish Academy*, *1913–1976)*. http://edil.qub.ac.uk/.

Eska, Joseph F. 2009. The emergence of the Celtic languages. In Martin J. Ball & Nicole Müller (eds.), *The Celtic Languages*, 22–27. 2nd edn. Abingdon/New York: Routledge.

Favereau, Francis 2016. *Geriadur Bras. Brezhoneg-galleg & français-breton*. Montroulez: Skol Vreizh. https://geriadurbrasfavereau.monsite-orange.fr/.

Furtwängler, Anja et al. 2020. Ancient genomes reveal social and genetic structure of Late Neolithic Switzerland. *Nature Communications* 11. no. 1915.

George, Ken (ed.). 2020. *An Gerlyver Meur. Cornish-English, English-Cornish Dictionary*. 3rd edn. Llandysul: Gomer Press.

Gorrochategui, Joaquín 1995. The Basque language and its neighbors in antiquity. In José Ignacio Hualde, Joseba A. Lakarra & Robert L. Trask (eds.), *Towards a History of the Basque Language*, 31–63. Amsterdam/Philadelphia: John Benjamins.

Gorrochategui, Joaquín 1996. Miscellanea Iberica. *Zeitschrift für celtische Philologie* 48. 1–16.

GPC = *Geiriadur Prifysgol Cymru (GPC) Online*. Aberystwyth: University of Wales Centre for Advanced Welsh and Celtic Studies 2014. https://welsh-dictionary.ac.uk/gpc/gpc.html.

Greene, David 1982. Varia I. *Ériu* 33. 161–164.

Haak, Wolfgang et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522. 207–211.

Hamp, Eric P. 1990. The pre-Indo-European language of Northern (Central) Europe. In Tom L. Markey & John A.C. Greppin (eds.), *When Worlds Collide. The Indo-Europeans and the Pre-Indo-Europeans. The Rockefeller Foundation's Bellagio Study and Conference Center Lake Como, Italy*, *February 8–13*, *1988*, 291–309. Ann Arbor: Karoma Publishers.

Hayden, Deborah & David Stifter 2022. Ogam and Trees. *OG(H)AM Blog Post* December 2022. https://ogham.glasgow.ac.uk/index.php/2022/12/20/ogam-and-trees/.

Hewitt, Steve 2007. Remarks on the Insular Celtic/Hamito-Semitic question. In Raimund Karl & David Stifter (eds.), *The Celtic World. Critical Concepts in Historical Studies. Volume IV. Celtic Linguistics*, 230–286. London/New York: Routledge.

Hogan, Edward 1910. *Onomasticon Goidelicum Locorum et Tribuum Hiberniae et Scotiae*. Dublin: Hodges & Figgis.

Huld, Martin E. 1990. The linguistic typology of the Old European substrata in North Central Europe. *Journal of Indo-European Studies* 18(3–4). 389–423.

Isaac, Graham R. 2003. Varia: I. Some Old Irish etymologies, and some conclusions drawn from them. *Ériu* 53. 151–155.

Isaac, Graham 2007. Celtic and Afro-Asiatic. In Hildegard L.C. Tristram (ed.), *The Celtic Languages in Contact. Papers from the Workshop within the Framework of the XIII International Congress of Celtic Studies*, *Bonn*, *26–27 July 2007*, 25–80. Potsdam University Press.

Iversen, Rune & Guus Kroonen 2017. Talking Neolithic: Linguistic and archaeological perspectives on how Indo-European was implemented in Southern Scandinavia. *American Journal of Archaeology* 121(4). 511–525.

Jasanoff, Jay 1978. Observations on the Germanic Verschärfung. *Münchener Studien zur Sprachwissenschaft* 37. 77–90.

Jasanoff, Jay H. 2017. The Old Irish *f*-future. In Bjarne Simmelkjær Sandgaard Hansen, Adam Hyllested, Anders Richardt Jørgensen, Guus Kroonen, Jenny Helena Larsson, Benedicte Nielsen Whitehead, Thomas Olander & Tobias Mosbæk Søborg (eds.), *Usque ad Radices. Indo-European studies in honour of Birgit Anette Olsen*, 325–337. Copenhagen: Museum Tusculanum Press.

Jordán Cólera, Carlos 2019. *Lengua y Epigrafía Celtibéricas*. 2 vols. (Monografías de Filología Griega 29). Prensas de la Universidad Zaragoza.

Koch, John T. 2020. *Celto-Germanic. Later Prehistory and Post-Proto-Indo-European vocabulary in the North and West*, Aberystwyth: University of Wales. URL: https://www.wales.ac.uk/Resources/Documents/Centre/2020/Celto-Germanic2020.pdf.

Kroonen, Guus 2012. Non-Indo-European root Nouns in Germanic: Evidence in support of the agricultural substrate Hypothesis. In Riho Grünthal & Petri Kallio (eds.), *A Linguistic map of prehistoric Northern Europe*, 239–260. Helsinki: Société Finno-Ougrienne.

Kroonen, Guus 2013. *Etymological dictionary of Proto-Germanic* (Leiden Indo-European Etymological Dictionary Series 11). Leiden/Boston: Brill.

Kuiper, Franciscus B.J. 1956. The etymology of ἄνθρωπος. In Hermann Kronasser (ed.), ΜΝΗΜΗΣ ΧΑΡΙΝ. *Gedenkschrift Paul Kretschmer 2*. *Mai 1866 – 9. März 1956*. 2 Bde, 211–226. Wien: Wiener Sprachgesellschaft.

Kuiper, Franciscus B.J. 1995. Gothic *bagms* and Old Icelandic *ylgr*. *North-Western European Language Evolution* 25. 63–88.

Lazaridis, Iosif et al. 2013. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513. 409–413.

LEIA = Joseph Vendryes, Édouard Bachellery & Pierre-Yves Lambert, 1959–. *Lexique étymologique de l'irlandais ancien*. Dublin – Paris: DIAS – CNRS [so far: A, B, C, D, M N O P, R S, T U].

LIV = Martin Kümmel & Helmut Rix, 2001. *LIV. Lexikon der indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen. Zweite, erweiterte und verbesserte Auflage*. Wiesbaden: Reichert.

Mac Eoin, Gearóid 2007. What language was spoken in Ireland before Irish? In Hildegard L.C. Tristram (ed.), *The Celtic Languages in Contact. Papers from the Workshop within the Framework of the XIII International Congress of Celtic Studies*, *Bonn*, *26–27 July 2007*, 113–125. Potsdam: Potsdam University Press.

Mallory, James P. 2013. The Indo-Europeanization of Atlantic Europe. In John T. Koch & Barry Cunliffe (eds.), *Celtic From the West 2. Rethinking the Bronze Age and the Arrival of Indo–European in Atlantic Europe*, 17–40. Oxford: Oxbow Books.

Marstrander, Carl 1908. Über irisches *loscann* und einige andere indogermanische Namen der Kröte. In Magnus Bernhard Olsen & Sophus Bugge, *Sproglige og historiske afhandlinger viede Sophus Bugges minde*, 240–246. Kristiania: H. Aschehoug & Co.

Matasović, Ranko 2009. *Etymological Dictionary of Proto-Celtic*. Leiden/Boston: Brill.

Matasović, Ranko 2012. The substratum in Insular Celtic. *Journal of Language Relationship* 8. 153–159, 165–168 (response to Tatyana Mikhailova).

Matasović, Ranko 2020. Language of the bird names and the pre-Indo-European substratum. In Romain Garnier (ed.), *Loanwords and Substrata. Proceedings of the Colloquium Held in Limoges (5th–7th June, 2018)*, 331–344. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

McCone, Kim 1996. *Towards a relative chronology of Ancient and Medieval Celtic sound change*. Maynooth: Department of Old and Middle Irish, St. Patrick's College, Maynooth.

McCone, Kim 2005. Mögliche nicht-indogermanische Elemente in den keltischen Sprachen und einige frühe Entlehnungen aus indogermanischen Nachbarsprachen. In Gerhard Meiser & Olav Hackstein (eds.), *Sprachkontakt und Sprachwandel. Akten der XI. Fachtagung der Indogermanischen Gesellschaft, 17.–23. September 2000, Halle an der Saale*, 395–435. Wiesbaden: Reichert.

McManus, Damian 1988. Irish letter-names and their kennings. *Ériu* 39. 127–168.

Menard, Martial 2021. *Devri – Le dictionnaire diachronique du breton*. Roazhon. URL: http://devri.bzh/.

Meyer, Kuno 1891. Loanwords in Early Irish. *Revue Celtique* 12. 460–469.

O'Brien, Michael A. 1962. *Corpus Genealogiarum Hiberniae*. Dublin: Dublin Institute for Advanced Studies.

Ó Corráin, Donnchadh 2017. *Onomasticon Goedelicum*. By Edmund Hogan, revised and corrected by Donnchadh Ó Corráin. https://www.dias.ie/celt/celt-publications-2/onomasticon-goedelicum/.

Olalde, Iñigo et al. 2018. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555. 190–196.

Ó Murchadha, Diarmuid 1997. *The annals of Tigernach. Index of names* (Irish Texts Society Subsidiary Series 6). London: Irish Texts Society.

O'Rahilly, Thomas F. 1946. Ir. *aobh*, *aoibheall*, etc. W. *ufel*, *uwel*. Gaul. *ēsus*. *Ériu* 14. 1–6.

Ó Riain, Pádraig 1985. *Corpus Genealogiarum Sanctorum Hiberniae*. Dublin: Dublin Institute for Advanced Studies.

Ó Riain, Pádraig, Kevin Murray, Emma Nic Cárthaigh & Diarmuid Ó Murchadha (eds.), 2003–. *Historical dictionary of Gaelic placenames. Foclóir stairiúil áitainmneacha na Gaeilge*. Vols. *A*, *B*, *C–Ceall Fhursa*, *Ceall Ghabhann–Cláiríne*, *Clais–Cnucha*, *Cóbh–Cutloch*, *D–Drongán*. London: Irish Texts Society.

Patterson, Nick et al. 2022. Large-scale migration into Britain during the Middle to Late Bronze Age. *Nature* 601. 588–594.

Pedersen, Holger 1909. *Vergleichende Grammatik der keltischen Sprachen.* 1. Band. *Einleitung und Lautlehre*. Göttingen: Vandenhoeck & Ruprecht.

Pedersen, Holger 1950. Old Irish *ainder*, 'a young woman'. *Journal of Celtic Studies* 1. 4–6.

Prósper, Blanca María 2002. *Lenguas y religiones prerromanas del occidente de la Península Ibérica*. Salamanca: Ediciones Universidad de Salamanca.

Prósper, Blanca María 2008. Lusitanian. A non-Celtic Indo-European language of Western Hispania. In Juan Luis García Alonso (ed.), *Celtic and other languages in ancient Europe*, 53–64. Salamanca: Ediciones Universidad Salamanca.

Rasmussen, Jens Elmegård 1990. Germanic *Verschärfung*: tying up loose ends. In Henning Andersen & Konrad Koerner (eds.), *Historical Linguistics 1987. Papers from the 8th International Conference on Historical Linguistics (Lille, 31 August – 4 September 1987)*, 425–441. Amsterdam/Philadelphia: John Benjamin.

Salmons, Joe 1992. Northwest Indo-European vocabulary and substrate phonology. In Roger Pearson (ed.), *Perspectives on Indo-European Language, Culture and Religion. Studies in honor of Edgar C. Polomé*. Vol. 2, 265–279. McLean, VA: Institute for the Study of Man.

Scheungraber, Corinna 2020. *Altgermanische und altkeltische Theonyme: die epigraphische Evidenz aus der Kontaktzone. Ein Handbuch zu ihrer Etymologie*. Innsbruck: Institut für Sprachwissenschaft.

Schrijver, Peter 1997. Animal, vegetable and mineral: some Western European substratum words. In Alexander Lubotsky (ed.), *Sound Law and Analogy. Papers in Honor of Robert S. P. Beekes on the occasion of his 60th birthday*, 293–316. Amsterdam: Rodopi.

Schrijver, Peter 2000. Varia V. Non-Indo-European surviving in Ireland in the first millenium A.D. *Ériu* 51. 195–199.

Schrijver, Peter 2001. Lost languages in northern Europe. In Christian Carpelan, Asko Parpola & Petteri Koskikallio (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations. Papers presented at an international symposium held at the Tvärminne Research Station of the University of Helsinki. 8–10 January, 1999*, 417–425. Helsinki: Suomalais-Ugrilainen Seura.

Schrijver, Peter 2002. Irish *ainder*, Welsh *anner*, Breton *annoar*, Basque *andere*. In David Restle & Dietmar Zaefferer (eds.), *Sounds and Systems: Studies in Structure and Change. A Festschrift for Theo Vennemann*, 205–219. Berlin/New York: Mouton de Gruyter.

Schrijver, Peter 2003. Early Developments of the vowel systems of North-West Germanic and Saami. In Alfred Bammesberger & Theo Vennemann (eds.), *Languages in Prehistoric Europe*, 195–226. Heidelberg: Universitätsverlag Winter.

Schrijver, Peter 2005. Varia I. More on non-Indo-European surviving in Ireland in the first millennium A.D. *Ériu* 55. 137–144.

Schrijver, Peter 2011. La langue hattique et sa pertinence possible pour les contacts linguistiques préhistoriques en Europe occidentale. In Coline Ruiz Darasse & Eugénio R. Luján (eds.), *Contacts linguistiques dans l'occident médiiterranéen antique*, 241–255. Madrid: Casa de Velázquez.

Schrijver, Peter 2014. *Language Contact and the Origins of the Germanic Languages*. New York – London: Routledge.

Schrijver, Peter 2015. Pruners and trainers of the Celtic family tree: the rise and development of Celtic in the light of language contact. In Liam Breatnach, Ruairí Ó hUiginn, Damian McManus & Katharine Simms (eds.), *Proceedings XIV International Congress of Celtic Studies Maynooth 2011*, 191–219. Dublin: Dublin Institute for Advanced Studies.

Schrijver, Peter 2016. Sound change, the Italo-Celtic linguistic unity, and the Italian homeland of Celtic. In John T. Koch & Barry Cunliffe (eds.), *Celtic from the West 3. Atlantic Europe in the Metal Ages: Questions of Shared Language*, 489–502. Oxford: Oxbow Books.

Schrijver, Peter 2018. Talking Neolithic: the case for Hatto-Minoan and its relationship to Sumerian. In Guus Kroonen, James P. Mallory & Bernard Comrie (eds.), *Talking Neolithic: Proceedings of the workshop on Indo-European origins held at the Max Planck Institute for Evolutionary Anthropology, Leipzig, December 2–3, 2013*, 336–374. Washington, DC: Institute for the Study of Man.

Sims-Williams, Patrick 2020. An alternative to 'Celtic from the East' and 'Celtic from the West'. *Cambridge Archaeological Journal* 30(3). 511–529.

Stifter, David 2009. A rule for *z*-deletion in Irish? *Ériu* 59. 159–164.

Stifter, David 2010. The invisible third: the Basque and Celtic words for 'swallow'. *Ériu* 60. 145–157.

Stifter, David 2019a. An apple a day. . . *Indogermanische Forschungen* 124. 171–218.
Stifter, David 2019b. Further to Avidic: 'geese' in Insular Celtic. *Historische Sprachforschung* 132. 312–316.
Stifter, David 2020. Old Irish etymology through the ages. *Language and History* 63/1. 24–46.
Stifter, David 2023. With the back to the ocean. The Celtic maritime vocabulary. In Kristian Kristiansen, Guus Kroonen & Eske Willerslev (eds.), *The Indo-European Puzzle Revisited. Integrating Archaeology, Genetics, and Linguistics*, 172–192. Cambridge: Cambridge University Press.
Stifter, David 2023b. The rise of gemination in Celtic. *Open Research Europe. Linguistic Diversity collection* 2 2.2023. doi 10.12688/openreseurope.15400.1.
Van Sluis, Paulus, Anders Richardt Jørgensen, Guus Kroonen 2023. European prehistory between Celtic and Germanic: the Celto-Germanic isoglosses revisited. In Kristian Kristiansen, Guus Kroonen & Eske Willerslev (eds.), *The Indo-European Puzzle Revisited. Integrating Archaeology, Genetics, and Linguistics*, 192–243. Cambridge: Cambridge University Press.
Villanueva Svensson, Miguel 2015. The development of **oū* in Baltic. *Indogermanische Forschungen* 120/1. 313–328.
Weiss, Michael 2012. Italo-Celtica: Linguistic and cultural points of contact between Italic and Celtic. In Stephanie W. Jamison, H. Craig Melchert & Brent Vine (eds.), *Proceedings of the 23rd Annual UCLA Indo-European Conference*, 151–173. Bremen: Hempen.
Weiss, Michael 2022. Italo-Celtic. In Thomas Olander (ed.), *The Indo-European Language Family. A Phylogenetic Perspective*, 102–113. Cambridge University Press.
White, Nora 2013. *Ogham in 3D*. Dublin Institute for Advanced Studies. https://ogham.celt.dias.ie.
Witczak, Krzysztof T. 2005. *Język i Religia Luzytanów. Studium historyczno-porównawcze*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
Wodtko, Dagmar 2017. *Lusitanian. Language | Writing | Epigraphy* (AELAW Booklet 4). Zaragoza: Prensas de la Universidad de Zaragoza.
Zair, Nicholas 2018. The shared features of Italic and Celtic. In Jared S. Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics. An International Handbook*, 2030–2037. Berlin/New York: De Gruyter Mouton.
Ziegler, Sabine 1994. *Die Sprache der altirischen Ogaminschriften*. Göttingen: Vandenhoeck & Ruprecht.
Zimmer, Stefan 2020. Celtic, Germanic and HARIGASTI TEIWA;. In Thomas L. Markey & Luka Repanšek (eds.), *Revisiting Dispersions. Celtic and Germanic ca. 400 BC – ca. 400 AD. Proceedings of the International Interdisciplinary Conference held at Dolenjski muzej, Novo mesto, Slovenia. October 12–14, 2018*, 48–80. Washington, D.C.: Institute for the Study of Man.

Part IV: **The Mediterranean**

Andrew Wigman

7 A European substrate velar “suffix”

1 Introduction

Numerous Latin words end in *-ex, -icis* and *-ix, -icis*. It has long been noticed that many of the words with this particular synchronic *k*-suffix 1) are without a good Indo-European etymology and 2) fall into some specific semantic categories (cf. Ernout 1946: 133–63, Leumann 1977: 375–6, Weiss 2020: 326–7). A neutral examination of the corpus shows they contain animals (*dentex* ‘porgy (fish)’, *fulix*/*fulica* ‘coot’, *ībex* ‘mountain goat’, *laurex* ‘rabbit fetus’, *murex* ‘purple-fish (mollusk)’, *natrix* ‘serpent’, *sōrex* ‘shrew’), biting insects (*cīmex* ‘bedbug’, *culex* ‘gnat’, *pūlex* ‘flea’), trees (*īlex* ‘holm-oak’, *larix* ‘larch’, *salix* ‘willow’, *vitex* ‘chaste-tree’), plants and plant parts (*cārex* ‘reed, sedge’, *caudex* ‘trunk, stem’, *cortex* ‘bark’, *felix*/*filix* ‘fern’, *frutex* ‘shrub, bush’, *rumex* ‘sorrel’, *ūlex* ‘heather’), and body parts (*famex* ‘bruise’, *panticēs* ‘guts’, *pōdex* ‘anus’, *pollex* ‘thumb’, *rāmicēs* ‘lungs; hernia’, *varix* ‘varicose vein’). But there are also materials (*pūmex* ‘pumice’, *silex* ‘flint’), people (*paelex* ‘concubine’, *rupex* ‘boor, clown’, *senex*[1] ‘old (man)’), and others (*apex* ‘summit, top’, *calix* ‘goblet’, *fornix* ‘vault’, (*h*)*irpex* ‘harrow’, *imbrex* ‘roof tile’, *latex* ‘fluid, liquid’, *vertex*/*vortex* ‘whirl’). Some bases on which this morphological feature occurs are certainly inherited and several others can more or less plausibly be linked to an Indo-European root (e.g. Ernout 1946, Martinet 1955, Schrijver 1991: 148–54, Olsen 2009 esp. on *senex*; Ettmayer 1926: 23, Ernout 1946, Specht 1947: 40–1, Weiss 2020: 326–7 on others like *cortex* ‘bark’, *dentex* ‘porgy’, *pūlex* ‘flea’, *pōdex* ‘anus’, *vertex*/*vortex* ‘whirl’; Matasović 2016 on some plant and tree names). But many are suspected of being pre-Latin loans (Ettmayer 1926: 23, Terracini 1929: 212–14, Bertoldi 1937: 157, Gerola 1942: 364, Alessio 1944a: 104, Hubschmid 1953: 84, Leumann 1977: 375, de Vaan 2008: 299).

At the same time, numerous Greek words ending in -αξ, -ακος are also etymologically obscure and pertain to semantic categories similar to the Latin words in *-ex*/*-ix* (Chantraine 1933: 376–83, Ernout 1946: 140–1, Beekes 2014: 31)[2]: animals (ἀσπάλαξ/σπάλαξ ‘blind-rat’, κόραξ ‘raven, crow’, δέλφαξ ‘sow’, μύαξ ‘purple-fish’, πόρταξ ‘calf’, σκολόπαξ ‘woodcock’, σκύλαξ ‘puppy’, ὕραξ ‘shrew’), trees (σμῖλαξ ‘holm-oak’,

1 An unusual case, since the oblique forms lack the velar.

2 A less strict approach by the same authors also considers thematic forms, several of which fall into the same categories: plants (αἴσακος ‘branch of myrtle or laurel’, πιστάκη ‘pistachio tree’, ἀμάρακον ‘marjoram’), animals (πίθηκος ‘ape’, ψίττακος ‘parrot’, ἐρίθακος ‘robin’), minerals (σανδαράκη ‘sandarac/realgar’), people (παλλακή ‘concubine’, cf. Lat. *paelex*).

https://doi.org/10.1515/9783111337920-007

στύραξ ‘storax tree’), plants (δόναξ ‘pole-reed’, ὄμφαξ ‘unripe grape’, θρίδαξ ‘lettuce’), body parts (μάσταξ ‘mouth, jaws’, μύσταξ/βύσταξ ‘upper lip, moustache’), as well as minerals/materials (ἄνθραξ ‘charcoal’, βῶλαξ ‘clod, dirt’, κλῶμαξ ‘heap of stones’), people (ἄναξ ‘lord, king’, κόλαξ ‘flatterer’, μεῖραξ ‘youth’, τίταξ ‘ruler’, φύλαξ ‘guard’), and others (θώραξ ‘cuirass’, κάμαξ ‘pole, shaft’, κόρδαξ ‘cordax dance’, λάρναξ ‘chest, coffin’, μύλαξ ‘millstone’, πῖδαξ ‘fountain, spring’, πίναξ ‘plank, board’, πύνδαξ ‘bottom of a vessel’, σχίδαξ ‘split wood’). Some, like σχίδαξ (to σχίζω ‘to split’) and μεῖραξ (cf. Skt. *márya-* ‘young man’) clearly contain inherited bases. But many others are suspected of having originated in a non-IE pre-Greek language (Nehring 1925, Chantraine 1933: 376–83, Beekes 2014: 32, 44).

2 Finding positive evidence for a substrate origin

Beyond the unetymologizability of (many of) these words, the main argument in favor of their being of substrate origin is that they belong to semantic fields that may be subject to borrowing. But a lack of a good Indo-European root etymology and belonging to a certain semantic field are only circumstantial arguments when it comes to the substrate origin of a word (cf. Polomé 1989: 54–5, Salmons 1992: 267, Schrijver 1997: 295), let alone a suffix. It is well known that a suffix **-k* is reconstructible for Indo-European (in many forms and functions, cf. Brugmann 1906: 472–506), and Chantraine (1933: 278–9) demonstrates that the suffix -αξ is relatively productive in Greek. He calls it *chimérique* to try to distinguish between its primary and secondary occurrences, which hints at a valid caveat: an inherited suffix can be added to a non-IE root just as easily as a non-IE suffix can be added to an inherited root. In fact, Nehring (1925: 153) suggested “neben diesem indogermanischen Suffix [-αξ] läßt sich ein gleichlautendes nichtindogermanisches Suffix nachweisen, das mit weiteren nichtindogermanischen *k*-Suffixen verwandt ist.”

To truly be able to say that a piece of morphology is of non-IE origin, there must be something about that morphology itself that proves this. Nehring (1925: 173–4) did try to establish the -αξ suffix functions differently than the inherited *k*-suffix—Indo-European morphology functioning in non-Indo-European ways may therefore not be Indo-European morphology after all. But there is even stronger evidence.

The strongest indication that an element, be it a lexical item or a derivational morpheme, is of non-IE origin (in the absence of attestation in a non-IE language) is when it is shared by two or more daughter languages but cannot be reconstructed through known sound laws to one unitary pre-form. In the case of sub-

strate borrowings, these words entered the Indo-European daughter languages after the dissolution of the parent language (PIE). Due either to pre-existing dialectal variation within the substrate language(s) and/or to the process of nativizing foreign phonemes, the substrate elements entered the separate daughter languages in different forms. Thus, their reconstruction results in what looks like irregular correspondences from a PIE perspective.

An illustrative example is Lat. *ervum* ‘bitter vetch’. It is a good semantic and formal match for Gr. ὄροβος ‘bitter vetch’, which itself cannot be separated from Gr. ἐρέβινθος ‘chickpea’. The latter has the *-inthos* suffix that had been celebrated since at least Kretschmer (1896: 401–9) as the Pre-Greek suffix *par excellence*. But the most important part of the comparison is that Greek, along with Arm. *aṙowoyt* ‘alfalfa’ (Thorsø, this volume) require a reconstruction with $^{*}b/^{*}b^{h}$ whereas Latin, along with Proto-Germanic **arwīt-* ‘pea’ (cf. Kroonen 2013: 37) require $^{*}\underset{\wedge}{u}$. There is no way to achieve a Proto-Indo-European reconstruction for these forms that does not entail an unprecedented $^{*}b/\underset{\wedge}{u}$ alternation. Thus, even without considering the fact that this lexeme belongs to the semantic category of agriculture, we can already be quite certain that it was borrowed from a non-Indo-European language.

There are a number of cases in which an irregular correspondence occurs between daughter languages’ reflexes of the *k*-suffix, suggesting that, at least in these cases, it is not the inherited PIE *k*-suffix, but rather one that existed amongst the pre-Indo-European substrate language(s) of Europe.

3 Three irregular Italo-Germanic correspondences

(1) ‘fern’

Lat. *filix*, *felix* ‘fern, bracken’	< PIt. **felik-*	as if < PIE $^{*}b^{h}$*elik-*
Gr. βλῆχνον, βλῆχρον ‘male fern’	< PGr. **blēk^{h}-n/r-*	as if < PIE **blēg^{h}-n/r-*
Da. *bregne*, etc. ‘fern, bracken’	< PGm. **brekna*(*n*)-	as if < PIE $^{*}b^{h}$*reg-n-*

In Latin, *felix* can be considered the original, with *filix* having arisen via assimilation from the *i* of the next syllable (cf. Leumann 1977: 101). It is often compared to reflexes of PIE $^{*}b^{h}$*el-* ‘henbane’ like G *Bilsenkraut* and Ru. *belená* (Petr 1896: 209, Pokorny 1959: 120, Schrijver 1999: 37–8, de Vaan 2008: 220), but I am not convinced that henbane is superficially similar enough or has similar enough medicinal properties to ferns to justify a comparison. Better comparanda are the Greek and Germanic words that have the exact same meaning. The alternation in aspi-

ration between Greek and Latin/Germanic occurs elsewhere (cf. Gr. βάσκανος 'sorcerer, slanderer' against Lat. *fascinus* 'charm, spell, amulet'; Hesychian βάσκιοι 'bundle' against Lat. *fascis* 'id.') as does the *l*/*r* alternation to which the Germanic form attests (cf. Gr. λείριον 'lily' against Lat. *līlium* 'id.'). The Greek and Germanic forms have an additional suffix, which may have contributed to the alternation in vocalization of the first syllable. If this comparison is accepted, then it is important to note that the reconstructed **k* of Latin *felix* corresponds to a voiced **g* for the Germanic forms. It is conceivable that an original **g* was devoiced in the Latin nominative singular and later leveled to the oblique forms, but this is not regular (cf. *rēx*, *rēgis* 'king').

The Greek comparanda seem to reconstruct to **gʰ* rather than **k*. Beekes (2014: 37) follows Furnée (1972: 132, fn. 64, 65) in noticing cases of velar aspiration before the *n*-suffix, indicating that it is not an inherited **n*. Not all examples are as good as others. For example, Gr. ἀράχνη 'spider' on comparison with Lat. *arāneus* 'id.' probably owes its χ to a suffix **-snā*. The comparison between Gr. δαῦκος 'umbelliferous plant' and Gr. δαύχνα 'laurel tree' (cf. Beekes 2010: 306) is semantically dubious. But several other cases are possible: Gr. κύλιξ 'drinking cup' against Gr. κυλίχνη 'small cup', Gr. πέλιξ 'bowl' against Gr. πελλίχνη 'id.'. Perhaps the *n*-suffix has aspirated what would otherwise have been a *k*-suffix; but as the aspiration also occurs in the variant βλῆχρον with an *r*-suffix, this seems uncertain.[3]

(2) 'shrew'

Lat. *sōrex* 'shrew'	< PItal. **sōrVk-*	as if < PIE **sōrVk-*
Gr. ὕραξ 'shrew'	< PGr. **surak-*	as if < PIE **surak-*
OSw. *surk* 'mole, vole, shrew'	< PGm. **s(w)ur(V)ka-*	as if < PIE **sur(V)g-*

The traditional explanation of Lat. *sōrex* and Gr. ὕραξ is a connection with a PIE root **suer-* 'to resound' (Walde & Hofmann 1938 II: 563, Pokorny 1959: 1049–50), based on reports of singing shrews (for example in Pliny *Nat.Hist*. 8.82 where they interrupt the auspices). Vine (1999: 572–3) treats it as an example of Cowgill's Law in the environment of **(-)Tu̯oR-* > **(-)Tu̯uR-*. Thus one can reconstruct for the Latin and Greek **ō* ~ **o* ablaut rather than a non-IE **ō* ~ **u* alternation. For the Germanic forms,[4] there are phonological problems with assuming a loan from

3 In fact, another interpretation is possible. Perhaps the *n*-suffix was added to a base ending in **-ks*. Lat. *arāneus* and Gr. ἀράχνη would represent *n*-suffixed **araks-* while Lat. *avēna* 'oats' < **awe(C)sna-* might represent **aweks-na*. Its **-ek* corresponds irregularly with the **-iǵ(ʰ)-* reconstructed for PBalt. **aviža?-* 'oats' and the **-iḱ-* reconstructed for PSl. **ovьsъ* 'id.' (cf. most recently Kroonen et al. 2022).

4 Masc. *surker*, neut. *surk*; Sw. *sork*. Also found in ODa. *syrcha mych* 'rat excrement?'.

Lat. *sōrex* (Kock 1909: 84, Hellquist 1922: 827). More likely then, it is an independent form, and if from **suer-*, then from a zero-grade. But an inherited origin leaves the voicing alternation of the velars unexplained. Thus we have another example where a **k* reconstructed for the Latin velar suffix corresponds to a voiced **g* for Germanic. It is theoretically possible that the root is composed of inherited material, but this seems unlikely because the velar element does not appear to be functioning as a suffix. Instead, it looks like the lexeme entered Germanic with the velar element attached.[5]

(3) ‘coot’

Lat. *fulica* ‘coot’	< PIt. **fulik-*	as if < PIE **bʰulik-*
ScG *bolachdan* ‘coot’	< PC **bo/ula/okk-*	as if < PIE **b(ʰ)o/ula/oK-*
OHG *belihha* ‘coot’	< PGm. **balikōn-*	as if < PIE **bʰa/olig-*

Lat. *fulica*, also occurring as *fulix*, is traditionally linked to **bʰel-* ‘shining, white’, allowing a connection with Gr. φαληρίς ‘coot’, Hesychian φαλός· λευκός (Pokorny 1959: 118–20, Beekes 2010: 1550). But this relies on the interpretation of a dialectal development of *u* < **bʰol-* (suggested by Walde & Hofmann 1938 I: 559, correctly labeled *ad hoc* by de Vaan 2008: 248). The obvious comparison with OHG *belihha* is in fact not regular, as it is another case in which a reconstructed Lat. **k* corresponds to a voiced **g* for Germanic. Van Sluis (this volume) has identified a comparandum in ScG *bolachdan*. In the region where the word is attested, the reflexes of OIr. *-cht* and *-cc* merge into [xg], thus behind *bolachdan* is PC **bo/ula/okk-agno-* which reconstructs to a non-IE geminate **kk*. This is a third example of a correspondence between Latin **k* and Germanic **g*.[6]

5 This demonstrates that at least some of the Latin words in *-ex, -icis* could originally have been from **-ax* with regularly expected vowel weakening (cf. Leumann 1977: 375). Weiss (2020: 326) even writes that, due to the possibility of remodeling on the basis of nouns like *artifex*, *artificis*, “it is not possible to draw any firm conclusions about what the pre-weakened vowel(s) of *-ek-* was (were).” Theoretically, given the possibility of analogy in both directions, it seems possible that some of the words in *-ix, -icis* could originally have had the same vocalism as well.

6 A reviewer has suggested that such apparent irregularities in voicing in Germanic could be the result of the rise, simplification, and analogical spread of stop variants triggered by Kluge’s Law (cf. esp. Kroonen 2011: 41–84). This cannot be ruled out per se, but I believe it is unlikely. In PGm. **brekna(n)-*, the nasal that would have triggered Kluge’s Law is still present, and the Greek accentuation demonstrates that this is an environment in which Kluge’s Law would not have operated. For PGm. **s(w)ur(V)ka-* there is no indication that the word was an *n*-stem, even within Germanic. In PGm. **balikōn-*, the vowel is not of a length that should have resulted in the shortening of a Kluge’s Law geminate. Additionally, there are other consonantal and vocalic alternations between the words that an IE explanation like Kluge’s Law does not do away with.

4 Discussion of suffix status

The preceding discussion identified three cases in which a Latin word with an *-ex* or *-ix* suffix corresponds to Germanic forms where the velar element reconstructs to a voiced **g*. But this is not the case for all examples of the *-ix*/*-ex* suffix in Latin that have Germanic comparanda. One such example in which this does not occur is Lat. *salix* 'willow'. It too is suspected to be of substrate origin (de Vaan 2008: 536, Matasović 2009: 319, Kroonen 2013: 424), but its correspondence with PC **salik-*, PGm. **salihōn-*, and (potentially) PGr. **helikā-* can be reconstructed as different ablaut grades of a root of the shape **selH-ik-*. In any case, none of the attested comparanda demonstrate an irregular correspondence in the reconstructed **-ik-* suffix.

This leads to an important question. Leumann (1977: 375) writes about the Latin words with a *k*-suffix, after a short list of a few clearly derivational cases, quite concisely, "sonst unableitbar, daher Suffixfunktion unbestimmbar." Because the roots are mostly otherwise unknown in Latin, the function of the *-ex*/*-ix* suffix is difficult to determine. However, given that consideration, we must also admit that its function *as* a suffix is difficult to determine. Since we do not know the identity of the European substrate language(s), can we be certain that the *-ex*/*-ix* suffix with irregular correspondences identified here is actually a suffix? Or is this an example of projecting our understanding of how Proto-Indo-European functioned onto the substrate language?

There may in fact be a few indications against the interpretation of a suffix. In the example of (1) 'fern', Latin *felix* can reconstruct to **bʰelik-*, but the Greek and Germanic forms had an additional *n*-suffix attached. Perhaps this means that the velar element, rather than being a suffix itself, was part of the root. A combination of suffixes, albeit less peculiar, would also have to be proposed for (3) 'coot', where Scottish Gaelic has added a diminutive suffix and Germanic has turned it into an *n*-stem. Thus, perhaps rather than a suffix, this is a case of a substrate phoneme whose reflex appears as PIt. **k*, PGm. **k*, PGr. **k* (and **kʰ*), and PCelt. **kk*. But if this is a morphological feature of a foreign language, then it seems equally likely that some cases of a substrate root+suffix would have been interpreted by some Indo-European daughter languages as a root to which other derivational morphology could be added whereas in others its status as a root+suffix was preserved. In any case, the correspondence of the velars amongst the daughter languages seems quite clearly to demonstrate that this item was borrowed into the separate daughter languages after the dissolution of PIE. The correspondences throughout the

daughter languages, and *not* upon their further reconstruction into PIE, are of an unvoiced, sometimes geminate and perhaps sometimes aspirated, **k*.[7]

5 Conclusion

In the end, it seems Chantraine was correct in calling this phenomenon *chimérique*. The subset of the Latin lexicon consisting of words ending in the synchronic velar suffix *-ix*/*-ex* is from (at least) two sources.[8] At least three of the words (*felix*, *sōrex*, and *fulica*) are from a substrate language of Europe where the PGm. reflex of the velar is **k* (as if < PIE **g*). One, *salix*, is demonstrably from a different source in which the PGm. reflex of the velar is **h* (as if < PIE **k*) and for which there are no indications beyond its limited distribution and semantics that it is of substrate origin. If it is inherited, it may contain the same suffix as words like *senex* ‘old’ whose root is widely attested in other branches.

Thematic velar suffixes and athematic velar suffixes with a long vowel have good parallels in other Indo-European branches and have relatively well-understood sources. Cf. **-ḱo-* in Lat. *iuvencus* ‘young bull/heifer’, the same formation as Skt. *yuvaśáḥ* ‘young’, Go. *juggs* ‘young’, OIr. *óac* ‘hero’ (Matasović 2009: 436, Kroonen 2013: 274, Weiss 2020: 215); **-ko-* in Lat. *ēsca* ‘food’ < **ed-s-ka*, the same formation as Lith. *ėskà* ‘fodder’ (de Vaan 2008: 185, Weiss 2020: 314). The latter suffix added to an *i*-stem base may have resulted in the suffix **-iko-* which results in the reflexes Lat. *-icus*, Gr. -ικός, PGm. **-iga-* (cf. Weiss 2020: 314; on the thematics, see further Fruyt 1986). Latin athematic *-ōx* is from **h₃okʷ-* ‘to see, look’ (Weiss 2020: 326). Latin athematic *-āx* originated as **-k* added to nominals in *-eh₂* and was extended to verbal roots (Weiss 2010: 66–8, Weiss 2020: 325). It also exists as thematized **-āko-*, which in Celtic is very productive. Latin *-īx* derives from the inherited *devī*-suffix **-ih₂-*, probably with the addition of **-k-* (Schrijver 1991: 148–54, Weiss 2020: 325), though arguments can be made that it is the result of laryngeal hardening (Olsen 2009).

7 The velar suffix has a potential match in Arm. *kᶜowpič* < **koubig-* and the irregularly related PGm. **habuka-* ‘hawk’ (here too from PIE **g*), PSl. **kobьcь*, **kobuzъ*, **kobъzъ* ‘falcon’ (see Thorsø, this volume). A word *capys* ‘bird of prey’ appears in Latin, and since it seems to lack the velar element, it might be evidence that the element is indeed a suffix. But because the word appears only in writings that attribute it to Etruscan and it is not continued in the Romance languages, its evidentiary value as a Latin word is questionable. The further lack of any attestation in Celtic and Greek makes it difficult to verify as part of the same phenomenon as that identified in *felix*, *sōrex*, and *fulica*.

8 Chantraine (1933: 378) suggested the same for the Greek words in -αξ.

On the other hand, words with the *-ix*/*-ex* suffix (athematic and with a short vowel) are virtually isolated to Latin and those with -αξ to Greek. In Latin, the most promising non-isolated case is *natrix* 'sea serpent' against OIr. *nathir* < **natrik*-, but this is complicated by the Brythonic forms reconstructing to **natrī*- and Germanic forms similarly lacking the velar element. In Greek, Kölligan (2017: 369–70) suggests that no cases of Greek formations in -αξ beside a velar element in another branch need be interpreted as anything but individual parallel developments. The best cases are between Latin and Greek, but there are remarkably few. Beyond Lat. *sōrex* ~ Gr. ὕραξ, there is Lat. *mūrex* ~ Gr. μύαξ 'murex'. More uncertain are Lat. *salix* ~ Myc. *e-ri-ka*, Arcad. ηελικης 'willow' and Lat. *īlex* ~ Macedonian (Hesychian) ἴλαξ 'holm oak'.[9]

Given this situation,[10] is it possible that, given a plethora of inherited suffixes of the shape **-Vko*- and **-V̄k*-, an influx of substrate words ending in a velar could have led to their interpretation as containing a suffix **-Vk*- that could then be have been extended to native bases? In Latin, this could have been further encouraged by the existence of nominal compounds in *-fex* < *facere*. In any case, it cannot be ruled out that a velar suffix of the shape **-Vk*- was inherited; perhaps its appearance on the word *senex* demonstrates this. After all, it cannot be ruled out that the *-ix* suffix of the *salix* type represents an avatar of the inherited velar suffix due to the identical pre-forms that can be reconstructed for all comparanda. In this case, the substrate words with the suffix of the *felix-sōrex-fulica* type would easily have been nativized by interpreting them as containing this suffix. The merging of at least two sources of *-ix*/*-ex* suffix would certainly have produced a larger group of lexical material from which the suffix could secondarily be spread, leading to the chimerical situation described by Chantraine.

References

Alessio, Giovanni. 1944. Suggerimenti e nuove indagini sul problema del sostrato "mediterraneo". *Studi Etruschi* 18. 93–157.

Beekes, Robert S.P. 2010. *Etymological Dictionary of Greek*. Leiden: Brill.

Beekes, Robert S.P. 2014. *Pre-Greek: phonology, morphology, lexicon*. Leiden: Brill.

9 Lat. *latex* 'liquid, fluid' ~ Gr. λάταξ 'drop of wine' is complicated by the fact that the Greek oblique forms have γ rather than κ. Lat. *paelex* 'concubine' matches Gr. πάλλαξ 'youth', but this is potentially a backformation from thematic Gr. παλλακή 'concubine' (Beekes 2010: 1147).

10 In his Master's Thesis (Κόραξ καὶ ἀλώπηξ: Time-depth in Indo-European velar stem suffixes), Simon Poulsen sums it up nicely when he writes, "It seems that the elements need to be old, but the forms are paradoxically always young" (p. 62).

Bertoldi, Vittorio. 1937. Contatti e conflitti di lingue nell’antico Mediterraneo. *Zeitschrift für romanische Philologie* 57(2). 139–169.
Brugmann, Karl. 1906. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*. Vol. 2, Part 1. 2nd edn. Strassburg: Karl J. Trübner.
Chantraine, Pierre. 1933. *La formation des noms en grec ancien*. Paris: Librairie ancienne Honoré Champion.
de Vaan, Michiel. 2008. *Etymological Dictionary of Latin and the Other Italic Languages*. Leiden: Brill.
Ernout, Alfred. 1946. *Philologica*. Paris: Klincksieck.
Ettmayer, Karl. 1926. Der Ortsname “Luzerne”. *Indogermanische Forschungen* 43. 10–39.
Fruyt, Michèle. 1986. *Problèmes méthodologiques de dérivation à propos des suffixes latins en . . .*cus. Paris: Klincksieck.
Furnée, Edzard J. 1972. *Die wichtigsten konsonantischen Erscheinungen des Vorgriechischen*. The Hague/Paris: Mouton.
Gerola, B. 1942. Substrato mediterraneo e latino. *Studi Etruschi* 16. 345–68.
Hellquist, Elof. 1922. *Svensk etymologisk Ordbok*. Lund: C. W. K. Gleerups.
Hubschmid, Johannes. 1953. Sardische Studien. *Romanica Helvetica, Vol. XLI*. Bern: A. Francke.
Kock, Axel. 1909. *Svensk Ljudhistoria*. Part 2. Lund/Leipzig: C.W.K. Gleerup/Harrassowitz.
Kölligan, Daniel. 2017. Gr. σάλαξ, σκύθραξ und die griechischen Nomina auf *-ak-*. In Ivo Hajnal, Daniel Kölligan & Katharina Zipser (eds.), *Miscellanea Indogermanica. Festschrift für José Luis García Ramón zum 65. Geburtstag*, 369–82. Innsbruck: IBS.
Kretschmer, Paul. 1896. *Einleitung in die Geschichte der griechischen Sprache*. Göttingen: Vandenhoeck und Ruprecht.
Kroonen, Guus. 2011. *The Proto-Germanic* n-*stems*. Amsterdam/New York: Rodopi.
Kroonen, Guus. 2013. *Etymological Dictionary of Proto-Germanic*. Leiden: Brill.
Kroonen, Guus, Anthony Jakob, Axel Palmér, Paulus van Sluis & Andrew Wigman. 2022. Indo-European Cereal Terminology suggests a Northwest Pontic homeland for the core Indo-European languages. *PLoS ONE* 17(10). e0275744.
Leumann, Manu. 1977. *Lateinische Grammatik: Erster Band, Lateinische Laut- und Formenlehre*. Munich: C.H. Bech.
Martinet, André. 1955. Le couple *senex-senatvs* et le “suffixe” *-k-*. *Bulletin de la Société de Linguistique de Paris* 51(1). 42–56.
Matasović, Ranko. 2009. *Etymological Dictionary of Proto-Celtic*. Leiden: Brill.
Matasović, Ranko. 2016. Latin tree and plant names with the suffix *-ix*, *-ex*. In Nikolai N. Kazansky (ed.), *Indo-evropejskoe jazykoznanie i klassičeskaja filologija* Vol. 20, 700–709. St. Petersburg: Nauka.
Nehring, Alfons. 1925. Griech. τίταξ, τιτήνη und ein vorgriechisches *k*-Suffix. *Glotta* 14(1/2). 153–92.
Olsen, Birgit Anette. 2009. Martinet's rule of laryngeal hardening - a reappraisal. In Stephanie W. Jamison, H. Craig Melchert & Brent Vine (eds.), *Proceedings of the 21st Annual UCLA Indo-European Conference*, *October 30th and 31st, 2009*, 207–22. Bremen: Hempen.
Petr, V.J. 1896. Slavische Etymologien. *Beiträge zur Kunde der indogermanischen Sprachen* 21. 207–17.
Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern/Munich: Francke.
Polomé, Edgar. 1989. Substrate Lexicon in Germanic. *NOWELE* 14. 53–73.
Salmons, Joe. 1992. Northwest Indo-European vocabulary and substrate phonology. In Roger Pearson (ed.), *Perspectives on Indo-European language, culture and religion: studies in Honor of Edgar C. Polomé*. Vol. 2, 265–279. Washington, DC: Institute for the Study of Man.
Schrijver, Peter. 1991. *The Reflexes of the Proto-Indo-European Laryngeals in Latin*. Amsterdam/Atlanta: Rodopi.

Schrijver, Peter.1997. Animal, vegetable and mineral: some Western European substratum words. In Alexander Lubotsky (ed.), *Sound Law and Analogy*, 293–316. Amsterdam: Rodopi.

Schrijver, Peter. 1999. On Henbane and Early European Narcotics. *Zeitschrift für celtische Philologie* 51 (1). 17–45.

Specht, Franz. 1947. *Der Ursprung der indogermanischen Deklination*. Neudruck. Göttingen: Vandenhoeck & Ruprecht.

Terracini, Benvenuto. 1929. Su alcune congruenze fonetiche fra Etrusco e Italico. *Studi Etruschi* 3. 209–48.

Vine, Brent. 1999. On "Cowgill's Law" in Greek. In Heiner Eichner & Hans Christian Luschützky (eds.), *Compositiones Indogermanicae in Memoriam Jochem Schindler*, 555–600. Prague: Enigma.

Walde, Alois & Johann Baptist Hofmann. 1938–54. *Lateinisches etymologisches Wörterbuch*. Heidelberg: Winter.

Weiss, Michael. 2010. *Language and Ritual in Sabellic Italy: The Ritual Complex of the Third and Fourth* Tabulae Iguvinae. Leiden/Boston: Brill.

Weiss, Michael. 2020. *Outline of the Historical and Comparative Grammar of Latin*. 2nd edn. Ann Arbor/ New York: Beech Stave Press.

Cid Swanenvleugel

8 Prefixes in the Sardinian substrate

1 Introduction

1.1 The Sardinian substrate

The goal of the present study is to explore what inferences we can make about the prehistoric linguistic situation of Sardinia on the basis of purely linguistic material. I will study the occurrence of the sequences **θ(i)-* and **ka-* in the lexicon of the modern Sardinian varieties, which may originally have been morphemes in a language (or multiple languages) preceding the arrival of Latin on Sardinia (Guarnerio 1904: 57–58, 58 fn. 2; Wagner [1950] 1997: 263). By means of a detailed analysis of their geographical attestation as well as their distribution within the lexicon, I believe there is good evidence that these two elements were not only morphemes of one and the same pre-Roman language (§ 5.1), but were perhaps even variants of the same morphemes (§ 5.2). Besides making a small step forward in our understanding of the enigma of the linguistic substrate of Sardinia, this study aims to contribute to a methodology of linguistic substrate research that is firmly based on the linguistic facts found in the language under study, rather than on lexical comparisons between geographically and/or genealogically distant languages. The identification of grammatical processes in substrate languages can hopefully aid us in determining their original extent and affiliation.

It has long been recognized that a considerable portion of the lexicon of Sardinian, the Romance language native to much of the Mediterranean island of Sardinia, is impossible to etymologize on the basis of inherited Latin material or of loans from other known languages. Dozens of such words are found across the Sardinian dialects, referring mostly to the island's native fauna (e.g. *katranákka* 'tick', *lírčis* 'viper'; DES, I: 319, II: 31), flora (e.g. *aθánda* 'poppy, *kòstiɣe* 'maple'; DES, I: 151, 392), geography (e.g. *ğára* 'basalt plateau', *tónneri* 'rock outcropping'; DES, I: 603, II: 495) and (agri)cultural practices (e.g. *kuɣústa* 'the sheaves of wheat chosen to be sown next year', *ispéli* 'traditional acorn bread'; DES, I: 422, 676). It is evident that many of these words must have entered the local Romance varieties from the language or languages that preceded Latin (Wagner 1932; Hubschmid 1953: 15; Serra 1960). The total body of these etymologically obscure and possibly pre-Roman linguistic elements on Sardinia is most frequently referred to as "Palaeo-Sardinian" (e.g. Blasco Ferrer 2017: 67). This does not necessarily imply that we are dealing with a single language (or even language family).

https://doi.org/10.1515/9783111337920-008

Not included in the concept of Palaeo-Sardinian are two non-native languages of Sardinia that are thought to have preceded Latin on the island as well: Phoenician (Wagner 1932: 208–210; 1997: 150–161; Serra 1960; Paulis 1990) and Greek (Wagner 1997: 162–165; but see Paulis 1983: 71–72, who argues that the linguistic influence of ancient Greek in Sardinia was marginal). These were spoken by Punic and Greek colonists, the former along the south coast, the latter in Olbia in the north east. Before Latin was ultimately adopted by all Sardinians, a process that may have taken centuries, Latin, Greek, Phoenician and one or more Palaeo-Sardinian languages may have been spoken side by side.[1] The sociolinguistic details of this pre-Roman multilinguism are impossible to recover however. While Phoenician and Greek on Sardinia count as pre-Roman substrate languages in the strict sense, I will focus here on the Palaeo-Sardinian substrate — i.e. the linguistic elements Sardinian Romance that cannot easily be attributed to any known attested language.

1.2 History of Sardinian substrate research

The Sardinian non-inherited lexicon has often figured in discussions about the prehistoric linguistic situation in the Mediterranean region.[2] Scholars of the so-called "Italian school" saw evidence for the identification of several large, linguistic layers pre-dating the spread of Latin (and other modern languages) in the Mediterranean basin. Their hypothesized strata connect linguistic features from the Alps to northern Africa and from the Iberian Peninsula to the Caucasus (e.g. Bertoldi 1928; 1937a; Alessio 1941). Besides the methodological issue that much of the hypothesized connections are based in large part on similarities between toponyms, especially in northern Africa (cf. also Serra 1960), where our knowledge about the pre-Phoenician linguistic situation is very scarce (but cf. Múrcia Sànchez 2011), many of the studies discuss individual etymologies rather than offering an in-depth discussion of the Sardinian material as a whole.[3] Exceptions to this are the overviews of the material by Wagner (1932; 1997: 254–289), and the studies conducted by Johannes Hubschmid, whose *Sardische Studien* (1953) is dedicated entirely to the pre-Roman lexical elements found in Sardinian. Hubschmid

1 On the romanization process of Sardinia, see Wagner (1928), Blasco Ferrer (1989), Mastino (1993), Lupinu (2005), and Virdis (2018).

2 For an extensive overview of the history of research into linguistic pre-Roman substrate in general, see Devoto (1955) and Russell Craddock (1969).

3 The pre-Roman toponyms of Sardinia on the other hand, have received a great deal of attention (e.g. Terracini 1927; Serra 1960; Paulis 1987; Wagner 1997: 259–260; Wolf 1998).

takes a more critical approach than e.g. Alessio and Bertoldi when it comes to the interpretation of the Romance forms, and relies more on lexical than toponymic evidence. Like the Italian school, he posits several linguistic strata with a large geographical distribution in order to account for the non-inherited lexicon of Sardinian and other Mediterranean languages. Still, his linguistic connections between Sardinian words and forms found in Basque, Ibero-Romance, Occitan, Italo-Romance and Berber do sometimes require a certain dose of imagination,[4] and several of them must be rejected after a closer look into the details of the non-Romance comparanda.[5] Even though the posited substrate layers of both Hubschmid and the Italian school remain very vague and large-scale concepts (e.g. Hubschmid's (1953: 122) "Eurafrican" and "Ibero-Caucasian" substrates) and at times methodologically questionable, their works are of great value if only for the wealth of linguistic evidence they document, and some of their hypotheses may in fact prove to be close to the truth.

A rather agnostic stance of the pre-Roman elements in the Sardinian language is taken by Max Leopold Wagner, often hailed as the "Father of Sardinian linguistics". He has written two good overviews of the material and literature about the pre-Roman linguistic elements in Sardinia up to that point; cf. Wagner (1932; 1997: 254–289). Wagner himself is very cautious not to confuse cognates with chance similarities and limits himself to providing a description of the internal Sardinian practicalities rather than comparing the Sardinian lexicon to other languages. Although Wagner does not explicitly express his own view on the identity or affiliation of the substrate, he is mildly favorable to some of the Basque and Berber connections made by other researchers.

4 Cf. for example his comparison of Srd. *éni* 'yew tree' (for which Bertoldi (apud Wagner 1941: 45) proposes an earlier **ágini* and connection to Bsq. *agin* 'id.') with Brb. *θiqqi, θaqqa, θagga* etc. 'juniper' (Hubschmid 1953: 25–26). Even if we accept a meaning shift, the formal comparison is based solely on the comparison of Basque *-g-* (which is reconstructed but not actually attested for Sardinian) and Berber *-qq-*, and is therefore too weak in my opinion.

5 For instance, Hubschmid (1953: 33–35) attributes Srd. *mátta* 'tree', Bsq. *mata* 'tree stump' and Brb. (Senhaja and Tarifit) *θamáṭṭa* 'pile of sheaves to be threshed', as well as many similar forms in Romance languages, to a common substrate layer present in a very large area. The Berber forms however, which coexist with Brb. *rmaṭṭa* 'id.' (Serhoual 2002: 329), are rather to be analysed as loans from Maghrebian Arab. *maṭṭa* 'small isolated heaps of sheaves' (Sīnāṣir 1994: 1844; Maarten Kossman p.c.). Moreover, Corominas (1980: III: 874–5) proposes to derive all Romance forms of this type from late Lat. *matta* 'mat', making a substrate explanation unnecessary. Bsq. *mata* 'tree stump' may be from Gasc. *mato* 'tuft of grass' vel sim. A similar case is the connection between Srd. *sakkáyu*, Bsq. *segaila*, Cat. *segall* and Gasc. *segalho* 'one-year old lamb/kid', proposed by Wagner (1932: 225–226) and Hubschmid (1953: 70) to be of pre-Roman origin. However, Wagner (1952a: 418) later prefers to see the Sardinian form as a Catalan loan, and Corominas (1980: V: 188–189) shows that the Basque form is in fact a loan from Bearnese.

A more recent hypothesis about the nature of the Sardinian substrate is by Massimo Pittau (1995). A crucial difference between his approach and those of earlier scholars is that Pittau, rather than suggesting cross-Mediterranean links and positing abstract linguistic layers, has a much more specific and well-defined view on Palaeo-Sardinian origins and affiliation. In his view, Palaeo-Sardinian (or *Nuragic*) was a sister language to Etruscan. At some point in the Bronze Age, both would have arrived in their eventual locations from Anatolia, where their closest linguistic relative would have been Lydian.[6] While his theory is firmly grounded in ideas on peoples' origins written down by classical authors, many of the comparisons between the Sardinian lexicon and Etruscan are rendered dubious by the fact that the Etruscan evidence mainly consists of personal names. While these obviously constitute a large part of the Etruscan linguistic corpus, they are usually devoid of any semantic information, rendering the ensuing connections to Sardinian appellatives purely speculative.

Similarly explicit on his ideas about Palaeo-Sardinian affiliation is Eduardo Blasco Ferrer, who argues in a series of articles for a language family comprising Palaeo-Sardinian and Basque (Blasco Ferrer 2010; 2011a; 2011b; 2017). He criticizes the approaches to the linguistic substrate adopted by all of his predecessors, and he argues that the Comparative Method is of no use when it comes to the study of unknown languages. While some of his reservations have merit, the approach he himself promotes instead, his so-called "structural method", does not quite offer an improvement when it comes to its results. He envisages Palaeo-Sardinian as an agglutinative language, and recognizes various recurring monosyllabic morphemes in the many pre-Roman toponyms found in Sardinia. He then goes on to compare these newly identified morphemes to Basque lexical items. It may be clear that the exclusive use of substrate toponyms, whose semantics have often been completely lost, means that any phonological similarities to Basque can never be backed up by semantic correspondence. So, although several more or less compelling lexical correspondences between Sardinian and Basque have been adduced by others — e.g. Srd. *golóstri*, Bsq. *gorosti* 'holly' (Bertoldi 1929: 261, fn. 3) and Srd. *ğáɣaru* 'hunting dog', Bsq. *txakur* 'dog' (Wagner 1932: 232) —, Blasco Ferrer's work does not ultimately succeed in convincingly establishing Palaeo-Sardinian as a relative to Basque (cf. Loporcaro 2013 for a more in-depth discussion of Blasco Ferrer's methodology).

6 Pittau holds the Sardinian substrate as well as Etruscan to be Indo-European languages (1995: 198). Even if there is some debate possible on the "Indo-European-ness" of Etruscan (the *communis opinio* being that it is not), it is in any case not closely related to the Anatolian branch of Indo-European, to which Lydian belongs (Wallace 2008: 1, 217–218).

1.3 Continued relevance of the Sardinian substrate

Despite its high profile in substrate research among scholars studying Romance, the Sardinian substrate lexicon is almost completely ignored in substrate research in the Indo-European context. While the occupation with linguistic substrates among Indo-Europeanists is arguably rather more recent than among Romanists, it is at least peculiar that such comparanda as Srd. *golóstri*, Gr. κήλαστρος 'holly', or Srd. *kóstike*, Gr. ἄκαστος 'maple' have not been included in etymological works on the topic, such as Beekes & van Beek (2010: 50, 685–686). Over the past few decades, there has been an increasing interest in Europe's pre-Indo-European linguistic history and a rapid development of insights into the genetic composition of the modern and ancient populations of Europe. As it has been shown that Sardinians are among the populations that were genetically least impacted by "Steppe DNA" in all of Europe (Chiang et al. 2018; Calò et al. 2021; Raveane et al. 2022), which arrived through Roman intermediation as late as ca. 300 CE (Fernandes et al. 2020), it appears that Sardinian linguistic (pre)history is more relevant than ever when it comes to pre-Indo-European Europe. In light of these new developments, I believe it is timely to revisit the problem of the Sardinian substrate taking into account the methodological considerations developed in Indo-European linguistics in recent decades.

The aim of the present study is neither to provide Palaeo-Sardinian with a relative (living or dead), nor to suggest etymological connections between non-inherited Sardinian words and substrate words in other languages. Because Sardinian is a living language, whose dialectal diversity and historical development have been amply documented (most notably in Wagner's Sardinian etymological dictionary, = DES), we have a very fine-grained view of potentially pre-Roman words, their irregular correspondences and their geographical distribution on the island. It is my goal to use this as an advantage and to explore what a detailed study of the internally Sardinian linguistic material may tell us about the island's prehistoric linguistic situation.

As a discussion of the entire corpus of etymologically obscure Sardinian elements would greatly exceed the limits of this study, I will focus on two apparent prefixes present in the Sardinian lexicon: **k(V)-* and **θi-*. As we will see, both elements occur on various words of obscure etymology, and we occasionally find variants of the same word with and without either prefix. Intriguingly, they do not exclusively occur in purported Palaeo-Sardinian words, but were even sometimes added to inherited Latin etyma. This can be observed, for instance, in the Sardinian reflexes of Lat. *lacerta* 'lizard': across the various dialects, we find: *ka-*

lužèrta and *θilikèrta*, next to un-suffixed *aliγèrta* (DES, II: 545–546).[7] In this study, I will explore the properties of these two elements. While both **k(V)-* and **θi-* were identified as an added feature already by Guarnerio (1904: 57–58, 58 fn. 2), **θi-* has received much more attention as a possible substrate feature than **k(V)-* (even Guarnerio mentions "*ka-*" only once in a footnote), and it has on multiple occasions been compared to the Berber feminine prefix *t-* by Wagner (1922: 253; 1932: 223–224; 1997: 263).[8,9] An alternation involving **k(a)-* has also been identified in non-inherited Greek words, although never in relation to Sardinian (Beekes 2014: 21). Although the comparisons to Berber and Greek are outside the scope of the present study, the Sardinian evidence for both prefixes **θi-* and **k(V)-* itself offers a promising possible road into the study of the morphology of the substrate language.

1.4 Methodological considerations

The major methodological challenge in linguistic substrate research is the fact that the comparative method can only be used to a very limited extent. Lexical items are identified as candidates for a potential substrate origin on the basis of their non-adherence to the known sound correspondences of a certain language or variety. Attributing these lexical items to a linguistic substrate is in part the result of strict application of the *Ausnahmslosigkeit der Lautgesetze*, and of the preference to explain deviations from this principle by means of linguistic contact with unattested languages over allowing for exceptions to sound laws and/or irregular developments due to contaminations, folk-etymologies, sound-symbolisms etc. (cf. Schrijver 1997: 294–296; Stifter, this volume). As there can be no doubt that these processes do occur in the diachronic development of any language, they must be taken into account as potential explanatory devices when dealing with words of non-straightforward etymology (cf. Beccaria 1995; Flaksman 2017: 16–18). The plausibility of any contamination/folk-etymology/sound-symbolism etc. must be evaluated for every case individually after thorough formal and semantic scrutiny (e.g. Schuhmann 2016: 383). Naturally, the likelihood of any folk-etymology (for instance) will be estimated differently by different people, partly depending on their inclination towards or away from accepting the influence of a non-accepted language to account for formal irregularity, over language-internal irreg-

7 For the phonetic details, see § 2.1.

8 Note that all instances of *-θ-* are misprinted as *-p-* in Wagner (1922).

9 Alinei's (1984: 29–30) suggestion to explain every instance of **θi-* as a being from *θíu, θía* 'uncle, aunt' is unconvincing.

ular processes driven by semantic associations such as contaminations or folk-etymology (e.g. Alinei 1984). In this regard, it must be noted that for several of the Sardinian words hypothesized to be of pre-Roman origin that are discussed below, alternative etymologies involving inheritance from Latin and/or folk-etymological contaminations have been proposed, with varying degrees of plausibility (cf. especially § 2, § 3.1, § 3.3–3.5, § 4.6). While none of them are without problems (in which case it would be irrelevant to discuss them), some cases of formal irregularity cannot fully be ruled out to be the result of language-internal processes rather than a substrate origin. Where to draw the line between a plausible and an implausible etymology is a question that remains important to discuss, as is the question when positing a substrate origin is or is not to be preferred over a problematic etymology. That being said, in light of the purposes of the current volume, I will opt for a formally and semantically critical approach when it comes to language-internal irregular processes, in order to explore the explanatory potential of assuming a substrate origin for words with a poorly understood etymology, following the methodological considerations outlined for instance in Schrijver (1997), and in this volume's introduction.

In what follows, I will discuss the evidence for positing **k(V)-* and **θi-* as morphological elements belonging to one or more pre-Roman languages spoken on Sardinia. The presentation of the material is structured as follows. For every word that shows evidence of the presence of a prefix, the various attested formation types are listed in a reconstructed "Proto-Sardinian" form. The determination of these types are the result of my own analysis, based on the lexical data provided in the appendix. The used data minimally includes forms attested in Wagner's *Dizionario etimologico sardo* (DES). Whenever the *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS) contains additional evidence, this is listed below the forms from DES. Next is listed any material from other relevant sources that discuss the etymon in question. The places of attestation are numbered from north to south. Finally, I give the lexical evidence found in Puddu's (2020) Sardinian dictionary, which contains a wealth of material, but regrettably does not offer any information on the place of attestation. The linguistic material is accompanied by a map of Sardinia with the geographical distribution of the forms and a discussion of the evidence. In § 2, I treat all instances of etyma showing evidence for both **k(V)-* and **θi-*; in § 3 words with evidence for **k(V)-* are discussed, saving the remaining instances of **θi-* for § 4. As a caveat, it must be noted that the geographical information in DES is generally detailed, but not consistently for every word, and the same goes for other sources. Whenever a form is given for a dialect group, rather than for a specific location, it is represented as a patterned area according to formation type.

For the classification of the Sardinian dialects, I use Virdis' (1988) division into four dialect groups: Campidanese in the southern half of the island, Logudorese in the north and northwest, Nuorese in the center east, and Arborense in the center west. Among these, the Nuorese dialects are generally considered to be phonetically and lexically the most conservative (Wagner 1997: 342; Virdis 1988: 910–911; but cf. Bolognesi & Heeringa 2002). The Arborense dialects can be considered transitional dialects between Logudorese in the north and Campidanese in the south. Besides these true Sardinian dialects, there are several other Romance languages spoken on Sardinia. Along the north coast, we find Sassarese in the west and Gallurese in the east, both of which are more closely related to Corsican (Ledgeway 2016: 208). In and around the city of Alghero in the northwest, a dialect of Catalan is spoken. Finally, on the islands of the Sulcis archipelago off Sardinia's southwestern coast, we find Tabarchino, a Ligurian dialect. Even though these languages are not particularly closely related to Sardinian, their dialects spoken on Sardinia potentially contain the same type of evidence for the linguistic situation in pre-Roman Sardinia as the true Sardinian dialects. Sassarese and Gallurese frequently exhibit the same non-inherited lexical items as the Logudorese and Nuorese dialects. Since the distribution of pre-Roman elements on Sardinia (and in fact anywhere) cannot *a priori* be expected to line up with modern dialect borders, no information on dialect classifications is included in the maps. For more details on the classification and geographic distribution of the language varieties found on Sardinia, see Virdis (1988). For a detailed account of the phonetic characteristics of the Sardinian dialects, see Wagner (1907) and especially (1941). For a general overview of the Sardinian language and its history, see Blasco Ferrer (1984) and Wagner (1997).

2 Alternation between **kV-* and **θi-*

This section includes two words that present potential evidence for the presence of both **kV-* and **θi-*. It concerns the words for 'lizard' and for 'celery'. The case of 'lizard' (§ 2.1) is especially interesting, as **kV-* and **θi-* appear to have been added to inherited Lat. *lacerta* 'id.'.

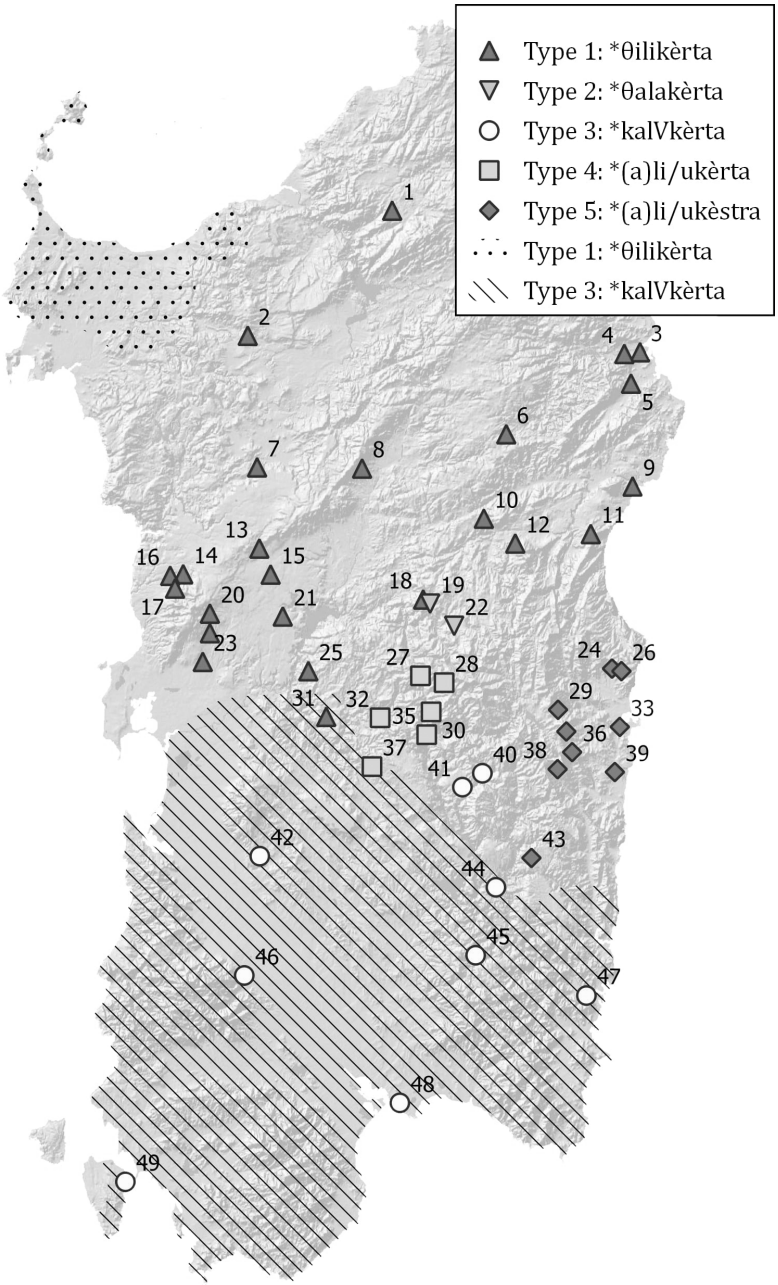


Figure 1: Lizard.

2.1 **ka/θi-lakèrta* 'lizard'

Type 1: **θilikèrta*
Type 2: **θalakèrta*
Type 3: **kalVkèrta*
Type 4: **(a)li/ukèrta*
Type 5: **(a)li/ukèstra*

All forms represented in Figure 1 are descendants of Lat. *lacerta* 'lizard'. The Sardinian material can roughly be divided into three groups: forms with addition of **θi-* or **θa-* (types 1 and 2),[10] forms with addition of **ka-* (type 3), and the remaining forms (types 4 and 5). Of the first group, almost all attested forms appear to regularly go back to **θilikèrta*. The existence of **θilikèrta* << Lat. *lacerta* makes an analysis of *θi-* as some kind of prefix rather straightforward. Despite this word's inherited origin, it is hard to explain *θi-* in terms of inherited Romance material. The two forms of type 2 need a slightly different pre-form: *θala?èrta* (19: Gavoi), *θalaèrta* (22: Fonni) contain initial **θa-* rather than **θi-*. In the case of *attaliɣèlte* (15: Scano di Montiferro) this is perhaps due to assimilation to the initial *a-*, especially in light of the existence of forms like *attiliɣèrta* (13: Macomer). The development would be as follows: *sa θilikèrta* >> *s' attiliɣèrta*[11] >> *s' attaliɣèlte*. For *θala?èrta* (19: Gavoi) and *θalaèrta* (22: Fonni) though, it is unlikely that the article could explain the *-a*-vocalism of the second syllable. A possible solution would be to assume that the *-la-* is simply the direct continuation of the *la-* in Lat. *lacerta*. This would be remarkable however, since there is not a single dialect elsewhere on the island that directly preserves *-la-*. Even the apparently unprefixed forms have *-li-*: e.g. *[s'] aliɣèrta* (27: Tonara, 28: Desulo).[12] Nevertheless, an original **θi-lakèrta* could probably yield attested *θala(?)èrta*, by assimilation of pretonic *θi-* to the vocalism of the next syllable. So it seems that "Proto-Sardinian"

10 Sardinian **-θ-* is the regular reflex of inherited Latin *-cį-* and *-tį-*. Roughly speaking, it is reflected in the Nuorese dialects as *-θ(θ)-* (an interdental fricative), in Logudorese as *-t(t)-* and in Campidanese as *-(t)ts-*, with some smaller areas having *-č-* and *-s(s)-* (Wagner 1941: 106–109). E.g. Lat. *facio* 'I do', *puteus* 'well' > Nuor. *fáθθo, púθθu*; Log. *fátto, púttu*; Camp. *fáttsu, púttsu*. Other cases of **-θ-* are found in loanwords, onomatopoeic formations and words of potentially pre-Roman origin, in all of which the distribution of the reflexes *-θ-*, *-t-* and *-ts-* is the same (Wagner 1941: 109–118). In Italian-based orthographies, Camp. *-ts-* is also represented as *-z(z)-* (e.g. Marcialis 2005) or *-tz-* (e.g. Puddu 2020). It should be noted that **-θ-* is above all a graphical device representing the modern correspondence of *-θ-*, *-ts-*, *-ss-*, *-č-* (Wagner 1941: 107–108).

11 Initial *a-* frequently appears in front of original **θ-* in Logudorese dialects (cf. the examples in Wagner 1941: 110).

12 With the initial *a-* reinterpreted from the feminine article *sa* (Wagner 1941: 213–14).

*θilikèrta involves two changes: the addition of initial *θi- to inherited *lacerta*, and the replacement of *-la-* with *-li-*. Vowel changes in this word's pretonic syllable are well-attested in French and Italian dialects too, where we find words for 'lizard' going back to *licerta as wel as *lucerta (REW 4821; FEW V: 117, 118 fn. 6; Battisti and Alessio 1952: 2276). For Sardinian *θilikèrta too, the sequence *-li-* may thus be of Vulgar Latin origin. Influence from the many Palaeo-Sardinian words starting in *θili-; cf. *θilikúku* (§ 4.1), *θilipírke* (§ 4.2), etc. is also imaginable however.

The forms containing *ka-* (type 3) are most easily explained from an original *kalVkèrta, again with vowel alternations in the pretonic syllable. These forms are found in the Campidanese dialects in the south of Sardinia, and many show various kinds of assimilations and metatheses. *Karǧilètta* (40: Seui) and *kožuètta* (49: Sant'Antioco) have been metathesized respectively from *kaliǧèrta and *koužètta, both of which can go back to *kali/ukèrta. In several forms the last two syllables *-èrta were replaced by *-èḍḍa*, which is the inherited diminutive suffix (< Lat. *-ella*). It is difficult to say to what extent the coexistence of forms containing *-la- (e.g. *kalažèḍḍa*, 44: Escalaplano), *-li- (e.g. *kaʔižèḍḍa*, 47: Muravera) and *-lu- (e.g. *kalužèrta*, 41: Sadali) are to be attributed to Vulgar Latin variation between *lacerta*, *licerta and *lucerta that must be assumed on the basis of other Romance languages (REW 4821; FEW V: 117, 118 fn. 6), as pretonic assimilations and dissimilations are very common across Sardinian (Wagner 1941: 21–33). The remaining forms, that have neither *θi- nor *ka- (types 4 and 5), can go back to *lukèrta or *likèrta. The initial *a-* is due to reanalysis of the vowel of the feminine definite article.[13] In the Ogliastran dialects in the east of Sardinia, *-èrta* has been replaced with *-èstra, -éstru, -ésti* (type 5): e.g. *liǧèstra* (33: Tortolì) << *liǧèrta < *likèrta.* This is reminiscent of the *-st(r)-* found in other Sardinian words of uncertain origin, such as *golóstri, golóstiu* 'holly' and *giḍḍóstru/a* 'tree-heath', which has been identified as a suffix originally belonging to a substrate (Bertoldi 1930; Van Sluis, this volume). In our lizard word, this suffix may have been added secondarily due to its relative frequency in the Sardinian lexicon, similarly to the replacement of *-èrta* with diminutive *-èḍḍa* seen in *kalažèḍḍa* (44: Escalaplano) and *kaʔižèḍḍa* (47: Muravera). Proposals to explain the addition of *ka- in this word include Bertoni's (1913: 167, fn. 4) suggestion that the reflexes of *lacerta* were contaminated with those of *colūbra* 'snake, serpent', while Böhne (1950: 109)

13 *Arǧiléstru* (34: Arzana, 36: Lanusei) is masculine, and thus takes the masculine article *su*. It is however safe to assume that its masculine gender is a secondary development, and that the initial *a-* was added when it still exhibited the feminine gender as inherited from Lat. *lacerta.* Variation between masculine and feminine gender for this word is already attested in Latin (FEW, V: 117a).

suggests of a contamination with *kalái* 'to descend'. Böhne's (1950) idea seems semantically rather far-fetched. Bertoni's (1913) hypothesis is semantically attractive, but the vocalism does not quite match in all locations (e.g. Escalaplano (44) *kalažèḍḍa, kalažètta* 'lizard' next to *kolóru* 'snake'; AIS 452).

To sum up, in order to explain the Sardinian reflexes of Lat. *lacerta*, we need to assume at least three different base forms: **θi-la/ikèrta*, **ka-la/i/ukèrta* and **li/ukèrta*. The variants of this word clearly show the secondary addition of **θi-* and **ka-*, neither are easily explained in Romance terms. As for the geographic distribution of the various types, we find types 1 and 2 (in **θi/a-*) in the northern part of the island, roughly corresponding to the Logudorese and Nuorese dialect areas (as well as the Corsican dialect of Gallurese). Type 3 (in **ka-*) is found in the south, with type 4 (in **(a)-*) wedged in between types 1 and 3. Type 5, with replacement of **-èrta* with *-èstra/-éstru/-ésti* is restricted to the Ogliastran dialects in the southeast.

2.2 **ku-/θeli-/θur-gús-* 'celery, fool's watercress'

Type 1: **kugúsa*
Type 2: **θurgúsa*
Type 3: **θelikúsa*
Type 4: **(a)θikúsa*/**(a)tikúsa*

DES (I: 422; II: 555) cites type 1 only for Mamoiada (3), and type 2 for Bitti, Nuoro and Orgosolo (1, 2, 4). Paulis adds type 2 for Orani. All of these are in the mountainous center of the island (cf. Figure 2). The forms from Mamoiada, Bitti and Nuoro all refer to species of the genus *Apium*. For the form from Orgosolo, DES gives the meaning 'poison hemlock (*Conium maculatum*)'. This meaning is not represented by Puddu, who only has 'fool's watercress' and 'celery'. The lexical forms themselves have in common that they end in *-gúsa*. If they belong together, we find an alternation between initial *ku-* vs. *θur-* (*θru-* is due to metathesis).[14]

On top of the forms presented by DES, Paulis (1992: 141) lists *turguṡòne* for Logudorese and *tseliɣúṡa* for "South Logudorese", which corresponds to Virdis'

14 Paulis (1992: 141–142) prefers to treat *θurgúṡa* and *kuɣúṡa* as unrelated words. *Kuɣúṡa* would be identical to *kukúsa* 'flake'. The problem is that this is expected to turn up as ***kuʔúṡa* in Mamoiada. He proposes a connection between *θurgúṡa* 'wild celery' and *orgòsa, urgúsa* 'source of water'. *Orgòsa* is itself of obscure origin, so this connection remains speculative. I think it is rather unattractive to separate *θurgúṡa* and *kuɣúṡa*, which both contain **-gúsa* and both mean 'wild celery'.

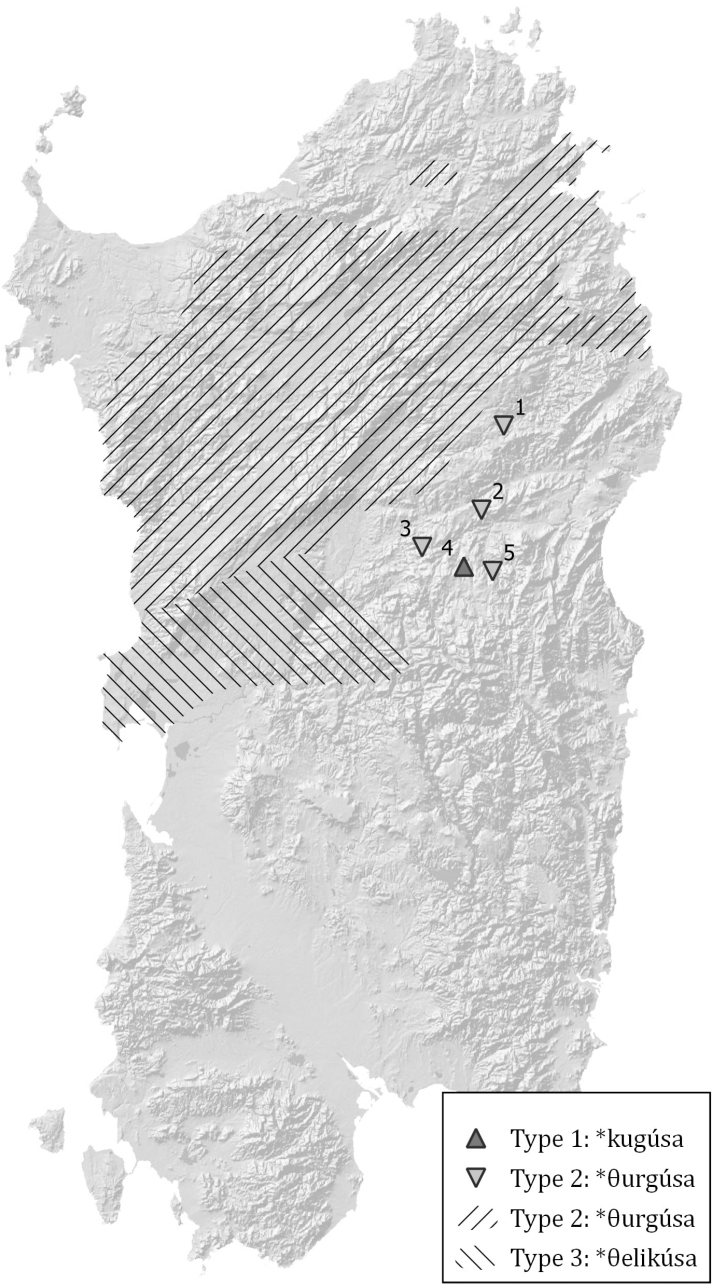


Figure 2: Celery, fool's watercress.

Arborense variety (Virdis 1988: 912). *Tseliɣúśa* 'celery' also has *-gúsa* but starts in *tzeli-*, which may be the same as *θili-* found in *θilikèrta* and other words (§ 2.1; § 4). Puddu has many forms, most of which can be compared to the forms attested in DES and Paulis. An exception to this is *atigúsa* 'celery', which could be a Logudorese reflex of **θi-* (cf. Macomer *attiliɣèrta* < **θilikèrta*). If all of these forms are taken into consideration, we seem to be dealing with an alternation between initial *ku-*, *θur-*, **θili-* and **θi-*. Except for the *θur-*, this matches up with the alternations we see in the words for 'lizard' and 'skink' (§ 2.1; § 4.1). Whether *θur-* is some variant of **θili-*, or a different morpheme altogether in the source language, is difficult to establish on the basis of one word. The Logudorese and Arborense forms listed by Paulis and Puddu (*turgusòne* and *tseliɣúśa*) show that the distribution of this family of words for plants of the genus *Apium* must be rather larger than the attestations in DES would suggest. Unfortunately, because of a lack of more specific geographical information, we cannot use these forms to say anything about the geographical distribution of any of the prefixes discussed.

One difficulty in the comparison of the forms in DES to those listed by Paulis and Puddu is the reconstruction of the velar obstruent. The Nuorese forms (1–5) all need an original **-g-*, since **-k-* would either have been preserved as such, or yielded ***-ʔ-* in Mamoiada and Orgosolo (Wagner 1941: 71); the same goes for the *-g-* in Logudorese *turgusòne*, where **-k-* is also expected to be preserved post-consonantally (1941: 175). In Arborense *tseliɣúśa* and probably also *atigusa* (if it is Logudorese), the *-ɣ-/-g-* should go back to **-k-*, since intervocalic **-g-* would have disappeared in these dialects (Wagner 1941: 79). Perhaps the presence of *-g-* in the Logudorese and Arborense dialects is due to influence from forms like Camp. *kuɣúttsula* 'wild artichoke' < Lat. *cucutium* 'hood, cap', as (cf. Paulis 1992: 142) suggests. Note however that *kuɣúsa* (4: Mamoiada) can for the same formal reasons not be the same word as *kukúśa* 'tuft' as Paulis (1992: 142) suggests.

3 Further evidence for **kV-*

This section contains all the other cases with potential evidence for **kV-* as a prefixed element. All of its representatives refer to the natural world (i.e. flora and fauna). Here too we find one potential instance of addition of **kV-* to a word of inherited Latin origin, in *kampíŋgu* 'pine tree' (§ 3.1).

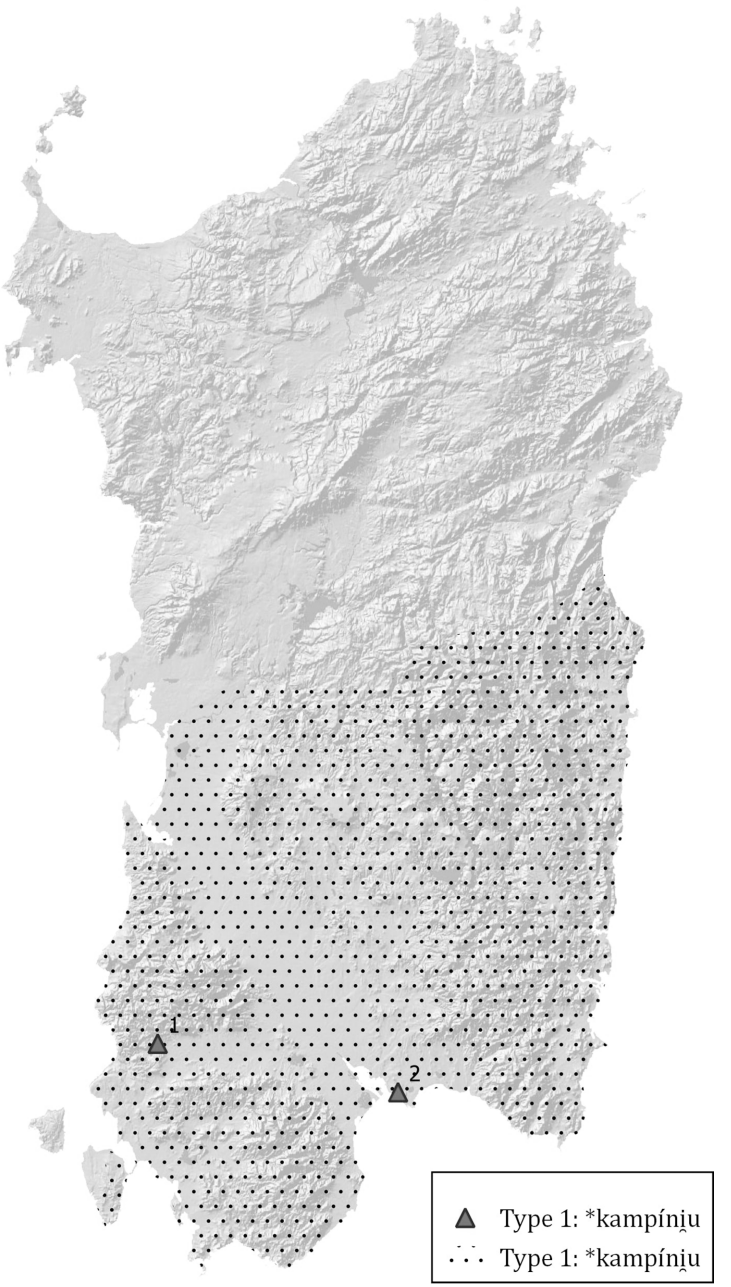


Figure 3: Pine tree.

3.1 **ka(m)-pinu̯* ‘pine tree’

Type 1: **kampínu̯*

This cluster of words (Figure 3) refers to various species of pine-tree (*Pinus*), along with words of the type *pínu* and *oppínu*, respectively from Lat. *pīnus* and *sappīnus* ‘kind of pine/fir’.[15] The element *-pínğu* is the regular outcome of **pínu̯* < Lat. *pīneus* ‘of the pine’. The variation between **pínu̯* ~ **kampínu̯* thus looks like another instance of *kV-* added to a Latin root. It is not clear whether the *-m-* was originally part of the prefix, or arose through secondary *m*-insertion, like in *sampunáre* < *sapunáre* ‘to wash’ (Wagner 1941: 223).

A completely different interpretation is given by Paulis (1992: 443), who prefers to explain *kampínğu, kompínğu* as a semantic extension of original **cōnus pīneus* ‘pine cone’. If this is right, this word obviously does not constitute an example of **ka*-prefixation. Although Paulis’ etymology is semantically very attractive, there are some objections to it. One is the fact that Lat. *cōnus* ‘(pine) cone’ is as such not preserved in Sardinian – although *kampínğu* etc. could be the result of a very early compound. The other issue is the phonetic development. We would have to assume loss of the final vowel of **kónu* (**kónu pínu̯* > **kónpínu̯* > **kompínu̯*), and subsequently the shift from pretonic *-o-* to *-a-*. This latter development, as well as the opposite development (pretonic *-a-* > *-o-*) quite well-attested so it does not pose any problem to Paulis’ etymology (cf. Wagner 1941: 26, 28, 32). I have however not been able to find any good example of vowel syncope on the boundary between two parts of a compound, as is needed to derive *kampínğu* from **kónu-pínu̯* < **cōnus pīneus*. Potential counter-examples to this accent development include the early Sardinian compounds *kenápura, čenáβara* etc. ‘Friday’ < Lat. *cēna pūra* (DES, I: 328), Srd. *issára, insára* ‘then’ < Lat. *ipsa hōra*, Camp. *nottèsta* ‘this night’ < Lat. *nocte istā* (Wagner 1941: 5), which all show preservation of the vowel on the word boundary and even retraction of the stress onto it.[16] Unfortunately, none of these examples have a first part ending in *-u*, as in supposed **kónu-pínu̯*. We do find Srd. *kústu* ‘this’, *kúḍḍu* ‘that’ < **eccú-istu*, **eccú-illu*,

15 Srd. *oppínu* is due to reanalysis of the *s-* in Lat. *sappīnus* as the definite article (DES II: 272).

16 Wagner (1941: 5) explains the accentuation of *kenápura, issára* < **ipsá-hora* and *nottèsta* < **nocté-ista* as the result of this compound being in the ablative case, so < **cēnā-pūrā, *ipsā-hōrā, *nocte-istā*. I do not think this would make a difference, as penultimate long *-ū-* in **cēnā-pūrā* and long *-ō-* in **ipsā-hōrā* would be stressed anyway, in accordance with Latin accentuation rules (e.g. Weiss 2009: 110). Moreover, the ablative case does not explain the accent placement in **nottèsta* ‘this night’ < **nocté-istā*, as the ablative *-e* of the third declension was short, and the penultimate syllable of *nocte-istā* would have been heavy and received the stress.

but here the accent retraction is explained by Wagner (1941: 5) as a retraction from the definite article onto a preposition also seen in *déssu* 'of the' < **de ipsu*. Although the evidence is slim, I am not fully convinced that Paulis' (1992) analysis of *kampínğu* 'pine tree' as the regular reflex of Lat. *cōnus pīneus* 'pine cone' works phonologically.

Although we lack specific information about the geographic distribution of this word (only the specifically mentioned locations of Cagliari and Iglesias are marked on the map above), DES attributes it to "general Campidanese". The Campidanese dialects cover more or less the southern half of Sardinia (represented

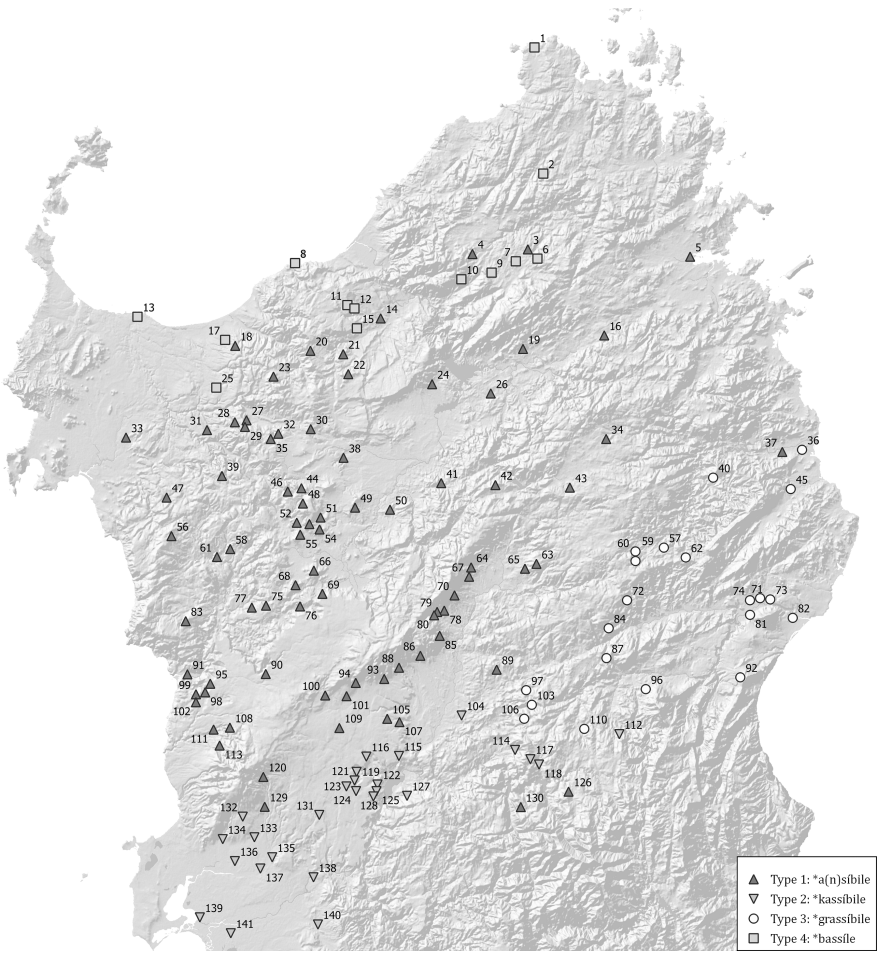


Figure 4: Marten.

by the dotted area on the map), thus matching the distribution of the type **kalukèrta* of the words for 'lizard' (§ 2.1). In light of the possibility that this word could go back to **cōnus pīneus* 'pine cone', its value as evidence for a **kV*-prefix relies on one's acceptance of this hypothesis.

3.2 **k-assíbile* 'marten'

Type 1: ****a(n)síbile***[17]
Type 2: ****kassíbile***
Type 3: ****grassíbile***
Type 4: ****bassíle***

Representatives of this word are found all over the northern half of Sardinia (Figure 4), and with considerable variation. This word cannot be explained as an inherited item, nor is there evidence that it is a more recent loan. Other names for the marten on Sardinia are *mártsu, márču*, supposedly from Vandalic **marþu-* (DES, II: 82), and *skírru, isbírru* etc. < **squiriolus* << Lat. *sciūrus* 'squirrel' (squirrels do not occur natively on Sardinia) (Wagner 1934: 486). The attested forms listed above vary in two respects: their onset and whether they end in *-íle* or *-íβile.* In the forms cited by Wagner, intervocalic *-β-* is present in *síβile* (type 1), *kassíβile* and *ʔassíβile* (type 2), and *grassíβile* (type 3).[18] All locations where *-íβile* is attested preserve intervocalic **-b-* and **-v-* as *-β-*, whereas these sounds are lost in most other dialects (Wagner 1941: 97–98). This means that **-íbile/-ívile* may be reconstructed as the original ending for this etymon. The fact that we also find *-íle* in some places where **-b/v-* is expected to be preserved (such as Bitti (60) *grassíle* etc.) can be explained by influence from neighboring Logudorese dialects, which have *assíle*.

In the onset, we find four major types: those that start in a vowel (type 1), those that start in *k-* (type 2),[19] those that start in *gr-* (type 3), and those that start

17 The data for this lemma are largely taken from Wagner's article from 1934, which goes into detail about the names of the marten on Sardinia. The map given below is thus an adaptation of the maps in Wagner (1934: 483–489). Words designating marten with a different etymology (such as *mártu*, *isbírru* etc.) are given on Wagner's maps, but are left out here. In order to clearly represent the great number of attestations, the map is zoomed in on the northern half of the island and the symbols used are somewhat smaller than on the other maps.

18 The *-β-* in *kassíβi* (135, 137, 138, 140) is the regular reflex of *-l-* in many Campidanese varieties, and is thus not relevant here (Wagner 1941: 120–121).

19 Including forms with *ʔ-*, which is the regular outcome of **k-* in the Barbaricine dialects (Wagner 1941: 69).

in *b-* (type 4). The latter onset (*b-*) is restricted to Gallurese and Sassarese; it can be explained as secondary, by analogy to the personal name *Basilio* (DES, I: 139). For the phenomenon of referring to animals by a personal name, cf. Srd. *marɻáne, marğáne* etc. 'fox' < *Mariane* (DES, II: 75). A secondary origin is also likely for the forms starting in *gr-*, which are restricted to the Nuorese dialects spoken in the island's mountainous interior (toward the east coast). Wagner (DES, I: 139) explains the forms like *grassíβile* as a folk etymology to *grássu pílu* 'greasy hair'. As it is clear that *grassí(βi)le* must be secondary, the question is whether it developed from original *kassí(βi)le* or from *assí(βi)le*. Although secondary *gr-* is not uncommon in the Nuorese dialects, it is most often the result of a metathesis involving *-r-*; e.g. (Nuoro) *graðanánka,* (Fonni) *kraðanákka* vs. (Bitti, Siniscola) *katranákka* 'tick' (DES, I: 319; cf. also Pittau 1972, 47, 53). Nevertheless, there are at least two examples where an original **k-* must have secondarily (and irregularly) developed into *gr-*:

– (Orgosolo) *granaðèllas* << *kanaðèllas* 'chalice used for water or wine during a mass' (DES, I: 279)
– (Bitti) *gríndalu* << *kíndalu* 'swift (textile)' (DES, I: 339)

On the other hand, I have not been able to find examples of secondary *gr-* to words starting with a vowel. This suggests that the forms of the type *grassíβile, grassíle* were most likely reshaped from original *kassíβile, kassíle*. This leaves us with the forms whose onsets start in *k-*, and those whose onsets start in a vowel.

The alternation between initial **k-* and **Ø-* is difficult to explain by means of inner-Sardinian processes. It is true that the Sardinian dialects contain many examples of words that start in an etymologically "incorrect" stop, or that have lost their original onset, due to consonant lenition of initial stops after a word ending in a vowel, and subsequent reanalysis and hypercorrection of pausa forms. For example, when *berdòne* 'cork' (< **quadrōne*) is accompanied by the definite article, this gives *sa erdòne*.[20] In some dialects, this syntactically conditioned form becomes the base form. In yet other dialects, a new base form is created with *g-*, which would also be lost intervocalically. The result is a seemingly irregular coexistence of *berdòne, gerdòne* and *erdòne* (Wagner 1941: 82, 207). Similar processes also occur in the case of voiceless stops, which become voiced fricatives between vowels. However, in order to explain *assíle* 'marten' from *kassíle*, we would need to assume two subsequent reanalyses of this type:

– **kassíle,* **su ɣassíle* >> **ɣassíle,* **su assíle* >> *assíle, su assíle*

20 Original *d-*, *g-* and *v-* behave identical to *b-* (Wagner 1941: 80–84).

Parallels to this double reinterpretation are very rare. One example is Latin *crux* 'cross', which is attested already in Old Sardinian as *gruke*, and which was hence reanalyzed as Log. *rúɣe,* Camp. *rúži*, which are now the normal pausa forms in most dialects. The other example is *urtéḍḍu* 'knife' < Lat. *cultellus*, with the variants *gurtéḍḍu* and *burtéḍḍu* (Wagner 1941: 206). The rareness of this development makes it unattractive to use it as an explanation for the coexistence of *kassíle* next to *assíle* etc.; especially since in the cases of *rúɣe* 'cross' and *urtéḍḍu* 'knife', the form with *k-* was completely lost, as opposed to well-attested *kassíle*. In the absence of a better explanation we must conclude that *kassíle* and *assíle* really do represent an original alternation between initial **ka-* and initial **a-*. The form *[sa] síβile*, from Fonni, must be explained as a reanalysis of the initial vowel as part of the feminine definite article: **s'assíβile* >> *sa síβile*.

The bulk of the forms can now be reconstructed as **(k)assíβile*.[21] Attested deviations from these reconstructions are *ansíle*, *essíle, asseíle,* and *assaíle* (all under type 1). Epenthesis of *-n-*, like in *ansíle*, is rather common; cf. Mores *lánsana* vs. Log. *lássana* 'mustard' < Lat. *lapsana* (Wagner 1941: 221). The same goes for shifts in the quality of pretonic vowels, as in the case of *essíle* (Wagner 1941: 24). The existence of *asseíle* (89: Orotelli) and *assaíle* (120: Santu Lussurgiu), which are found in non-contiguous areas, is more difficult to explain. Perhaps *-eí-* and *-aí-* are dissimilations of *-íi-* < **-íβi-*, but the stress placement is unexpected. The form *grassímene*, mentioned by Puddu, are perhaps influenced by neuter nouns ending in *-mene* (e.g. *númene* 'name', *fámene* 'hunger' etc.).[22] I do not know what the origin of *grassímile* could be, but it too seems secondary.[23] This word provides a clear case of alternation of initial *k-*, despite some uncertainties regarding the phonetic details.

On the map, we can clearly see that the forms of type 1 (**assíbile*) are largely restricted to the Logudorese dialects, with two Barbaricine outliers in Fonni *[sa] síβile* (126) and Ovodda *assíle* (130). Type 2 (**k-assíbile*) occurs to the south of types 1 and 3, in the transitional dialects between Logudorese and Campidanese, and in some Barbaricine dialects (which have *ʔassíβile*). Type 3 (*grassíβile, grassíle*) is found in the Nuorese dialects, and type 4 (**bas(s)íle*) in Gallurese and Sassarese in the North. Even though the entire southern half of the island uses

21 From a Latin perspective this **-β-* can represent either *-b-* or *-v-*, the distinction between which is lost in intervocalic position in all dialects (Wagner 1941: 97–98).

22 Suggested to me by an anonymous reviewer.

23 Perhaps Srd. *mèle* 'honey' plays some role; cf. Camp. *búkk'e mèli* 'weasel' (litt. 'honey mouth') (Wagner 1934: 11). One could also imagine Lat. *mēles* 'marten, badger' as a possible source of influence, but this word is itself not attested in Sardinian. However, in both scenarios we run into problems with the forms' stress patterns.

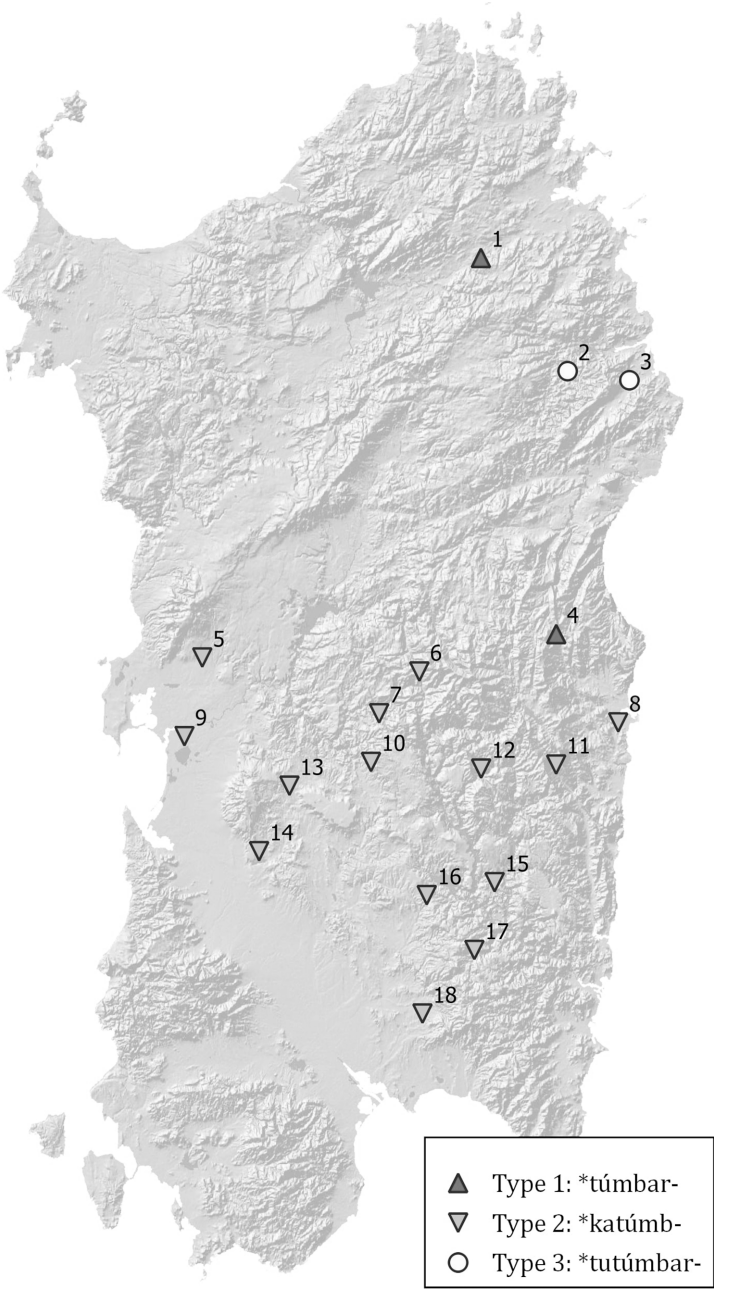


Figure 5: Mullein.

different words for marten, the distribution of type 2 (with **k-*) roughly corresponds to that of **kalVkerta* (§ 2.1) and **kampíṇiu* (§ 3.1).

3.3 **ka-tumb-* 'mullein'

Type 1: **túmbar-*
Type 2: **katúmb-*
Type 3: **tutúmbar-*

In the meaning 'great mullein (*Verbascum thapsus*)', we find forms like *kaðúmbu* (type 2) next to *túmbara* (type 1), the first of which invariably refers to 'great mullein' (cf. Figure 5). *Túmbara* (2) 'great mullein' however has a near-homonym *túmbaru* (1) that denotes 'sow thistle'. Mullein and sow thistle are two rather different, unrelated plants. They do share some characteristics, like yellow flowers and their use in traditional medicine. It is more attractive to treat *túmbara* and *túmbaru* as the result of a semantic shift than to assume the existence of two unrelated but homonymous formations referring to plant species. This is what is also proposed by Paulis (1992: 61, 354–355), who argues that *tutúmbaru* (2: Lodè, 3: Siniscola) 'spurge (*Euphorbia sp.*)' was inherited from Lat. *tithymalus* (which in turn is from Gr. τιθύμαλ(λ)ος 'milkweed (Euphorbia peplus)' Beekes & van Beek 2010: 1483–1484). This *tutúmbaru* would have been extended semantically, on the one hand, to 'sow thistle (*Sonchus oleraceus*)', due to both plants' "milk-like" sap (cf. e.g. the Dutch names *wolfsmelk* 'spurge' and *melkdistel* 'sow thistle'), and on the other hand to 'mullein (*Verbascum thapsus*)', because of both plants' use as a piscicide (cf. also Heizer 1953: 263). There are several issues with this scenario. Firstly, Lat. *tithymalus* 'spurge' appears to generally be continued in Sardinian as **θiθímbalu*; e.g. Log. *titímbalu*, Nuor. *θiθímbalu* (DES, II: 489; Paulis 1992, 170). This does not correspond regularly to *tutúmbaru* found in Lodè (2) and Siniscola (3). The same goes for *túmbaru* 'sow thistle', which Paulis explains as the result of influence from Log. *túmbaru* 'empty beehive', on account of the hollow stem of the sow thistle, citing Wagner's (DES, II: 532) hypothesis that the word for 'beehive' is a loan from It. *túbolo* 'tube'. Regardless of the likelihood of the semantic and phonetic developments needed for *túbolo* 'tube' >> *túmbaru* 'empty beehive', it is important to note that *túmbaru* itself does not mean 'tube'. Moreover, spurges (*Euphorbia sp.*) do not have hollow stems. It is thus very unlikely that this word for 'empty beehive' has had anything to do with the botanical terms discussed here, and the forms *tutúmbaru* 'spurge' and *túmbaru* 'sow thistle' remain unresolved. I believe Paulis is right in attributing the similarity between *tutúmbaru* 'spurge', *túmbaru* 'sow thistle' and *túmbara* 'mullein' to popular association of these plants because of similar uses or botanical traits, but it seems more likely that it was *túmbara* 'mullein'

that influenced the outcome of Lat. *tithymalus* 'spurge' in Lodè (2) and Siniscola (3). For *kaðúmbu* etc., Paulis (1992: 355) considers two different etymologies. On the one hand he compares it to the Hesychian gloss καράμβαν 'shepherd's staff',[24] and to Skt. *kaḍamba-* 'staff', which in turn has Dravidian comparanda. On the other hand he proposes *kaðúmbu* to be a contraction of **kánn'e ðúmbu* 'reed of the "*túmbu*" (one of the three pipes of the traditional Sardinian musical instrument *launèḍḍas*). For the Sanskrit (and Hesychian) comparison, the geographical as well as the semantic distance are simply too great for it to be of any use. The second comparison seems unlikely on basis of the fact that *launèḍḍas* are made of reed, which is not at all similar to mullein. The best option is still to treat this word as being of possible pre-Roman origin, and to analyze it within Sardinian before jumping to far-flung comparisons.

In the *kaðúmbu* type, we find a large variety of forms. Most of these can be explained by secondary insertion of *-r-*. In the case of this specific word, this may have been due to influence from *kardu* 'thistle', but non-etymological *-r-*epenthesis is a common phenomenon in Sardinian in general; cf. (Dorgali) *mastrikáre* << LLat. *masticare* 'to chew', (Mores, Ploaghe) *braśíle* << *baśíle* 'basil' (Wagner 1941: 225–227). This allows us to see *kaðúmbu* and its diminutive *kaðúmbulu* as the base form of this type.[25] The intervocalic *-ð-* in this word must go back to an original *-t-* (Wagner 1941: 69), leading to a reconstruction **katúmbu*. Hubschmid (1953: 29–33) postulates a pre-Roman **kátano-* 'great mullein', which would be cognate to **kátapo-* 'asphodell (*Asphodelus sp.*)' and **katóukʲo-* 'gorse (*Ulex europaeus*)', all derived from a root **kat-*. This requires the assumption that in some pre-Roman language spoken in the Mediterranean, **-ano-*, **-apo-* and *-óukʲo-* coexisted as derivational affixes, for which I think the evidence presented by Hubschmid is too slim. Moreover, it is not clear how **kátano-* would have yielded Srd. **katúmbu*, nor does it explain the existence of *túmbara* (4: Urzulei) 'mullein'.

This leaves us with **katúmbu* next to *túmbaru/a*. What looks like a suffix *-ara/u* occurs with some frequency in the non-inherited Sardinian lexicon, e.g. *(a)θáṇḍa* vs. *tzándh-ara*[26] 'poppy' and perhaps *bít(t)a* 'fawn' vs. *bíttara* 'female mouflon' (DES, I: 211–212; Wagner 1997, 270; Puddu 2020, 270).[27] This allows us to identify a

24 "καράμβας· ῥάβδον ποιμενικήν, ἣν Μυσοὶ συκαλόβον", 'καράμβας: shepherd's staff, which the Mysians call συκαλόβον' (Hesychius Alexandrinus 2020: 523).

25 Some of the other differences are due to local dialectal developments, such as the development of **-l-* in *kaðúmbuŕu* (S. Nicolò Gerrei, Dolianova), *kaðúmbuʔu* (Siurgus) < *kaðúmbulu*.

26 Puddu's spelling *tzándh-* corresponds to Wagner's *θáṇḍ-*.

27 This suffix has been analyzed as a collective or pluralizing suffix (Bertoldi 1937a: 164–165; 1937b). This is corroborated by the attestation of certain toponymic forms, such as *Mandara*, which is modern day *Mandas*. In this case, the *-as* in *Mandas* appears to be a romanisation of the original collective/plural suffix *-ar(a)*; another possible instance is *Gennor* next to later *Gennos* (Terracini 1927: 139).

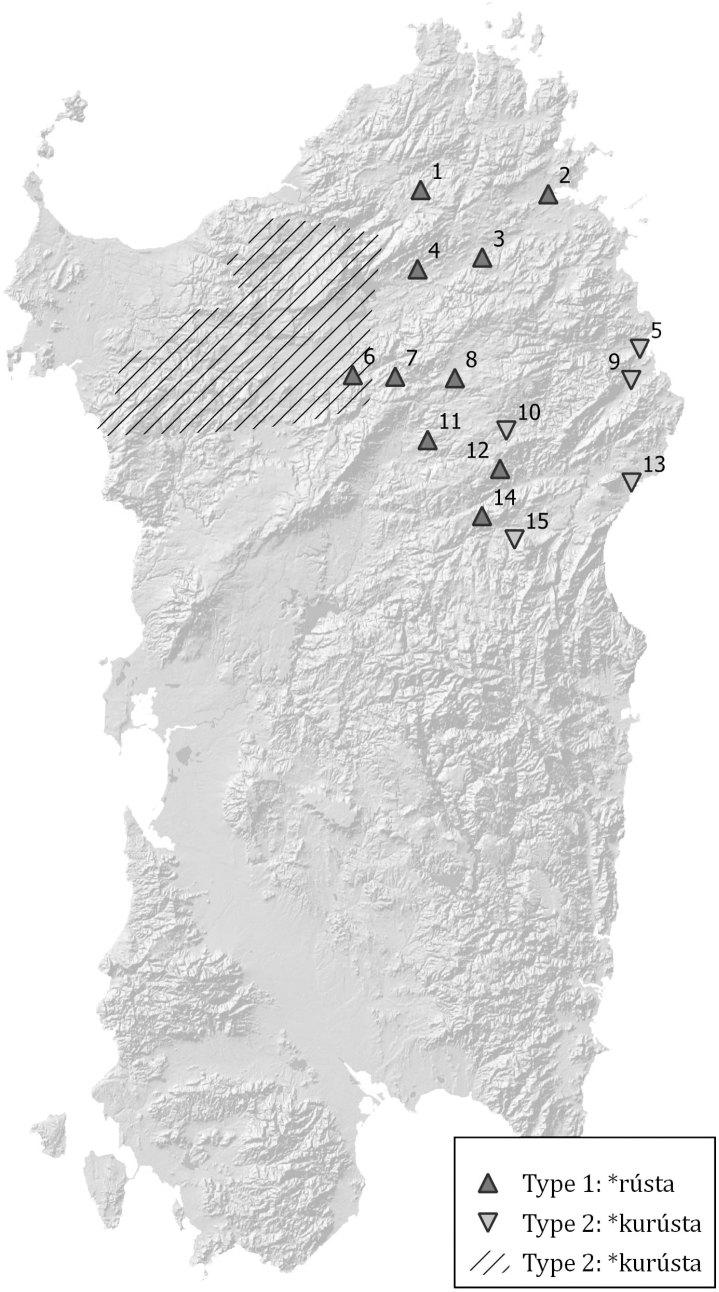


Figure 6: Bed bug, pest.

stem *-tumb-* in both **ka-túmb-* and *túmb-ara*, showing that this word too has an alternation between forms with and without initial *ka-*. We do however not find a direct alternation between **katúmbu* and ***túmbu*. The form *túmbu* does exist, but it refers to 'thyme' (DES, II: 533). On the map, we can see that the forms with **ka-* (type 2) are again found in the island's southern portion, as was the case in the previously discussed words.

3.4 **ku-rústa* 'bed bug, pest'

Type 1: ****rústa***
Type 2: ****kurústa***

This word generally refers to 'bed bug', but in some places was extended to 'pest', and more specifically 'fox'. The bulk of the forms represented in Figure 6 can be reduced to two bases: **kurústa* and **rústa*, neither of which can be explained in terms of Romance material.[28] The vocalism *-o-* instead of *-u-* in *koròsta* is difficult to explain however. The DES notes that *koròsta* is attested by Casu in North Logudorese. But Casu (2011: coròsta) does not specify any information of this kind, so it is difficult to establish what the exact provenance of this word could be. Note that in North Logudorese we would also expect *-st-* to yield *-łt-*, which would most likely have been written as ***koròlta* (see fn. 28). This form and its localization must thus be regarded as rather uncertain. Paulis (1990: 614) wants to derive Srd. *kurústa* and *rústa* 'bed bug' from Lat. *crusta* 'crust', possibly influenced by *kúrma, kúruma* 'rue (*Ruta graveolens*)'. On top of Wagner's rejection of this etymology already proposed by Rolla (1894: 55, 58) on semantic grounds, there are some formal objections to deriving *kurústa, rústa* from Lat. *crusta*. The regular outcome of this would have been ***krústa* (cf. Srd. *krás* < Lat. *cras* 'tomorrow'). Paulis argues that *rústa* could have behaved like Lat. *crux* 'cross', whose onset consonant was voiced early on to *gruke*, eventually yielding Nuor. *rúke*, Log. *rúɣe*, Camp. *grúži*. However, there is only a handful of words that, like *crux*, show initial voicing early enough to be continued as such in all Sardinian dialects (Wagner 1941: 206). Moreover, neither *crusta* nor **grusta* can easily explain the attested form *kurústa*. Paulis' suggestion that this word would have undergone influence from *kúrma, kúruma* 'rue' (because this plant was presumably used to fight bed bugs), does not solve much, because of the differences in the stress pattern. Combined with Wagner's observation in DES

28 *Rúłta* and *rúlta* are the same form. This cluster, [łt] is the regular outcome of *-st-* (as well as *-lt-* and *-rt-*) in North Logudorese and Sassarese (Wagner 1941: 191; Virdis 1988: 907).

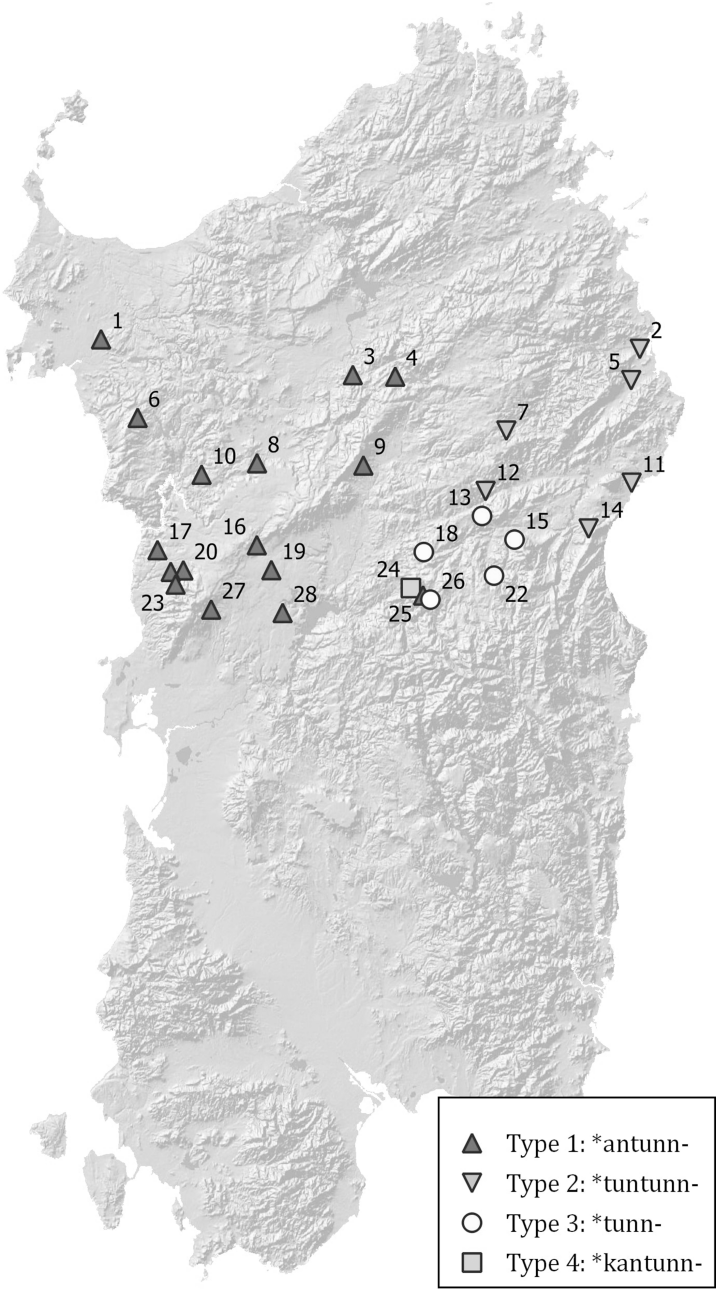


Figure 7: Mushroom.

that Srd. *kurústa, rústa* do not refer to 'bed bugs' everywhere, which makes both a derivation from Lat. *crusta* 'crust' and influence from *kúruma* 'rue (used to fight bed bugs)' semantically more difficult, I think Lat. *crusta* can in the end not be the source of this word. I rather prefer to follow Wagner's suggestion that this may be a word of pre-Roman origin. Taken at face value, the coexistence of *kurústa* and *rústa* provides a clear example of *k*-alternation.

Geographically, the forms of type 2 (**ku-rústa*) are restricted to the Nuorese dialects in the northeast. The forms of type 1 occur to the west of these, in the Logudorese dialects and in the Nuorese dialects of Orune (12) and Nuoro (14). This occurrence of forms with **kV-* in the Nuorese dialects is in contrast with the word for 'lizard', where **ka-lVkèrta* is restricted to Campidanese. It does agree with the distribution of the word for marten on the other hand, where we have Nuor. *grassíβile, grassíle* << **kassíβile* (§ 3.2).

3.5 **(k)antunn-* 'mushroom'

Type 1: ***antunn-**
Type 2: ***tuntunn-**
Type 3: ***tunn-**
Type 4: ***kantunn-**

All of the forms represented in Figure 7 seem to contain a root *-tunn-*. Four types of this can be distinguished, on the basis of the sequences before *-tunn-*. All Logudorese dialects have type 1: **antunn-*. In the Baronie region, we find reduplicated forms of type 2: **tuntunn-*. Type 3 *tunn-* (without any "prefix") is found in the dialects of Nuoro and southward. Finally, *kantúnna* (type 4) is only found in Olzai, close to the border between the types *antunn-* and *tunn-*. Type 2 may perhaps be explained as a reduplication of type 3, although the circumstances under which this happened are unclear. The single instance of type 4, from Olzai, is in close proximity to type 1 *antúnna* from Ollolai. Given the isolated attestation of *kantúnna*, it cannot be ruled out that it is a local and secondary deformation of *antúnna*, in which case it does not tell us anything about the *k(V)-*prefix (cf. Paulis 1992: 19). This leaves us with the opposition between type 1 in *antunn-* and type 3 in *tunn-*. The presence of *-n-* in all three seemingly "prefixed" forms is reminiscent of the *-m-* in *kampínğu* 'pine tree' (§ 3.1) << Lat. *pīneus*. However, this nasal is not present in *kaðúmbu* < **ka-túmb-* (§ 3.3). The simplest explanation for this discrepancy is irregular nasal insertion, which is rather common in Sardinian (Wagner 1941: 219–223).

Even if this word cannot inform us about the **k(V)-* and **θi-*prefixes, it is interesting to note that the isogloss separating the forms of type 3 from the other types lines up more or less with the dialect boundary between the Logudorese and the Nuorese dialects. Paulis (1992: 19–20) wants to derive the words for 'mushroom' discussed here from Lat. *autumnus* 'autumn'. But even if the phonetic variation can be explained away — the expected outcome would be Nuor. ***atúnnu*, Log. ***aðúnnu* —, I think the semantic development from 'autumn' to 'mushroom' without any identifiable derivational devices is too big a leap. Yet, even if these forms do not directly go back to Lat. *autumnus*, influence of this word is perhaps imaginable. Note that the Sardinian words for 'autumn' (Nuor. *atóndzu*, Log. *atúndzu*, Camp. *atúnğu, atónğu* etc.) cannot be inherited and are likely modelled after Sp. *otoño* or It. *autunno* 'autumn'. Lat. *autumnus* thus does not have regular descendants in the Sardinian dialects.

3.6 **kV-* as a prefix

Of the words discussed above, the words for 'lizard' (**ka/θi-lVkèrta*; § 2.1), 'marten' (**k-assíbile*; § 3.2) and 'bed bug' (**ku-rústa*; § 3.4) offer good evidence for an alternation of forms with and without initial **k(V)-*. In the cases of 'fool's watercress/celery' (**ku-gúsa*; § 2.2) and 'mushroom' (**k-antúnn-*; § 3.5) we find an alternation as well, but the evidence is slimmer. Nevertheless, the variation shows that **k(V)-* cannot have been part of the lexical stem of the words in question, which means that it may have been some kind of morpheme. The variants with and without **k(V)-* show little to no differentiation in meaning, suggesting that it was not a lexical derivational device, but rather a morphosyntactic morpheme, e.g. marking definiteness, grammatical number, or noun class. Another possibility are semantically neutral derivations like diminutives etc. Most of the words discussed have no straightforward Latin etymology and all of them refer to animals and plants native to Sardinia, making the identification of **k(V)-* as belonging to some pre-Roman language spoken on Sardinia attractive.

This identification as a substrate morpheme makes it all the more interesting that it is exhibited by least one inherited word: *lacerta* 'lizard'; A second instance might be Lat. *pīneus* 'pine tree', but only if *kampínğu* does not go back to Lat. **cōnus pīneus* 'pine cone' (cf. § 3.1). This tells us that the speakers of the language in which **k(V)-* was productive must have been in direct contact with speakers of Latin. The question is how and why these two inherited words were extended with this element. I think it is unlikely that Latin speakers borrowed **k(V)-* as a morpheme, as it does not explain why it was extended only to the words for 'lizard' and 'pine'. Nor is it likely that **k(V)-* was added to these words by analogy to

other words starting in **k(V)-*, since the amount of possible models is rather low and it again fails to explain the restriction of this element to only two inherited words. Rather, we should consider the scenario that the words for 'lizard' and possibly 'pine tree' were in fact also borrowed from the substrate. Especially the case of 'lizard', with the irregularly corresponding types **kalVkèrta, *θilikèrta, *alikèrta, *lVkèstra* (§ 2.1), checks all the boxes for being considered a substrate word, except for the fact that we know its Latin predecessor. In order to explain **kalVkèrta* and **kampíṇu* as substrate loans, we need to assume that these words were borrowed from Latin into the native language of Sardinia, where **k(V)-* was somehow attached, after which it was borrowed back into the local variety of Romance. Although this scenario is rather complex, I think it better explains the attested situation than assuming a grammatical loan that was productive only twice. It is also more attractive than the alternative scenario in which the native language of Sardinia had its own word for 'lizard' and 'pine' starting in **ka-*, which were perceived as similar to *lacerta* and *pīneus* by Latin speakers, causing a blend of the forms.[29] In this case, we would expect a more varied outcome of such a folk-etymological process; the reflexes of **kalVkèrta* and **kampíṇu* are rather uniform in comparison to many of the other words discussed.[30] Regardless of which scenario is to be preferred, I believe it is inevitable to assume a loan from the substrate into Latin.

As for the exact function of this morpheme, we have established that it is most likely not a semantically loaded derivational device. Rather, we may try to explain it as an element that originally belonged to the morphosyntactic realm. A parallel case could be the many Arabic words borrowed into various Romance languages, variably with and without the Arabic article *al-*. E.g. the Italian words *carciofo* and *articiocco* 'artichoke', both of which go back to Arab. *(al)-xaršūf(a)* 'artichoke' (cf. also Sp. *alcachofa* vs. Cat. *carxofa* 'artichoke'). Whereas the former form was borrowed into Italian without the article *al-*, the latter was borrowed with it (FEW, XIX: 68–69; Bramon 1987; Schrijver 1997, 293; Dworkin 2012, 107). In the case of Sardinian, this specific morpheme need not have functioned exactly like an article — other types of referent tracking or even case marking would probably behave simi-

29 If this scenario is to be preferred, this could suggest some relation between the substrate language(s) spoken on Sardinia and (one of) the pre-Indo-European language(s) that Latin came into contact with in continental Europe — neither Lat. *lacerta* nor *pīnus* are of clear IE origin (de Vaan 2008: 321, 467). As of yet there is hardly any evidence for this however.

30 Note that even if we add the reflexes of **θilikèrta*, it is only the non-inherited part (i.e. the prefix) that shows significant variation, rather than the lexical stem of the words. Only the forms of the type *luǧèstra* etc. could perhaps be viewed as the result of a folk-etymological process, but it is precisely these forms that do not contain any prefix.

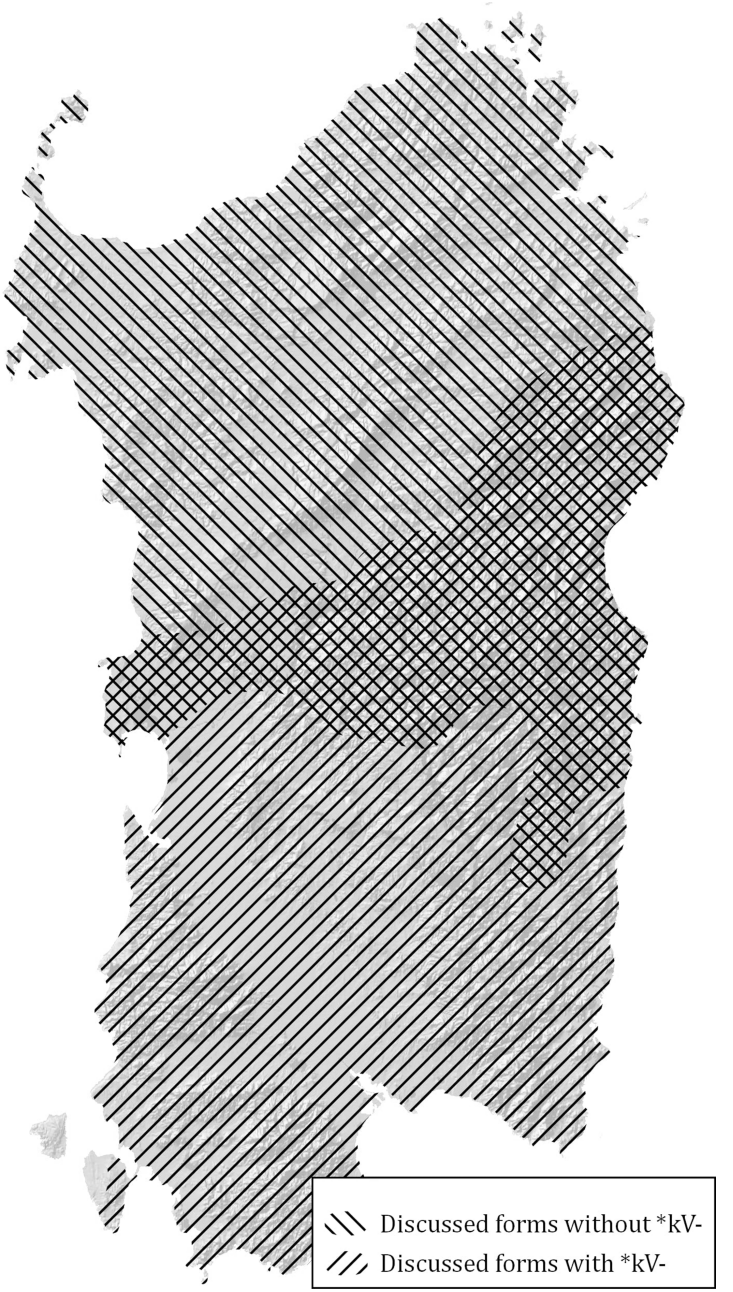


Figure 8: Variants with and without **k(V)-* discussed in § 2 and § 3.

larly — but the alternation together with the apparent absence of meaning alteration do point to a morphosyntactic identification of **k(V)-*.

The exact vowel of this prefix is somewhat uncertain. In three cases we find **ka-* (*kalužèrta, kampínğu, kaðúmbu*), in two cases **ku-* (*kurústa, kugúsa*) and in at least one case only **k-* (*kassíle* next to *assíle*; § 3.2).[31] The cases of *ku-* both have an *-u-* in the next syllable as well, suggesting that it could be due to assimilation from **ka-*. Among the three clear instances of *ka-* (i.e. *kalužèrta, kampínğu, kaðúmbu*), two are built from Latin words. If the hypothesis that they were borrowed from Latin into the pre-Roman language of Sardinia and hence back into Sardinian Romance is correct, we would not expect the vowel of **k(V)-* to behave differently from non-Latin words. Yet, the evidence is insufficient to draw any final conclusions about the quality of the vowel in **k(V)-*.

Geographically, **k(V)-* does not seem to be present everywhere on Sardinia. If we compare the distribution of the various forms that show presence of **k(V)-*, as opposed to their variants lacking **k(V)-*, the picture in Figure 8 emerges.

The forms with **k(V)-* (as opposed to those without it) have a predominantly southern distribution. In the west, the border of the **kV-*prefixed forms roughly runs along the modern dialect border between Logudorese and Campidanese. In the east, three forms (**ka-lVkèrta, *ka-túmb-* and **ka(m)-pínḭu*) are restricted to Campidanese, while **k-assíβile* (and its by-form *grassíβile*) and **ku-rústa* are found in Nuorese as well. It is in any case striking that there is virtually no evidence for forms with **k(V)-* in the Logudorese dialects in the northern portion of the island. The only possible candidate is *koròsta* 'bed bug'. DES (I: 438) has this word as North Logudorese and cites Casu, but its provenance is somewhat uncertain (cf. § 3.4). As we will see later on, this distribution is in contrast with the distribution of the **θi-*prefix discussed in the next section.

4 Further evidence for **θi-*

The following section discusses the remaining cases with potential evidence for **θi-* as a prefix. Other than for **kV-*, it also contains some words beginning in **θi-* without a prefix-less counterpart, if they have been proposed to contain it in earlier literature.

31 *Kantúnna* vs. *antúnna* could also be an instance of only **k-*, but this remains an uncertain case (§ 3.5).

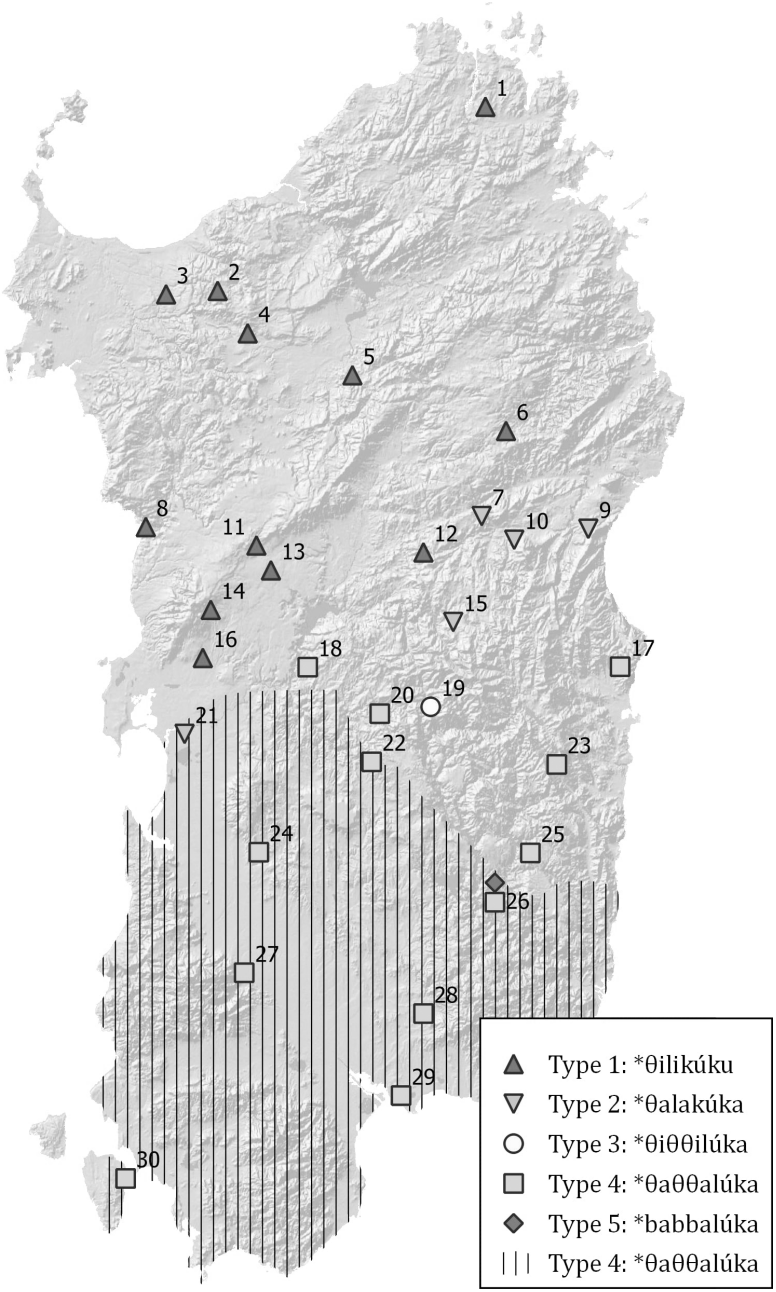


Figure 9: Skink, slug.

4.1 *θilikúku ‘skink, slug’

Type 1: ***θilikúku***
Type 2: ***θalakúka***
Type 3: ***θiθθilúka***
Type 4: ***θaθθalúka***
Type 5: ***babbalúka***
Type 6: ***alikúku***

The majority of the words represented in Figure 9 can be separated into two main groups. To the first group belong all forms of type 1 *θilikúku and type 2 *θalakúka. These forms bear an obvious similarity to *θilikèrta* and *θala?èrta* ‘lizard’ (§ 2.1), and are thus suspect of containing the *θi*-prefix. The second group includes the forms of type 3 *θiθθilúka and type 4 *θaθθalúka, that are found in the Campidanese dialects. Despite the superficial phonetic similarities between *θili-/θalakúku and *θaθθalúka, and despite their shared double meaning ‘skink; slug’, it is very difficult to reconcile the two groups formally. The types 2 *θalakúka and 3 *θiθθilúka occur in a strip between the more common types *θilikúku and *θaθθalúka. It is therefore attractive to explain *θalakúka (type 2) as originally being *θilikúku with influence from forms like *θaθθalúka, and vice versa for *θiθθilúka (type 3). This scenario is supported by the fact that most forms of type 2 are feminine like type 4, and by the Dorgalese form *θaθθalakúkka* (9) in the AIS (449), which seems to be a complete blend of the two types. If this is right, we are left with two seemingly unrelated words: *θilikúku and *θaθθalúka, both of which mean ‘skink’ as well as ‘slug’. This homonymy between a species of lizard and an invertebrate is rather understandable in light of skinks’ small, almost vestigial, legs, which give them the looks of short, fat snakes or indeed slugs. In the meaning ‘slug’, the AIS (461) further gives *babbalúɣa* (26), in which the first part is related to *babbói̯, bobbói̯* ‘generic term for invertebrates’ (DES, I: 161). Puddu (2020) also lists *alicúcu* ‘kind of long snail, but without a shell’ (type 6), which is however absent from DES and AIS.

The coexistence of *alikúku and *θilikúku in the meaning ‘slug’ is good evidence for the identification of *θi- in this word as a separable morpheme, like in *θilikèrta.* However, as mentioned before, Puddu does not supply us with information on the provenance of the listed forms. This makes it difficult to use this word for the analysis of the distribution of this feature. The part *-kúku* is reminiscent of several formations with the meaning ‘slug’ containing the element *kokk-*: *kokkói̯, kokkóile, kokkorói̯, kokkolóḍḍe* etc. (DES, I: 356), but any direct connection is problematic on phonetic grounds. There are several other phonetically similar words for slugs and snails, such as *kòrra* ‘snail’, *sittsi-ɣórru* ‘slug, snail’ << Lat. *cornus* (DES, I: 387, II: 422) and *kròka, krokkèḍḍa* ‘(various species of) snail’ (DES, I: 407).

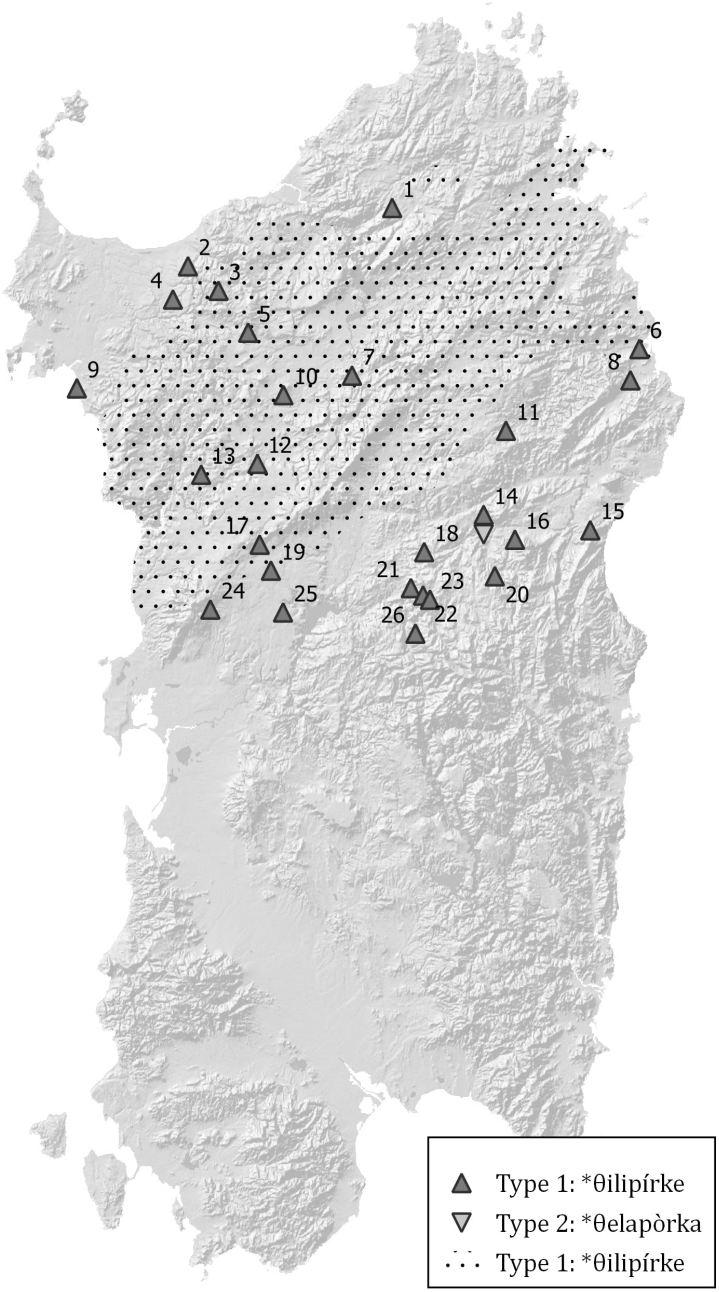


Figure 10: Grasshopper.

It is very possible that the different words for 'snail' and 'slug' influenced each other.

4.2 *θilipírke 'grasshopper'

Type 1: *θilipírke
Type 2: *θelapórka

All forms of type 1 can be reconstructed as "Proto-Sardinian" *θilipírke or *θilipílke (cf. Figure 10).[32] The form *tsilibrínkidi* (25: Norbello) has been influenced by the verb *brinkáre* 'to hop' (DES, II: 547), and *tsimpilíye* (26: Ovodda) must be due to metathesis. In Nuoro (14) we find *θilipírke* 'grasshopper' next to *θelapòrka* 'large grasshopper'.[33] The vocalism in *θelapòrka* is similar to that of e.g. *θelakúkku* 'skink' (§ 4.1). Influence by *pòrka* 'sow, female pig' (referring to this grasshopper's larger size) cannot be excluded.

Wagner (DES, II: 546–7) rightly rejects Alessio's (1939: 14–15) connection to Gr. ἀττέλαβος 'edible locust', which has been suggested to be a Semitic or Egyptian loan in Greek (cf. Beekes & van Beek 2010: 166). While there may be a superficial phonetic similarity between the Sardinian and the Greek word, there is no explanation for the Sardinian ending in *-írke/-ílke. Moreover, if the *θi- in Sardinian is a prefix like in *θilikèrta*, the comparison is reduced to Srd. *-lipírk- vs. Gr. *attelab-, which are rather too different. Wagner instead follows Schuchardt (1907: 17), who sees Srd. *θilipírke* as an internally Sardinian onomatopoeic formation. I do not think this is a satisfying solution either, as the sequence *θi(li-)* also recurs in words for animals that do not sound like grasshoppers at all (e.g. *θilikèrta* 'lizard', *θilikúku* 'skink, slug'). In my opinion, it is more reasonable to assume that this word was borrowed from a pre-Roman language on Sardinia.

32 This word is not attested for the dialect of Baunei, which preserves the distinction between *-rk- and *-lk- (cf. Wagner 1941: 176–77). *Ziliblich* (8) from Algherese (a Catalan dialect spoken in the Sardinian city of Alghero) is difficult to interpret. According to Blasco Ferrer's historical phonology (1984: 67–68) *-bl- should have yielded **-br- (e.g. Algh. *dóbra* vs. Cat. *doble*). Nor do syllables of the type *-CVLC- metathesize, as opposed to syllables of the type *-CVRC- (Blasco Ferrer 1984: 93–94). In short, the form *ziliblich* looks like it is not supposed to exist. A few lines after mentioning *ziliblich*, Schuchardt (1907: 17) mentions the form *silibrich volador* 'dragonfly', also from Alghero. This form is probably identical to *ziliblich*, and is the expected outcome of *θilipírk-. I do not know whether it could also reflect *θilipílk-, so the original liquid here remains uncertain.

33 The notion 'large grasshopper' perhaps refers to the great green bush-cricket (*Tettigonia viridissima*) as opposed to smaller kinds of grasshoppers like locusts (*Acrididae*) or pygmy grasshoppers (*Tetrigidae*).

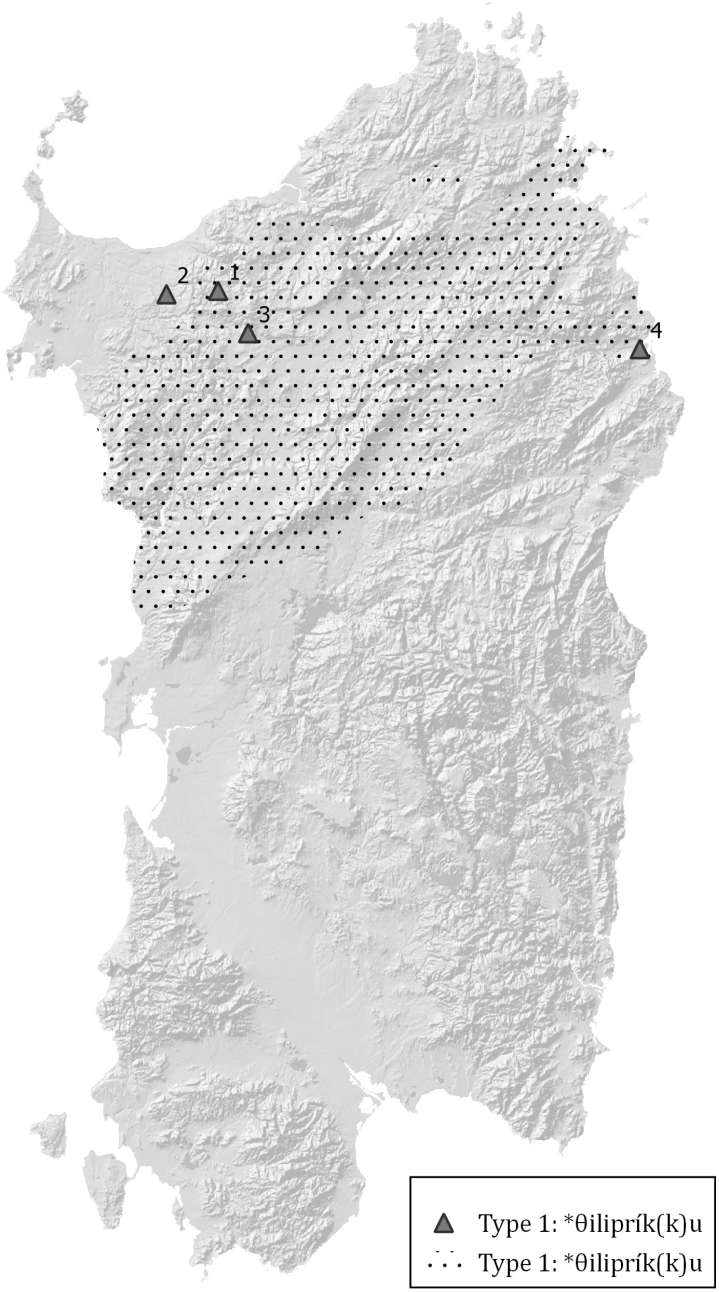


Figure 11: Kestrel.

4.3 *θiliprí(kk)u ‘kestrel’

Type 1: ***θiliprík(k)u***

The words for ‘kestrel’ represented in Figure 11 all occur in the northern part of Sardinia — the dotted area on the map above reflects the mention in DES of *tilibríkku* as the general Logudorese form. The forms go back to **θiliprík(k)u/*θiliprí(g)u*. There is no alternation of **θi-* in this word, but the morphological similarity to the etyma discussed above is obvious. DES (II: 483) cites Sass. *tir’ibíkku* ‘kestrel’, which is identical to the Sassarese attestation in AIS for ‘grasshopper’ (there spelled as *tiribíkku*; § 4.2). Since neither of the sources have this form in the other meaning, it is possible that they were confused either in DES or in AIS.

4.4 *θilingròne ‘earthworm’

Type 1: ***θiling(ul)òne***
Type 2: ***θilíng(ul)a/u***
Type 3: ***θulung(ul)òne***

This word for ‘earthworm’ is widely distributed on Sardinia (cf. Figure 12). All of the forms listed above are built from the base **θiling-* (type 1) or **θulung-* (type 3). In both types, most forms have been suffixed with **-(u)l-*, as is apparent from e.g. *tilinğòne* < **θiling(u)lòne* and *θulungròne* < **θulung(u)lòne*. It is possible that the presence of **-(u)l-* is original, and that it has been lost in the few dialects that do not have it. In the Campidanese dialects, this is regular (cf. Wagner 1941: 161). Most forms also contain the very common suffix *-òne*, with the exception of the forms in type 2. As for the vocalism, it is attractive to explain the difference between *-i-i-* (type 1) and *-u-u-* (type 3) as the result of some assimilation or dissimilation of pretonic vowels. Even though the Nuorese dialects generally preserve the original pretonic vowel distinctions, it is exactly there that we find *θilingròne* (Nuoro) next to *θurunkròne* (Bitti). As the forms without *-òne* (e.g. *tilínğa*) have stressed *-í-*, it is reasonable to assume that this must be the original vowel. The *-u-* vocalism of type 3 may be due to assimilation to the rounded vowel of stressed *-òne*. We may thus reconstruct this word as **θiling(u)l-(òne)*. There is no direct evidence for the identification of **θi-* as a genuine prefix in this word, but it patterns well with the other discussed instances of **θi-* both morphologically and semantically.

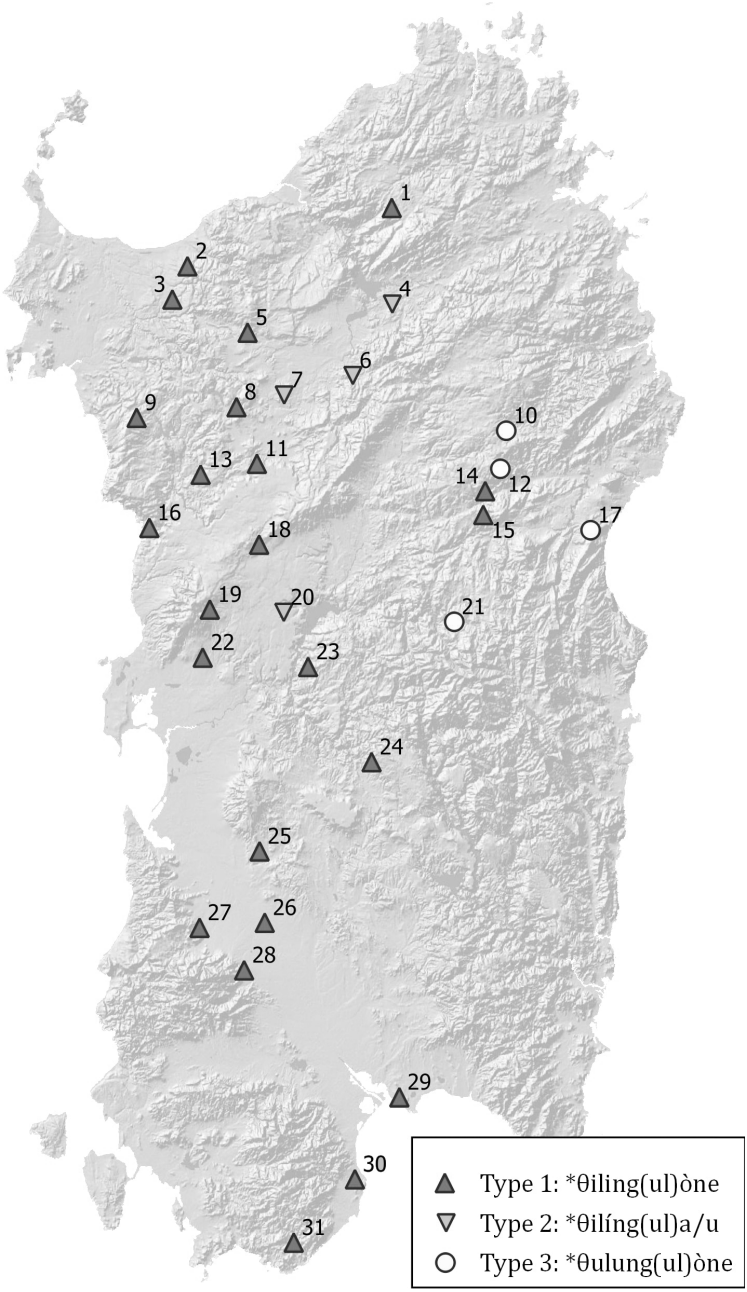


Figure 12: Earthworm.

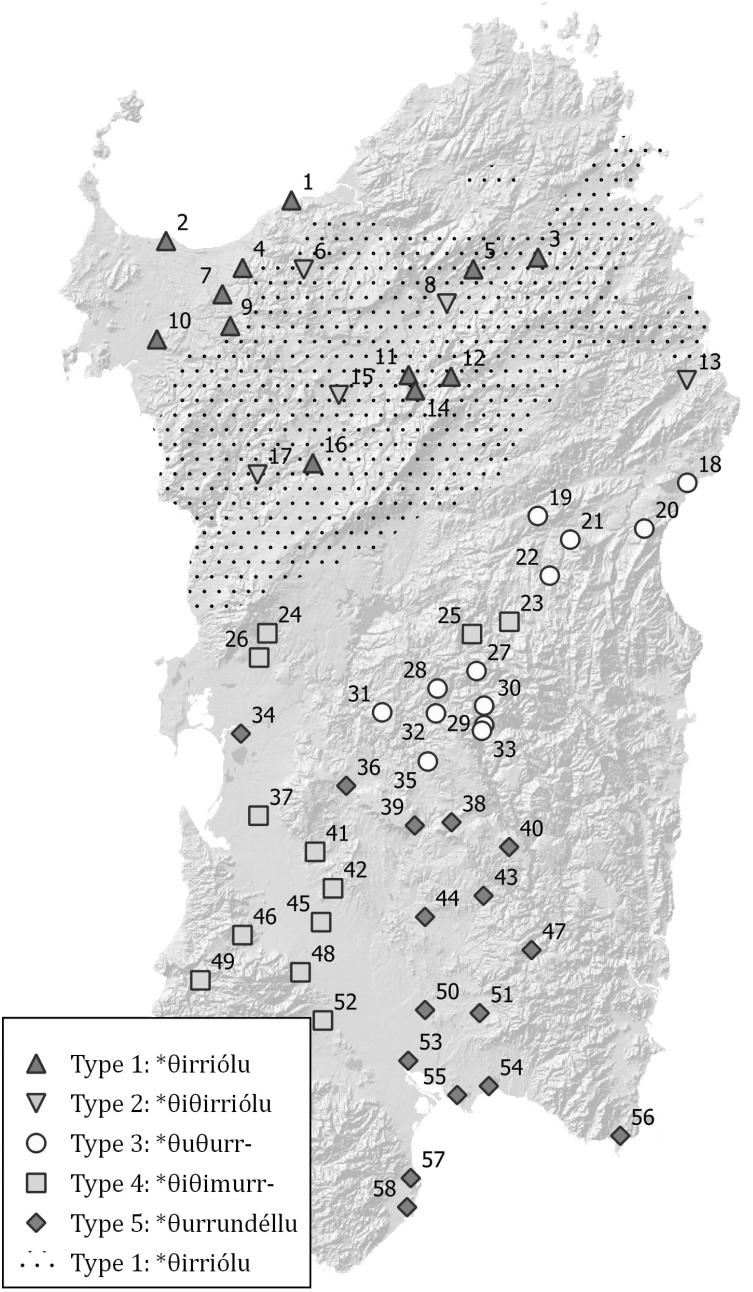


Figure 13: Bat.

4.5 *θiθirr̰iólu ‘bat’

Type 1: *θirriólu
Type 2: *θiθirriólu
Type 3: *θuθurr-
Type 4: *θiθimúrr-
Type 5: *θurrundéllu

The words for ‘bat’ on Sardinia are diverse (cf. Figure 13). There are additional types that are not etymologically related to the forms discussed here and which have therefore been left out; cf. Terracini (1959) for an overview. Even though the five types presented show considerable variation, some of it may be explained as the result of folk etymologies etc. In type 3, most forms need *θuθurr-. However, in Gadoni (33) we find *tsintsurréḍḍu*, with *θiθurr-.[34] The forms with *θu-* are probably due to assimilation of *θiθurr- > *θuθurr-. DES (II: 556) notes about *tirr̰iólu, θiθirr̰iólu* etc. (types 1 and 2) that they seem to have been influenced by *θirr̰iare* ‘to screech’.[35] Since most other forms contain the base *θurr-, it is attractive to assume that types 1 and 2 originally had *θurr- as well. The suffix *-i̯ólu* is the regular Sardinian outcome of *-eolus/-iolus*, the Latin diminutive suffix after stems in *-i* and *-e* (Weiss 2009: 280). Type 2 (*θiθirr̰iólu*) is identical to type 1 (*θirr̰iólu*), but with initial *θi-.

In type 4 (*sittsimurréḍḍu* etc.), the DES assumes influence from *múrru* ‘grey’, which is not implausible. There is however also *múrru* ‘muzzle’, which could have caused the same influence. The question is what the underlying form was. It seems as if, rather than replacing *-θurr-, *-murr-* was “inserted” into original *θiθurr- >> *θiθimurr-. I do not know how common the insertion of an extra syllable is typologically, but the alternative is that *-murr-* replaced *-θurr-, in which case we would need to assume *θiθimúrr- << *θiθiθúrr-, which seems unlikely. In *tsurrundéḍḍu* (Type 5), Terracini (1959: 15) and Guarnerio (1904: 68, 259) see influence of Lat. *hirundō* ‘swallow’. Wagner (DES, II: 556) expresses his doubts about this because all words for ‘bat’ have a geminate *-rr-* rather than single *-r-* in Lat. *hirundō*. But, the reflex of *hirundō* in the Campidanese dialects where we find

34 The diminutive suffix *-éḍḍu* < Lat. *-ellus* is highly productive (Wagner 1952b: 106). Sardinian *-éri* entered the language in loans from Tuscan (with *-eri, -iere*) and Catalan (with *-er*), after which it became productive as a nominalizing suffix (Wagner 1952b: 75–77). It also occurs in some Sardinian words of unknown etymology, e.g. *tserpeḍḍéri* ‘kestrel’, *ninn̰éri* ‘dog rose’ and indeed *θuθurréri* ‘bat’. Whether it should be analysed as the Tuscan/Catalan suffix in these cases is uncertain.

35 In the same types, many forms have been compounded with *pèḍḍe* ‘skin’.

tsurrundéḍḍu ‘bat’ is in fact *arrúndili* ‘swallow’ (with secondary *a-* by regular prothesis before initial *r-*) (DES, II: 370), removing Wagner’s objection.

By “undoing” the different folk etymologies in the different words for ‘bat’, we are left with the following pre-forms for the various types: 1) **θurr-*, 2) **θi-θurr-*, 3) **θi/u-θurr-*, 4) **θi-θurr-*, 5) **θurr-*. This is another clear instance of alternation between presence and absence of **θi-*. One could perhaps argue that the additional **θi-* in types 2, 3 and 4 could be the result of a reduplication, but as **θi-* also occurs in other words where it cannot be explained by reduplication (i.e. all of the other words with **θi-* discussed in this study), I do not think that this is an attractive scenario.

As for the distribution of the forms, type 3 (**θiθurr-*) occurs in the Nuorese and Barbaricine dialects. Type 4 (**θiθimurr-*) is restricted to the western Campidanese dialects, but is intersected by type 5 (**θurrundéḍḍu*), probably due to a more recent expansion of the latter (cf. Terracini 1959: 15). Types 1 and 2 (respectively **θirṛiólu* and **θiθirṛiólu*) occur next to each other in the Logudorese dialects. The intricate distribution of all of these forms is probably at least partially the result of more recent expansions of the folk-etymological forms that lie at the bases of types 1, 2, 4 and 5 (cf. Terracini 1959). Being the *lectio difficilior*, type 3 is probably most original, allowing us to reconstruct **(θi-)θurr-* ‘bat’.

4.6 **θúklu* ‘neck’

Type 1: ****θúk(u)lu***

This is the normal Sardinian word for neck. It is present in all Sardinian dialects (hence the absence of a map), but not in the Sardo-Corsican dialects spoken on the island (cf. Sass. *lu ggóḍḍu*, Gall. *lu kóḍḍu*, AIS 118). Moreover, all forms can go back regularly to a “Proto-Sardinian” **θúk(u)lu*. If it were not for our extensive knowledge of Latin, the internal regularity and the omnipresence in the Sardinian dialects would make this a perfect candidate for an inherited word, but **θúk(u)lu* is of course not the Latin word for ‘neck’. Wagner rejects all previous attempts of etymologizing this word and settles on deriving it from Lat. *iugulum* ‘collarbone’, with a prefixed **θ-* (DES, II: 553). Besides the prefixation of a non-inherited and probably pre-Roman element **θ(i)-*, two assumptions are needed for this derivation. First is the semantic development from ‘collarbone’ to ‘neck’, which is unproblematic. Second is the devoicing of *-g-*: *iugulum* >> *iuculum* >> **θ-iuculum* > **θúkulu*. This too is rather trivial in Sardinian, where many words get the ending **-k(u)lu* secondarily on the basis of the frequent inherited suffix *-culus* (cf. Wagner 1952b: 29–30). Blasco Ferrer (1999: 67) instead believes Srd. **θúk(u)lu* ‘neck’ to regularly go back to

Lat. *sucula* 'screw/mechanism for lifting a wine/oil press' (cf. Lewis and Short 1879: sucula). Not only is this semantically quite a leap, his argument that *θ-* would be the regular outcome of **s-* is solely based on the form *suglu* he adduces from Baunei. However, not only is the form *θúɣulu* with *θ-* documented for Baunei in the AIS (118) and DES (II: 553), the *s-* in *suglu* could be from the neighbouring Ogliastran dialects, where *-s-* is in fact the regular outcome of **-θ-* regardless of its origin (Wagner 1941: 107). Blasco Ferrer's (1999) etymology must thus be rejected.

If Wagner's etymology is correct, this means that *θúkru* etc. 'neck' is another possible candidate for secondary *θ(i)-*prefixation of a word inherited from Latin. However, Sardinian *-θ-* is the outcome of *-t̯i̯-* or *-k̯i̯-* (Latin <ci>) before a vowel, so in the case of *iugulum*, an initial **t-* or **k-* would work equally well. Additionally, we have no information about what the regular outcome of **θju-* or **θiju-* would be, since a sequence **t̯i̯u-/*k̯i̯u-* did not occur in Latin. We are thus forced to choose between **t̯i̯úkulu/*k̯i̯úkulu* that is phonetically warranted but morphologically opaque, and **θi̯úkulu* which is phonetically uncertain but gives us at least some parallels when it comes to its derivation (even though these parallels still require an explanation involving a substrate language). For now it seems best to regard this word as no more than a possible example of secondary **θi-*addition.[36]

4.7 **θi-* as a prefix

The sequence **θi-* has been discussed much more than **k(V)-* as a possible substrate prefix. Clear cases of alternation between presence and absence of this potential prefix are **θilikèrta* 'lizard' (vs. **kalVkèrta* and Lat. *lacerta*; § 2.1), **θilikúku* 'skink' (vs. *alikúku*, for which the place of attestation is not known; § 4.1), *tzeligúsa* and *atigúsa* 'celery' (with unknown place of attestation, but next to *kugúsa, θurgúsa*; § 2.2), **θiθirr-*, **θuθurr-* 'bat' (vs. **θirr-*, **θurr-*; § 4.5) and possibly **θuk(u)lu-* 'neck' (if from Lat. *iugulum* 'collar bone'; § 4.6). Among these, the words for 'lizard', 'celery' and 'bat' provide strong evidence that **θi-* does indeed behave in the same way as **k(V)-*. The other words — **θilipírke* 'grasshopper', **θilibríkku/*θilibrígu* 'kestrel' and **θilingròne* — do not show any alternation, but have been identified as instances of the **θi-*prefix due to obvious similarities. Like in the case of **k(V)-* the presence of this element on words of obscure origin, most of which refer to native Sardinian animals and plants, make a substrate origin very likely.

36 I agree with my anonymous reviewer that it might be easier to regard Srd. **θúg(u)lu* 'neck' as a pre-Roman word as a whole, rather than Wagner's complicated scenario.

What is more is that there is no regular Latin source for the sequence **θi-*. In Sardinian, the phoneme *-θ-* (and its dialectal variants *-t-* and *-ts-* etc.) is the regular outcome of (Vulgar) Latin **-t̂i̯-* and **-k̂i̯-* (e.g. Srd. *púθθu* 'well' < **puti̯u* < Lat. *puteus*; Srd. *fáθθo* 'I do/make' < Lat. *faci̯ō*) (Wagner 1941: 109–111). This means however, that Srd. *θi-* should go back to something like **t̂i̯i-* or **k̂i̯i-*, neither of which occurs in Latin. Although *-θ-* is also found as the Sardinian adaptation of various foreign phonemes in borrowed words (e.g. Gr. -θ-, -τζ-, It. *-z(z)-*, Sp. *-ch-* etc.), none of the words discussed above can be explained as a loan from a known language. This is good additional evidence for the identification of *θi-* as a substrate element, and it gives us a glimpse into the phonology of the source language. As the situation is very similar to that discussed for **k(V)-* (§ 3.6), I believe that for **θi-* as well, a morphosyntactic analysis is most plausible. And in light of *θilikèrta* << *lacerta*, it too must have belonged to a language that was in direct contact with Latin.

For the (partially) inherited word *θilikèrta* 'lizard', the same considerations apply as to *kalužèrta* etc. (§ 2.1) 'lizard' and *kampínğu* 'pine tree' (§ 3.1). That is, I think it is unlikely that *θi-* was borrowed as a morpheme and attached to an inherited word in only one instance. More than for *kalužèrta* however, one could argue that the addition of *θi-* to Lat. *lacerta* could have been by analogy to other (non-inherited) words denoting small animals, such as *θilikúku* 'skink', *θilipírke* 'grasshopper' and *θilingròne* 'earthworm'. This would also explain the fact that we find *θi-likèrta*, rather than ***θi-lakèrta*. For **θúk(u)lu* 'neck', it is much more difficult to envisage the mechanism by which an inherited **i̯úkulu* << *i̯ugulum* would have been extended with **θ(i)-*. Analogy to other words with *θi-* is very unlikely, as none of them refer to body parts, and it lacks the characteristic sequence **θili-* that was adopted by the word for 'lizard'. In light of the train of assumptions needed to make an etymology in *iugulum* 'collarbone' work, I am very hesitant about seeing **θúk(u)lu* as an instance of **θi-* as a separate morpheme at all.

When we compare the geographical distribution of **θi-* to that of **k(V)-* (§ 3.6), we find a rather different situation, shown in Figure 14.

This map represents all discussed forms plausibly containing **θi-* as opposed to their variants lacking it. The most striking contrast is that forms with the **θi-* prefix are overwhelmingly found in the northern part of the island, whereas the **k(V)-*prefix is restricted to the south. Only the words for 'earthworm' and 'bat' have representatives on the Campidanian plain in the southwest, but the southeast is devoid of **θi-*. Equally striking is the sharp distinction between the Ogliastra and Barbagia regions in the east of the island when it comes to the present of this element. In the west, the situation is much more less clear-cut. This situation is almost opposite to that of the forms with **k(V)-*, where we find a wider transi-

Figure 14: Variants with and without **θi-* discussed in § 2 and § 4.

tional zone in the east than in the west. One point of caution is that this map excludes forms like **θaθθalúkka* 'skink' (§ 4.1), which are found all across the southern dialects. Due to these forms' obvious resemblance to northern **θilikúku*, one might wonder whether **θa-* or **θaθθa-* is perhaps a (southern) variant of **θi-*, but it is outside the scope of this study to entertain this hypothesis.

5 Conclusion

5.1 The relation between **kV-* and **θi-*

The question that remains is what the exact relation is between **k(V)-* and **θi-* discussed in this study. We can make several observations. One is that both show a similar alternation between absence and presence on multiple words, without any apparent change in meaning. This suggests that they did not belong to the lexical stem of the words in question. Rather, both were some separable morpheme with a low semantic load. Attractive candidates are morphosyntactic morphemes like articles etc., as already proposed by Pittau (1995: 197).[37] Their similar distributions would also suggest a similar original function. Another common factor is that both prefixes are attached to one word that is clearly inherited from Latin: *lacerta* 'lizard'. Besides that, **k(V)-* is found in *kampínğu*, which probably contains Lat. *pīneus*, and **θi-* might be present in *θúkru, θúɣulu* etc. 'neck' which could be from *iugulum* (which is much less clear however; cf. § 4.6). Regardless of the exact process by which these elements came to be attached to Latin words, it strongly suggests that the speakers of the language(s) in which **k(V)-* and **θi-* were productive morphemes were in direct contact with speakers of the Sardinian variety of Latin (rather than a situation in which **k(V)-* and **θi-* were common but non-productive elements in the donor language). However, this does not necessarily imply that the two prefixes belonged to one and the same language at the same time.

In fact, possible evidence against **k(V)-* and **θi-* existing next to each other in the same pre-Roman linguistic stratum is found in their geographical distribution. Whereas **k(V)-* predominates in the south and is completely absent from the northwest, **θi-* predominates in the north and is completely absent from the

37 I regard Pittau's connection to the Etruscan demonstrative pronoun *ta* as unproven for now, as **ta-* could not have yielded Srd. *θi-*. Our limited knowledge of Etruscan (especially its lexicon) makes it difficult to substantiate any links between it and any pre-Roman elements found on Sardinia (cf. § 1.2).

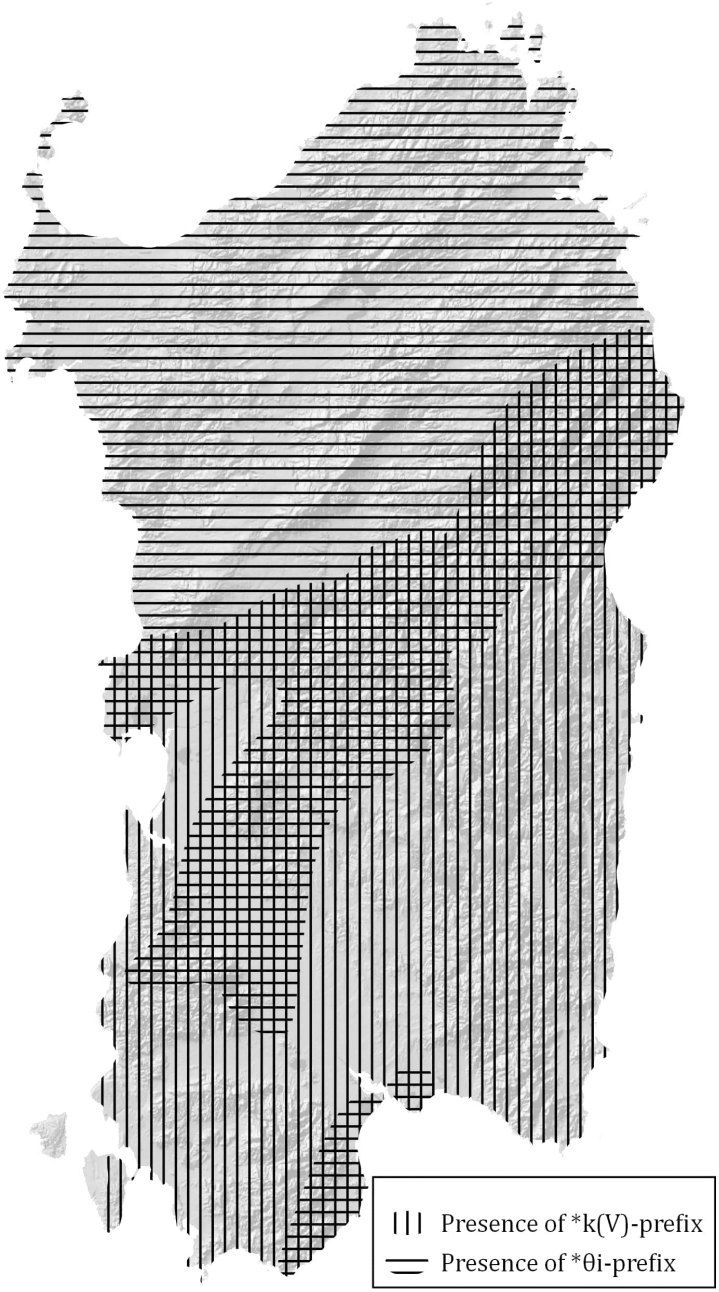


Figure 15: Distribution of **k(V)-* and **θi-*.

southeast (cf. Figure 15). Nevertheless, in the words where we find variation between absence and presence of both prefixes, we do find the prefix-less forms also in the areas where only the other prefix is attested. For example, forms like *k-assíle* 'marten' are found in the southern half of Sardinia where *ka-* is common, but their un-prefixed variant *assíle* is well-attested in the north, where we do not have any evidence for **k(V)-*. And vice versa, *θurrundéḍḍu* 'bat' is attested in the southeast, where *θi-*prefixes appear to be absent, while its prefixed counterparts *θuθurréḍḍu*, *titirr̥ólu* etc. exist in the north. So even though the prefixes themselves seem to be restricted geographically, the lexical stems to which they are attached are clearly not. This strongly suggests that the language varieties from which the forms with **k(V)-* and **θi-* were borrowed into the Romance language of Sardinia must at least have been related to each other.

5.2 **kV-* and **θi-* and the Sardinian substrate

I think there are three possible scenarios that could explain the relation between **k(V)-* and **θi-*. These scenarios follow from the observations that 1) both must have belonged to a language that was in contact with Latin, 2) they probably had a very similar function, and occurred on the same lexical items (although never together), but 3) they show different geographical distributions. The second observation suggests that **θi-* and **kV-* were mutually exclusive, which may point to them being variants of some kind of the same morpheme. It is this variation that requires an explanation.

The first option is that **k(V)-* and **θi-* did belong to the same language and had similar, but slightly different functions. In the case that their function was that of an article or something similar, the difference between **k(V)-* and **θi-* may have been one of gender/noun class, proximity etc. In that case, the geographical discrepancy between **k(V)-* and **θi-* could be an artefact of the borrowing process, with Romance speakers in the north preferring one form and those in the south preferring another and a transitional zone in between.

The second possible scenario is that **k(V)-* and **θi-* were one and the same morpheme, but phonologically different due to dialectal differences in the Pre-Roman language to which they belonged. Recall that, in terms of inherited material, **θi-* would go back to something like **k̯i̯-/*t̯i̯-*. We might speculate that one dialect of pre-Sardinian had a morpheme /ka/ > [ka],[38] while another had /ki/ > [či] or [kʲi], which were borrowed respectively into Romance Sardinian as **ka-* and **k̯i̯-* > **θi-*. Note that a development like *ki-* > *či-/kʲi-* must have taken place in

38 On the original vowel of **k(V)-*, see § 3.6.

the substrate language, since the Logudorese and Nuorese dialects of Sardinian are in fact the only living Romance varieties that did not palatalize inherited *-k-* before front vowels (e.g. Srd. *iskíre* < Lat. *scīre*).[39]

The third possibility is that **k(V)-* and **θi-* were originally one and the same morpheme, but separated in time rather than in space. If we assume that the original shape of the morpheme was **k(V)-*, this could have been borrowed into the Latin of the earliest Roman settlers of Sardinia. If the substrate language underwent a phonological shift at some point by which this specific sequence was palatalized to **čі-/k^{j}i*, this may have been borrowed by later generations of Latin speakers as **kі̯i-* vel sim., eventually developing into Srd. **θi-*.[40] This scenario has the benefit over the second scenario that it may be able to explain the zone where we find overlap between **k(V)-* and **θi-* as the result of prolonged contact between Latin and the indigenous language. The cline we see, with **k(V)-* in the southeast, **θi-* in the northwest and a strip from southwest to northeast where both areas overlap, would suggest that Latin spread from the south, and that the substrate language was maintained longest in the north. This is consistent with the romanization process of Sardinia as envisaged by Virdis (2018), according to which the northern parts of the island were romanized later than the south, by a somewhat more innovative variety of Latin, which was at the same time more archaizing.[41] A drawback of this scenario is that we need to assume that Lat. *lacerta* was borrowed into the substrate language, spread throughout the entire island, and was then borrowed back into Sardinian Romance on at least two independent occasions (i.e. "early" **kalakèrta* >> Srd. *kalužèrta* vs. "late" **čilikèrta* >> Srd. *θilikèrta*).

The reality that caused the presently observed situation is probably much more complex than any of the three scenarios sketched above. Moreover, none of them are mutually exclusive. We could for example be dealing with two morphemes that were grammatically distinguished in one area, but homophonous in another, or with different dialectal forms with a subsequent spread of one of them. The fact that the zone where **k(V)-* and **θi-* seem to overlap is also the area where

39 It is of course also possible that the phonetic realization [θ] (as found in Nuorese) and/or [ts] (as found in Campidanese) were already present in the pre-Roman language. This does not change anything about the possibility that **θi-* was a palatalized variant of **k(V)-*. For a discussion of the original value of the phoneme **-θ-*, cf. Wagner (1941: 107–108).

40 The exact phonetic value of these stages is of course impossible to determine with absolute certainty. As mentioned before, it is also possible that this phoneme was borrowed into Sardinian Latin already as some kind of fricative **-θ-* or affricate **-ts-/*-tθ-* (cf. Wagner 1941: 107–108).

41 As opposed to the views of Wagner (1928), Blasco Ferrer (1989) and Lupinu (2005), who argue for an initial wave of romanization that reached all parts of the island, after which the south continued to be influenced by more innovative stages of Latin. Cf. also Mastino (1993: 514), who rather sees the various stages of romanization as a function of the distance to the coast.

we find transitional dialects between Logudorese and Campidanese (the Arborense area in Virdis 1988: 905–906) make it probable that some of the distributions of the words discussed might be due to more recent dialectal spreads.

The material presented provides a good example of the open methodological questions in substrate research discussed in § 1.4. If one prefers to accept the various alternative etymologies that have been proposed for words displaying **kV-* (§ 2, 3), involving irregular developments of inherited material as the result of contaminations or folk-etymologies etc., the case for **kV-* (more so than for **θi-*, but cf. Alinei (1984: 29–30)) as an identifiable pre-Roman morpheme is seriously weakened. I however believe that it is more economical to explain seven examples of poorly understood **kV-* alternation as the result of a single substrate phenomenon than to propose seven cases of irregular development for each instance individually. Regardless, I hope the present discussion will be able to further the debate on methodology regarding linguistic substrate research.

In summary, there is good evidence for the identification of two elements, **k(V)-* and **θi-*, as variants of the same morpheme, either morphologically conditioned or due to dialectal or diachronic differentiation, in the language directly preceding Latin on the island of Sardinia. While their relative geographical distributions allow for some tentative speculation about the linguistic make-up of pre-Roman Sardinia, this study only treated a small part of the non-inherited portion of the Sardinian lexicon. Further investigation into the remaining linguistic material is bound to provide deeper insights into the linguistic history of Sardinia and possibly its surroundings.

Appendix A: phonetic transcription of Sardinian

β [β]; č [ʧ]; ć [tɕ]; ḍ [ɖ]; ð [ð]; é [e]; è [ɛ]; ğ [ʤ]; ɣ [ɣ]; i̯ [j]; ɬ [ɬ]; ṇ [ɳ]; ñ [ɲ]; ó [o]; ò [ɔ]; ʀ [ʀ]; š [ʃ]; ṡ [z]; θ [θ]; χ [χ]; y [j]; ʔ [ʔ]; ʀ [ʀ]; š [ʃ]; ṡ [z]; θ [θ]; [χ]; y [j]; ʔ [ʔ].[42]

42 This transcription is only used for sources that use a phonetic transcription themselves, such as Wagner (1934; DES), AIS, Terracini (1959) and Paulis (1992). Where different sources use different transcriptions for a certain sound, I have chosen to use the IPA character. Sources that use an orthography based on Italian orthography, such as Marcialis (2005) and Puddu (2020), have not been transcribed. In these sources, <tz> is generally used for [θ] and [ts], and the voiced fricatives [β], [ð] and [ɣ] are represented with <b, d, g>. Puddu (2020) further uses for [ɖ].

Appendix B: lexical data

Forms represented in § 2.1: ‘lizard’

DES (II: 545):

Type	Forms
1: ****θilikèrta***	***tiriyètta*** (Sass.), ***tsirikèlta*** (1: Tempio Pausania), ***tiliyèʧʧa*** (2: Ploaghe), ***tiliyèrta*** (3: Posada, 4: Torpè, 8: Bono, 16: Sennariolo, 17: Cuglieri, 20: Santu Lussurgiu) ***θilikèrta*** (5: Siniscola, 6: Bitti, 9: Orosei, 10: Nuoro), ***tiliyèlta*** (7: Bonorva), ***θilikètta***, ***tilikètta*** (11: Dorgali), ***θiliʔèrta*** (12: Oliena, 18: Ollolai), ***attiliyèrta*** (13: Macomer), ***ateliyèlta***, ***attaliyèlte*** (15: Scano di Montiferro), ***tsiliyèrta*** (14: Borore, 21: Norbello), ***tsiliyètta*** (25: Busachi, 31: Samugheo) ‘Tyrrhenian wall lizard (*Podarcis tiliguerta*)’
2: ****θalakèrta***	***θalaʔèrta*** (19: Gavoi) ‘id.’
3: ****kalVkèrta***	***kalužèrtula*** (general Camp.), ***karǧilètta*** (40: Seui), ***kalužèrta*** (41: Sadali), ***kalažèḍḍa*** (44: Escalaplano), ***kaʀišèḍḍa*** (45: San Nicolò Gerrei), ***kaʔižèḍḍa*** (47: Muravera), ***kožuètta*** (49: Sant’Antioco) ‘id.’
4: ****(a)li/ukèrta***	*[s’]* ***aliyèrta*** (27: Tonara, 28: Desulo), ***lulǧètta*** (30: Aritzo), ***anǧulètta*** (32: Meana), ***aliyèlta*** (35: Gadoni), ***aližèrta*** (37: Laconi) ‘id.’
5: ****(a)li/ukèstra***	***luǧèstra*** (24: Triei, 26: Baunei), ***ǧiléstru*** (29: Villagrande Strisaili), ***liǧèstra*** (33: Tortolì), ***arǧiléstru*** (34: Arzana, 36: Lanusei), ***luǧǧèsti*** (38: Gairo), ***čilèstra*** (39: Barì Sardo), ***aliǧèstra*** (43: Perdasdefogu) ‘id.’

AIS (449):

Type	Forms
1: ****θilikèrta***	***[sa] tiliyèrta*** (13: Macomer), ***tsiliyèrta*** (20: Santu Lussurgiu), ***tsiaʀinyèrta*** (23: Milis) ‘id.’
2: ****θalakèrta***	***θalaèrta*** (22: Fonni) ‘id.’
3: ****kalVkèrta***	***[sa] ɣawižètta*** (42: Mogoro, 46: Villacidro), ***[sa] ɣalužèrtula*** (48: Cagliari), ***[sa] ɣalažètta*** (44: Escalaplano) ‘id.’
4: ****(a)li/ukèrta***	
5: ****(a)li/ukèstra***	***lučèstra*** (26: Baunei) ‘id.’

Puddu (2020: *atalighèlte, cabexèta, lucésti, talaèrta, tilichèrta*):

Type	Forms
1: ****θilikèrta**	*atelighèlta, aterighèlta, atilighèlta, atilighèrta, atulighèrta, tilichèrta, tilichèta, tilighèlta, tilighèrta, tilighèta, tiligrèta, tirighèta, tzilighèrta* ‘lizard’
2: ****θalakèrta**	*atalighèlte, atolighèlta, talaèrta* ‘id.’
3: ****kalVkèrta**	*cabexèta, cabixèta, calaxèrta, calixértula, calixèta, caluxerta, caluxértula, caluxétula, carixeta, cauxeta* ‘id.’
4: ****(a)li/ukèrta**	*luxèta* ‘id.’
5: ****(a)li/ukèstra**	*lucésti, lugésti, lugèstra* ‘id.’

Forms represented in § 2.2: ‘celery, fool’s watercress’

DES (I: 422, II: 555):

Type	Forms
1: ****kugúsa**	***kugúṡa*** (4: Mamoiada) ‘fool’s watercress (*Apium nodiflorum*), wild celery (*Apium graveolens*)
2: ****θurgúsa**	***θurgúsa*** (1: Bitti), ***θrugúsa*** (2: Nuoro) ‘wild celery (*Apium graveolens*)’, ***θurgúsa*** (5: Orgosolo) ‘poison hemlock (*Conium maculatum*)’
3: ****θelikúsa**	
4: ****(a)θikúsa**/****(a)tikúsa**	

Paulis (1992: 141):

Type	Forms
1: ****kugúsa**	
2: ****θurgúsa**	***turguṡòne*** (Log.), ***θurgúsa*** (3: Orani) ‘wild celery (*Apium graveolens*)’
3: ****θelikúsa**	***tseliɣúṡa*** (South Log./Arborense) ‘id.’
4: ****(a)θikúsa**/****(a)tikúsa**	

Puddu (2020: *tirgúsa, tzurgúsa*):

Type	Forms
1: ****kugúsa***	*cugúsa,* ‘fool’s watercress (*Apium nodiflorum*), celery (*Apium graveolens*)’
2: ****θurgúsa***	*tirgúsa, trugúsa, turgúsa, tzurgúsa, tzurvúsa* ‘fool’s watercress (*Apium nodiflorum*)’; *turgusòne, tzurgúsa* ‘celery (*Apium graveolens*)’
3: ****θelikúsa***	*tzeligúsa* ‘celery (*Apium graveolens*)’
4: ****(a)θikúsa/**(a)tikúsa***	*atigúsa* ‘celery (*Apium graveolens*)’

Forms represented in § 3.1: ‘pine tree’

DES (I: 278):

Type	Forms
1: ****kampiṇ̲u***	***kompínğu*** (1: Iglesias), ***kampínğu*** (Camp., 2: Cagliari) ‘pine’

Puddu (2020: *campíngiu*):

Type	Forms
1: ****kampiṇ̲u***	*campíngiu, compíngiu, cumpíngiu, cumpríngiu* ‘pine’.

Forms represented in § 3.2: ‘marten’

Wagner (1934, 483–89; DES, I: 139):

Type	Forms
1: ****a(n)síbile***	***assíle*** (3: Luras, 5: Olbia, 14: Perfugas, 16: Monti, 18: Sennori, 19: Berchidda, 20: Nulvi, 21: Martis, 22: Chiaramonti, 23: Osilo, 24: Tula, 26: Oschiri, 29: Cargeghe, 30: Ploaghe, 33: Olmedo, 34: Alà dei Sardi, 35: Florinas, 37: Torpè, 38: Ardara, 41: Ozieri, 42: Pattada, 43: Buddusò, 44: Siligo, 46: Banari, 47: Putifigari, 48: Bessude, 50: Ittireddu, 51: Bonnanaro, 55: Cheremule, 61: Monteleone, 63: Nule, 64: Bultei, 65: Benetutti, 66: Giave, 67: Anela, 68: Cossoine, 69: Bonorva, 70: Bono, 75: Pozzomaggiore, 76: Semestene, 78: Bottida, 79: Burgos, 80: Esporlatu, 83: Montresta, 85: Illorai, 86: Bolotana, 88: Lei, 90: Sindia, 91: Bosa, 94: Bortigali, 95: Suni, 98: Flussio, 99: Magomadas, 100: Macomer, 102: Tresnuraghes, 105: Dualchi, 107: Noragugume, 108: Scano di Montiferro, 109: Borore, 111: Sennariolo, 129: Bonarcado, 130: Ovodda), ***asíli*** (4: Aggius), ***ansíle*** (27: Muros, 28: Ossi, 31: Usini, 32: Codrongianus, 39: Ittiri, 49: Mores, 52: Thiesi, 53: Borutta, 54: Torralba, 58: Romana, 77: Padria, 113: Cuglieri), ***essíle*** (56: Villanova Monteleone, 93: Silanus, 101: Birori), ***asseíle*** (89: Orotelli), ***assaíle*** (120: Santu Lussurgiu), ***[sa] síβile*** (126: Fonni) ‘marten’
2: ****kassíbile***	***kassíβile*** (104: Ottana), ***ʔassíβile*** (112: Orgosolo, 114: Olzai, 117: Ollolai, 118: Gavoi), ***kassíle*** (115: Sedilo, 116: Aidomaggiore, 119: Domusnovas Canales, 121: Norbello, 122: Soddì, 123: Abbasanta, 124: Ghilarza, 125: Zuri, 127: Bidonì, 128: Boroneddu, 131: Paulilatino, 132: Seneghe, 133: Milis, 134: Narbolia), ***kassíβi*** (135: Bauladu, 137: Tramatza, 138: Villanova Truschedu, 140: Siamanna, 141: Oristano), ***kassí*** (136: San Vero Milis), ***kassíli*** (139: Cabras) ‘id.’
3: ****grassíbile***	***grassíle*** (36: Posada, 40: Lodè, 45: Siniscola, 57: Onanì, 59: Garofai, 60: Bitti, 62: Lula, 71: Irgoli, 72: Orune, 73: Onifai, 74: Loculi, 81: Galtellì, 82: Orosei, 92: Dorgali, 110: Mamoiada), ***grassíβile*** (87: Nuoro, 96: Oliena, 97: Oniferi, 103: Orani, 106: Sarule) ‘id.’
4: ****bassíle***	***bassíli*** (1: Santa Teresa Gallura, 6: Calangianus, 7: Nuchis, 9: Tempio Pausania), ***bassíle*** (2: Luogosanto, 10: Bortigiadas, 12: Bulzi, 15: Laerru), ***basíli*** (8: Castelsardo, 13: Porto Torres, 17: Sorso, 25: Sassari), ***vassíli*** (11: Sedini) ‘id.’

Puddu (2020: *ansíle, cassí, grassíbile*):

Type	Forms
1: ****a(n)síbile**	*ansíle, assaíle, assíle, essíle, issíle, assíbile, síbile* ‘marten’
2: ****kassíbile**	*cassí, cassíbi, cassíle, cassíli, cassívile* ‘id.’
3: ****grassíbile**	*grassíbili, grassíle, grassímene, grassímile* ‘id.’
4: ****bassíle**	*bassíli, vassíli* ‘id.’

Forms represented in § 3.3: ‘mullein’

DES (I: 261–262):

Type	Forms
1: ****túmbar-**	***túmbaru*** (1: Monti) ‘sow thistle (*Sonchus oleraceus*)’, ***túmbara*** (4: Urzulei) ‘great mullein (*Verbascum thapsus*)’
2: ****katúmb-**	***kaðúmbu*** (5: Milis, 9: Oristano, 13: Usellus, 14: Mogoro), ***kardúmmulu*** (6: Tonara), ***kaðúmbulu*** (7: Meana, 9: Laconi), ***karúmbulu*** (8: Tortolì), ***karćúmbulu*** (11: Gairo), ***kardúmbulu*** (12: Seui), ***kaðrúmbulu*** (15: Escalaplano), ***kaðúmbuʔu*** (16: Siurgus), ***kadúmbuʁu*** (17: San Nicolò Gerrei, 18: Dolianova)
3: ****tutúmbar-**	

Paulis (1992: 61):

Type	Forms
1: ****túmbar-**	
2: ****katúmb-**	
3: ****tutúmbar-**	***tutúmbaru*** (2: Lodè, 3: Siniscola) ‘spurge (*Euphorbia sp.*)’

Puddu (2020: *cadrúmbulu*):

Type	Forms
1: ****túmbar-**	*túmbara* ‘mullein (*Verbascum*)’; *túmbaru* ‘sow thistle (*Sonchus oleraceus*)’
2: ****katúmb-**	*cadrúmbulu, cadúmbu, cadúmbulu, cadúmburu, cardúmbulu, cardúmmulu, carúmbulu, codúmbu, corómbu, carciúmbulu*, ‘mullein (*Verbascum*)’
3: ****tutúmbar-**	*tutúmbaru* ‘sow thistle (*Sonchus oleraceus*)’

Forms represented in § 3.4: ‘bed bug, pest’

Wagner (DES, I: 438):

Type	Forms
1: ****rústa**[43]	***rústa*** (1: Luras, 2: Olbia, 3: Monti, 4: Berchidda, 7: Pattada, 8: Buddusò, 11: Nule, 12: Orune, 14: Nuoro), ***rúɬta*** (6: Ozieri) ‘bed bug; noxious animal, fox’
2: ****kurústa**	***kurústa*** (5: Posada, 9: Siniscola, 10: Bitti, 13: Orosei), ***[sa] ʔurústa*** (15: Oliena), ***koròsta*** (North Log.) ‘bed bug’

Puddu (2020: *coròsta, rúlta*):

Type	Forms
1: ****rústa**	*rúlta, rústa* ‘bed bug, pest, fox’
2: ****kurústa**	*coròsta, corústa, curústa* ‘bed bug’

43 Cf. also AIS 473.

Forms represented in § 3.5: ‘mushroom’

DES (II: 533–34):

Type	Forms
1: ****antunn-**	***antúnnu*** (1: Olmedo, 3: Ozieri, 4: Pattada, 6: Villanova Monteleone, 10: Padria, 20: Scano di Montiferro, 23: Cuglieri, 21: Sennariolo, 17: Tresnuraghes), ***antúnna*** (8: Bonorva, 9: Bono, 16: Macomer, 19: Borore, 25: Ollolai, 27: Santu Lussurgiu, 28: Norbello) ‘mushroom’
2: ****tuntunn-**	***tuntunníu*** (2: Posada, 14: Dorgali), ***tuntúnnu*** (5: Siniscola, 7: Bitti, 11: Orosei, 12: Lollove) ‘id.’
3: ****tunn-**	***túnn̰ju*** (13: Nuoro), ***tuññíu*** (15: Oliena), ***túnna*** (18: Orani), ***tunníu*** (22: Orgosolo, 26: Gavoi) ‘id.’
4: ****kantunn-**	***kantúnna*** (24: Olzai) ‘id.’

Puddu (2020: *túgnu, cantúnna, antúnna, tontonníu*):

Type	Forms
1: ****antunn-**	*antúnna, antúnnu* ‘mushroom’
2: ****tuntunn-**	*tontonníu, tuntúnna, tuntunnína, tuntunníu, tuntúnnu* ‘id.’
3: ****tunn-**	*túgnu, túnna, tunnío, tunníu, túnniu* ‘id.’
4: ****kantunn-**	*cantúnna* ‘id.’

Forms represented in § 4.1: ‘skink, slug’

DES (II: 546):[44]

Type	Forms
1: ****θilikúku***	***attiliɣúɣu*** (2: Osilo, 5: Ozieri), ***tirigúru*** (3: Sassari), ***tiliɣúɣu*** (4: Ploaghe, 11: Macomer, 8: Bosa), ***θilikúkku*** (6: Bitti, 12: Orani), ***attaliɣúɣe*** (Sennori) ‘skink (*Chalcides ocellatus tiligugu*)’; ***tsirikúkku***[45] (1: Tempio Pausania), ***tiliɣúɣu*** (11: Macomer, 14: Santu Lussurgiu), ***tsiliɣúɣu*** (13: Borore), ***θilikúkku*** (Orani), ‘slug’
2: ****θalakúka***	***θalakúkku*** (7: Nuoro), ***θalakúkka*** (9: Dorgali), ***θelakúkka*** (10: Oliena), ***θalakú?a*** (15: Fonni), ***tsaβaɣúɣu*** (21: Oristano) ‘skink’
3: ****θiθθilúka***	***sintsilúɣa*** (19: Aritzo) ‘skink’
4: ****θaθθalúka***	***tsattsalúɣa*** (22: Laconi), ***tsantsallúɣa*** (20: Meana), ***satsaúɣa*** (27: Villacidro), ***sattsalúɣa*** (29: Cagliari, general Camp.), ***sossòɣa*** (Sulcis reg.) ‘skink (*Chalcides ocellatus*)’; ***sattsalúɣa*** (18: Busachi), ***tsattsalúɣa*** (22: Laconi), ***sassalúɣa*** (23: Gairo), ***sattsaɣúʀa*** (28: Dolianova) ‘slug’
5: ****babbalúka***	
6: ****alikúku***	

44 Besides the forms listed, Marcialis (2005: 17, 65, 69) also has *attilingia* (Pattada), *azzacauga* (without location), *tiligulu* (Bosa) and *zaraclua* (Olzai). Although all have a similar structure, they are probably not directly cognate with **θilikúku*. *Attilingia* from Pattada probably belongs together with words for ‘earthworm’, cf. *tilíŋga* ‘earthworm’ (Oschiri). *Azzacauga* is difficult to judge, especially in absence of its location. We might speculate that it goes back to **θakalúka*, metathesized from **θalakúka* etc., but the loss of the *-l-* (which also occurs in Oristanese *tsaβaɣúɣu*) is difficult to reconcile with the preservation of the voiceless stop *-k-* (spelled *-c-*). *Tiligulu* is cited by Guarnerio (1904: 66) as “*tilígulu*” (with preparoxytone accent), which makes a connection to **θilikúku* difficult, unless some analogy to the diminutive suffix *-ulu* is involved. Guarnerio’s (1904: 61) proposal of *zaraclua* (Olzai) as a metathesis from **zaraluge* is not convincing phonologically, nor does it suffice to fully explain this form.

45 The DES (II: II: 546) gives *tsilikúkku* for Tempio Pausania and cites AIS (461). However, AIS has *tsirikúkku*.

AIS (449, 461):

Type	Forms
1: ****θilikúku**	***[su] tsiaʁiɣúɣu*** (16: Milis) ‘skink (*Chalcides ocellatus tiligugu*)’
2: ****θalakúka**	***[sa] θaθθalakúkka*** (9: Dorgali) ‘skink’
3: ****θiθθilúka**	
4: ****θaθθalúka**	***[sa] θaθθalúkka*** (17: Baunei), ***[sa] sattsaúɣa*** (24: Mogoro), ***[sa] ṡassalúɣa*** (25: Perdasdefogu), ***[sa] sattsalúɣa*** (26: Escalaplano), ***[sa] soṡòɣa*** (30: Sant’Antioco) ‘skink’
5: ****babbalúka**	***[sa] babbalúɣa*** (26: Escalaplano) ‘slug’
6: ****alikúku**	

Puddu (2020: *alicúcu, telacúcu, satzalúga, sosòga, tzabagúgu*):

Type	Forms
1: ****θilikúku**	*tilicúcu, tiligúgu, attaligúghe* ‘skink (*Chalcides ocellatus*), slug’
2: ****θalakúka**	*telacúcu, talacúca, tzalacúca, tzalacrúca, tzelacúca, tzabagúgu* ‘slug’
3: ****θiθθilúka**	*sintzilúga* ‘id.’
4: ****θaθθalúka**	*satzalúga, satzaúga, tzantzallúga, sosòga, sossòga* ‘skink, slug’
5: ****babbalúka**	*baballúca, baballúga, babbalòca, babbalúca, babbaúga* ‘slug’
6: ****alikúku**	*alicúcu* ‘slug’

Forms represented in § 4.2: ‘grasshopper’

DES (II: 546–47):

Type	Forms
1: ****θilipírke**	***[s’] attaliβíχχe*** (2: Sennori), ***attilibíske*** (3: Osilo), ***tiliβíχχe*** (5: Ploaghe, 10: Mores), ***tilipírke*** (6: Posada, 15: Dorgali), ***θilipírke*** (8: Siniscola, 11: Bitti, 14: Nuoro, 15: Dorgali, 18: Orani), ***tilibíɬke*** (12: Bonorva), ***atteriβílke*** (13: Padria), ***θilibrí?e*** (16: Oliena), ***attiliβílke*** (17: Macomer), ***tsilibríɬke*** (19: Borore), ***θiliprí?e*** (20: Orgosolo, 22: Ollolai, 23: Gavoi), ***tsiliprí?e*** (21: Olzai), ***tsilibrínkidi, zilibrínkiḍḍi*** (25: Norbello), ***tsimpilíɣe*** (26: Ovodda) ‘grasshopper’
2: ****θelapórka**	***θelapòrka*** (14: Nuoro) ‘large grasshopper’

AIS (466):

Type	Forms
1: ****θilipírke**	***[lu] tiribíkku*** (4: Sassari), ***[su] ðiliβírke*** (17: Macomer), ***[su] tsiliβríkke*** (24: Santu Lussurgiu) ‘grasshopper’
2: ****θelapórka**	

Marcialis (2005: 65):

Type	Forms
1: ****θilipírke**	***zilibricu*** (1: Tempio Pausania), ***tilibrirche***, ***tilibische***, ***tilibrichi***, ***tiliprie***, ***tilipricu*** (Log.), ***tilipische*** (7: Ozieri), ***ziliblich*** (9: Alghero)
2: ****θelapórka**	

Puddu (2020: *tilibílche, tzimpilíghe, tzilibrínchidi, telapòrca*):

Type	Forms
1: ****θilipírke**	*tilibílche*, *ateribílche*, *silipírche*, *tilibírche*, *tilibísche*, *tilipírcu*, *tilipríghe*, *tilipísche*, *tiribrícu*, *tzilibíche*, *tzimpilíghe*, *tzilibrínchidi* ‘grasshopper’
2: ****θelapórka**	*telapòrca* ‘big grasshopper’

Forms represented in § 4.3: ‘kestrel’

DES (II: 483):

Type	Forms
1: ****θilipírke**	***tilibríkku*** (Log.), ***čiliβríu*** (North Log.), ***attiliβríu*** (1: Osilo), ***tiribíkku*** (2: Sassari), ***tilibríu*** (3: Ploaghe, 4: Posada) ‘kestrel (*Falco tinnunculus*)’

Marcialis (2005: 65):

Type	Forms
1: ****θilipírke**	***tilibricu*** ‘kestrel (*Falco tinnunculus*)’

Puddu (2020: *tilibbríu*):

Type	Forms
1: ****θilipírke**	*tilibbríu, atalibbríu, tilibrícu, tilibríu, tzilibbríu* ‘kestrel (*Falco tinnunculus*)’

Forms represented in § 4.4: ‘earthworm’

DES (II: 555–56):

Type	Forms
1: ****θiling(ul)òne**	***tsirinòni*** (1: Tempio Pausania), ***attaltingòne***[46] (2: Sennori), ***tiringòni*** (3: Sassari), ***tilingòne*** (5: Ploaghe, 11: Bonorva), ***attilingòne*** (8: Thiesi), ***attolingòne*** (9: Villanove Monteleone), ***atterringòne*** (13: Padria), ***θilingròne*** (14: Lollove, 15: Nuoro), ***tilingròne*** (18: Macomer), ***tsilingòne*** (19: Santu Lussurgiu, 23: Busachi), ***tsiaʀingòi*** (22: Milis), ***tsiringòne*** (24: Laconi), ***tsirringòi*** (25: Mogoro), ***tsiringòi*** (26: San Gavino Monreale), ***sittsiringòni*** (27: Guspini), ***tsiringòi*** (28: Villacidro), ***tsirringòni*** (30: Sarroch), ***tserringòni*** (31: Domus de Maria) ‘earthworm (*Lumbricus terrestris*)’
2: ****θilíng(ul)a/u**	***tilíngа*** (4: Oschiri, 7: Mores), ***attulíngа*** (6: Ozieri), ***tsilíngu*** (20: Norbello) ‘id.’
3: ****θulung(ul)òne**	***θurunkròne*** (10: Bitti), ***θulunkòne*** (12: Orune), ***θulungròne*** (17: Dorgali, 21: Fonni) ‘id.’

AIS (457):

Type	Forms
1: ****θiling(ul)òne**	***[lu] tsiriñòni*** (1: Tempio Pausania)
2: ****θilíng(ul)a/u**	
3: ****θulung(ul)òne**	

46 Like this in DES. I suspect that the *-lt-* in *attaltingòne* is a typographical error for *-l-*. If it is the real form, I have no explanation for it.

Marcialis (2005: 64–65, 114):

Type	Forms
1: ****θiling(ul)òne**	***tilingòne*** (16: Bosa) ‘earthworm (*Lumbricus terrestris*)’; ***zirringoni*** (29: Cagliari) ‘various kinds of marine worms, fish bait’
2: ****θilíng(ul)a/u**	
3: ****θulung(ul)òne**	***sullungòne*** ‘earthworm’

Puddu (2020: *atalingòne, atulíngia, tilingiòne, tzuluncòne*):

Type	Forms
1: ****θiling(ul)òne**	*atalingòne, aterignòne, atirigliòne, atolingiòne, sitziringòni, tilingiòne, tilingòne, tilingròne, tilingròre, tiringòni, tzerringòni* ‘earthworm (*Lumbricus terrestris*)’
2: ****θilíng(ul)a/u**	*atulíngia, tilìgna, tzilíngu* ‘id.’
3: ****θulung(ul)òne**	*tulungròne, tzuluncòne, tzulungròne, turruncròne, tzuruncròne* ‘id.’

Forms represented in § 4.5: ‘bat’

DES (II: 556):

Type	Forms
1: ****θirriólu**	***tirr̩ióla*** (Log.), ***tsirr̩ióla*** (4: Sennori), ***tsirr̩iolu[βèḍḍe]*** (10: Olmedo), ***tirr̩ióla [malaβèḍḍe]*** (11: Ozieri), ***tirr̩ióla [‘e βèḍḍe]*** (12: Pattada) ‘bat’
2: ****θiθirriólu**	***θiθirr̩ióla*** (13: Siniscola) ‘id.’
3: ****θuθurr-**	***θuθuréḍḍu*** (19: Nuoro), ***θuθuréri*** (20: Dorgali, 21: Oliena, 22: Orgosolo), ***tsuntsuréḍḍu*** (27: Tonara, 31: Samugheo, 32: Meana), ***tsuntsuríttu*** (28: Atzara, 30: Aritzo), ***tsintsuréḍḍu*** (33: Gadoni) ‘id.’

(continued)

Type	Forms
4: **θiθimurr-*	***tsintsimurréḍḍu*** (Camp.),[47] ***tsintsimúrru*** (23: Fonni, 25: Ovodda), ***sattsamurréḍḍu*** (48: Villacidro) 'id.'
5: **θurrundéllu*	***tsur(r)undéḍḍu*** (Camp.) 'id.'

AIS (448):

Type	Forms
1: **θirriólu*	***[la] tsirr̤òra*** (Sassari) 'bat'
2: **θiθirriólu*	
3: **θuθurr-*	***[su] tsutsurríttu*** (Laconi) 'id.'
4: **θiθimurr-*	***[su] sintsimurréḍḍu*** (Milis) 'id.'
5: **θurrundéllu*	

Terracini (1959: fig. I):

Type	Forms
1: **θirriólu*	***tsirr̤ólu*** (1: Castelsardo), ***tsirr̤óra*** (2: Porto Torres, 7: Sassari), ***[su] tirr̤ólu*** (3: Monti, 9: Tissi), ***tirr̤ólu[βèḍḍe]*** (14: Nughedu San Nicolò), ***tsirr̤ólu[βèḍḍe]*** (16: Bonorva) 'bat'
2: **θiθirriólu*	***titirr̤ól' [impeḍḍáðu]*** (5: Berchidda), ***tsitsirr̤ólu*** (6: Nulvi), ***tintirr̤ólu[βèḍḍe]*** (8: Oschiri), ***činčirr̤ólu*** (15: Mores), ***tintirr̤ólu*** (17: Padria), ***θiθirr̤ólu*** (Siniscola) 'id.'
3: **θuθurr-*	***θuθurréri*** (18: Orosei, 22: Orgosolo), ***[su] tθutθuréḍḍu*** (19: Nuoro), ***tuturréri*** (20: Dorgali), ***[su] tθutθurréri*** (21: Oliena), ***tsuntsurréḍḍu*** (27: Tonara, 32: Meana), ***tsuntsurríttu*** (29: Belvì, 35: Laconi) 'id.'

47 The attribution in DES of this form to "Campidanese rustico" would suggest its presence in much of southern Sardinia. However, Terracini's (1959: fig. I) documentation of the words for bat shows however that it is rather limited to the western portion of the Campidanese dialects. I have therefore chosen not to mark all of Campidanese for this type on the map.

(continued)

Type	Forms
4: ****θiθimurr-**	***sintsimurréḍḍu*** (24: Bonarcado, 26: Milis), ***sittsimurréḍḍu*** (37: Marrubiu, 41: Mogoro, 42: Sardara, 45: San Gavino, 46: Arbus, 48: Villacidro, 49: Fluminimaggiore,), ***[su] sattsemurréḍḍu*** (52: Vallermosa) 'id.'
5: ****θurrundéllu**	***tsurrundéḍḍu*** (34: Oristano, 36: Usellus, 38: Isili, 39: Gesturi, 40: Orroli, 43: Siurgus, 44: Guasila, 47: San Nicolò Gerrei, 50: Monastir, 51: Dolianova, 53: Assemini, 54: Quartu Sant'Elena, 55: Cagliari, 56: Villasimius, 57: Sarroch, 58: Pula,) 'id.'

Marcialis (2005: 67):

Type	Forms
1: ****θirriólu**	
2: ****θiθirriólu**	
3: ****θuθurr-**	***tutturreri*** 'bat'
4: ****θiθimurr-**	
5: ****θurrundéllu**	

Puddu (2020: *satzamurrédhu, terriolubèdhe, titirrìola, tzinzurédhu, tzurrundédhu*):

Type	Forms
1: ****θirriólu**	*terriolubèdhe, tirriolubèdhe, tirrolupèdhe* 'bat'
2: ****θiθirriólu**	*titirrìola* 'id.'
3: ****θuθurr-**	*tzintzurédhu, tzintzurrédhu, tzuntzurédhu, tzuntzurrédhu* 'id.'
4: ****θiθimurr-**	*satzamurrédhu, sitzimurrédhu, tzitzimurrédhu, tzintzimorrédhu* 'id.'
5: ****θurrundéllu**	*tzurrundédhu, itzurrundédhu* 'id.'

Forms represented in § 4.6: ‘neck’

DES (II: 553):

Type	Forms
1: ****θúk(u)lu***	***θúkru*** (Fonni, Mamoiada), ***θrúk(k)u*** (Bitti, Orune, Nuoro, Orosei, Siniscola), ***θúl?u*** (Orgosolo), ***θúčču*** (Orani), ***θúɣulu*** (Baunei), ***θúɣru*** (Triei, Talana), ***θrúɣu*** (Villagrande Strisaili), ***tsúkkru*** (Olzai), ***trúɣu*** (general Log.: e.g. Bono, Macomer, Seneghe), ***túğu*** (North Log.: e.g. Bonorva, Mores, Buddusò, Oschiri, Sennori), ***túğu*** (Berchidda), ***túɣu*** (Ploaghe, Osilo), ***(b)utúɣu*** (Olmedo), ***tsrúɣu*** (Bonarcado, Bauladu, Cruccuris), ***tsúɣu*** (general Camp.) ‘neck’

AIS (118):

Type	Forms
1: ****θúk(u)lu***	***[su] tsrúɣu*** (Milis), ***[su] tsúɣu*** (Santu Lussurgiu), ***[su] θrúχu*** (Dorgali), ***[su] čúɣu*** (Desulo), ***[su] súɣu*** (Perdasdefogu), ***[su] zúɣu*** (San Antioco) ‘neck’

Puddu (2020: trúcu, tzúgulu):

Type	Forms
1: ****θúk(u)lu***	*butúgiu, butúju, ciúgu, itzúgu, súgru, trúcu, trúgu, túciu, túcru, túju, túxu, tzúciu, tzúgulu* ‘neck’

References

AIS = Jaberg, Karl,Jakob Jud & Paul Scheuermeier. 1928. *Sprach- und Sachtatlas Italiens und der Südschweiz*. Zofingen: Ringier.

Alessio, Giovanni. 1939. I nomi della cavalletta in Italia. *Archivio Glottologico Italiano* 31. 13–84.

Alessio, Giovanni. 1941. Fitonimi mediterranei. *Studi Etruschi* 15. 177–224.

Alinei, Mario. 1984. *Dal totemismo al cristianesimo popolare: Sviluppi semantici nei dialetti italiani ed europei*. Alessandria: Edizioni dell’Orso.

Battisti, Carlo & Giovanni Alessio. 1952. *Dizionario etimologico italiano*, vol. 3. Firenze: Editore G. Barbèra.

Beccaria, Gian Luigi. 1995. *I nomi del mondo: Santi, demoni, folletti e le parole perdute*. Torino: Einaudi Editore.

Beekes, Robert & Lucien van Beek. 2010. *Etymological dictionary of Greek*. Leiden: Brill.

Beekes, Robert. 2014. *Pre-Greek: Phonology, morphology, lexicon*. Edited by Stefan Norbruis. Leiden: Brill.

Bertoldi, Vittorio. 1928. Antichi filoni nella toponomastica mediterranea incrociantisi nella Sardegna. *Revue de Linguistique Romane* 4. 222–250.
Bertoldi, Vittorio. 1929. *Alica* 'farina bianca' e formazioni affini. *Studi Italiani in Filologia Classica* 7. 251–263.
Bertoldi, Vittorio. 1930. Gallico **cabrostos* 'ligustro' (da **cabros*, il corrispondente gallico del greco κάπρος, ecc.). *Revue Celtique* 47. 184–196.
Bertoldi, Vittorio. 1937a. Contatti e conflitti di lingue nell'antico Mediterraneo. *Zeitschrift für Romanische Philologie* 57. 137–169.
Bertoldi, Vittorio. 1937b. Plurale mediterraneo in residui fossili. In Jacobus van Ginneken, *Mélanges de Linguistique et de Philologie Offerts à Jacq. van Ginneken*, 157–169. Paris: Librairie C. Klincksieck.
Bertoni, Giulio. 1913. Denominazioni del 'ramarro' (*Lacerta viridis*) in Italia. *Romania* 42: 161–73.
Blasco Ferrer, Eduardo. 1984. *Grammatica storica del catalano e dei suoi dialetti con speciale riguardo all'algherese*. Tübingen: Gunter Narr Verlag.
Blasco Ferrer, Eduardo. 1989. Il latino e la romanizzazione della Sardegna: Vecchie e nuove ipotesi. *Archivio Glottologico Italiano* 74. 5–89.
Blasco Ferrer, Eduardo. 1999. La *s-* latina e la fine di una crux etimologica: Sardo *thucru*/*zugu* 'collo'. *Sardegna Mediterranea* 5. 65–70.
Blasco Ferrer, Eduardo. 2010. Paleosardo, paleobasco, iberico. In Eduardo Blasco Ferrer (ed.), *Paleosardo: le Radici Linguistiche della Sardegna Neolitica*, 89–136. Berlin: De Gruyter.
Blasco Ferrer, Eduardo. 2011a. Methode gegen Zufall. Prinzipien und Erkentnisse der Substrataforschung am Beispiel der Toponomastik im Mittelmeer. *Indogermanische Forschungen* 116(4). 271–299.
Blasco Ferrer, Eduardo. 2011b. Il sostrato paleosardo: Fine d'un rebus. *Zeitschrift für Romanische Philologie* 127(1). 63–110.
Blasco Ferrer, Eduardo. 2017. Paleosardo: sostrati e toponomastica. In Eduardo Blasco Ferrer, Peter Koch & Daniela Marzo (eds.), *Manuale di linguistica sarda*, 67–84. Berlin: De Gruyter.
Böhne, Rudolf. 1950. *Zum Wortschatz der Mundart des Sárrabus*. Berlin: Deutsche Akademie der Wissenschaften.
Bolognesi, Roberto & Wilbert Heeringa. 2002. De invloed van dominante talen op het lexicon en de fonologie van Sardische dialecten [The influence of dominant languages on the lexicon and the phonology of Sardinian dialects]. *Gramma/TTT* 9. 45–83.
Bramon, Dolors. 1987. Aglutinación y deglutinación del artículo en los arabismos del castellano y del catalán. *Vox Romanica* 46. 138–179.
Calò, Carla Maria, Giuseppe Vona, Renato Robledo & Paolo Francalacci. 2021. From old markers to next generation: Reconstructing the history of the peopling of Sardinia. *Annals of Human Biology* 48(3). 203–12.
Casu, Pietro. 2011. *Vocabolario sardo logudorese italiano*. Edited by Giulio Paulis. http://vocabolariocasu.isresardegna.it.
Chiang, Charleston W.K., Joseph H. Marcus, Carlo Sidore, Arjun Biddanda, Hussein Al-Asadi, Magdalena Zoledziewska, Maristella Pitzalis, Fabio Busonero, Andrea Maschio, Giorgio Pistis et al. 2018. Genomic history of the Sardinian population. *Nature Genetics* 50(10). 1426–1434.
Corominas, Joan & José A. Pascual. 1980. *Diccionario crítico etimológico castellano e hispánico*. Madrid: Editorial Gredos.
DES = Wagner, Max Leopold. 1960. *Dizionario etimologico sardo*. Heidelberg: Carl Winter Universitätsverlag.
Devoto, Giacomo. 1955. Le fasi della linguistica mediterranea. *Studi Etruschi* 23. 217–228.

Dworkin, Steven N. 2012. *A history of the Spanish lexicon: A linguistic perspective*. Oxford: Oxford University Press.

Fernandes, Daniel M., Alissa Mittnik, Iñigo Olalde, Iosif Lazaridis, Olivia Cheronet, Nadin Rohland, Swapan Mallick, Rebecca Bernardos, Nasreen Broomandkhoshbacht, Jens Carlsson et al. 2020. The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. *Nature Ecology and Evolution* 4(3). 334–345.

FEW = Wartburg, Walther von, Margarete Hoffert, Carl Theodor Gossen, Jean-Pierre Chambon & Jean-Paul Chauveau. 1922–2003. *Französisches Etymologisches Wörterbuch: eine Darstellung des Galloromanischen Sprachschatzes*. Bonn: Klopp Verlag.

Flaksman, Maria. 2017. Iconic treadmill hypothesis: The reasons behind continuous onomatopoeic coinage. In Angelika Zirker, Matthias Bauer, Olga Fischer & Christina Ljungberg (eds.), *Dimensions of Iconicity*, 15–38. Amsterdam: John Benjamins Publishing Company.

Guarnerio, Pier Enea. 1904. Postille sul lessico sardo. *Romania: Recueil Trimestriel Consacré à l'Etude des Langues et des Littératures Romanes* 33. 50–70.

Heizer, Robert F. 1953. Aboriginal fish poisons. *Smithsonian Insitution, Bureau of American Ethnology Bulletin* 151. 225–284.

Hesychius Alexandrinus. 2020. *Hesychii Alexandrini lexicon*. Edited by Kurt Latte & Ian C. Cunningham, vol. 2. Berlin: De Gruyter.

Hubschmid, Johannes. 1953. *Sardische Studien*. Bern: A. Francke AG. Verlag.

Ledgeway, Adam. 2016. Italian, Tuscan, and Corsican. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford Guide to the Romance Languages*, 206–227. Oxford: Oxford University Press.

Lewis, Charlton T. & Charles Short. 1879. *A Latin dictionary. Founded on Andrews' edition of Freund's Latin dictionary*. Oxford: Clarendon Press.

Loporcaro, Michele. 2013. Sulla Sardegna preromana. *Romanische Forschungen* 125. 54–65.

Lupinu, Giovanni. La romanizzazione linguistica della Sardegna. In Attilio Mastino (ed.), *La Sardegna e la sua Storia*, vol. 2, 193–197. Nuoro: Edizioni Il Maestrale.

Marcialis, Efisio. 2005. *Vocabolari*. Edited by Eleonora Frongia. Cagliari: Cooperativa Universitaria Editrice Cagliaritana.

Mastino, Attilio. 1993. Analfabetismo e resistenza: Geografia epigrafica della Sardegna. In Alda Calbi, Angela Donati & Gabriella Poma (eds.), *L'Epigrafia Del Villaggio*, 457–536. Faenza: Fratelli Lega Editori.

Múrcia Sànchez, Carles. 2011. *La llengua amaziga a l'antiguitat a partir de les fonts gregues i llatines* [The Amazigh language in antiquity from Greek and Latin sources]. Barcelona: Promociones y Publicaciones Universitarias.

Paulis, Giulio. 1983. *Lingua e cultura nella Sardegna bizantina: Testimonianze linguistiche dell'influsso greco*. Sassari: L'Asfodelo Editore.

Paulis, Giulio. 1987. *I nomi di luogo di Sardegna*. Sassari: Carlo Delfino Editore.

Paulis, Giulio. 1990. Sopravvivenze della lingua punica in Sardegna. In Attilio Mastino (ed.), *L'Africa Romana: Atti del VII Convegno di Studio, Sassari, 15–17 Dicembre 1989*, vol. 2: 599–639. Sassari: Edizioni Gallizzi.

Paulis, Giulio. 1992. *I nomi popolari delle piante in Sardegna*. Sassari: Carlo Delfino editore.

Pittau, Massimo. 1972. *Grammatica del sardo-nuorese: Il più conservativo dei parlari neolatini*. 2nd edn. Bologna: Casa Editrice Prof. Riccardo Pàtron.

Pittau, Massimo. 1995. *Origine e parentela dei Sardi e degli Etruschi: Saggio storico-linguistico*. Sassari: Carlo Delfino editore.

Puddu, Mario. 2020. *Ditzionàriu in línia de sa limba e de sa cultura sarda* [Online dictionary of the Sardinian language and culture]. 2020. https://ditzionariu.nor-web.eu/it/.

Raveane, Alessandro, Ludovica Molinaro, Serena Aneli, Marco Rosario Capodiferro, Luciana de Gennaro, Linda Ongaro, Nicola Rambaldi Migliore, Sara Soffiati, Teodoro Scarano, Antonio Torroni et al. 2022. Assessing temporal and geographic contacts across the Adriatic Sea through the analysis of genome-wide data from Southern Italy. *Genomics* 114.

REW = Meyer-Lübke, Wilhelm. 1911. *Romanisches etymologisches Wörterbuch*. Heidelberg: Carl Winter's Universitätsbuchhandlung.

Rolla, Pietro. 1894. *Secondo saggio di un vocabolario etimologico sardo*. Cagliari: Tipo-Litografia Commerciale.

Russell Craddock, Jerry. 1969. *Latin legacy versus substratum residue: The unstressed "derivational" suffixes in the Romance vernaculars of the western Mediterranean*. Berkeley/Los Angeles: University of California Press.

Schrijver, Peter. 1997. Animal, vegetable and mineral: Some Western European substratum words. In Alexander Lubotsky (ed.), *Sound law and analogy: Papers in honor of Robert S.P. Beekes on the occasion of his 60th birthday*, 293–316. Amsterdam: Rodopi.

Schuchardt, Hugo. 1907. Zur romanischen Wortgeschichte. *Zeitschrift für Romanische Philologie* 31(1). 1–35.

Schuhmann, Roland. 2016. Where is the substrate in the Germanic lexicon? In Bjarne Simmelkjær Sandgaard Hansen, Benedicte Nielsen Whitehead, Thomas Olander & Birgit Anette Olsen (eds.), *Etymology and the European lexicon*, 377–384. Wiesbaden: Reichert Verlag.

Serhoual, Mohammed. 2002. *Dictionnaire tarifit-français*. Tétouan: Université Abdelmalek Essaâdi.

Serra, Giandomenico. 1960. L'action du substrat lybique sur la structure des mots de la structure de la langue sarde. *Orbis: Bulletin International de Documentation Linguistique* 9 (2): 404–18.

Sīnāṣir, Zakīyah 'Irāqī. 1994. *Le dictionnaire Colin d'arabe dialectal marocain*, vol. 7. Rabat: al-Manahil.

Terracini, Benvenuto. 1927. Osservazioni sugli strati più antichi della toponomastica sarda. In *Il convegno archeologico in Sardegna, giugno 1926*, 137–150. Reggio nell'Emilia: Officine grafiche reggiane.

Terracini, Benvenuto. 1959. L'ALI e i problemi della dialettologia italiana – note di geografia linguistica: Le denominazioni del pipistrello in Sardegna. *Bollettino dell'Atlante Linguistico Italiano* 3–4. 10–21.

Vaan, Michiel de. 2008. *Etymological dictionary of Latin and the other Italic languages*. Leiden: Brill.

Virdis, Maurizio. 1988. Sardisch: Areallinguistik. In Günter Holtus, Michael Metzeltin & Christian Schmitt, *Lexikon der romanistischen Linguistik*, vol. 4, 897–912. Tübingen: Max Niemeyer Verlag.

Virdis, Maurizio. 2018. Sociolinguistica storica nella Sardegna medievale: Aspetti, problemi, affioramenti: fra dialettologia e filologia. In Giulio Paulis, Ignazio Putzu & Maurizio Virdis (eds.), *Il sardo medioevale: Tra sociolinguistica storica e ricostruzione linguistico-culturale*, 11–34. Milan: Franco Angeli.

Wagner, Max Leopold. 1907. *Lautlehre der südsardischen Mundarten: Mit besonderer Berücksichtigung der um den Gennargentu gesprochenen Varietäten*. Halle (Saale): Max Niemeyer Verlag.

Wagner, Max Leopold. 1922. Los elementos español y catalán en los dialectos sardos. *Revista de Filología Española* 9. 221–65.

Wagner, Max Leopold. 1928. La stratificazione del lessico sardo. *Revue de Linguistique Romane* 4. 1–61.

Wagner, Max Leopold. 1932. Über die vorrömischen Bestandteile des Sardischen. *Archivum Romanicum* 15. 207–247.

Wagner, Max Leopold. 1934. Weitere sardische Tiernamenstudien. *Archivum Romanicum* 18. 1–18; 481–494.

Wagner, Max Leopold. 1941. *Historische Lautlehre des Sardischen*. Halle(Saale): Max Niemeyer Verlag.

Wagner, Max Leopold. 1952a. Pro domo. *Romanische Forschungen* 64. 416–420.

Wagner, Max Leopold. 1952b. *Historische Wortbildungslehre des Sardischen*. Bern: A. Francke AG. Verlag.
Wagner, Max Leopold. 1997 [1950]. *La lingua sarda: Storia, spirito e forma*. Edited by Giulio Paulis. Nuoro: Ilisso Edizioni.
Wallace, Rex E. 2008. *Zikh rasna: A manual of the Etruscan language and inscriptions*. Ann Arbor: Beech Stave Press.
Weiss, Michael. 2009. *Outline of the comparative grammar of Latin*. Ann Arbor: Beech Stave Press.
Wolf, Heinz Jürgen. 1998. *Toponomastica barbaricina: I nomi di luogo dei comuni di Fonni, Gagoi, Lodine, Mamoiada, Oliena, Ollolai, Olzai, Orgòsolo, Ovodda* Nuoro: Comunità Montana del Nuorese.

Lotte Meester

9 Substrate stratification: An argument against the unity of Pre-Greek

It has long been suspected that a significant part of the Greek lexicon consists of material which cannot be inherited from Proto-Indo-European, and is borrowed from one or several Pre-Greek substrate languages (cf. Pott 1853: 451; Kretschmer 1896: 401–409). This is not in itself a problematic position: we know from historical and archaeological sources that other people lived in Greece before the Greeks arrived, and that other languages were spoken in the area (see Katičić 1976, Chapter 2). The classical author Herodotus states in chapter 1.57 of his *Histories* that a people he calls the Pelasgians "spoke a language which was not Greek". It would therefore not be unexpected if such non-Greek, and potentially non-Indo-European speech communities left traces on the language of the newcomers.

However, linguists disagree on almost every other aspect of the situation. What words in Greek can be considered substrate borrowings? From how many different substrate languages were these words borrowed? What was the linguistic affiliation of the substrate(s), and what did their geographical distribution look like? Over the course of the twentieth century, very different theories have been put forward to answer these questions (see esp. Verhasselt 2009; 2011); so far none of them appears to be commonly accepted.

This lack of consensus on Pre-Greek is not only problematic from a linguistic point of view, it also deprives us of crucial opportunities to recover the pre-Indo-European prehistory of Europe, which is also relevant within the fields of archaeology and genetics. The spread of languages and vocabulary can provide insights into population movements, contacts and cultural developments which are not directly evident from the archaeological or palaeogenomic record. For that reason, it seems worthwhile to take another critical look at the material with our knowledge from these other fields in mind.

One of the few points that the major theories on Pre-Greek have in common is that they tend to focus on the material found within Greek and the surrounding Mediterranean area (cf. Meillet 1908; Beekes 2014). While this may seem a sensible approach when reconstructing a Greek substrate language, much can be gained by considering the questions surrounding Pre-Greek within the broader context of pre-Indo-European Europe and the archaeological and genetic evidence that has surfaced over the past few years. After all, since we now know that Europe – including Greece – was mainly inhabited by a genetically homogenous farming population prior to the steppe invasions (e.g. Skoglund 2012; Skour-

https://doi.org/10.1515/9783111337920-009

tanioti et al. 2023), it is not unlikely that the language spoken in Greece before the arrival of the Indo-Europeans has ties to the rest of Europe.

In this study I focus on two main questions: firstly, does the substrate layer underlying the Greek language consist of one or of several different languages, and secondly, how does what we know about Greek and Pre-Greek compare to the substrate situation that we can distinguish in the rest of Europe? As we will see, these two questions may actually be closely related.

1 Introduction

Substrate words can generally be recognized by their irregular behaviour or formal characteristics, e.g. unexpected word formations or irregular sound correspondences with apparent cognates in related languages (cf. Meillet 1908; see Introduction). This has long been a challenge for research into the subject: if any irregularity can occur, falsification of the results becomes impossible, as the comparative method is based on the regularity of sound change. This means there is no secure methodological framework to distinguish between accidental similarities and actual connections.

However, over the last decades, several scholars have demonstrated, across multiple papers, that the irregularity of substrate words need not result in a licence for uninhibited etymological speculation; as it turns out, it is often quite possible to look for regularity *within* the irregularity. An important example is European *a*-prefixation (cf. Battisti 1959: 155, fn. 1), particularly as defined by Schrijver (1997). According to the latter author, some pre-Indo-European substrate words show two alternants: one with an initial *a*-, and one without such an *a*-, but with an extra vowel in the first syllable. For example, we find Old High German *amsala* 'blackbird' < **a-msl-* next to Latin *merula* 'blackbird' < **mesal-*. Since this alternation is not isolated, but a recurrent feature of European substrate words, it possibly reflects a morphological feature of the underlying donor language (Schrijver 2011). In recent years, a dozen or so *a*-prefixed lexemes have been identified (see Iversen and Kroonen 2017; Schrijver 2018), with many others doubtlessly still to be discovered. Such patterns show that substrate words, although irregular within the languages we know, can be regular in their irregularity, which allows us to distinguish between trivial and less trivial characteristics and recognize features of the underlying languages with more certainty.

Attempts to find an underlying regularity in the Greek substrate material have yielded rather divergent results over the years. I will not attempt to discuss all the literature on the subject, but rather limit myself to the general picture and

the most relevant materials; a more complete overview can be found in Verhasselt (2009; 2011).

1.1 Pre-Greek as an Indo-European language

For decades, scholars have attempted to explain the Pre-Greek substrate as an Indo-European language. For example, Kretschmer assumed in his later works that Pre-Greek, which he called 'Pelasgian', belonged to a sister branch of Proto-Indo-European (1925; 1939). Georgiev (1941-1945) likewise assumed that Pelasgian was an Indo-European language, for which he devised a set of sound laws, e.g. the vocalization of resonants to $^{*}uR$ and satem reflexes for the palatovelars (also cf. van Windekens 1960; Heubeck 1961).

Of course, it is by no means unthinkable that unattested Indo-European languages and dialects have existed. Considering how many IE languages died out leaving behind just a few words, it would be naïve to expect that none of them vanished *before* its speakers adopted writing. It is also certainly possible that Greek borrowed words from such unattested sister languages. The Pelasgian theory, however, has suffered from what Mihaylova (2012) calls "the incontrollable extent of the etymological imagination of its adherents". For some cases – Mihaylova lists 23 candidates – Georgiev's sound laws might work, and if there are convincing etymologies among them which paint a coherent picture, a small part of the Pre-Greek substrate layer could be explained in this way. However, for most of the material, an Indo-European origin simply does not hold.

Another Indo-European origin theory was proposed by Palmer (1980), who argued that Pre-Greek was an Anatolian language. This connection was drawn based on suffixes (e.g. -ινθ-, which Palmer connects to *-anda-*) found in toponyms both in Anatolia and in Greece (see also Kroonen, this volume). A major problem with this theory – beside the fact that it cannot explain the vast majority of lexical items found as Greek substrate words – is that even the supposedly identical suffixes are never found with the same root in Greek and Anatolian.

1.2 Pre-Greek as a non-Indo-European language

Since assuming an IE origin for Pre-Greek has turned out to be problematic, other scholars have maintained that the substrate language(s) underlying Greek may not be IE at all. An important proponent of this line of research was Furnée, who in his dissertation (1972) suggested that substrate words in Greek come from two different non-IE languages, a Kartvelian language and another, unrelated Medi-

terranean language that he calls "Aegean" (see also Furnée 1979). However, most of the links drawn between Greek words and Kartvelian roots are doubtful at best. As the Kartvelian component of Pre-Greek would be chronologically very close to Proto-Kartvelian, Furnée assumes that the forms we find in Greek are closer to this protolanguage than the attested Kartvelian languages. As a result, he sometimes bases his reconstruction of Proto-Kartvelian on the Greek borrowings rather than on the attested material in Kartvelian. This runs a high risk of circularity. In other words, Furnée shapes the Kartvelian protolanguage to be a fitting donor language to Greek, rather than showing us how the Pre-Greek material fits into the known facts of Kartvelian. Although there are some attractive comparisons between the two languages, the reliable material is too scarce to support a significant Kartvelian layer in Greek.

Nonetheless, Furnée's work represents a landmark in the study of Pre-Greek, if only because of the unprecedented scale of his research. His dissertation comprises nearly four thousand words. These are labelled Pre-Greek based on phonological irregularities within Greek, certain suffixes which he assumes to be non-IE, and semantics. This corpus was the main source of inspiration for Beekes, who was the most productive scholar on the subject of Pre-Greek over the past few decades. In his *Etymological Dictionary of Greek* (2010), Beekes marks around two thousand words as (possibly) Pre-Greek based primarily on the features observed by Furnée. Moreover, in 2014 he published a small monograph on Pre-Greek, containing a word list of some eleven hundred putative Pre-Greek words in Greek. Again, many of the Pre-Greek markers employed are based on Furnée's findings, although Beekes discards part of Furnée's material for lack of evidence.

Both Beekes' dictionary and his later book on Pre-Greek have been met with fierce criticism (e.g. Meissner 2013; De Decker 2015; Nikolaev 2018): the phonological alternations Beekes allows for Pre-Greek are too broad, enabling him to lump virtually anything together; his assumption that the Pre-Greek substrate layer must have consisted of one single language is not sufficiently supported; Indo-European etymologies are discarded too easily. The general impression of the critics seems to be that Beekes is overly zealous in finding Pre-Greek forms, while it would be preferable to assume an Indo-European origin whenever possible.

It cannot be denied that Beekes' collection of Pre-Greek forms has its shortcomings. Some of the forms given may rather be of IE origin. In other cases, forms he connects are formally or semantically so far apart that it is doubtful whether they can go back to the same preform (e.g. βράκαλον vs. ῥόπαλον 'club, cudgel'). However, this does not necessarily diminish the material's use for the research into Pre-Greek. In my view, Beekes' book should not be seen as a definitive collection of Pre-Greek substrate words, but rather as another starting point: a corpus of potentially non-inherited words that await further analysis. Even if

the most likely explanation is not that of a substrate origin, like in the handful of cases discussed by De Decker (2015), and even if some suffixes have been incorrectly assigned to the Pre-Greek substrate, we are still left with hundreds of words that cannot possibly have an Indo-European provenance. In any case, these must be explained otherwise. For all their valid criticism of specific forms or methodological problems, the reviews by Meissner, De Decker and others offer surprisingly little acknowledgement of this problem. Beekes' overview may be maximalistic, but at this point in the study of Pre-Greek, accumulating a corpus of all potentially relevant material is a helpful undertaking. Further selection and analysis are necessary next steps, but they would be impossible without access to the complete material in the first place.

In the following sections, I critically review the most important linguistic features of Pre-Greek as suggested by Beekes. I have chosen to proceed with caution: if I encounter words of possible Indo-European origin, I discard them in my assessment.[1] The following analysis is therefore based on the majority of the material that is definitely not Indo-European, and which can therefore be considered part of a Pre-Greek substrate.

2 The unity of Pre-Greek

2.1 Substrate features vs. Borrowing features

In my analysis of the Pre-Greek material, the first point of focus is the question of Pre-Greek unity. Beekes' conclusion that the Greek substrate must have been a single language has been criticized for not being sufficiently supported by the evidence (see the discussion below). Before I can offer my conclusions on this point, it is important to first sketch the arguments Beekes presents to defend his thesis.

In short, Beekes reasons as follows. Like Furnée, he assumes that we can recognize Pre-Greek borrowings by their irregular formal variations, e.g. alternations between voiced, voiceless and aspirated stops, prenasalization, vowel alternations, prothetic *a-*, *s-*, *k-* and *t-*, variation between dentals and liquids and several other phonemes (Beekes 2014, Chapter 2). However, whereas Furnée assumes that these

1 For example, I did not take into account νυκτάλωψ 'with night vision', despite Beekes' claim that the suffix -ωπ- proves a Pre-Greek origin; or ἄσφαλτος 'asphalt', which could be a negated form to the root of σφάλλεσθαι 'to fall', i.e. the material that keeps walls from collapsing. The PIE etymologies could be incorrect, but the existence of reasonable doubt is enough reason not to consider them when searching for Pre-Greek patterns.

variations are due to "expressive" alternations in the substrate language, Beekes argues they are rather the consequence of a phonetic mismatch between Greek and Pre-Greek. When Greek speakers encountered sounds which their own language did not allow, they could render those in several ways. For example, nasals sometimes occurring before stops (σαλάμβη, σαλάβη 'vent-hole') could be a reflection of Pre-Greek prenasalized stops. As we find similar alternations occurring throughout his list of substrate words, Beekes concludes that all these forms must have been borrowed from the same donor language, and that Pre-Greek was therefore the only major substrate language underlying Greek.

In principle, a mismatch between two languages' phonological systems is a better way to account for unexpected phonological variations than Furnée's expressiveness (especially as it is unclear why rather non-emotional words such as names for tools or geographical elements should undergo expressive changes). However, Beekes' reconstruction of Pre-Greek is based on the basic assumption that the majority of the variation must reflect some specific phonological feature in the donor language. Unfortunately, this may be somewhat too optimistic.

A quick look at other known borrowing situations shows us that some of the "characteristic" Pre-Greek features are quite common in general in borrowing situations. Aspiration, differences in voicing, nasalization and vowel variation can occur even when both variants are distinguished in the donor language. The Algonquian loanword *moccasin* is found with extensive vowel variation, e.g. *maccaseene, makissin, mogason, moccasin, mogozeen*. Although this variation may to a certain extent reflect actual dialect differences within Algonquian, cf. Ojibwe *makkisin* vs. Massachusett *mokussinash*, it is probably at least partially due to different adaptations of a foreign phonemic system. In other words, it appears that some of the variations observed by Beekes might not be caused by specific phonological ambiguities of Pre-Greek, but rather by the borrowing process itself.

I believe we have reason to assume this was also the case for Pre-Greek. Take the apparently free alternation between voiced, voiceless and aspirated stops (e.g. γαλλερίας, καλλαρίς and χελλαρίης 'name of a fish'). According to Beekes, this shows that Pre-Greek knew no distinction between these categories and had only one stop series (2014: 14). However, less than two hundred entries in his corpus of more than a thousand show such variation. If the outcome of the single Pre-Greek stop in Greek were entirely random, we might expect to see variation in most of the forms, not in a mere 20% of them, although it must be admitted that it is difficult to make statistical predictions like this. Nevertheless, this suggests to me that there probably *were* different stop series in Pre-Greek, which did not match the Greek phonology perfectly, but which usually had at least a preferred outcome in Greek.

Moreover, borrowing mechanisms are not the only way in which donor forms might be obscured. We should not forget that Pre-Greek must have known dialectal differences too: especially more trivial changes, such as vowel quality, may easily have been caused by this type of variation.

Finally, in reviewing Beekes's analysis, we should pay attention to the frequency with which certain variations appear. Some (e.g. σκ / σσ) occur only a handful of times, but are treated by Beekes as equally significant for the Pre-Greek phonology as alternations which occur dozens or hundreds of times. It is important to remember that 'accidents' happen in language, e.g. due to blends or other lexical contaminations. I would therefore suggest to use variations as diagnostic clues only when they are well-attested throughout the corpus, here too, there is no clear line between statistically significant and insignificant variations.

In other words, much of the variation we see in the Greek substrate data may not be a direct reflection of the phonology of the underlying donor language, and can therefore not be used to support an argument in favour of a single Pre-Greek substrate language. This is all not to say that *all* substrate characteristics we see in Greek must be trivial. There are some features of Pre-Greek loanwords that did not occur in other borrowing events examined in preparation for this study, and which can therefore shed more light on the underlying donor phonology in specific cases. The most important of these features are *a*-prefixation, *k*-prefixation, *s*-mobile and some suffixes, most importantly -ινθ-. Due to their rarity in other borrowing situations, it seems unlikely that these arose purely as a consequence of phonological mismatches.

2.2 The substrate features of Pre-Greek

So what happens if we consider the unity question of Pre-Greek again on the basis of these non-trivial characteristics, rather than taking any small phonological alternation in the data at face value?

As already observed by Beekes, vowel and consonant alternations occur throughout the corpus of Greek substrate forms. The more characteristic features mentioned above, however, show a rather interesting pattern if we look not just at their formal attributes, but also at the semantics of the lexemes in which they appear.

2.2.1 A-prefixation

We find some 65 cases of *a*-prefixation in Greek substrate lexemes.[2] Some of these show an associated zero grade in the next syllable when the word has a prefix, but not without the *a*, e.g. ἀστραπή ~ στεροπή 'lightning' and ἄγλις ~ γέλγις 'garlic'. Others do not show this pattern: for example ἀρωδιός ~ ῥωδιός 'heron', ἄκαστος ~ κάστον 'maple' and ἄσταχυς ~ στάχυς 'ear of corn'.

The semantic distribution of this feature across the Greek substrate lexicon seems significant. Nineteen lexemes with *a*-prefixation denote plants and crops, and ten more denote animal names. In smaller numbers, we see the prefix occur in the semantic fields of landscapes, food and the human body. However, it is *never* found in the lexemes with meanings related to jewellery, armour, social hierarchy, specific professions, musical instruments and performing arts, sculpture, and other such 'cultural' fields.

2.2.2 K-prefixation

Beside the *a*-prefix, we find thirteen cases of a *k(a)*-prefix in Greek substrate words. Examples are κίχλη ~ ἴχλα 'thrush', κάρυον 'nut' ~ ἄρυα 'walnuts', κόγχναι ~ ὄγχνη 'pear' and κασκάνδιξ 'onion' ~ σκάνδιξ 'chervil'.

In a few cases, it appears that the *ka*-prefix is not so much a *ka*-prefix, but rather a combination of a *k*-prefix and an *a*-prefix. For example, we find καπᾶνα, ἀπήνη and πήνα, all in the meaning 'wagon'. More speculatively, we could see the same pattern in κάχληξ 'small stones' ~ ἄχλαξ 'id.' ~ χάλαζα 'hail'. Because the number of examples is so small, it is hard to speculate with any certainty on the nature of this apparent *k*-prefix, although some interesting observations can be made even within this small corpus.

Beside the cases where we can see *ka*- alternate with *a*-, the large majority of forms which appears to have only a *k*-prefix show an initial *a*- in the prefixless form (10 of 13 examples). This is more than we would expect if the *k*- could be added to any word, and suggests that the initial *k*- could be linked to the *a*-prefix: it may therefore well be the case that some of these forms are actually *a*-prefixed forms which are not attested without the prefix. It is possible that the initial **k*- in Greek was not a morphological feature, but the result of a phonetic mismatch. It could originally have been a post-velar consonant in the substrate language,

2 The exact count depends somewhat on the variation one is willing to accept; for example, we find one case with a prefixed η rather than an α. 65 is the number of "pure" *a*-prefixed forms.

which did not match the phonetics of Greek and was rendered either as *k*- or as zero. Speculatively, we could consider the possibility that the *a*-prefix was actually a *ḫa*-prefix, which would be interesting in view of Schrijver's suggestion of linking it to the Hattic prefix *ha*- (2011: 246-249, 254; see also Swanenvleugel, this volume).

Relevant for this analysis is the fact that the semantic distribution of the *k*-prefix aligns with the distribution of the *a*-prefix described above: all words showing this feature denote plants, animals and simple technology. We also find a single verb, καλινδέομαι ~ ἀλινδέομαι 'to roll about, wallow'.

2.2.3 S-mobile

Pre-Greek loanwords show an initial *s*-mobile in thirty-three lexemes in Beekes (2014), e.g. σμῖλαξ ~ μῖλαξ 'taxus'; σφάγνος ~ φάγνος 'salvia'; στρύχνον ~ τρύχνον 'nightshade' and σκόνυζα ~ κόνυζα 'fleabane'. The *s*-mobile can be combined with an *a*-prefix as well, in which case the *a*- always precedes the *s*: examples are ἀσκαμωνία ~ σκαμ(μ)ωνία ~ κάμων 'scammony' and ἀσκάλαβος ~ σκαλαβώτης ~ καλαβώτης 'lizard'.

Again, the majority of these are terms for plants, animals and body parts: 24 occurrences can be found in these categories. The rest are verbs and adjectives such as σμοιός 'bad'; all rather general words that cannot be linked to specific cultural traits or activities.

The only possible exception to the general semantic pattern is σκινδαψός ~ κινδαψός, which is defined by Beekes (2009: s.v.) as "name of a four-stringed musical instrument with thorn-like appendices; name of an ivy-like plant". However, Dawkins (1936) argues that σκινδαψός must have been a thorny plant rather than some type of ivy. If this is indeed the case, it seems more logical to me that the instrument with thorn-like parts would be named after the plant than the other way around, as the plant was presumably around before the instrument. Even this may then be a plant name rather than a cultural term.

2.2.4 The suffix -ινθ-

The suffix -ινθ- is one of the most recognizable substrate suffixes in Greek (see Kroonen, this volume). It is often found in plant names (e.g. ἀψίνθιον 'wormwood', ἐρέβινθος 'chickpea' and ὑάκινθος 'hyacinth'), but we also find it in other semantic domains (e.g. ἀσάμινθος 'bathtub', λαβύρινθος 'labyrinth', μέρμινθα 'string' and σαλαμίνθη 'spider'). The function of the suffix is unclear. It has been

interpreted as a collective suffix (cf. Moreschini 1984: 108), but it could equally well be an individualizing suffix, so that ἐρέβινθος would denote one chickpea rather than a mass of peas. This would perhaps fit the case of κηρός 'wax' vs. κήρινθος 'beebread', as wax is generally a mass substance, whereas bee-bread is taken from the combs in separate little lumps. However, the lexical evidence is too scarce for any clear conclusions.

The suffix is found both with and without the nasal, e.g. in μέρμινθα ~ μέρμῑθος 'string', γάλινθοι ~ γάλιθοι 'chickpeas' and τέρμινθος ~ τρέμιθος 'turpentine tree'. In some cases, the nasal is "compensated" by vowel length, but this is certainly not always the case. Moreover, we find a few potential cases in which the stop is not aspirated, such as κικριβιντίς 'purslane'.

This variation within Greek makes it hard to accurately count the relevant instances. It is clear, however, that again the majority of forms containing -ινθ- refer to flora and fauna, with a few exceptions for tools, such as μέρμινθα 'thread', πείρινθα 'basket' and πλίνθος 'brick, flint stone'. Two interesting semantic outliers are λαβύρινθος 'labyrinth' and ἀσάμινθος 'bathtub'. I have considered that both labyrinths and a particular type of richly decorated bathtubs are associated with the Minoan civilization (cf. Meissner 2013: 14), but whether that would be reason to assume any link between this substrate suffix and the Minoans is not something I can fully assess here.

2.2.5 Other suffixes

A similar distribution is found for a few other suffixes which occur in suspected Pre-Greek lemmas. -ζα is found in forms such as μάνυζα 'garlic' and ἄρπεζα 'hedge', and occurs in the same semantic fields as the *a*-prefix. The suffix -ιγγ- is not very common, but generally occurs in animals, body parts and landscape features.

Finally, a noteworthy point is that we frequently find several of these features in one and the same lexeme. For example, the suffix -ιγγ- sometimes alternates with -ινθ-, e.g. in μήρινθος 'thread' ~ σμῆριγξ 'hair' and ἔλμινθος ~ ἔλμιγγος 'worm'. We also find the initial *s*- in combination with some of these suffixes, as in μήρινθος ~ σμήρινθος 'thread'. And as already mentioned above, initial *a*- and *s*- occur together as well, as with ἀσκάλαβος ~ σκαλαβώτης ~ καλαβώτης 'lizard'. All of this suggests that these features may originally have belonged to the same language.

In other words, we are dealing with a cohesive cluster of features showing a semantic distribution that is both clear-cut and strikingly similar across all of them: they occur, some rather frequently, in some semantic subsets of the Greek

substrate lexicon, and are entirely absent from the substrate data in other sizeable semantic fields.

This raises the question of how to explain this pattern if we assume that the Greek substrate was a single language. We would have to assume that these features – which do not seem to have a particular semantic function as far as we can gather from the material at hand – would for some reason only occur with words for flora, fauna, human bodies, tools, landscapes and a handful of verbs. Reconstructing them as noun class markers does not seem very sensible, considering the presence of similar alternations in some verbs. Moreover, we would have to accept that this language has multiple noun class markers for the semantic fields mentioned above and none which can be identified for any other semantic field, including jewelry, clothing, the arts, etc., as we find no similar prefix-like patterns there. This would be rather unusual; when a language has a gender or noun class system, generally *all* nouns in the language are classified into one of these categories.

Considering this, it seems more plausible that the Pre-Greek lexical material in reality hails from multiple non-Indo-European source languages. One of these contains features such as initial *a, k* and *s* and the ινθ-suffix, and provided terms for flora, fauna, body parts and landscape features. One or multiple other languages were the source(s) of more cultural terms such as dances, jewelry, titles, etc.

Not only is this a more attractive scenario from a linguistic point of view, it also corresponds better to what is known from the Pre-Indo-European history of the area. We know, based on genetic and archaeologic evidence, that Greece has been the setting of multiple migrations and contact situations. It would, at the very least, not be surprising if these waves of movement left multiple linguistic layers behind: for example, a layer of agriculturally relevant words from the Neolithic farmers who migrated from Anatolia through Greece into the rest of Europe, and later layers with newer influences in the cultural field.

3 Pre-Greek and the European substrate

The final question is then whether the linguistic situation in the rest of Europe can provide us with any support for the assumption of a multi-layered Greek substrate.

Let us first consider how Beekes unites his idea of a homogeneous Pre-Greek with the undeniable existence of apparent comparanda elsewhere in Europe. As mentioned above, he assumed that his version of Pre-Greek must have been “one

language, or a group of closely related dialects or languages" (2014: 45), and that this language was spoken only in Greece, with perhaps some offshoots throughout the Mediterranean to account for words which have been borrowed into Latin as well. However, he does reconstruct a handful of Greek substrate words not as Pre-Greek, but rather as part of a larger European substrate which has also left a handful of traces in Greek (Beekes 2000). In order to maintain this distinction, Beekes has to come up with a solution for forms which look like part of his Pre-Greek, but which do appear to have counterparts in the rest of Europe, too. He applies two main mechanisms to account for these cases:

1. Borrowing from Pre-Greek. For example, beside ῥάφανος (also ῥάφυς, ῥάπυς) 'radish', we also find Lat. *rāpum*, OHG *ruoba*, Lith. *ropė́* 'turnip' < **rāp-*, and OCS *rěpa* 'turnip' < **rē/aip-*. Beekes (2010: s.v.) notes: "Since the variation π ~ φ and the suffix -αν- are evidently Pre-Greek features, the word may originally be of Pre-Greek stock; thence the European cognates cited above were borrowed."
2. European substrate words borrowed by Pre-Greek. For example, in his introduction on Pre-Greek, Beekes (2014:45) writes: "Another matter is that (non-Indo-European) loanwords from old Europe may have entered Greece [. . .]. Moreover, these may have already been adopted in Pre-Greek, as is suggested by ἐρέβινθος, which has a Pre-Greek suffix, but a root which is attested (with some variation), as a substrate word, in other European languages." These other attestations are Lat. *ervum* 'bitter vetch' and PGm. **arwīt-* 'pea', whence e.g. Du. *erwt*. In Greek itself we also find ὄροβος 'chickpea', which appears to be the same root with another suffix.

Unfortunately, both of these explanations are problematic. For ῥάφανος, all cognates mentioned by Beekes could theoretically have been borrowed from Pre-Greek, but it is unclear why so many languages in Europe would want to borrow the name for a crop that has been known throughout the continent since the first agriculturalists introduced it. More problematic is another comparandum which was discovered recently, namely W *erfin* 'turnip' < PC **arb-īn-* (Iversen & Kroonen 2017: 518). This form appears to exhibit Schrijver's *a*-prefixation (with associated zero grade of the root), which is assumed to be a feature of a European-wide substrate. If we assume all turnip words in Europe were borrowed from Pre-Greek, we cannot account for the Celtic material.

In the case of ἐρέβινθος, Beekes assumes the -ινθ-suffix appears exclusively in Greece. However, Kuiper (1956) has convincingly argued that the **-īt-* suffix of PGm. **arwīt-* may in fact be exactly the same suffix (cf. also Kroonen 2012 and

Kroonen, this volume).[3] If this is correct, we can no longer maintain a scenario in which the suffix was added only within Greece. Instead, we must assume that the word spread over Europe with the suffix attached. To complicate matters even further, we also find a possible Kartvelian cognate in the Caucasus: Geo. *erevindi* 'chickpea'. If not a Greek loan, this would suggest that the word – including the suffix – may originate in Anatolia or the Near East, which further discourages the idea that the formation is a combination of a European substrate word and an exclusively Pre-Greek suffix. [4]

Especially ἐρέβινθος, with its very recognizable suffix, makes it problematic to draw any sharp distinction between Pre-Greek and the wider European substrate. If we assume, like Beekes, that the suffix was exclusively Pre-Greek, we have to assume that Pre-Greek was spoken across a larger area in Europe to explain the Germanic form. If we rather assume that the suffix belonged to a European-wide substrate, the other Pre-Greek features which are associated with it, such as *s*-mobile and other suffixes occurring in the same words, are *also* likely to be European. In other words, regardless of how we explain the matter, the line between both presumed substrates blurs quickly.

The matter becomes much easier to solve, however, if we discard Beekes' assumption that Pre-Greek must be a single, Mediterranean language. If we are dealing with multiple substrate layers in Greek, we can accept that some of the parts Beekes considers "typically Pre-Greek" are rather European in origin, whereas others are exclusively Mediterranean. Indeed, if we consider the Greek substrate lexemes that have apparent cognates elsewhere in Europe, two observations connect those forms to the semantic subset that we identified above.

Firstly, although I looked for cognates for all lexemes in Beekes' data, I have again only been able to identify them for forms with semantic meanings related to nature, tools and the human body. The many Greek substrate words denoting clothing, festivals, art, and other such cultural concepts never show up in the rest of Europe, as far as I have been able to find.

Secondly, if we consider the Greek substrate words shared with the rest of Europe, we find all of the characteristic features mentioned above in this list. For example:

3 The semantic shift to 'pea' in Germanic is not surprising if we take into account that chickpeas do not grow in the north of Europe; it makes sense that their name would be transferred to another, similar crop. For peas, it is important to mention they were usually eaten dried in prehistoric times; they therefore looked more like chickpeas than our modern, green peas do.

4 This form could be borrowed from Greek. However, that makes it challenging to explain Georgian forms showing a suffix *-indi* which do not have an attested Greek equivalent, e.g. *rapindi* 'laurel'.

– Initial *a-*: Gr. ῥάφανος 'radish', Lat. *rāpum*, OHG *ruoba,* W *erfin* 'turnip'
  Gr. ἐρῳδιός, ἀρωδιός, ῥωδιώς, Lat. *ardea*, Serb. *róda* 'heron'
– Initial *s-*: Gr. ταῦρος, Lat. *taurus*, PGm. **þeura-*, **steura-* 'bull'
– Initial *k(a)-*: Gr. κάχληξ, ἄχλαξ, κόχλαξ 'small stones', PGm. **hagla-* 'hail'
– -ινθ-suffix: Gr. ἐρέβινθος, PGm. **arwīt-*

The link with the rest of the continent seems to suggest another aspect that separates my suggested layers of the Greek substrate. Again, we see a clear semantic distribution between two sub-categories of the substrate in question: flora, fauna, tools and landscape terms appear to have cognates in the rest of Europe, whereas clothes, arts and similar cultural terms remain confined to Greece or the Mediterranean area at most.

Again, I find it hard to explain this divide if we work with the hypothesis that Pre-Greek was a single substrate language. If that were the case, we would have to assume that either the migrating speakers of this substrate abruptly lost their entire culture-specific lexicon by the time they reached the Balkans, *or* that the rest of Europe was highly prone to borrowing terms for crops and plants they already knew, but not terms for more recently innovated cultural items. Both scenarios seem implausible to me.

This leads me to suggest that we are simply dealing with (at least) two languages here, one of which has a Europe-wide distribution, whereas the other(s) only left traces in Greece and perhaps some of the neighbouring areas. The fact that this "European" layer ties in perfectly with an already postulated layer within the Greek substrate, both in terms of semantics and formal characteristics, further supports this hypothesis.

4 Conclusion

Summarizing the different points brought up above, Beekes' argument for a single, uniform Pre-Greek can only be maintained if all the variations in Greek words are assumed to reflect a single, specific phonological system. However, as we saw in Section 2, features such as vocal variation and voicing alternations in consonants are rather common in borrowing situations, regardless of the specific donor language. In other words, the fact that we see the same alternation in ῥάπυς ~ ῥάφυς 'turnip' and in κλανίον ~ χλανίαι 'bracelet(s)' only suggests that both words were borrowed, but not necessarily that their sources are identical.

If we set trivial features aside, we are left with at least one rather clear split in the Pre-Greek lexicon as collected by Beekes. One group contains *a*-prefixes,

s-mobiles and suffixes such as -ινθ-, -ζα and -Vγγ-; these features are regularly combined within the same words, which suggests a shared origin. Words with these features denote plants, animals, landscape features, body parts, tools and perhaps some verbs; other semantic categories are excluded. These are also the exact same semantic categories for which we occasionally find cognates in the rest of Europe. It therefore seems most economical to assume that this complex was borrowed from one single, Europe-wide donor language or language continuum.

The remaining part of the lexicon is less clearly defined or more heterogeneous. I have only found a small number of returning characteristics, none of which interact with the features described above. These are the suffixes -υβ- and -ικ-, among others, which are infrequent. It is possible that I overlooked some shared features, but this need not be the case as we have no clear reason to assume that this second category is made up of loanwords from a single donor language, either. Quite the opposite is true: considering how many cultures have influenced Greece through the millennia, it appears reasonable that words from a larger number of languages may have found their way into the Greek lexicon over time.

We know, for example, that Greek has borrowed words from languages spoken in Anatolia (e.g. ἀξίνη 'axe', either borrowed from or borrowed from the same source as Akk. *haṣṣinu* and Aram. *ḥaṣṣīnā*, 'id.') and the Middle East (e.g. κάννα 'reed' cf. Ass. *qanū* 'id.'). These are classified by Beekes as Pre-Greek, because he assumes borrowing through Pre-Greek (2010: xv), but they should be considered a separate category in the Greek substrate vocabulary. The few convincing Kartvelian loans referred to above form another part of the puzzle. Yet another subcategory may be the set of words which have apparent cognates only in Latin, without showing any of the typical "European-wide" features. Examples are κάπια 'onions' which appears related to Lat. *cēpa* 'onion', and ῥᾶξ, ῥώξ 'grape' as a pseudo-cognate to Lat. *racēmus* 'grape.' Forms like these could belong to a separate Mediterranean substrate (cf. Meillet 1908).

Add to these influences the possibilities of *Wanderwörter* from the east and forms borrowed from unattested Indo-European languages, each of which would be expected to bring its own characteristic features, and it becomes clear that the Greek substrate lexicon must consist of multiple highly diverse layers. My suspicion is that the cluster of substrate loans I have described above with a European-wide distribution and clear formal characteristics is simply one of the many elements that constitute the borrowed lexicon in the Greek language as we know it.

To acquire a better understanding of the layers of the Greek substrate lexicon, it would be interesting to compare the tentative distinction in this study with archaeological and genetic data, to see whether we can link influxes of genetic or cul-

tural elements to specific linguistic features. If supported by sufficient lexical data, this may allow us to connect substrate layers of Greek to historic events, migrations or contact situations, providing us with new insights on the prehistoric linguistic landscape of both Greece and sub-Indo-European Europe in general.

References

Battisti, Carlo. 1959. *Sostrati e parastrati nell'Italia preistorica*. Firenze: Felice Le Monnier.

Beekes, Robert S.P. 2000. European substratum words in Greek. In Michaela Ofitsch & Christian Zinko (eds.), *125 Jahre Indogermanistik in Graz*, 21–31. Graz: Leykam.

Beekes, Robert S.P. 2010. *Etymological dictionary of Greek*. Leiden/Boston: Brill.

Beekes, Robert S.P. 2014. *Pre-Greek: phonology, morphology, lexicon*. Edited by Stefan Norbruis. Leiden/Boston: Brill.

Dawkins, Richard McGillivray. 1936. The semantics of Greek names for plants. *The Journal of Hellenic Studies* 56(1). 1–11.

De Decker, Filip. 2015. A new book on Pre-Greek. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 12. 1–22.

Furnée, Edzard. 1972. *Die wichtigsten konsonantischen Erscheinungen des Vorgriechischen*. The Hague: Mouton.

Furnée, Edzard. 1979. *Vorgriechisch-Kartvelisches: Studien zum ostmediterranen Substrat nebst einem Versuch zu einer neuen pelagischen Theorie*. Leuven-Louvain: Edition Peeters.

Georgiev, Vladimir I. 1941–1945. *Vorgriechische Sprachwissenschaft* I-II. Sofia: University Press.

Heubeck, Alfred. 1961. *Praegraeca: Sprachliche Untersuchungen zum vorgriechisch-indogermanischen Substrat*. Erlangen: Verlag Universitätsbund Erlangen.

Iversen, Rune & Guus Kroonen. 2017. Talking Neolithic: Linguistic and archaeological perspectives on how Indo-European was implemented in southern Scandinavia. *American Journal of Archaeology* 121(4). 511–525.

Katičić, Radoslav. 1976. *Ancient languages of the Balkans*. The Hague/Paris: Mouton.

Kretschmer, Paul. 1896. *Einleitung in die Geschichte der griechischen Sprache*. Göttingen: Vandenhoeck & Ruprecht.

Kretschmer, Paul. 1939. Die vorgriechischen Sprach- und Volksschichten. *Glotta* 28. 231–278.

Kroonen, Guus. 2012. *Etymological Dictionary of Proto-Germanic*. Leiden/Boston: Brill.

Kuiper, F.B.J. 1956. "The etymology of ἄνθρωπος." ΜΝΗΜΗΣ ΧΑΡΙΝ, Gedenkschrift Paul Kretschmer 2. Mai 1866-9. März 1956, I, Wien, pp. 211–226.

Meillet, Antoine. 1908. De quelques emprunts probables en grec et en latin. *Mémoires de la Société de Linguistique de Paris* 15. 161–164.

Meissner, Torsten. 2013. Review of Beekes 2010. *Kratylos* 58. 1–32.

Mihaylova, Biliana. 2016. The Pre-Greek substratum revisited. In Bjarne Simmelkjaer Sandgaard Hansen, Benedicte Nielsen Whitehead, Thomas Olander & Birgit Anette Olsen (eds.), *Etymology and the European Lexicon*: *Proceedings of the 14th Fachtagung der Indogermanischen Gesellschaft, 17–22 September 2012, Copenhagen*, 307–317. Wiesbaden: Reichert Verlag.

Moreschini, Adriana Quattordio. 1984. *Le formazioni nominali greche in* -nth-. Roma: Edizioni dell'Ateneo.

Nikolaev, A. 2018. Greek etymology in the 21st century [Conference handout]. *Society for Classical Studies 149th Annual Meeting*. Boston, Massachusetts.

Palmer, Leonard Robert. 1980. *The Greek language*, London/Boston: Faber and Faber.

Pott, August Friedrich. 1853. *Die Personennamen insbesondere die Familiennamen und ihre Entstehungsarten: auch unter Berücksichtigung der Ortsnamen*. Leipzig: F.A. Brockhaus.

Schrijver, Peter. 1997. Animal, vegetable and mineral. Some Western European substratum words. In Alexander Lubotsky (ed.), *Sound Law and Analogy*. *A Festschrift for Robert Beekes*, 293–316. Amsterdam: Rodopi.

Schrijver, Peter. 2007. *Keltisch en de buren: 9000 jaar taalcontact*. Utrecht: University Press.

Schrijver, Peter. 2011. La langue Hattique et sa pertinence possible pour les contacts linguistiques préhistoriques en Europe occidentale. In Coline Ruiz Darasse & Eugenio R. Lujan (eds.), *Contacts linguistiques dans l'occident méditerranéen antique*, 241–255. Madrid: Casa de Velásquez.

Schrijver, Peter. 2018. Talking Neolithic: the case for Hatto-Minoan and its relationship to Sumerian. In Guus Kroonen, Bernard Comrie & James P. Mallory (eds.), *Talking Neolithic: Proceedings of the workshop on Indo-European origins held at the Max Planck Institute for Evolutionary Anthropology*, *Leipzig*, *December 2-3*, *2013*, 336–374. Washington, D.C.: Institute for the Study of Man.

Skoglund, Pontus et al. 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science 336*(6080). 466–469.

Skourtanioti, Eirini et al. 2023. Ancient DNA reveals admixture history and endogamy in the prehistoric Aegean. *Nature Ecology & Evolution* 7. 290–303.

Verhasselt, Gertjan 2009. The Pre-Greek linguistic substratum. An overview of current research. *Etudes Classiques* 77. 211–239.

Verhasselt, Geertjan. 2011. The Pre-Greek linguistic substratum. A critical assessment of recent theories. *Etudes Classiques* 79. 257–283

Van Windekens, Albert Jan. 1960. *Études pélasgiques*. Louvain: Publications universitaires.

Guus Kroonen

10 For the *nth* time: The Pre-Greek νϑ-suffix revisited

1 Introduction

As early as in the middle of the 19th century, the νϑ-suffix was suggested as a marker of "Pre-Greek" vocabulary. Pott (1853: 451) identified ἄψινθος 'wormwood', ἴονθος 'young beard', κήρινθος 'bee bread', μήρινθος 'thread', πείρι(ν)ς 'wagon box' and ὑάκινθος 'hyacinth' as "Pelasgian", i.e. as belonging to the non-Indo-European language spoken in Greece before the arrival of the Greeks. When Kretschmer accepted this in his *Einleitung in die Geschichte der griechischen Sprache* (1896: 401–409), it seemed to have become the canonical view, at least for a while (cf. Fick 1905: 153; Meillet 1908: 162; Cuny 1910: 154–156; Lafon 1934; Alessio 1941: 202).

There is no lack of attempts in later scholarship to reclaim an Indo-European origin for the suffix. Georgiev (1941) famously revamped Pott's Pelasgian into a submerged Indo-European branch (see also van Windekens 1960; Carnoy 1960). This led him to derive the -ινϑ-suffix from the PIE suffix *-*uent*- (cf. Τίρυνς, -υνθος < *Τίρινς, -ινθος < *τύρσι-ϝιντ-ς, i.e. "having a tower" (1941: 18)). Merlingen (1955: 7) took -ινϑ- from PIE *-*ent*- and -υνθ- from *-*n̥t*-. From an inherited perspective, the same element -ινϑ- has even been etymologized as the direct Greek continuation of the PIE root *h_2i-n-d^h- 'to burn' (Lanszweert 1994: 90–92). However, such manipulations have never gained serious traction (cf. Bonfante 1953; Schachermeyr 1955: 240; Hester 1957), and more recent studies generally endorse Pott's and Kretschmer's original view (Katičić 1976: 16–55; Moreschini 1984; Morpurgo Davies 1986: 110–111; Guidi 1990; Stumfohl 1990: 32; Beekes 2014).

Of crucial importance are the Greek toponyms with the νϑ-suffix, cf. Ἀράκυνθος, Ζάκυνθος, Ζώμινθος and indeed also the athematic Τίρυνς (Figure 1). Together, they form a direct link between the intrusive elements in the Greek lexicon and the indigenous linguistic landscape (Fick 1905: 154). This link is underlined by the striking fact that, although most placenames in -νϑ- are semantically obscure, some have direct counterparts in the appellatives, cf. Ἄψινθος ~ ἄψινθος 'wormwood', Κήρινθος ~ κήρινθος 'bee bread', Ὄλυνθος ~ ὄλυνθος 'wild or sterile fig'. The general consensus is that these toponyms, too, are non-Indo-European in origin (Kretschmer 1896: 401–409; Meissner 2013: 12–14). Naturally, it is possible that the Greeks themselves named these places after they had adopted Pre-Greek vocabulary. In fact, the very tradition of naming places after natural phenomena,

https://doi.org/10.1515/9783111337920-010

without any derivational modifications, may have been adopted from Pre-Greek (Fick 1905: 154). However, since the νθ-suffix is so closely associated with Cretan culture, most notably with King Minos' Λαβύρινθος (cf. Meissner 2013: 14), it seems preferable to see both toponyms and appellatives as parallel Pre-Greek loans in a situation of close language contact, possibly involving widespread bilingualism, in the not too distant prehistory of Greek.

A difficult question is whether the distribution of Pre-Greek toponyms in -νθ- extended beyond the Greek-speaking world, the answer being crucial to establishing the original range of the donor language. The νθ-suffix has famously been identified in the placenames in *-nd-* of Anatolia (Pauli 1885: 44–52; Schwyzer 1953: 60–61; Hester 1957; Carnoy 1960), cf. *Purusẖanda-*, and this area as the original center of the Pre-Greek language (Kretschmer 1896; Haley 1928). In later scholarship, the focus was shifted also to the Balkans and Italy, where many additional toponyms in *-nt-* and *-nd-* are found (Kretschmer 1925; cf. also Deroy 1957: 149). Although it does not *a priori* seem unlikely that Pre-Greek was spoken also outside Greece, the material suffers from an insurmountable delimitation problem (Meissner 2013: 12–13 fn. 8). Even when the roots of these toponyms resemble Greek lexical bases, without their original meanings it is impossible to establish any etymological links (cf. Laroche 1961: 57). Moreover, the *nt*-suffix itself is of such generic nature that it likely is unrelated to the νθ-suffix most of the time, and in some cases even demonstrably Indo-European in origin. For Italy, Calabrian Κόκυνθος seems to have a fair chance to be a "Pre-Greek exonym" (cf. Kretschmer 1925: 104), but Ferentum is an originally Etruscan settlement and Tarentum is clearly "post-Greek", i.e. it continues Greek Τάρας, gen. Τάραντος.

The ambiguity of toponyms is equally problematic for the school of thought that takes words in -νθ- to be vestiges of an Indo-European stratum (Meyer 1886; Renfrew 1998), whether a lost *satəm* branch (Georgiev 1941), Luwian (Palmer 1958; Szemerényi 1974: 152; Woudhuizen 2016: 316 fn. 4) or Lydian (Heubeck 1961). Proponents of an Anatolian affiliation have pointed to the suffixes Hittite *-ant-* and Luwian *-and-* as possible sources, but this is problematic (Beekes 2008; de Hoz 2004). While the presence in Asia Minor of toponyms in *-nd-* may show that the distribution of Pre-Greek once extended beyond Greece, into Anatolia, it does not follow from this that the original distribution of Anatolian conversely extended beyond the Aegean, into Crete and the Greece mainland (Meyer 1886). Like the Italian placenames in *-nt-*, the West Anatolian placenames in *-nd-* are intrinsically ambiguous, and while some can be analyzed as Anatolian in origin, others may be related to Pre-Greek (Bilgiç 1945; Yakubovich 2010: 9–11).

The non-Indo-European nature of the lexical elements in -νθ- in Greece is in any case implied by the structure of the roots to which the suffix is attached. These roots exhibit a strong tendency to consist of open syllables (Chadwick *apud* Hester 1957:

108; Hester 1968: 221), the sole exceptions of word-medial clusters being -σκ- (Κόσκυνθος, Ῥήσκυνθος), -ψ- (Ἄψυνθος, ἄψινθος), -ρμ- (τέρμινθος) and -λμ- (ἔλμι(ν)ς). Word initially, the only clusters known to occur are πρ- (Πρεπέσινθος, Προβάλινθος), σμ- (Σμίνθη, σμήρινθος). Thus, although the picture may be obscured by the paucity of the evidence, the donor language of these words was evidently different in structure from Proto-Indo-European, which abounded in complex consonant clusters both word-initially and word-medially. Unless we are willing to postulate the ancient presence of a dialect that had evolved beyond recognition in Greece (Renfrew 1998: 259), there is little to suggest that "Pelasgian", or at least the language that served as the source for Greek words in -νθ-, was a lost Indo-European branch.

Figure 1: Aegean toponyms ending in -νθ- (data from Schachermeyr (1955: 246), Carnoy (1960: 323–327) and Lindner (1995)).

2 Research questions

Given the monumental research history of the problem of the νθ-suffix (and Pre-Greek in general, see Verhasselt 2009), it is hard to imagine that much can be added to our understanding of it. However, since the νθ-suffix plays a prominent role in the study of the non-Indo-European lexical elements in Greek, the issue cannot be left untouched when engaging in the study of pre-Indo-European substrates.

The main questions that need to be addressed are concerned with the delimitation of the νθ-suffix. In other words, what are the criteria by which the suffix can reliably be identified? The difficulties at answering this question become immediately apparent when collecting the evidence. Due to the ambiguities in the material it is not immediately evident where to draw the line between relevant and irrelevant evidence. Multiple gradations of relevance must be distinguished for several divergent layers of evidence concerning the νθ-suffix.

Within Greek, the largest lexical category evidently consists of thematic nouns that have the νθ-suffix. The importance of this category is underscored by a relative multitude of cases. It includes many classical examples in -ινθος, like λαβύρινθος 'maze' and ὑάκινθος 'hyacinth', but also those in -υνθος, cf. κολοκύνθη 'gourd' and ὄλυνθος 'wild or sterile fig'. In addition, a small number of athematic nouns in -ι(ν)ς and -υνς can be identified, cf. ἕλμι(ν)ς 'intestinal worm' and Τίρυνς 'Tiryns'.

It is much less clear, however, what the relevance is of monosyllabic stems ending in -νθ-. Next to appellatives, βρίνθος (βρένθος) 'a water bird', γόνθος 'shell', γρόνθος 'fist', μίνθα/μίνθη 'mint', μίνθα/μίνθος 'excrement', ὄνθος 'dung', πλίνθος 'brick', σκινθός 'diver', σμίνθα/σμίνθος/σμίς 'mouse', there are some toponyms, Κύνθος, Σίνθος, Σμίνθη, Ψίνθος. Since little is known about the phonotactics of the language in which the νθ-suffix originated, it would be premature to *a priori* exclude these items, because of an unverified premise that monosyllabic stems cannot be segmented any further. On the other hand, without a distinguishable root, it is difficult to establish the suffixal nature of -νθ- in these words.

The aforementioned thematic and athematic nouns in -νθ- raise another question. Some of the examples are known to oscillate between variants with and without a nasal in their suffixes, cf. μέρμῖς, obl. μέρμινθ- ~ -ῖθ- 'thread' and τέρμινθος, τερέμινθος ~ τρέμιθος 'turpentine tree'. Although such variation is not by itself problematic in non-inherited vocabulary, it does raise the question whether words that exclusively appear without a nasal in the suffix should also be accepted as potential evidence of the νθ-suffix. Here we may mention κρῖ, κρῖθή 'barley', λέκιθος 'gruel' and λήκυθος 'oil flask': it cannot *a priori* be excluded that these words originally had nasal variants as well, e.g. **κρινθή, **λέκινθος and **λήκυνθος, that were somehow lost. Then again, without an ac-

tually attested nasal, it is impossible to verify whether the suffixes are variants of the νθ-suffix or in fact something else.

The relevance of words ending in -αθ- is even less clear. This category of words is relatively large and consists at least of ἀσπάλαθος 'spinous shrub', γαβαθόν 'cup', γεργαθός (and variants) 'creel', κάλαθος 'type of basket', κέρπαθος 'incense', κύαθος 'ladle for drawing wine', λάπαθον 'monk's rhubarb (*Rumex patientia*)', λάπαθος 'pitfall', μάραθον 'fennel (*Foeniculum*)', ψίαθος (ψίεθος) 'rush-mat'. It also includes the name of the island Κάρπαθος. Quite possibly, these words, mainly phytonyms, contained a different suffix -θ-, like ἄρκευθος 'juniper', στροῦς ~ στρουθός 'sparrow' and the placenames Κικύνηθος and Πάρνης, gen. Πάρνηθος (cf. Alessio 1941: 218–220). However, the doublet ἄμμος, ψάμμος ~ ἄμαθος, ψάμαθος 'sand' is well known to exhibit an alternation with and without -αθ-, ostensibly confirming the suffixal nature of the element at least in this word. Theoretically, these forms in -αθ- can therefore all be analyzed as having an νθ-suffix in which a nasal was lost. However, while etymologically obscure words ending in -ανθ- are extant, cf. κανθός 'corner of the eye', σκάνθος 'some bird', no examples exist in which -αθ- alternates with -ανθ-.[1] As a result, the parallelism with -ινθ- and -υνθ- must be considered incomplete (cf. Brandenstein 1936: 31). While it can potentially be saved, perhaps, by projecting the suffix back into Proto-Greek, assuming that -αθ- continues *-*ā̃t*^{h_4} < *-*n̥d*^{h_4}, with vocalized *n* (cf. Moreschini 1984: 106), the question is whether additional evidence can be furnished for the assumption that the suffix once contained a nasal.

The strongest evidence in support of -νθ- as a suffix, in any case, comes from lexemes exhibiting an alternation between forms with and without it (cf. Deroy 1957: 193). In Greek, such doublets are few, but the alternation κηρός 'beeswax' ~ κήρινθος 'bee bread' is an unambiguous case. Another example is the doublet ὄροβος 'bitter vetch' ~ ἐρέβινθος 'chickpea', even though its vowel alternation is not understood. These cases, too, pose a delimitation problem, however, albeit on a different level, because their suffixal alternation is not confined to Greek, but in fact resurfaces in some of the other Indo-European branches. Gr. κηρός ~ κήρινθος has an irregular comparandum in Lith. *korỹs*, and ἐρέβινθος cannot be disconnected from Lat. *ervum* 'bitter vetch' < **erwo-* ~ OHG *arawīz* < **orwīd-* (cf. Fick 1905: 153; Cuny 1910: 155; Kuiper 1956: 217). The presence of at least two cases with an almost pan-European distribution raises the daunting question whether other

1 The doublet ἴον ~ ἴανθος 'violet' (Deroy 1957: 193) does not involve a Pre-Greek suffix, but rather the inherited noun ἄνθος 'flower' (Frisk 1972: 1,704; Chantraine 1999: 465–466; Beekes 2010: 573).

manifestations of “Pre-Greek” words can be identified in the Indo-European languages of Europe, and perhaps beyond.

3 Evidence

In order to establish the full extent of the corpus of -νθ-, both within and outside Greek, and exclude all uncertain and unreliable cases, the relevant material is reviewed here. In the present section, the most promising evidence is analyzed to assess the relevance of the various lexical items. Each item is evaluated so as to establish the probability that it contains the νθ-suffix. The evidence is divided into the following categories:

++ Cases displaying an alternation with and without an νθ-suffix
+ Cases that plausibly contain an νθ-suffix
? Cases that possibly but not demonstrably contain an νθ-suffix
– Rejected cases

The corpus is compiled along the lines of the discussion of the problem in the previous section. Only words potentially having the suffixed -ινθ- and -υνθ- are considered, and additionally those in -ῑθ- and -ῡθ-. Words in -αθ- are left out of consideration unless an alternation is found similar to that of the νθ-suffix.

3.1 Potential cases with a Greek manifestation

3.1.1 Garlic 1 [?]

Gr. ἄγλῑς, -ιθος ‘clove of garlic’ (Ar., Hp.). Within Greek, this word cannot be separated from γέλγῑς, -ῑθος ‘(clove of) garlic’, but the correlation is unclear (Chantraine 1999: 12; Beekes 2010: 13). By invoking the non-Indo-European *a*-prefix it is possible to set up an original doublet **a-ggl-* : **gegl-* (Kroonen 2012a; Schrijver 2018: 362).

Outside Greek, it has a correspondence in Lat. *ālium* (Pol.+), *allium* (inscr.) ‘garlic’. If the form *ālium* is primary, it could go back to PIt. **aχalio-*, or if the later *allium* is given preference, PIt. **adlio-* (< **ag(d)lio-*?). Whichever is correct, the Italic form cannot be derived, without problems, from the Greek, because 1) the onset αγλ- matches (the pre-stages of) neither Lat. *āl-* nor *all-*. Furthermore, the Latin word does not show a trace of the suffix -ῑθ-. This would make the compari-

son comparable to the case of ἐρέβινθος ~ Lat. *ervum*, although the parallelism would be more complete if Greek had **ἄγλινθος or Latin ***ervium*.[2]

A third correspondence is found in Berber, cf. Awjila *agílum*, Ghadamsi *ağelum* 'garlic', which might be derived from Proto-Berber **agVlum* (Marijn van Putten, *p.c.*). This looks close to the reconstructed Italic proto-forms, but it does not seem possible to assume a direct loan from Latin into Berber.

3.1.2 Loom-weight [?]

Gr. ἀγνύς, -ῡθος 'loom-weight' (Plu., Poll., Hdn.). This word, generally assumed to be of obscure origin (Frisk 1972: 1:13; Chantraine 1999: 12; Beekes 2010: 14), contains the suffix -ῡθ-.

3.1.3 Linnet [+]

Gr. αἴγινθος (Dionys.), αἴγιθος (Arist.) 'a kind of bird, possibly linnet' shows an alternation between the suffix variants -ινθ- and -ῑθ- (Oštir 1930: 14; Furnée 1972: 285). The *o* of the variant αἰγίοθος (Arist.) is obscure, as is the etymology of the word itself (Chantraine 1999: 30; Beekes 2010: 32).

3.1.4 Cake 1 [?]

Gr. ἄμιθα 'a kind of meat, condiment' (Anacr.) could have a nasal-less variant of the suffix -ινθ-, but if somehow related to ἄμης, -ητος m. 'a kind of milk-cake' (Chantraine 1999: 75; Beekes 2010: 88), the suffix could be of a different nature.

3.1.5 Juniper [?]

A well-known irregular alternation exists between Gr. ἄρκευθος (Hp.) and Cypr. ἄργετος (Hsch.) 'juniper (*Juniperus macrocarpa*)'. The element -ευθ- looks similar

2 A Greek variant *ἄγλίον can perhaps be postulated as a possible source. An unsuffixed form has been assumed, albeit tacitly, on the basis of Hsch. ἀγλίδια·σκόροδα. Rather than interpreting this as having an interchange θ ~ δ (Furnée 1972: 194; Beekes 2010: 13), it can synchronically be analyzed as a diminutive to a root ἀγλ(ι)- (Frisk 12). For a similar proportion, cf. κηρός : κηρίον : κηρίδιον : κήρινθος.

to the suffix -ῡθ-, but beyond the potentially parallel κέλευθος 'road', ostensibly from κελεύω 'to command', it is isolated in Greek and it may have a different origin (Chantraine 1999: 109). The Spartan name Ἄρκεσος (Plu., Pel.) could reflect a form *ἄρκεθος (with regular θ > σ) that is closer to the Cypriot variant reported by Hesychius.

Outside Greek, Latv. *ẽrcis* 'juniper' invites the conclusion that -ε(υ)θ- is indeed suffixal and that both the Greek and the Latvian words were borrowed from related donor languages that shared a base **ark-* (Beekes 2010: 132). Ru. *rakíta*, Cz. *rokyta*, SCr. *ràkita* 'brittle willow' < PSl. **orkyta* is usually adduced here as well. The Balto-Slavic forms have alternatively been connected to Lat. *arcus* 'bow', however, and Go. *arhvazna* 'arrow' (de Vaan 2008: 52).

3.1.6 Bathtub [+]

Myc. *a-sa-mi-to*, Gr. ἀσάμινθος 'bathtub' (Hom.) contains the ινθ-suffix (Schwyzer 1953: 61; Deroy 1957: 182–183; Chantraine 1999: 122). Its ultimate etymology is unknown. Not much can be said for derivation, through an Indo-European substrate, from **akámn-to-* "made of stone" (Georgiev 1941: 45), or from a hypothetical Pre-Greek element **s₂am-* 'wet' (Alessio 1943). The segmentation into **ἀ-σάμινθος "without sand" (Deroy 1957: 182–185) is too fantastical.

Since ἀσάμινθος ostensibly adds the ινθ-suffix to a base **as₂am-* (von Blumenthal 1930), various interpretations take this element as a borrowing from the east. Garnier and Sagot identify it as Lydian **asamv* 'seat' (2020: 186). Semantically more straightforward are proposals starting from Semitic loans, i.e. from either Akk. *assammu* 'goblet' (Furnée 1972: 45–46; Beekes 2010: 146) or Akk. *namsū* 'washing bowl, tub' (Reece 2002), assuming that the ινθ-suffix was added after the borrowing. The assumption of direct borrowing from Akk. *namasītu* 'bathtub' (Szemerényi 1971: 657) only works if the native Akkadian suffix was *reinterpreted* as -ινθ-.

3.1.7 Wormwood [+]

Gr. ἄψινθος 'absinth (*Artemisia absinthium*)' (Aris.) as well as the more derived ἀψινθία and the ultimately prevailing ἀψίνθιον contain the suffix -ινθ- (Schwyzer 1953: 61; Chantraine 1999: 152). A variant in -υνθ- has been proposed on the basis of the placename Ἄψινθος ~ Ἄψυνθος. In the root, formal irregularities are exhibited by ἀσπίνθιον (Hsch.), apparently with metathesis, and ἀπίνθιον (E.M.), with no *s*, but these forms are isolated.

3.1.8 Magical plant [?]

Gr. βάλλις 'a magical plant' (attributed by Pliny to Xanthus) is said by Herodianus to have been inflected as δέλλῑς, -ῑθος 'wasp' (see below).

3.1.9 Bison [+]

Gr. βόλινθος 'bison' (Arist.) is widely assumed to contain the suffix -ινθ-. The often cited inner-Greek comparison with βόνασ(σ)ος 'bison' (Arist.), perhaps to be segmented into a root βόν- and a suffix -ασ(σ)ος, is uncertain (Beekes 2010: 225), because it builds on the assumption that βόλινθος was dissimilated from *βόνινθος (Frisk 1972: 1,259; Chantraine 1999: 184). As a result, it is not possible to substantiate a doublet of one variant with and one without the suffix -ινθ-.

Outside Greek, Chechen, Ingush *bula* 'wisent' (< PNakh **bula*, obl. **bulan-*, Peter Schrijver, *p.c.*) and Udi *bele* 'cattle' have been adduced as possible comparanda (Bengtson 2017: 80). However, without a clear representation of the νθ-suffix in these languages, the comparison remains uncertain. Given the shortness of the element, a chance resemblance cannot be excluded, cf. E *bull* (Merlingen 1962a: 28, 47).

3.1.10 Cow dung [?]

Gr. βόλβιτον, βόλβιτος (Thphr.), βόβλιτον (Eust.), βόλιτον (Crat.) 'cow dung' has the variants βόλβιθος (PMag. Par.) and βόλυνθον (Hsch.), the latter of which points to the presence of a suffix -υνθ-. Notably, this βόλυνθον lacks the second β. It is tempting to speculate that this loss is somehow correlated with the rounded vowel of the suffix. However, since the form is only attested as a Hesychius gloss, and not necessarily primary (Chantraine 1999: 180), not much can be concluded from it.

The non-Indo-European origin of the lexical base is implied by the aforementioned irregularities (Beekes 2010: 225), although these previously have been ascribed to the register in which the word occurs (Frisk 1972: 1,249; Chantraine 1999: 184). The inclusion of Alb. *bajgë*, dial. *balgë* 'dung of cow, horse' leads to the reconstruction of a "Balkan" element **bolg*^{w}- (Demiraj 1997: 86–87). The validity of this reconstruction is limited, however, as it only accounts for a subsection of the Greek variants, the remaining forms requiring **bol-*.

3.1.11 Chickpea 1 [+]

Gr. γάλινθοι, γάλιθοι, γέλινθοι, γέρινθοι 'chickpeas' (Hsch.) form a cluster of irregular variants alternating between *a* ~ *e* and *l* ~ *r*. All of these variants exhibit the suffix -ινθ- (Chantraine 1999: 208). The forms γάλινθοι and γάλιθοι furthermore alternate between the nasal and non-nasal variants of this suffix (Beekes 2010: 258).

3.1.12 Garlic 2 [?]

Gr. γέλγις,-ῑθος '(clove of) garlic' (Thphr.) has the suffix -ῑθ- (Beekes 2010: 265). The word has been compared to ἄγλῑς (Frisk 1972: 1,295; Furnée 1972: 194–195), assumed that it arose by metathesis from a reduplicated form *γέγλις (Chantraine 1999: 214). The latter is similar to Akk. *gidlu* 'string of onions, garlic' (Kroonen 2012a), especially when the regular Greek change of *-δλ- to -γλ- is taken into consideration. If it is a loanword, the suffix may have been added in the Aegean. A caveat is, however, that the *i*-stem, attested pl. γέλγεις (Thphr.), appears early and only later gives way to forms in -ῑθ-. This may suggest that the latter is secondary (Monzó Gallo 2015: 99–103).

3.1.13 Butcher's broom [+]

Gr. γοργυνθία 'butcher's broom (*Ruscus aculeatus*)', potentially derived from *γόργυνθος, is found in Dioscorides' *De materia medica* (IV, 144) and contains an element -υνθ- (cf. Carnoy 1960: 324).

3.1.14 Wasp [?]

Gr. δέλλῑς, -ῑθος 'a kind of wasp' (Hsch., Hdn.) contains an athematic suffix -ῑθ- (Chantraine 1999: 260; Beekes 2010: 313). The formation served as the basis for δελλίθιον 'wasp nest' (Hsch.). The traditional comparison with Lith. *gélti* 'to sting' (Pokorny 1959: 470–471; Frisk 1972: 1,361), through *g^{w}el-nī-*, is difficult, as the involved PIE root is *g^{w}elH-*, with a laryngeal. The laryngeal may have been lost in a verbal formation *g^{w}elH-ie-* due to Pinault's law, but no such formation is known from Greek.

3.1.15 Intestinal worm [+]

Gr. ἕλμις (Arist., Thphr.), ἕλμινς (Hp.), gen. ἕλμινθος (Hp.), acc. ἕλμιθα (inscr.) 'intestinal worm' is an athematic stem in -νθ-. With a different suffix, the same base is found in ἕλμιγξ, -γγος (Hp.), apparently continued by MoGr. Cret. ὄρμιγγας (Thumb 1901: 100) and Capp. ὄρμιγκας (Fraenkel 1906: 290 fn. 3). This may be a diminutive, cf. the proportion of λᾶας 'stone' : λᾶϊγξ 'pebble' (Fraenkel 1906: 291; Monzó Gallo 2015: 128). Remarkably, ἕλμιγξ has a parallel diminutive *λίμιγξ as implied by MoGr. Capp. λιμίγγι. It seems to be based on the divergent Cypriot variant λίμι(ν)ς, attested as λίμινθες (Hsch.).

Under the assumption that the suffixes -ινθ- and -ιγγ- are secondary, ἕλμις has been derived from the PIE root **uel-*, continued in Greek by εἰλέω 'to roll, wind', with the suffix *-mi-* as in PIE **k*^{w}*rmi-* ~ **urmi-* 'worm' (Frisk 1972: 1,501; Chantraine 1999: 341). An *i*-stem is indeed attested as ἕλμεις (Dsc.). This account does not allow us to explain Cypr. λίμινθες and Capp. λιμίγγι (< *λίμιγξ), however, but by *ad hoc* metathesis and raising of the vowel. Additionally, metathesis has to be assumed for MoGr. λεβίθα (with dissimilation of μ > β), Mani λεμίθα, Tsak. λεμίσα,[3] all ostensibly from *λέμις.[4] Taken at face value, the various forms continue **s*$_1$*elm-*, **s*$_1$*lem-* and **s*$_1$*lim-*, an alternation that recalls that of τέρμινθος.

3.1.16 Chickpea 2 [++]

Gr. ἐρέβινθος 'chickpea' is generally accepted as a loan exhibiting the νθ-suffix (Schwyzer 1953: 61; cf. Frisk 1972: 1,549–550). It appears itself to have been borrowed, possibly twice, into Georgian: once as OGeo. *erbindi* (Beekes and Kuipers 1975: 85) from what looks like unattested *ἔρβινθος, and a second time as Geo. *erevindi*, possibly through *koine* Greek ἐρεβίνθιον (cf. MoGr. ρεβίθι).[5] Within Greek, the word alternates with ὄροβος 'bitter vetch', but the vocalic alternation is unclear.

Outside Greek, Lat. *ervum* 'vetch' < PIt. **erwo-* and ON *ertr*, OHG *arawīz* 'pea' < PGm. **arwīt-* are accepted (yet irregular) comparanda (Furnée 1979: 22). As early as Hehn (1885: 167), the Proto-Germanic element **-īt-* < **-īd-* has been compared to what in Greek surfaces as -ινθ-. At the same time, the root to which it is attached,

3 With the regular 4th c. BCE Laconian change θ > σ.

4 For the stem class transfer, cf. MoGr. όρνιθα 'hen' for original ὄρνις.

5 On a third variant, *erevandi*, cf. Lafon (1934: 34): "Son *a* exclut l'hypothèse d'un emprunt au grec ἐρέβινθος.". However, according to Beekes and Kuipers (loc.cit.) the form should be read *erevendi* or *erevindi*.

**arw-*, is closer to ὄροβος. Conversely, PIt. **erwo-* looks most similar to the root of ἐρέβινθος, but lacks the suffix. Thus, Greek, Latin and Germanic appear to attest to *different recombinations* of the same non-Indo-European base and suffix.

A hitherto overlooked comparandum is Arm. *aṙowoyt*, *aṙawoyt* 'alfalfa (*Medicago sativa*)' < **HrVbʰ-oud-* . Although its semantics are more peripheral, the word still designates a species of the legume family. If indeed connected, it represents yet another recombination of the same elements that are found in the European languages, quasi Greek **ὀρόβυνθος.

Outside Europe, a likewise overlooked term is reported for the Pamir languages: Shugni *rivand*, Rushani *ravand* 'wild chickpea (*Cicer songaricum*)', Yazghulami *raván* 'pea' (Baranov & Raykova 1928: 47 fn. 2; Edel'man 1971: 221; Morgenstierne 1974: 70). The formal and semantic similarities of these forms to Gr. ἐρέβινθος are difficult to ignore. If not a chance resemblance, these Pamir forms could possibly constitute an Iranian manifestation of the European term, viz. PIIr. **H(a)rab$^{(n)}$anT-*. However, the limited distribution could be suggestive of a late loan (cf. Steblin-Kamenskiy 1982: 41) and it is perhaps conceivable, given the history of the region, that the word was borrowed from Bactrian Greek.

3.1.17 Dance [+]

Gr. ἐσχάρινθον 'a dance at Sparta' (Poll.) is isolated and attested late. It contains the suffix -ινθ-, but the root is obscure. Inner-Greek derivation from ἐσχάρα 'hearth, house' (Schwyzer 1953: 526; Furnée 1979: 45) seems speculative (Frisk 1972: 1,577; Chantraine 1999: 380); even if the -ινθ-suffix was productive in Greek, ἔσχαρος 'a fish' would be an equally arbitrary derivational base.

3.1.18 Catnip [?]

Gr. καλαμίνθη (Hp, Aris., Hsch.), καλάμινθα (Phot.), καλάμινθος (Nic.). This word looks like a compound with μίνθη 'mint' (cf. Chantraine 1999: 483), but the first element is obscure (Frisk 1972: 1,760) and probably foreign. Derivation from κάλαμος 'reed', either by haplology from *καλαμο-μίνθη (Meyer 1896: 393) or by direct addition of the ινθ-suffix (Schwyzer 1953: 526), is semantically unmotivated (cf. Beekes 2010: 621).

3.1.19 Beeswax [++]

Gr. κηρός 'beeswax' (Hom.) alternates, within Greek, with κήρινθος 'bee bread' (Arist.). The latter is sometimes analyzed as a genuinely Greek derivation of the former (Frisk 1972: 1,844), but the νθ-suffix generally characterizes non-native vocabulary (Chantraine 1999: 526–527). A non-native origin is supported also by the clearly connected Lat. *cēra* f. 'wax' and Lith. *korỹs*, Latv. *kâre* 'honeycomb' (Beekes 2010: 689); while *cēra* may be a Greek loan, the Baltic forms point to a quasi PIE **keh*$_2$*-ro-*, which is irreconcilable with the vocalism of κηρός (which does not continue **κᾱρός), unless one is willing to invoke Eichner's law.

3.1.20 Gourd [+]

Gr. κολοκύνθη 'gourd, *Cucurbita maxima*' (Hp., Com., Arist.), also Att. κολοκύντη (Phryn.). The word ostensibly consists of an etymologically obscure element **kolok-* and the suffix -υνθ-. The formal resemblance to Skt. *kālinda-* n. 'watermelon' is coincidental (Frisk 1972: 1,902; Beekes 2010: 738). The habitually compared P *kālak*, *kalik* 'unripe melon', Kurd. *kalak* 'a small melon' is, too, very much unrelated.[6]

3.1.21 Cake 2 [+]

Gr. κόρυνθος 'cake' (Hsch.). An isolated word. The link to Gr. κόρυς 'helmet' (Chantraine 1999: 569) seems speculative. The epithet of Κόρυνθος Ἀπόλλων (Paus.) is sometimes interpreted as referring to a sacrificial cake, for which parallels appear to exist.

3.1.22 Helmet [?]

Gr. κόρυς, -υθος 'helmet' (Hom., Eur.). The vowel of -υθ- is short, which suggests that the suffix is not a non-nasal variant of -υνθ-, but simply had a disyllabic base *κορυθ-. The potentially derived κορυνθεύς 'rooster' (Hsch.) does show a root κορυνθ- containing this element, however. No further certain comparanda are ex-

6 As implied by Kurmanji *k'al*, *k'alak*, *k'alik* '(water)melon', these go back to an adjective attested as *k'al* 'unripe, green (e.g. of melons)' and Sorani *kâł* 'raw; unripe, immature'.

tant. The traditional derivation from the PIE root **ḱerh₂-* is uncertain (Chantraine 1999: 569), because it is difficult to explain the suffixation (Frisk 1972: 1,925–926).

3.1.23 Mask [+]

Gr. κυλίνθιον 'wooden mask' (Hsch.) possibly constitutes a derivation from *κύλι(ν)ς or *κύλινθος. The word is close to the semantically identical κύριθρον (Furnée 1972: 288). The extension of the latter seems parallel to that of MoGr. λεβίθρα 'intestinal worm', but the formations are from very different time periods.

3.1.24 Bundle [?]

Gr. κώμῡς, -ῡθος f. 'bundle of hay; *pl.* reeds' (Crat., Thphr.), 'laurel branch' (Hsch.) has the suffix -ῡθ- (Schwyzer 1953: 510; Beekes 2010: 814), but no variant -υνθ- is found. The connection with the PIE root **kem-* 'to squeeze' is questioned already by Pokorny (1959: 555).

3.1.25 Maze [+]

Myc. *da-pu₂-ri-to*, Gr. λαβύρινθος 'maze; coil' (Hdt., Str.). This word serves as a textbook example of the ινθ-suffix. The alternation Myc. *d* ~ Gr. *l* is irregular and comparable to that of δάφνη ~ λάφνη (Furnée 1972: 397–398). If *da-pu₂-ri-to* is to be interpreted as *δαφύρινθος (Judson 2017; Piquero Rodríguez 2017: 257–258), the equally irregular interchange of φ ~ β can be added to the set of non-native features. The etymology of this word is much researched, but remains obscure (Frisk 1972: 2,67; Chantraine 1999: 610–611; Beekes 2010: 819): Neither derivation from λάβρυς 'axe', with the ινθ-suffix (Alessio 1941: 216; Alessio 1943: 126), nor from λαύρα 'narrow street' (Güntert 1932; Deroy 1957: 173–176) can be accepted without reservation.

3.1.26 Coffin [+]

Gr. λαΐνθη 'stone chest, box, coffin' (Cyr.). Given the fact that a stone object is referred to in Cyril's glossary ("λάρναξ λιθίνη"), it is possible to speculate on derivation from λᾶας 'stone' (cf. Furnée 1972: 239 fn. 55), but it is uncertain whether this is the original *Benennungsmotiv*.

3.1.27 Dock [-]

Gr. λάπαθον 'monk's rhubarb (*Rumex acetosa*)' (Ep., Thphr.) has been connected to Lat. *lappa* 'burdock' under the assumption of a suffixal alternation similar to that of ψάμμος ~ ψάμαθος (Alessio 1941: 218–220; Deroy 1957: 193; Frisk 1972: 2,84). The divergent meanings of the words, however, do not permit any conclusions concerning their relatedness.

3.1.28 Chickpea 3 [+]

Gr. λεβίνθιοι 'chickpeas' (Hsch.). This word seems isolated, but given that similar alternations are exhibited by γέλινθοι ~ γέρινθοι, it is difficult to exclude the possibility that this word is some obscure variant of ἐρέβινθος (Schwyzer 1953: 61; Moreschini 1984: 44).

3.1.29 Gruel [?]

Gr. λέκιθος 'gruel of pulse or cereals; egg yolk' (Hp., Arist.). It is unclear whether -ιθ- is a variant of -ινθ- in this word. Regardless, the etymology of this word is obscure (Frisk 1972: 2,103; Chantraine 1999: 630; Beekes 2010: 847). The derivation from λέκος 'dish, pot, pan' (Grošelj 1952: 212) is possible, but requires an additional semantic shift from 'pot' to 'gruel'.

3.1.30 Flask [?]

Gr. λήκυθος 'oil flask, oil bottle' (Hom.+), Dor. (Epid.) λάκυθος. In absence of a variant **λήκυνθος the example remains uncertain. No nasal appears in Etr. *leχtumuza* 'a kind of small bottle, flask', apparently borrowed from Greek (cf. Whatmough 1997: 62). Other comparanda, such as λάγυνος 'flagon, pitcher, flask' and Sum. *lahan* 'bottle, flask' (Furnée 1972: 121), are too vague.

3.1.31 Thread 1 [-]

Gr. μέρμῑς, -ῑθος 'cord, string' (Hom.), also thematic μέρμῑθος (Hsch., Zonar.). This word contains the athematic suffix -ῑθ- (Schwyzer 1953: 510; Furnée 1972: 289). No evident etymology is at hand. Derivation from a PIE root **mer-* 'to bind' (Pokorny

1959: 733) is considered untenable (Chantraine 1999: 687). The connection with μήρινθος, under the assumption of "broken reduplication" (Frisk 1972: 2,211), is rendered implausible by the variant σμήρινθος.

3.1.32 Thread 2 [+]

Gr. μήρι(ν)ς, acc. μήρινθα (Orph.), thematic μήρινθος (Hom., Aris.) and σμήρινθος (Pl.) 'cord, (fish)line, string' point to a non-native root extended with the suffix -ινθ- (von Blumenthal 1930; Schwyzer 1953: 510; Frisk 1972: 2,230; Chantraine 1999: 697). The plausibly related doublet μῆριγξ 'bristle' (Hsch.) ~ σμῆριγξ '(animal) hair' (Lyc., Poll., Hsch.) adds the (diminutive?) suffix -ιγγ- to the same stem, for which ἔλμιγξ can be invoked as a parallel. Perhaps the meaning 'rope' developed from '(twined) animal hair'.

No certain outer-Greek connections are at hand. Hitt. *išmeri-* 'bridle, rein' has been adduced as another manifestation of the same non-Indo-European element (Furnée 1972: 289), but it may alternatively be derived from inherited $^{*}sh_2$*-mer-* (cf. Rieken 1999: 364 ff.). Deroy (1957: 191–192) reconstructs a Pre-Greek base **merw-*, but given the alternation μήρ- ~ σμήρ- < $^{*}s_1$*mēr-* ~ $^{*}s_2$*mēr-*, an initial sibilant seems required.

3.1.33 Mint [?]

Gr. μίνθα (Thphr.), μίνθη (Plu.) 'mint'. If Lat. *menta* 'id.' is not ultimately borrowed from Greek, e.g. through Oscan mediation, the mutually irreconcilable forms point to borrowing from an unknown source (Meillet 1908: 162; Hester 1964: 360; Frisk 1972: 2,241–242; Chantraine 1999: 704).

3.1.34 Excrement [?]

Gr. μίνθη, μίνθα (Hsch.), μίνθος (Thphr., Plu.) 'excrement'. This word is generally held to be etymologically obscure (Frisk 1972: 2,242; Chantraine 1999: 704). If it contains the suffix -ινθ-, a root **(s$_1$)m-* must be inferred, but there is no additional evidence on the basis of which such an element can be secured.

3.1.35 Chickpea 4 [+]

Gr. ὀδόλυνθοι 'chickpeas' (Hsch.). This word ostensibly adds the suffix -υνθ- to a foreign base (Beekes 2010: 1046). No certain etymology can be given, but it is possible to think of a compound with ὄλυνθος, created within the donor language. The interpretation as *ὁδο-όλυνθοι "road figs" (Strömberg 1944: 9) seems like a folk etymology (Chantraine 1999: 774).

3.1.36 Wild fig [+]

Gr. ὄλονθος, ὄλυνθος 'wild fig; sterile fig' (Hp., Hes., Thphr., Gal.) is widely acknowledged as having the suffix -νθ- (Schwyzer 1953: 61; Frisk 1972: 2,384; Chantraine 1999: 796). No outer-Greek comparanda can be given but the Latin glosses *bolunda* (Cyrillus 554; Cod. Harl. 3376: 7v.) and *bolundi* (Ampl. gl. 279) 'unripe fig(s)'. These have been explained as reflecting a borrowing from Doric Greek (Rönsch 1886; Alessio 1944: 138–139). Under the assumption that *b* substitutes ϝ (Biville 1990: 89–90), the reconstruction of the Greek root can be set to **wol-* (Beekes 2010: 1074). The alternative is to take *bolunda* as a parallel borrowing from an unknown source, in which case the suffix *-unda* has to be an independent, Latin reflection of the suffix -υνθ- (Furnée 1972: 198). For this, we may potentially compare *harundō* 'reed' and *hirundō* 'swallow' (see below). However, Lat. *menta* and **plenta* would then show conflicting outcomes of the same suffix.

3.1.37 Bird [-]

Gr. ὄρνῑς, -ῑθος 'bird; chicken' (Hom.+) can be derived from PIE **h₃er-on-* 'eagle' (Pokorny 1959: 325–326), probably through the feminine **h₃r-n-iH-s* (cf. also Chantraine 1999: 832). Although the word contains the potentially non-native suffix -ῑθ-, this seems secondary, like in δέλλῑς 'wasp' (Schwyzer 1953: 510; Beekes 2010: 1106). This type, which seems to have been extended also to ἄγλῑς, γέλγῑς and μέρμῑς, appears to be a vestige of the Proto-Indo-European *vr̥kī*-class of *iH*-stems (cf. Schwyzer 1953: 465).

3.1.38 Basket [+]

Gr. πείρι(ν)ς, -ινθος 'wagon box, trolley basket' (Hom., Hsch.) consists of a base **per-* combined with the non-native suffix -ινθ- (Chantraine 1999: 871; Beekes

2010: 1163). No etymology can be given. The reconstruction **peri-uend^{h}-* "herumgewunden" (Meringer 1907: 228) fails to account for the vocalism. The proposed link with the toponym Πέρινθος, Πείρινθος (Wörner 1876) is unwarranted.

3.1.39 Driver seat [+]

Gr. πύρινθον 'part of the vehicle, where the driver sits'. A proposed comparandum is σπυρίς, -ίδος 'basket', but the connection is invalidated by the divergent meanings and the variation between a root with and without *s*.

3.1.40 Laurel [++]

Gr. δάφνη (Hom., Hes.), δαύχνα (Alcm.) and λάφνη (Hsch.) 'laurel (*Laurus nobilis*)' are in irregular correspondence but undeniably related (Cuny 1910: 159 fn. 1; Furnée 1979: 22). The initial alternation *d* ~ *l* is reminiscent of that of λαβύρινθος 'maze' and although λάφνη is only found in Hesychius, it has an echo in Lat. *laurus* (Chantraine et al. 2009: 255). The variants δάφνη and δαύχνα reflect different proto-forms, viz. PGr. **dak^{wh}-nā-* and **dauk^{h}-nā-*, pointing to independent borrowing (Frisk 1972: 1,353). The reconstruction of a labiovelar **k^{wh}* for δάφνη (Beekes 2010: 306) is potentially supported by Lat. *laurus*, which can continue PIt. **lau(χ)-ro-*, but not **laf-ro-*.[7]

The Greek forms lack a variant with the suffix -ινθ-, but Geo. *rapindi* 'laurel' appears to have an equivalent of it, viz. *-indi* (Lafon 1934: 32–33). Given that the φ of the aforementioned Greek and Latin forms must go back to a labiovelar, *rapindi* is difficult to interpret other than as a borrowing from unattested Gr. **ῥάφινθος (cf. OGeo. *erbindi* 'chickpea' for a parallel). If correct, it implies a root variant **rag^{wh}-* that can be added to the other irreconcilable variants **dag^{wh}-* and **lag^{wh}-*.

3.1.41 Turpentine tree [+]

Gr. τέρμινθος (Hp., Thphr.), τρέμιθος (Nic.), τρίμινθος (Const. Porph.), τερέμινθος, τερέβινθος (Sept., Hell.) 'turpentine tree (*Pistacia terebinthus*)'. The variants found with these words, even when ignoring the late ones, are difficult to reconcile with

7 The origin of the suffixal alternation of *n* ~ *r* is unclear, but finds a parallel in βλῆχνον (Sch., Cyr.) ~ βλῆχρον (Dsc.), βλήχρα (Hsch.) 'fern'.

each other. Together, they point to an etymologically obscure base **t(e)r(e)m-* combined with the suffix -ι(ν)θ- (Schwyzer 1953: 61; Alessio 1944: 139; Furnée 1972: 289; Frisk 1972: 2,881; Chantraine 1999: 1107; Beekes 2010: 1469).

3.1.42 Stone [?]

Gr. πλίνθος 'brick' (Hdt., Arist.). This obvious loan is often segmented into an obscure root **πλ-* and a suffix -ινθ- (Meillet 1908: 162; Frisk 1972: 2,562–563; Chantraine 1999: 917–918; Beekes 2010: 1211). Southern Italian **plenta*, cf. Camp. *kyenda*, Cos. *kyenta* may be a parallel borrowing from the similar source (cf. Alessio 1944: 139), although it seems difficult to rule out an Oscan-mediated loan from Greek (cf. Gr. μίνθη ~ Lat. *menta* 'mint'). The comparison with OE *flint*, OHG *flinz* < PGm. **flinta-* 'rock, flint' (Torp 1909: 253; Orel 2003: 107; Güntert 1932: 22) is certainly incorrect (thus already Kretschmer 1934: 12): this formation can plausibly be derived from the Germanic verbal complex as represented by Sw. *flinta* 'to strike, hit' and Nw. *flunta* 'id.', e.g. through a specialization 'to chip'.

3.1.43 Spider [+]

Gr. σαλαμίνθη (Cyranid.) 'spider'. This hapax appears to add the ινθ-suffix to an obscure base **s₂alam-* (cf. Beekes 2014: 88). The attribution of σαλαμίνθη to Hesychius (Merlingen 1962a: 31) seems to be a mistake (Hester 1964: 364).

3.1.44 Mouse [?]

Gr. σμίνθος 'mouse' (Aesch., Lyc., Str.), σμίνθα (Hsch.) is widely acknowledged as a Pre-Greek word on the basis of its alleged νθ-suffix (Kretschmer 1943: 133; Merlingen 1962b: 35; Katičić 1976: 53; Beekes 2010: 1369). The Hesychian form σμίς appears to continue an athematic variant (already Döhring 1888: 14).

The word has no evident etymology and possibly originates in Asia Minor (Frisk 1972: 2,750; Chantraine 1999: 1028). If the suffix -ινθ- is indeed to be identified here, the monosyllabic base **s₂m-* would be similar to PSem. **ŝ_m* 'kind of mouse, rat' (Kroonen 2016: 59), cf. Akk. *ušummu*, *šummu* 'dormouse', Arab. *šīm* 'rat'. Without an actual suffixal alternation this segmentation cannot be verified, however. On the positive side, the alternation between athematic σμίς and thematic σμίνθος is reminiscent of that of μήρις ~ μήρινθος and can be adduced in support thereof.

3.1.45 Hyacinth [+]

Gr. ὑάκινθος (Hom., Xen., Saph., Thphr.) 'hyacinth' is one of the textbook examples of etymologically obscure words with the ινθ-suffix (Pott 1853: 451; Chantraine 1999: 1149–1150). Beyond Greek, a traditional comparandum is Lat. *vaccīnium* (Meillet 1908: 162), but this seems unwarranted (Frisk 1972: 1,952–953). The attestation of the word does not suggest that it referred to the hyacinth and rather points to a shrub with berries (Vander Kloet 1992). The proposed connection with Lat. *bāca*, *bacca* 'berry' (Deroy 1957: 185–189) does not convince either, because the derivational base of *vaccīnium* is *vacc-*, not *bacc-*. This rather suggests that the word was not merely influenced by (Walde & Hofmann 1938–1954: 2,722) but in fact derived, within Latin, from *vacca* 'cow' with a suffix **-īnio-* (for which cf. OIr. *áirne*, W *eirin* 'sloe(s)' < PC **agrīnio-*).

3.1.46 Sand [-]

Gr. ψάμμος (Hom., Hdt.) and ἄμμος (Xen., Pl.) alternate with ψάμαθος (Hom., Plu.) and ἄμαθος (Hom., A.R.) 'sand'. The variation has traditionally been explained as resulting from several blends of two originally unrelated yet inherited words, one belonging to ψῆν 'to rub; to crumble' (Frisk 1972: 1,84; Chantraine 1999: 69; Garnier 2006). Irregular manifestations of the same word in some of the other Indo-European languages are rather suggestive of a loan (Kuiper 1956: 208; Deroy 1957: 183; Furnée 1979: 209; Kuiper 1995: 67). Lat. *sabulum* 'sand' reconstructs to Pre-It. **sabʰ-lo-* or **sadʰ-lo-*. Arm. *awaz* 'sand, gravel' appears to continue a base **sabʰadʰ-*, either directly through thematic **sabʰadʰ-o-* (Ačaŕyan 1971–1979: 1,351) or – if the change **dʰ* > *z* is rejected – **sabʰadʰ-s*, through an intermediate form **awaj* with thematicization and lenition (Thorsø, this volume). Finally, ON *sandr*, OE *sand*, OHG *sant*, ostensibly from PGm. **sanda-*, is regular from **samdʰ-o-*, but MHG *sambt*, *sampt*, Bav. *samb(d)* (Schmeller and Fromman 1872: II,282), Yid. *zamd* 'sand' (cf. Gerzon 1902: 97) does not show the regular Germanic change **-md-* > **-nd-* and rather requires **samm(a)dʰ-o-*. In view of these problems, it seems preferable to postulate a non-native stem **sam(m)-adʰ-* ~ **sabʰ-adʰ-*, without the νθ-suffix, even if the theoretical possibility exists that at least Greek -αθ- continues a vocalized manifestation of it, quasi PIE **-n̥dʰ-* (Moreschini 1984: 106; Kroonen 2012b: 243 fn. 6).

3.2 Potential cases without a Greek manifestation

3.2.1 Acorn (-)

OCS *želudь* < *$g^{w}elh_2$-e/ond-* and Lat. *glāns*, *-andis* < *$g^{w}lh_2$-nd-* can be derived from an inherited *nd*-stem related to e.g. Lith. *gìlė* < *$g^{w}lH$-ieh$_2$-* (Pokorny 1959: 472–473). Gr. βάλανος is usually taken together with Arm. *kałin* < *$g^{w}lh_2$-en-o-* (Beekes 2010: 195). Although the *nd*-suffix is rare in Proto-Indo-European, it has a parallel in Ru. *želúdok,* SCr. *žèludac* 'stomach' < *$g^{h}elH$-ond-*, whose Greek cognate χολάδες f.pl. 'intestines' < *$g^{h}olH$-n̥d-* (Pokorny 1959: 435; Derksen 2008: 556) shows that there is no relationship with the νθ-suffix.[8] An Indo-European origin of this word is further supported by the apparently ancient compound with **dieu-* 'sky god': Gr. διοσβάλανος 'chestnut', Arm. dial. *tkołin* 'hazelnut' (Martirosyan 2010: 348–349) and – if not a calque from Greek – Lat. *iūglāns* 'walnut' (Walde and Hofmann 1938–1954: 1,727; Schrijver 1991: 273).

3.2.2 Reed [?]

Lat. *harundō* f. 'reed, cane, rod'. This etymologically isolated word (de Vaan 2008: 279) may continue PIt. **χar-o/und-*. It contains a suffix *-und-*, which superficially resembles that of *hirundo* and *bolunda*, *bolundus*. The suffix can alternatively be derived from PIt. **-uD-n-*, however (Thurneysen 1883), and without a Greek comparandum, i.e. **χάρινθος *vel sim.*, the nature of the suffix cannot be verified.

3.2.3 Swallow [?]

The suffix of Lat. *hirundō* 'swallow' is similar to that of *harundō* 'reed' and *bolunda*, *-us* 'fig', and can hence be suspected to provide a non-Greek parallel to the suffix -υνθ-. The Greek comparandum is χελῑδών, however, not **χέλινθος or **χέρινθος, so that no direct comparison can be made. Potentially closer is Alb. *dallëndyshe* f. 'swallow' (cf. Çabej 1976: I,105–106) < PAlb. **da(u)lanT-* < quasi PIE

8 Without the suffix, cf. OIr. *gaile* 'stomach' < *$g^{h}lH$-io-* and related formations in Celtic (Hayden and Stifter 2022).

*ǵʰo(u)l(H)-(o)nT-, but due to the nature of the Albanian sound changes, it is impossible to establish the exact proto-form of this suffix.[9]

3.2.4 Swan [?]

ON *ǫlpt*, OE *ielfetu*, OHG *elbiz* 'swan' < PGm. **albut-* alternates with Po. *łabędź*, SCr. *làbūd* < PSl. **olbǫdь* (Derksen 2008). Given the overall irregularities of the sound correspondences, derivation from PIE **h₂elbʰ-* 'white' (Suolahti 1909: 407) has to be rejected and borrowing from a non-Indo-European source seems plausible. Since the Germanic element **ut-* can be compared to that of **arwīt-*, and Sl. **-ǫd-* seems to correspond to it, it is possible to hypothesize an νθ-suffix for this word. However, it is difficult to see how the Slavic variant **lebedь*, cf. Ru. *lébed'*, Bulg. *lébed*, fits into the picture and without a direct Greek cognate, i.e. **ὄλοφυ(ν)ς, this remains difficult to prove.

4 Results

Having evaluated the evidence relevant to the nature and distribution of the -ινθ- and -υνθ- suffixes, the following can be observed.

Thematic terms in -ινθ- are by far the most numerous and most reliable examples of this suffix. The majority of these cases are attested with -ινθ- only, but some also have variants with -ιθ-. The origin of this variation is not fully clear, but it seems reasonable to assume that the nasal was lost sporadically in the dialects, thus foreshadowing the regular loss in Modern Greek in this position (cf. ρεβίθι 'chickpea' < ἐρεβίνθιον). Next to these, there are cases attested only with -ιθ-. These are ambiguous, however, since it is not possible to verify whether this element truly is a variant of -ινθ-, or in fact an unrelated suffix. For this reason, they are counted as uncertain.

Compared to the thematic evidence, athematic terms in -ι(ν)θ- are fewer and at the same time more diverse in nature. A key outcome is that the cases with -ιθ- may be secondary, i.e. they arose, within Greek, by the secondary introduction of a θ, by Schwyzer called a "bloßes Suffix", in inherited *iH*-stems. This analogical extension may have been triggered by the Ionic-Attic merger of the -ινθ- and -ῑ-

9 Formally close, but nevertheless unrelated to *hirundo* is G Swab. *Geiritz* 'lapwing', Swi. *giriz* 'lapwing, tern' < **gīrītja-*: this word is derived from the verb *gieren* 'shreek' with the suffix of G *Kiebitz* and *Stieglitz* (Suolahti 1909: 402).

-ινθ-	-ιθ-	base
αἴγινθος	αἴγιθος	**(w)aig-*
–	?ἄμιθα	–
ἀσάμινθος	–	**as$_2$am-*
ἄψινθος	–	**(w)aps-*
βόλινθος	–	**bol-*
γάλινθοι, γέλινθοι, γέρινθοι	γάλιθοι	**gal-* ~ **gel-* ~ **ger-*
ἐρέβινθος, *ἔρβινθος	–	**er(e)b-*
ἐσχάρινθον	–	–
κήρινθος	–	**kēr-*
κυλίνθιον	?κύριθρον	**kul-* ~ **kur-*
λαβύρινθος, *δαφύρινθος	–	**labur-* ~ **dab^{h}ur-*
λαΐνθη	–	**(w/s$_1$)la(w/s$_1$)-*
λεβίνθιοι	–	?**leb-*
–	?λέκιθος	?**lek-*
μήρινθος, σμήρινθος	–	**s$_1$mēr-* ~ **s$_2$mēr-*
πύρινθον	–	**pur-*
*ῥάφινθος	–	**rag^{wh}-*
σαλαμίνθη	–	**s$_2$alam-*
τέρμινθος	τρέμιθος	**t(e)r(e)m-*
ὑάκινθος	–	**(s$_1$)uwak-*

stems in the nominative case through the regular development of *-ινθ-ς > *-ινς > -ῑς.[10] As a result of different levelings, an analogical oblique stem -ῑθ- could be created next to an analogical nominative -ινς (cf. Fraenkel 1906: 290). The analogy occurs early, as demonstrated by e.g. ὄρνῑς, -ῑθος < **h$_3$r-n-iH-s*.

-ινθ-	-ιθ-	base
–	?ἄγλις	?**aggl-*
–	?βάλλις	–
–	?γέλγις	?**gegl-*
–	?δέλλις	–
ἔλμις, *λέμις, λίμις	–	**s$_1$elm-* ~ **s$_1$le/im-*
–	?μέρμις	–
μήρις	–	**s$_1$mēr-*
–	†ὄρνις	–
πείρις	–	**peir-*

10 Cf. OAtt. τρῖς ‘three’ < **trins*.

The partial merger of the types in -ινθ- and -ῑ- may conversely have motivated the removal of -ινθ- in some athematic nouns. At any rate, this suffix is not retained in secondary formations created with the suffix -ιγγ-. Thus, ἔλμιγξ ~ *λίμιγξ and μῆριγξ ~ σμῆριγξ, which look like diminutives comparable to λᾶϊγξ ('pebble'), were formed to ἔλμις ~ λίμις and μήρις. The first example is remarkable, because both the variants ἔλμις and λίμις have parallel diminutives. The second example is also remarkable, but rather because it is one out of just two stems in -ινθ- that occurs both as a thematic and athematic noun: μήρις : μήρινθος ~ σμήρινθος, the other one being the scarcely attested σμίς (Hsch.) for more current σμίνθος 'mouse'. The variation in anlaut of μῆριγξ ~ σμῆριγξ ostensibly points to parallel derivation from the thematic doublet, but may alternatively imply a lost athematic variant *σμήρις.

A striking imbalance in the material is that the cases in -υνθ- are much sparser than those in -ινθ-. It is not possible to retrieve the mechanism behind this distribution. However, it is conspicuous that the roots with this suffix all have *o*-vocalism and contain a (usually root-final) liquid. There is only one real counterexample to this distribution, and that is βόλινθος. Perhaps, then, the old explanation that it is dissimilated from **βόνινθος is correct after all.

-υνθ-	**-υθ-**	**base**
?βόλυνθον	–	*$bol(g^w)$-
γοργυνθία	–	*gorg-
κολοκύνθη	–	*kolok-
κόρυνθος	–	*kor-
–	?λήκυθος	–
ὀδόλυνθοι	–	*(w/s_1)odol-
ὄλυνθος (ὄλονθος)	–	*wol-

Athematic representations of the -υνθ- suffix are rarer still, even if the toponym Τίρυνς is included. Two out of three examples are attested with non-nasal -υθ- only, which makes them uncertain. As with the cases of the athematic stems in -ιθ-, there is a possibility that -θ- was added secondarily, within Greek, to original *uH*-stems. The one example that does have a variant -υνθ- conspicuously has an *o* in the root again.

-υνθ-	**-υθ-**	**base**
–	?ἀγνύς	–
(κορυνθεύς)	κόρυς	*kor-
–	?κώμυς	–

The mechanism behind the distribution is obscure. Theoretically, it is possible to speculate that the vowel of -υνθ- had a rounding effect on any preceding *a*'s and *e*'s in the donor language.[11] In view of the variation between ὄροβος and ἐρέβινθος, however, it is equally plausible that the suffix -ινθ- had a fronting effect, meaning that **ereb*- developed from **öröb*- by subsequent unrounding. A third possible explanation is that the variation between ἐρεβ- and ὀροβ- is primary, and that the distribution of -ινθ- and -υνθ- resulted from some form of vowel harmony. It is not certain that appellatives are representative of the original distribution, however, since the reality of this distribution is itself challenged by toponymical evidence, i.e. by Κόρινθος.

Of all the material analyzed, the examples that oscillate between bases with and without -ινθ- clearly are the most relevant to the problem investigated here, as they unambiguously demonstrate the suffixal nature of the element. Having excluded the suffixal variation of the type ἕλμι(ν)ς : ἕλμιγξ, which seems secondary, the evidence for such alternations remains limited but not insignificant. Within Greek itself, the most straightforward case is κηρός : κήρινθος. If, in addition, OGe. *rapindi* was borrowed from *ῥάφινθος, it forms a doublet with δάφνη ~ δαύχνα ~ λάφνη and Lat. *laurus*. This example is less transparent, however, due to the formal variation of the root and the presence of other suffixes in the variants without -ινθ-. Finally, there is the well-known case of ὄροβος ~ ἐρέβινθος, which also exhibits variants with and without the suffix in other Indo-European branches. In Greek terms, Italic points to **erw*- (≈ Gr. **ἔρεβος), Germanic to **orw-īd*- (≈ Gr. **ὀρόβινθος) and Armenian to **HrVbʰoud*- (≈ Gr. **ὀρόβυνθος). Moreover, there is the possibility of an (Indo-)Iranian correspondence with the suffix, viz. PIIr. **H(a)rabʰ(u)anT*-.

More generally, an important remaining question is what the evidence is for manifestations of the suffixes -ινθ- and -υνθ- outside Greek, in the other Indo-European languages. Major obstacles need to be overcome, however, in order to prove that such manifestations exist. In the discussions of the material, possible examples in Latin are *bolunda* 'fig', *harundō* 'reed' and *hirundō* 'swallow'. The first example has been interpreted as a loan from Doric *ϝόλυνθος, however, and the other two do not have correspondences in Greek. Likewise, Slavic **olbǫdь* and Germanic **albVt*- 'swan' appear to have similar, yet irregular suffixes that resemble -υνθ-, but in lack of a Greek correspondence, e.g. **ὅλοφυ(ν)ς, the nature of the suffixation cannot be substantiated. Gr. κηρός ~ κήρινθος does have a

11 Šorgo (2020: 458–459) proposes a development *ἔρεβος > *ἔροβος > ὄροβος, assuming rounding of ε to ο by an adjacent labial and subsequent vowel assimilation of *-ερο- to -ορο-. It is unclear, however, why ἐρέβινθος was not affected by the same series of developments.

compelling (yet irregular) comparandum in Lith. *korỹs*, Latv. *kâre*, but here the root appears without a suffix. In the end, the only case in which an etymological link reliably can be established between a Greek word in -νθ- and an equivalent in another Indo-European branch is that of ἐρέβινθος ~ PGm. **arwīt-* ~ Arm. *aṙowoyt* ~ PIIr. **H(a)rab$^{(h)}$anT-*. In conclusion, it is clear that Greek was most heavily impacted by the language of the νθ-suffix, but while the evidence for the other (European) branches being impacted is marginal, it is not zero.

5 Discussion

The traces of the "Pre-Greek" νθ-suffix in several other Indo-European languages than Greek raise the question where these languages could have absorbed this suffix.

One scenario that comes to mind is that the various manifestations of ὄροβος ~ ἐρέβινθος spread from the East Mediterranean to the rest of Europe as a *Wanderwort* after the arrival of the Indo-European subgroups. Yet there is an important objection to such a scenario. *Wanderwörter* are adopted over large distances across multiple established language communities together with the introduction of a new phenomenon. Legumes, however, had been introduced to Europe long before the Indo-European dispersal as part of the package of founder crops introduced by Anatolian farmers. In other words, without a cultural innovation, there was no clear motivation for the word to become a *Wanderwort*. More fatally, there is no direct evidence by which it can be substantiated that the word frog-leaped from branch to branch, i.e. none of the various manifestations can be characterized as direct borrowings from any of the known Indo-European branches. If the word indeed wandered, it must have done so prior to the Indo-European dispersal.

From this perspective, it is worth recalling the similarly large distribution of the Pre-Greek word for 'sand'. Although we have rejected this lexeme as a case of the νθ-suffix, it does exhibit an alternation, both within Greek and beyond, that is parallel to that of the νθ-suffix:

	Root	**Root + *-d^{h}-**
Germanic	–	**samm(a)d^{h}-*
Italic	**sab^{h}-*	–
Greek	**(p)samm-*	**(p)samad^{h}-*
Armenian	–	**sab^{h}ad^{h}-*

Since 'sand' is hardly a concept that spreads as a *Kulturwort*, it seems reasonable to assume, on semantic grounds, that it was adopted by the various Indo-European branches through direct language contact. This implies a scenario in which they borrowed the word after their dispersals into Europe, from the local languages with which they came into contact.

Exactly where this contact could have occurred cannot easily be established. It could have been anywhere between the steppe homeland and the various historically known locations in which the Indo-European languages of Europe were spoken. In a maximalist scenario, related non-Indo-European languages were spoken by farming groups across much of the continent. Most of the future homelands of Indo-European branches of Europe lie within the areas settled by the European farmers of the Linear Pottery culture (5500–4500 BCE) and the Funnel Beaker culture (4300–2800 BCE). Their farmers formed a genetically homogeneous population (Haak et al. 2010) and plausibly also a language continuum, at least initially (cf. Shennan 2018: 105). In the most minimalist scenario, various Indo-European groups adopted vocabulary exclusively from those farmers living directly west of the steppe homeland, i.e. in the West Pontic, Carpathian and Balkans regions. Prior to the Indo-European dispersal, this area had been settled by farmers from Greece through the closely related Starčevo–Körös–Criș culture (6200–4500 BCE) and Cucuteni–Trypillia culture (4800–3000 BCE), which in turn are rooted in the Sesklo culture (6850–4400 BCE) of Thessaly and Macedonia.

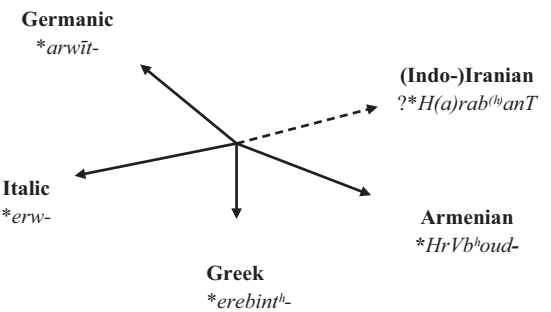


Figure 2: Correlation between the various manifestations of the non-Indo-European terms surfacing as ἐρέβινθος in Greek.

While the question cannot be answered with the present evidence, potential clues are offered by the Armenian and perhaps Iranian manifestations of the term that in Greek surfaces as ἐρέβινθος, and in Germanic as **arwīt-* (cf. Figure 2). Regarding Armenian **HrVbʰoud-*, the most economical scenario would seem that it was adopted in Southeast Europe, since the alternative of *in situ* adoption within Anato-

lia would remove the borrowing location further from that of the other branches. A similar line of reasoning can be applied to (Indo-)Iranian **H(a)rab$^{(n)}$anT-*. If this etymon was indeed borrowed from a non-Indo-European source related or connected to Pre-Greek, and not from Bactrian Greek, the contact likely took place in East Europe. Although this raises new, daunting questions regarding the time of borrowing and the archaeological context, the alternative of early (Indo-)Iranian borrowing beyond the steppe, in Asia, would be even more difficult to defend. Regardless, the possibility that Armenian possesses a manifestation of the same word that in Greek emerges as ἐρέβινθος suggests that Pre-Greek influences extended beyond Greece, to the north. As a consequence, it is difficult to avoid the conclusion that lexical features reminiscent of Pre-Greek impacted, at the very least, the wider Southeastern or East-Central European region to the west of the Pontic-Caspian steppe.

References

Ačaṙyan, Hračᶜeay. 1971–1979. *Hayeren armatakan baṙaran*. Yerevan: Erevani Hamalsarani Hratarakčᶜowtᶜyown.

Alessio, Giovanni. 1941. Fitonimi mediterranei. *Studi Etruschi* 15. 178–224.

Alessio, Giovanni. 1943. Gr. ἀσάμινθος 'vasca da bagno.' *Studi italiani di filologia classica* 20(2–4). 121–133.

Alessio, Giovanni. 1944. Suggerimenti e nuove indagini sul problema del sostrato "mediterraneo." *Studi Etruschi* 18. 93–157.

Baranov, Pavel Aleksandrovich & Ilariya Alekseevna Raykova. 1928. *Darvaz i yego kul'turnaya rastitel'nost'*. Tashkent.

Beekes, Robert S. P. 2008. Pre-Greek. *Glotta* 83. 13–17.

Beekes, Robert S. P. 2010. *Etymological dictionary of Greek*. Leiden; Boston: Brill.

Beekes, Robert S. P. 2014. *Pre-Greek: phonology, morphology, lexicon*. Leiden; Boston: Brill.

Beekes, Robert S. P. & Aert Hendrik Kuipers. 1975. Review of E. J. Furnée. 1972. Die wichtigsten konsonantischen Erscheinungen des Vorgriechischen. *Lingua* 36. 69–100.

Bengtson, John D. 2017. The anthropological context of Euskaro-Caucasian. *Iran and the Caucasus* 21. 75–91.

Bilgiç, Emin. 1945. Die Ortsnamen der "kappadokischen" Urkunden im Rahmen der alten Sprachen Anatoliens. *Archiv für Orientforschung* 15. 1–37.

Biville, Frédérique. 1990. *Les constructions absolues en français moderne*. Louvain-Paris: Edition Peeters.

Blumenthal, Albrecht von. 1930. Ägäisches. *Indogermanische Forschungen* 48(1). 50.

Bonfante, Giuliano. 1953. Review of Die Träger der kretisch-mykenischen Kultur, ihre Herkunft und ihre Sprache, I. Teil: Urgriechen und Urillyrier; Die Träger, etc. II. Teil: Italiker und Urillyrier; Die Sprache der Etrusker; Das Schicksal der indogermanischen o-Deklination im Etruskischen; Die sprachliche Zugehörigkeit der Etrusker; Vorgriechische Sprachwissenschaft; Vorgriechische Sprachwissenschaft; Contribution à l'étude de la toponymie grecque: noms de lieux prétendus préhelléniques, by V. Georgiev. *Classical Philology* 48(3). 205–207.

Brandenstein, Wilhelm. 1936. Die Sprachschichten im Bereich der Ägäis. In Helmut Arntz (ed.), *Germanen und Indogermanen: Volkstum, Sprache, Heimat, Kultur; Festschrift für Herman Hirt*, vol. 2, 25–44. Heidelberg: Carl Winters Universitätsbuchhandlung.

Çabej, Eqrem. 1976. *Studime gjuhësore*. Prishtinë: Rilindja.

Carnoy, Albert. 1960. Les suffixes toponymiques pré-grecs. *L'antiquité classique* 29(2). 319–336.

Chantraine, Pierre. 1999. *Dictionnaire étymologique de la langue grecque*. Paris: Klincksieck.

Chantraine, Pierre, Jean Taillardat, Alain Blanc, Olivier Masson, Jean Louis Perpillou & Charles de Lamberterie. 2009. *Dictionnaire étymologique de la langue grecque: histoire des mots*. Nouv. ed. Paris: Klincksieck.

Cuny, Albert. 1910. Les mots du fonds préhellénique en grec, latin et sémitique occidental. *Revue des Études Anciennes Année* 12(2). 154–164.

Demiraj, Bardhyl. 1997. *Albanische Etymologien* (Leiden Studies in Indo-European 7). Amsterdam; Atlanta, GA: Rodopi.

Derksen, Rick. 2008. *Etymological dictionary of the Slavic inherited lexicon*. Leiden/Boston: Brill.

Deroy, Louis. 1957. La valeur du suffixe préhellénique *-nth-* d'après quelques noms grecs en *-νθος*. *Glotta* 35(3/4). 171–195.

Döhring, Alfred. 1888. *Die Etymologie der sogenannten Gerundivformen*. Königsberg: Hartungsche Buchdruckerei.

Edel'man, Dzhoy Iosifovna. 1971. *Yazgulyamsko-russkiy slovar'*. Moskva: Nauka.

Fick, August. 1905. *Vorgriechische Ortsnamen als Quelle für die Vorgeschichte Griechenlands*. Göttingen: Vandenhoeck und Ruprecht.

Fraenkel, Ernst. 1906. *Griechische Denominativa in ihrer geschichtlichen Entwicklung und Verbreitung*. Göttingen: Vandenhoeck & Ruprecht.

Frisk, Hjalmar. 1972. *Griechisches etymologisches Wörterbuch*. Heidelberg: Winter.

Furnée, Edzard J. 1972. *Die wichtigsten konsonantischen Erscheinungen des Vorgriechischen*. De Gruyter.

Furnée, Edzard J. 1979. *Vorgriechisches-Kartvelisches: Studien zum ostmediterranen Substrat nebst einem Versuch zu einer neuen pelasgischen Theorie*. Leuven-Louvain: Edition Peeters.

Garnier, Romain. 2006. Nouvelles réflexions autour de gr. ψάμαθος. *Die Sprache* 46(1). 81–93.

Garnier, Romain & Benoît Sagot. 2020. New results on a centum substratum in Greek: the Lydian connection. In Romain Garnier (ed.), *Loanwords and substrata. Proceedings of the colloquium held in Limoges (June 5th-7th, 2018)*, 169–200. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Georgiev, Vladimir I. 1941. *Vorgriechische Sprachwissenschaft*. Sofia: Universitätsdruckerei.

Gerzon, Jacob. 1902. *Die jüdisch-deutsche Sprache: eine grammatisch-lexikalische Untersuchung ihres deutschen Grundbestandes*. Frankfurt am Main: Verlag von J. Kauffmann.

Grošelj, Milan. 1952. Etyma Graeca. *Živa antika* 2. 205–220.

Guidi, Michele. 1990. Greco λαβύρινθος: note di linguistica mediterranea. *Minos* 25. 175–193.

Güntert, Hermann. 1932. *Labyrinth: eine sprachwissenschaftliche Untersuchung*. Heidelberg: C. Winter.

Haak, Wolfgang, Oleg Balanovsky, Juan J. Sanchez, Sergey Koshel, Valery Zaporozhchenko, Christina J. Adler, Clio S. I. Der Sarkissian, et al. 2010. Ancient DNA from European early Neolithic farmers reveals their Near Eastern affinities. (Ed.) David Penny. *PLoS Biology* 8(11). e1000536.

Haley, J. B. 1928. The coming of the Greeks: I. The geographical distribution of Pre-Greek place-names. *American Journal of Archaeology* 32(2). 141–145.

Hayden, Deborah & David Stifter. 2022. The lexicography and etymology of OIr. *eclas*. *North American Journal of Celtic Studies* 6(2). 236–250.

Hehn, Victor. 1885. *The wanderings of plants and animals from their first home*. London: Swan Sonnenschein & Co.

Hester, David A. 1957. Pre-Greek placenames in Greece and Asia Minor. *Revue Hittite et Asianique* 15. 107–119.
Hester, David A. 1964. “Pelasgian”—A new Indo-European language? *Lingua* 13. 335–384.
Hester, David A. 1968. Recent developments in mediterranean «substrate» studies. *Minos* 9(1). 219–235.
Heubeck, Alfred. 1961. *Praegraeca: sprachliche Untersuchungen zum vorgriechisch-indogermanischen Substrat*. Erlangen: Verlag Universitätsbund Erlangen.
Hoz, Javier de. 2004. De cómo los protogriegos crearon el griego y los pregriegos lo aprendieron. In Pedro Bádenas de la Peña (ed.), *Lenguas en contacto: el testimonio escrito*, 35–56. Madrid: Consejo superior de investigaciones científicas.
Judson, Anna P. 2017. The mystery of the Mycenaean “labyrinth”: The value of Linear B pu_2 and related signs. *Studi micenei ed egeo-anatolici NS* 3. 53–72.
Katičić, Radoslav. 1976. *Ancient languages of the Balkans*. Berlin: Mouton de Gruyter.
Kretschmer, Paul. 1896. *Einleitung in die Geschichte der griechischen Sprache*. Göttingen: Vandenhoeck & Ruprecht.
Kretschmer, Paul. 1925. Das *nt*-Suffix. *Glotta* 14(1/2). 84–106.
Kretschmer, Paul. 1934. Lautverschiebung im Griechischen. *Glotta* 23(1/2). 1–17.
Kretschmer, Paul. 1943. Die vorgriechischen Sprach- und Volksschichten. *Glotta* 30. 88–218.
Kroonen, Guus. 2012a. On the etymology of Greek ἄγλις and γέλγις ‘garlic.’ *Journal of Indo- European Studies* 40. 289–299.
Kroonen, Guus. 2012b. Non-Indo-European root nouns in Germanic. In Riho Grünthal & Petri Kallio (eds.), *A linguistic map of prehistoric Northern Europe*, 239–260. Helsinki: Suomalais-Ugrilainen Seura.
Kroonen, Guus. 2016. Hittite *kapart-*/*kapirt-* ‘small rodent’ and Proto-Semitic *$^{*\text{ʕ}}$*kbr-t-* ‘mouse, jerboa.’ *Indogermanische Forschungen* 121(1). 53–62.
Kuiper, Franciscus B. J. 1956. The etymology of ἄνθρωπος. In Heinz Kronasser (ed.), *ΜΝΗΜΗΣ ΧΑΡΙΝ. Gedenkschrift Paul Kretschmer*, vol. 1, 211–226. Wien: Wiener Sprachgesellschaft.
Kuiper, Franciscus B. J. 1995. Gothic *bagms* and Old Icelandic *ylgr*. *North-Western European Language Evolution (NOWELE)* 25. 63–88.
Lafon, René. 1934. Mots « méditerranéens » en géorgien et dans quelques autres langues caucasiques. *Revue des Études Anciennes* 36(1). 32–46.
Lanszweert, René. 1994. Πάπυρος: ein mykenisches Schimpfwort? In Roland Bielmeier & Reinhard Stempel (eds.), *Indogermanica et Caucasica*. Berlin/Boston: De Gruyter.
Laroche, E. 1961. Études de toponymie anatolienne. *Revue hittite et asianique* 19. 57–98.
Lindner, Thomas. 1995. Griechische (incl. mykenische) Ortsnamen. In Ernst Eichler, Gerold Hilty, Heinrich Löffler, Hugo Steger & Ladislav Zgusta (eds.), *Namenforschung: Ein internationales Handbuch zur Onomastik*, vol. 1, 690–705. Berlin/New York: Walter de Gruyter.
Martirosyan, H. 2010. *Etymological dictionary of the Armenian inherited lexicon*. Leiden; Boston: Brill.
Meillet, Antoine. 1908. De quelques emprunts probables en grec et en latin. *Mémoires de la Société de Linguistique de Paris* 15. 161–164.
Meissner, Torsten. 2013. Forschungsbericht Griechische Etymologie: Rezension von Beekes 2009. *Kratylos* 58. 1–32.
Meringer, Rudolf. 1907. Zu ἅμαξα und zur geschichte des wagens: ein beitrag zur methode der etymologie. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der indogermanischen Sprachen* 40. 217–234.
Merlingen, Weriand. 1955. *Das “Vorgriechische” und die sprachwissenschaftlich-vorhistorischen Grundlagen*. Wien: Gerold & Co.

Merlingen, Weriand. 1962a. Zum "Vorgriechischen." *Balkansko ezikoznanie: Linguistique Balkanique* 4. 25–55.
Merlingen, Weriand. 1962b. Zum "Vorgriechischen" (Fortsetzung). *Balkansko ezikoznanie: Linguistique Balkanique* 5(2). 5–44.
Meyer, Georg. 1886. Die Karier: eine ethnographisch-linguistische untersuchung. *Beiträge zur Kunde der indogermanischen Sprachen* 10. 147–202.
Meyer, Gustav. 1896. *Griechische Grammatik*. 3. verm. aufl. Leipzig: Breitkopf & Härtel.
Monzó Gallo, Carlos. 2015. *El tema en –ι en griego y la clasificación nominal indoeuropea: enfoque tipológico y semántico*. Valencia: Universitat de València.
Moreschini, Adriana Quattordio. 1984. *Le formazioni nominali greche in* -nth-. Roma: Edizioni dell'Ateneo.
Morgenstierne, Georg. 1974. *Etymological vocabulary of the Shughni group*. Wiesbaden: Reichert.
Morpurgo Davies, Anna. 1986. The linguistic evidence: Is there any? In Gerald Cadogan (ed.), *The end of the early Bronze Age in the Aegean*, 92–187. Leiden: E. J. Brill.
Orel, Vladimir E. 2003. *A handbook of Germanic etymology*. Leiden; Boston: Brill.
Oštir, Karl. 1930. *Drei vorslavisch-etruskische Vogelnamen*. Ljubljana: Znanstveno društvo.
Palmer, Leonard R. 1958. Luvian and Linear A. *Transactions of the Philological Society* 57(1). 75–100.
Pauli, Carl Eugen. 1885. *Altitalische Forschungen*. Vol. II. Leipzig: J.A. Barth.
Piquero Rodríguez, Juan. 2017. *El léxico del griego micénico*. Madrid: Universidad Complutense de Madrid.
Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. 2 vols. Bern, München: A. Francke.
Pott, August Friedrich. 1853. *Die Personennamen insbesondere die Familiennamen und ihre Entstehungsarten: auch unter Berücksichtigung der Ortsnamen*. Leipzig: F.A. Brockhaus.
Reece, Steve. 2002. The Homeric ἀσάμινθος: Stirring the waters of the Mycenaean bath. *Mnemosyne* 55(6). 703–708.
Renfrew, Colin. 1998. Word of Minos: The Minoan contribution to Mycenaean Greek and the linguistic geography of the Bronze Age Aegean. *Cambridge Archaeological Journal* 8(2). 239–264.
Rieken, Elisabeth. 1999. *Untersuchungen zur nominalen Stammbildung des Hethitischen*. Wiesbaden: Harrassowitz.
Rönsch, Hermann. 1886. Das Subst. bolunda. *Romanische Forschungen* 2(2). 317–318.
Schachermeyr, Fritz. 1955. *Die ältesten Kulturen Griechenlands*. W. Kohlhammer Verlag. Stuttgart.
Schmeller, Johann Andreas & Georg Karl Fromman. 1872. *Bayrisches Wörterbuch*. München: Oldenbourg.
Schrijver, Peter. 1991. *The reflexes of the Proto-Indo-European laryngeals in Latin*. Amsterdam; Atlanta, GA: Rodopi.
Schrijver, Peter. 2018. Talking Neolithic: The case for Hatto-Minoan and its relationship to Sumerian. In Guus Kroonen, Bernard Comrie & James P. Mallory (eds.), *Talking Neolithic: Proceedings of the workshop on Indo-European origins held at the Max Planck Institute for Evolutionary Anthropology, Leipzig, December 2–3, 2013*, 336–374. Washington, D.C.: Institute for the Study of Man.
Schwyzer, Eduard. 1953. *Griechische Grammatik*. Vol. 1. Bd. Allgemeiner Teil. Lautlehre. Wortbildung. Flexion. München: Oscar Beck.
Shennan, Stephen. 2018. *The first farmers of Europe: An evolutionary perspective*. Cambridge University Press.
Šorgo, Aljoša. 2020. Characteristics of lexemes of a substratum origin in Proto-Germanic. In Romain Garnier (ed.), *Loanwords and substrata: proceedings of the colloquium held in Limoges, 5th-7th June 2018*, 427–472. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Steblin-Kamenskiy, Ivan Mikhaylovich. 1982. *Ocherki po istorii leksiki pamirskikh yazykov: Nazvaniya kul'turnykh rasteniy*. Moskva: Nauka.
Strömberg, Reinhold. 1944. *Griechische wortstudien: Untersuchungen zur Benennung von Tieren, Pflanzen, Körperteilen und Krankheiten*. Göteborg: Elanders boktryckeri aktiebolag.
Stumfohl, Helmut. 1990. Die Pelasger. *Almogaren* 21(1). 11–41.
Suolahti, Hugo. 1909. *Die deutschen Vogelnamen: eine wortgeschichtliche Untersuchung*. Straßburg: K.J. Trübner.
Szemerényi, Oswald. 1971. Review of Dictionnaire étymologique de la langue grecque: histoire des mots by Pierre Chantraine. *Gnomon* 43(7). 641–675.
Szemerényi, Oswald. 1974. The origins of the Greek lexicon: *Ex Oriente Lux*. *The Journal of Hellenic Studies* 94. 144–157.
Thumb, Albert. 1901. *Die griechische Sprache im Zeitalter des Hellenismus*. Straßburg: Karl J. Trübner.
Thurneysen, Rudolf. 1883. Urspr. dn, tn, cn im Lateinischen. *Zeitschrift für vergleichende Sprachforschung* 26. 301-314.
Torp, Alf. 1909. *Wortschatz der germanischen spracheinheit*. Göttingen: Vandenhoeck & Ruprecht.
Vaan, Michiel A. C. de. 2008. *Etymological dictionary of Latin and the other Italic languages*. Leiden/Boston: Brill.
Vander Kloet, Sam P. 1992. On the etymology of *Vaccinium L*. *Rhodora* 94(880). 371–373.
Verhasselt, Gertjan. 2009. The Pre-Greek linguistic substratum: An overview of current research. *Les Études Classiques* 77. 211–239.
Walde, Alois & Johann Baptist Hofmann. 1938–1954. *Lateinisches etymologisches Wörterbuch*. 3rd edn. 2 vols. Heidelberg: Carl Winter.
Whatmough, Margaret M. T. 1997. *Studies in the Etruscan loanwords in Latin*. Firenze: L.S. Olschki.
Windekens, Albert Joris van. 1960. *Études pélasgiques*. Louvain: Publications universitaires.
Wörner, Emil. 1876. ἡ πείρινς, πείρινθος und Πέρινθος. *Studien zur griechischen und lateinischen Grammatik* 9. 458–462.
Woudhuizen, Fred C. 2016. *Documents in Minoan Luwian, Semitic, and Pelasgian*. Amsterdam: Dutch Archaeological and Historical Society.
Yakubovich, Ilya. 2010. *Sociolinguistics of the Luvian Language*. Leiden/Boston: Brill.

Part V: **Anatolia & the Caucasus**

Rasmus Thorsø

11 Alternation of diphthong and monophthong in Armenian words of substrate origin

1 Introduction

The fact that most of the Classical Armenian lexicon is not inherited from Proto-Indo-European has been recognized ever since Hübschmann (1877) established that Armenian constitutes an independent, monophyletic branch within the Indo-European family, rather than being part of the Iranian branch. Aside from the late, more or less easily identifiable, loanwords from known sources (Middle Iranian, Syriac, Greek, Urartian etc.), a sizeable portion of the lexicon remains without comparanda. A large part of this lexicon may reflect loanwords from unattested languages of the Mediterranean and Asia Minor (Clackson 2017). However, the identification of this type of loanwords is still fraught with difficulty, as in most cases, it is hard to exclude that they actually represent unrecognized inherited words or borrowings from attested, but fragmentary, languages like Urartian or Luwian. Most previous research concerning the early influence of European substrate languages on the Armenian lexicon has principally focused on the role of the so-called Mediterranean substrate, which implies areal loanwords shared with Greek and Latin (cf. J̌ahowkyan 1987: 306–11). Martirosyan (2010: 805–7 *et passim*; 2013: 121–123) highlights the role of European substrates in a wider sense (cf. Beekes 1996), as demonstrated by substrate words shared with Germanic, Balto-Slavic, and Celtic. Yet, this line of research is still in an early stage.

As discussed elsewhere in the present volume and in the publications of, e.g., Kuiper (1956, 1995), Polomé (1986), Salmons (1992), and Schrijver (1997),[1] the most reliable method for identifying loanwords from unknown languages ("substrate words") is the identification of what may be termed "irregular comparanda" in other languages. The validity of such comparisons is significantly strengthened

1 This relatively young tradition of research into the substrate languages of the Indo-European branches, to which the present volume belongs, builds, in fact, upon methods pioneered by such scholars as Karel Oštir, Johannes Hubschmid and Edzard Furnée, but it is characterized by a more critical approach and a more rigorous methodology.

Acknowledgements: I thank Anthony Jakob and Guus Kroonen for valuable suggestions on earlier versions of this study.

https://doi.org/10.1515/9783111337920-011

when recurring alternations between phonemes can be identified. Theoretically, the genetic affiliation of the compared languages is irrelevant when following this method. If two entirely unrelated languages have borrowed words from the same source at roughly the same time, they should be identifiable based on recurring alternations as well. This is an avenue of research where future efforts may also bear fruit. When so far, research has focused on the substrate lexicon shared by the Indo-European languages, it is no doubt because the reconstruction of these is so relatively far advanced that we can identify alternations on a deeper chronological level. At the same time, we can use the comparative method for excluding the possibility that the comparanda in question are regular cognates. Strictly speaking then, it is only the irregular, but recurring, sound alternations (or "regular irregularities") which constitute positive evidence that the comparanda are in fact related and not simply lookalikes (cf. the Introduction to this volume).

This study presents three new proposals, and an additional uncertain case, for substrate words shared by Armenian and several other Indo-European languages of Europe, including the languages of the Mediterranean region (Greek and Latin) as well as the Indo-European languages of northern Europe (Germanic, Balto-Slavic, and Celtic). On the basis of their geographic distribution, they can thus tentatively be assigned to a "European" substratum in the sense of Beekes (1996, 2000). The most striking feature of these etyma is that they demonstrate a recurring vocalic alternation. That is, Armenian shows the reflex of a diphthong **ou*,[2] while other comparanda reflect a monophthong in the same position. The implications of this will be discussed further in § 3.

2 Material

2.1 *aṙowoyt, aṙawoyt* 'alfalfa, *Medicago sativa*'

This word is first attested in the Galen Dictionary, where it glosses Gr. μηδική (Greppin 1985: 76). Additionally, it appears in an Arabic-Armenian botanical dictio-

2 In the following, the reconstruction **ou* will refer to the source of Arm. *oy* (unstressed *ow*) disregarding the fact that PIE **eu* or **ou* are usually both considered sources of this diphthong (Meillet 1936: 44, Schmitt 1981: 52). Lamberterie (1982) and Olsen (2020) argue that the regular reflex of PIE **eu* is Arm. *iw*. This assumption solves several etymological problems (e.g. *hiwsem* 'weave', which can be equated with a root **seuk-* 'turn', Lith. *sùkti*), but is not accepted by all scholars. Whichever view one subscribes to, the **ou* implied in this study can be equated with the "**ou̯*$_2$" of Macak (2017: 1069), covering a diphthong that results in Arm. *oy*.

nary (ca. 9th century), where it glosses Arab. *ar-raṭbah* ‘red clover, alfalfa’ (Greppin 1996: 393). For later and more marginal spelling variants, see HAB (I: 265).

No widely accepted etymology exists, and the word is not discussed in most of the recent etymological handbooks (e.g. Solta 1960, Greppin 1983, J̌ahowkyan 1987, Martirosyan 2010).[3] However, a key to the etymology may have been found already by Dervischjan (1877: 29), who compares Gr. ἐρέβινθος ‘chickpea’ and OHG *araweiz* (a variant of *arawīz*) ‘pea’. These forms cannot be regular cognates, however. Greek β (< **b*) does not correspond with OHG *w* (< **u̯*), nor Arm. *w* (< **bʰ* or **p*). Neither do the vowels match, and the suffix Gr. -ινθ- has a nasal which is not found elsewhere. Consequently, these words are better analyzed as a complex of independent borrowings from a non-Indo-European language (cf. WH I: 419–20, Kuiper 1956: 217–9, Furnée 1972: 231, 273, Kroonen 2012: 242–4, Šorgo 2020: 434). The following quasi-Proto-Indo-European reconstructions can be posed:

– **ereb-indʰ-* ~ **orob-*: Gr. ἐρέβινθος ‘chickpea’, ὄροβος ‘bitter vetch, *Vicia ervilia*’
– **a/oru̯-īd-*: PGm. **arwīt-* ‘pea’ (OHG *arawīz, araweiz*, OS *erit*, ON *ertr*)
– **eru̯-o-*: Lat. *ervum* ‘pea’

The question is whether the Armenian material can be added to this complex as well. If we assume that the variants *aṙowoyt* and *aṙawoyt*, attested in the earliest sources, are primary, they presuppose quasi-PIE **HrVbʰ-oud-* or **HrVp-oud-*. Now it is possible that the borrowing took place when intervocalic *tenues* and *mediae aspiratae* were at the stage of fricatives in Pre-Armenian. Especially in view of the alternation **b* ~ **u̯* observed in the comparanda,[4] it is possible to surmise that the donor form(s) contained a bilabial approximant **β* or the like. This sound could thus have been substituted for the corresponding fricative in Pre-Armenian, i.e. a

3 Ačaṙyan (HAB I: 265) rejects the earlier etymology of Dervischjan (1877: 29) but cites Geo. dial. (Kakheti, Kartli) *alaverdi* ‘alfalfa’ as a loan from Armenian. The sound substitutions implied for this putative loan would be unexpected, however. Greppin (1992: 72–3) compares the Semitic root *rṭb* ‘fresh, green, juicy, tender’, assuming that the root entered Armenian from Semitic through an unknown medium. This requires an unexplained metathesis, however. Furthermore, Arab. *ar-raṭbah* may be found in the sense ‘alfalfa’, but this meaning is isolated among the Semitic languages and appears to be caused by a late lexicalization of the sense ‘fresh, green’. At the same time, the word must have been borrowed in Armenian before the lenition of PIE **bʰ* > *w*, i.e. before the introduction of Iranian loanwords. J̌ahowkyan (2010: 75) hesitantly reconstructs PIE “**orobʰ-*” and compares Ru. *rjabína* ‘rowan’ (etc.). This is far from compelling. The Slavic comparanda are formally and semantically distant, and they are usually considered to reflect a derivation of PSl. **erębъ* ‘partridge, grouse’ (ÈSSJa I: 73–4, Derksen 2000).

4 The same alternation appears in the clearly non-Indo-European lexeme represented by Lat. *faba*, Fal. *haba* ‘bean’ (< **bʰab-*), OPr. *babo*, OCS *bobъ* ‘id.’ (< **bʰabʰ-*) vs. ON *baun*, OHG *bōna* (< **bʰau̯-n-*), cf. Kroonen (2013: 55), Šorgo (2020: 435, 460–1).

form *(*V*)*ṙVβ*-. Such an input is better compatible with the evidence of the other languages, in particular Greek, and it would date the borrowing to sometime between the lenition of stops and the sound shift, which can probably be observed in the suffix *-*oud*- > -*oyt*- (cf. the chronology of Ravnæs 2005).[5] Another fact that suggests a relatively late borrowing is the presence of *ṙ* in place of *r*, the latter of which would be the regular reflex of PIE **r* between vowels. The trilled *ṙ* usually reflects the clusters **sr*, **rs*, and perhaps **rH* (cf. Macak 2017: 1061), but the comparanda do not indicate that any such cluster was present in the input form. It is thus possible that the borrowing took place late enough for the trilled *ṙ* to have already emerged as a Pre-Armenian phoneme.

The original vocalism cannot be determined with certainty, but a few observations can be made. As for the second vowel, the only option among Proto-Indo-European vowels that can be outright rejected is short **e*. Since no external evidence exists for **ē*, **ō*, **i*, or **u*, the most parsimonious option is **o*, leading to quasi-PIE **Hrobʰ/p*-. The rarer variant *aṙawoyt*, then, may be the result of assimilation. If we accept the rule that **o* yielded *a* in an initial, open syllable, a change that took place only after the loss of pretonic **i* and **u* (Pedersen 1900: 99, Grammont 1918: 223–225, Clackson 2020), we may assume that the initial vowel was **o.* We thus arrive at something like **oṙoβ*-, which comes formally close to Gr. ὄροβος 'bitter vetch'.[6] Even still, it cannot be ascertained whether the initial vowel was present in the input or results from the regular vowel prothesis that affects initial **r*- and **ṙ*- in inherited words, as well as most later loans (e.g. *aṙat* 'liberal, generous, abundant', cf. Parthian *rād* 'id.'). In Greek too, the initial vowel ἐ- may be caused by secondary prothesis, but the Germanic and Latin evidence shows that forms with initial vowels were also in circulation.

A crucial observation is the presence of a suffix **-oud*-, potentially related to Gr. -ινθ- and Gm. **-īt*- (cf. Kroonen, this volume). An archetypically non-Indo-European suffix, it is already known in several variants. The form **-oud*-, with a back vowel, comes close to the variant -υνθ-, in e.g. Gr. ὄλ-υνθος, ὄλ-ονθος 'winter fig' (perhaps related to the Lat. gloss *bolunda*; Beekes 2010: 1074) and especially

5 Another possibility is that PIE **bʰ* had lost its aspiration but without merging with older **b*, either because **b* had shifted to *p*, had maintained an inherited glottalic articulation (as assumed by proponents of the Glottalic Theory, cf. Kortlandt 1983: 98–9), or were simply non-existent in the language. At the same time, however, this would also require the assumption that the suffix was borrowed with a final voiceless (glottalic) stop, an alternant for which no evidence exists elsewhere.

6 This observation is also interesting in view of semantics, because both bitter vetch and alfalfa are crops primarily used for the feeding of ruminant animals and not generally consumed by humans, as opposed to (chick)peas.

the nasal-less variant -ῦθ- as in ἀγνύς, -ῦθος 'loom-weight' (Chantraine 1999: 12) and κωμύς, -ῦθος 'bundle, truss of hay' (Beekes 2010: 814). Another form of the suffix with final *-*d*- underlies PGm. **arwīt*-; cf. further **albut*- ~ **albet*- 'swan' (ON *ǫlpt*, OE *ielfetu*) vs. PSl. **olbǫdь* 'swan' (SCr. *làbūd*, Sln. *labǫ́d* 'swan') and PSl. **lebedь* 'swan', Ru. *lébed'* (Kroonen 2013: 20, Jakob, this volume). Thus, the only alternant that is unique to Armenian is the diphthong *oy*, pointing to quasi-PIE **ou*.

2.2 *artoyt* 'lark, skylark'

The earliest attestation of this word is also in the Galen Dictionary, where it glosses Gr. κορύδαλος (Greppin 1985: 62). It is widespread in the dialects (HAB I: 344). It has been compared to a set of European words for 'thrush', including Lat. *turdus*, Lith. *strãzdas*, SCr. *drôzd*, ON *þrǫstr* and OIr. *truit* 'thrush, blackbird' (Łapᶜancᶜyan 1961: 359, J̌ahowkyan 1967: 151). This comparison does not adhere to established sound laws, however.[7] A proto-form **trosdo*- would yield Arm. ***arost* while the zero-grade formation **trsdo*- would probably yield ***tᶜaṙt*.[8]

Rather, *artoyt* must reflect quasi-PIE **droud-V*-.[9] This form more closely resembles Gr. στροῦθος, στρουθός 'sparrow' (also 'ostrich; flounder'), which can reflect **stroudʰ*-.[10] The Hesychius gloss στροῦς· ὁ στρουθὸς καὶ ὄσπριον (sparrow/ostrich and pulse) appears to be an old root noun, which indicates that the input

7 Among these comparanda, Hamp (1981: 81), de Vaan (2008: 637), Matasović (2009: 392, 2020: 335) and others rather cite Arm. *tordik* 'thrush'. However, it is unlikely that *tord*- can reflect **dorzdʰ*-, which would rather give ***toṙt*- (cf. fn. 8), and the aspirated **dʰ* conflicts with the Germanic evidence. Most importantly, there is no reliable attestation of the form *tordik* in classical sources. It is found only once in a 19th century edition of Philo, where Ačaṙyan (HAB IV: 422) suspects it to have been added by the editor. Therefore, it may simply have been borrowed from It. *tordo* and furnished with the highly productive diminutive suffix *-ik* (Vahagn Petrosyan, p.c.).

8 There are no certain examples that show the outcome of **-rsd*- or **-rst*-. If Arm. *owṙtᶜ** 'rain', seen in *owṙtᶜem* 'fertilize' and *y-owṙtᶜi* 'irrigated, fertile', reflects **h₁urs-ti*- (cf. Skt. *vr̥ṣṭí*- 'rain'; Martirosyan 2010: 498–499), it shows that the merger of **rs* > *ṙ* also took place before a stop, but only after blocking the post-resonant sonorization of **t* > *d*. The regular outcome of **-sd*- is *-st*- (cf. *nist* 'seat' < **ni-sd-o*-) which suggests that any opposition with the voiced allophone **z* was neutralized, so that **-rsd*- would yield **-ṙt*-.

9 J̌ahowkyan (2010: 96–7) also gives the option of reconstructing **t* (**troud/t*-, **trud/t*-). This is, however, impossible, since **t* would undergo lenition, not metathesis, and we should thus expect ***aroy(t)*/*arow(t)*. The written variant *artowt* can be easily understood as a levelling after the oblique cases where unstressed *oy* becomes *ow*.

10 A variant with voiced onset and no **s*- may be seen in the personal name (gen.) Δρούθου (Furnée 1972: 182), but this is obviously circumstantial evidence.

form ended in a consonant, but the Armenian form must have been transferred to a vocalic class since otherwise, we would expect final **-ds* to yield ***-c*. Although the Greek and Armenian forms are not identical, they are formally and semantically similar enough that we can tentatively assume independent borrowings from a third source.

Turning back to the European words for 'thrush', these are usually traced to **trosd-* (*vel sim.*; cf. IEW 1096 **trozdos-*; Greppin *apud* EIEC: 582 **trosdo-*; Hamp 1981: 81 **(s)drosdʰ-*). However, there are several irregularities between these comparanda alone, rendering it unlikely that the etymon is inherited from PIE. First, Lat. *turdus* must reflect either **torsd⁽ʰ⁾o-*, which would show an irregular metathesis, or **trsd⁽ʰ⁾o-* (de Vaan 2008: 634–635), which would be a rare case of an *o*-stem with a root zero grade.[11] Second, ON *þrǫstr* can reflect PGm. **þrastu-* < **trosd-*, but the West Germanic forms OHG *thrōšca*, *drŏsca*, OE *þrysce* must reflect **þrusk(j)ōn-* (< **trus(T)-(s)k-*) with an unexpected *u*-vocalism (Kroonen 2013: 545) and possibly a suffix **-sk-*, which seems to be associated with animal names of substrate origin in Western Europe (see Stifter, this volume). Finally, all Slavic forms have an irregular initial **d-*.[12] These formal issues, coupled with the limited, but geographically contiguous, distribution of the word suggest that it has a non-Indo-European origin (cf. Matasović 2009: 392, 2020: 335). We are thus faced with two main groups of alternating forms denoting passerine birds. One group shows a sibilant before the root-final consonant, while the other does not:

1)
- **stroudʰ-*: Gr. στροῦθος 'sparrow'
- **droud-*: Arm. *artoyt* 'lark'

2)
- **trosd-*: ON *þrǫstr* 'thrush', OE *þræsce*; OIr. *truit, truid* 'thrush' (or < **trusd-*)[13]
- **tresd⁽ʰ⁾-*: OPr. (EV) *tresde* 'thrush'
- **strosd⁽ʰ⁾-*: Lith. *strãzdas*, Latv. *strazds* 'thrush, blackbird, starling'
- **drosd⁽ʰ⁾-*: Ru. *drozd*, SCr. *drôzd* 'thrush'

11 The best example of such a noun is of course **iugo-* 'yoke', which is both widespread, archaic (cf. Hitt. *yūk-* 'yoke') and derived from a well-attested verbal root **ieug-* (LIV^2 316). None of these criteria can be said to apply to **trsdo-*.

12 As a parallel for voicing in this environment, Smoczyński (2018: 1308) cites OCS *nozdri*, Lith. *nas(t)raĩ* 'nostrils', but these forms are irrelevant because they go back to **nas-ra-* with an epenthetic dental (cf. Smoczyński 2018: 843). Assuming assimilation (Vasmer 1955, 1: 372) is an *ad hoc* solution without clear parallels.

13 Related forms in British Celtic, viz. W *trydw*, OBret. *trot*, OCorn. *troet* 'starling', appear to be loans from Irish. If not, these forms (along with the Irish) continue PC **troddi-*, which would then represent yet another irregular alternant (cf. Stifter, this volume).

– **trusT-(s)k-*: OHG *drǒsca* (alternatively **trauᵒ*), OE *þrysce* 'thrush' (**þruskjōn-*)
– **t(o)rsd⁽ʰ⁾-*: Lat. *turdus* 'thrush'

Whether the forms of group 1 and 2 are ultimately related remains uncertain. However, at least one possible parallel for a substrate alternation **-VsC-* ~ **-VC-* can be adduced, and it shows a similar north(west)–south(east) distribution; consider the following words for 'barley' (IEW 446, Witczak 2003: 55–57, Martirosyan 2010: 199, Kroonen 2013: 175, Thorsø 2020, Šorgo 2020: 439):

1)
– **gʰrīdʰ-*: Gr. κρῑθή, Epic κρῖ (< **κρῖθ*; Chantraine 1999: 583)
– **gʰ(ə)rĭt-*: Arm. *gari*, gen.pl. *gareacᶜ*

2)
– **gʰersd-*: OHG *gersta*
– **ǵʰrsd-*: Alb. *drithë* 'cereals, grain'
– **gʰ(o)rsd-iu̯o-*: Lat. *hordeum* (or < **gʰord-*)

To summarize, the words for 'thrush', 'sparrow' and 'lark' in Germanic, Celtic, Balto-Slavic, and Italic show several formal irregularities and an alternation **-VsC-* ~ **-VC-* vis-à-vis Greek and Armenian, which shows that they are of non-Indo-European origin. Again, this example also exhibits the alternation between Arm. **ou*, this time found in Greek as well, with a monophthong elsewhere.

2.3 *kᶜowpič* 'male hawk or falcon'

This is a hapax found in the commentaries on Dionysius Thrax by Grigor Magistros and Yovhannēs Erznkacᶜi (Adonc 1915: 240), where names for male animals are discussed. It is said to designate the male of the *šahēn* 'peregrine falcon' and the *gawaz* 'hawk'.[14] Ačaṙyan (HAB IV: 593) records no etymologies, and the word is not cited in more recent etymological works. I propose a connection with the following forms in Germanic and Slavic (cf. Suolahti 1909: 359–362, Boutkan 1998: 125, Kroonen 2013: 97–98, Šorgo 2020: 440, Jakob 2023: 168).

– **ko/abʰ-ou̯ǵ-*: PSl. **kobuzъ* (Po. *kobuz* 'hobby', USrb. *kobušk* 'red-footed falcon')

14 *Ew bazēi arakan čowrak* [. . .] *isk šaheni ew gawazi kᶜowpič. Ew yaytni nšanakowtᶜiwn, zi očᶜ owrowkᶜ ayocᶜ hawowcᶜ lini kᶜowpič anown*: "And the *čowrak* is the male of the goshawk [. . .] but the male of the *šahēn* and the *gawaz* is the *kᶜowpič*. And the meaning is clear, for *kᶜowpič* is not the name of any other birds." (cf. Greppin 1978: 67).

– **ko/abʰ-ug/ǵ-* (or **ka/opúg-*): PGm. **habuka-* (ON *haukr*, OE *heafoc*, *hafoc*, *hafuc* 'hawk', OHG *habuh* 'hawk')
– **koub-(ig-i̯V-)*[15]: Arm. *kᶜowpič*

Additional Slavic forms reflect **kobьcь* (ORu. *kobecъ* 'hawk', SCr. *kóbac* 'merlin', Sln. *(s)kóbəc* 'sparrowhawk', Po. *kobiec* 'falcon'). This form could have replaced the suffix with **-ьcь*, whereas **-uzъ*, on the other hand, can hardly be explained as secondary (Jakob 2023: 168). The Germanic and Slavic comparanda show several signs of being non-Indo-European borrowings. First, the root structure **ka/obʰ-* is a disallowed Proto-Indo-European root structure, containing a *tenuis* and a *media aspirata*. If the occasional comparison with Lat. *capys*, *capus* 'falcon, hawk' (Suolahti 1909: 360, Kroonen 2013: 197) is valid, we are faced with an additional root variant **kap-*, which would at best match the Germanic comparanda through Verner's Law. However, the Latin form may also be an unrelated loan from Etruscan (WH I: 164, Ernout & Meillet 1951: 176). Taking PSl. **kobuzъ* at face value leads to an alternation of the suffixes **-ug/ǵ-* and **-ouǵ-*. A Proto-Indo-European ablaut **ou* : **u* would be highly unusual in a suffix, and this alternation thus supports the assumption that the Germanic and Slavic words were independent borrowings.

The addition of the Armenian comparandum suggests that the input of the Germanic and Slavic forms had the root syllable **o*, not **a*. In this way, we can reconstruct the main root alternants **koub-* ~ **kobʰ-*, yet again with an alternation of diphthong and monophthong, as well as **b* ~ **bʰ*. The final syllable *-ič* points to a suffix **-ig-i̯V-*, as opposed to the back-vocalic suffix **-ug-* ~ **-ouǵ-* seen in Germanic and Slavic. However, we cannot exclude the possibility of secondary influence by the suffix Arm. *-ič*, which appears to have been marginally productive in the pre-literary period, in particular in animal names, cf. *karič* 'scorpion', *xaṙnič* 'locust', *owtič* 'moth' (root *owt-* 'eat'?) and *daṙnič* 'endive' from *daṙn* 'bitter'; Greppin 1975: 96–97). This suffix also has a rarer variant *-owč*, cf. *parkowč* 'follicle, shell', probably from *parik* 'mermaid'. We could thus envisage an older **kᶜowpowč*, or even a simplex **kᶜowp*, although this remains hypothetical.

The scant and relatively late attestation of the word is not surprising, given its highly specialized semantics, which became limited to male individuals of specific hunting birds. The clear similarity with the Germanic and Slavic forms makes it likely that it reflects a loanword adopted when the ancestor of Armenian was still spoken in Europe.

15 If not simply **koubig-i-* with palalization of **-g-* before front vowel, as in *čmem* 'squeeze' < **gim-*, cf. OCS *žьmǫ* 'press'. However, we have no information on the stem type of the Armenian form. Synchronically, forms with the suffix *-ič* are always *a*-stems (*karič*, gen.-dat.pl. *karčacᶜ* 'scorpion').

2.4 *poytn* ‘pot’?

Finally, I will discuss an additional potential example, although it is fraught with more uncertainty. Arm. *poytn* (gen. *powtan*, var. *boytn* [Bible], *poyt* [Agathangelos]) ‘pot’ has long been compared to ON *pottr*, OE *pott* ‘pot’ ?< PGm. **putta-* (Petersson 1916: 254, IEW 99), under the assumption that these forms reflect **boud-no-* and **bud-no-*, respectively. However, the supposition of a Proto-Indo-European root **beud-* is fundamentally flawed, since it contains two *mediae.* The word may thus be better interpreted in a non-Indo-European context.

As suggested by PUr. **pata$_2$* ‘pot’ (Fi. *pata*, Meadow Mari *pot*, Khanty (VV) *put*, Mansi (Tavda) *pōt*, Hung. *faz-ék* ‘pot’, Selkup (Taz) *pot-* ‘put in a pot’; Zhivlov 2014: 120), we could be dealing with a *Wanderwort* with an East-West trajectory of spread.[16] On the other hand, PGm. **fata-* (ON *fat*, OE *fæt* ‘vat, barrel’) and Lith. *púodas* ‘pot’, which point uniformly to **podo-*, are perhaps more obvious candidates for an early borrowing of this etymon.[17] It is theoretically conceivable that the word was borrowed twice into Germanic, i.e. once before and after Grimm’s Law, but it is doubtful whether the form **putta-* existed in Proto-Germanic at all. ON *pottr* is not attested before the 14th century and therefore appears to be a Low German borrowing (de Vries 2000: 427). Consequently, the word is limited to West Germanic, where it is also attested late (13th century in both Old English and Middle Low German). It has been considered a loan from OFr. *pot* ‘pot’, presupposing **pottus* (Frings 1966: 111), but the opposite direction of borrowing cannot be excluded, as this form has no clear background within Italic. It is hardly a *littera*-variant of Lat. *pōtus* ‘drink’; the meanings do not match, and it seems that the *littera*-rule (i.e. *-V̄C-* > *-VCC-*) was limited to syllables with high vowels (Sen 2015: 65). Thus, **pott-* most likely represents a Western European substrate word adopted in both Romance and West Germanic, whence it later spread to Nordic (Ernout and Meillet 1951: 936; Hubschmid 1955: 158–160; FEW IX: 270).[18]

16 A borrowing from Baltic to Uralic can be excluded because all Uralic cognates are regular. The spread of the word could be correlated with the secondary spread of pottery from the Far East to Europe via Siberia, by way of the Pit-Comb Ware Culture, ca. 4000–2000 BCE (Gibbs & Jordan 2013; Isaksson et al. 2018).

17 Of course, while **podo-* is confined to a limited geographical area, there are no formal reasons to consider it a loanword.

18 LLat. *potus* ‘drinking cup (?)’ (a hapax in *The Life of St. Radegund* by Venantius Fortunatus, 6th c. AD) may be an early attestation of this word, if it was borrowed after the lenition of Lat. *-t-* had begun (Meyer-Lübke 1911: 502), but it is not fully excluded that this word rather means ‘drink’ and is identical to *pōtus* (see FEW IX: 271).

It remains uncertain whether these forms bear any connection with Arm. *poytn,* presupposing **boud(n)-*.[19] Nevertheless, no suitable context exists for a more recent spread of the **pott-* etymon from Western Europe, since Arm. *poytn* is attested already in the 5th century. Furthermore, since we may be faced with the same alternation of diphthong and monophthong as exhibited by the three other examples discussed in this study, it is conceivable that **boud-* is a variant of the foreign etymon **pott-*, which surfaces only in Western Europe, and perhaps **podo-*. Especially the latter form would entail that it was borrowed into Armenian from the same source as *kᶜowpič* (§ 2.3) which shows an identical distribution.

3 Evaluation

To summarize the material presented above, we find at least three, perhaps four examples of a recurring alternation in European substrate words, where Armenian contains the reflex of a diphthong **ou* in positions where a monophthong is reflected elsewhere. First of all, these examples provide additional evidence for the fact that Pre-Armenian was once spoken within in a sphere of Indo-European languages that had contact with at least one non-Indo-European language. Furthermore, the fact that a recurring alternation, with a consistent distribution, can be identified in these etyma ties them to the same temporal and geographic stratum.

Some deductions about the relative chronology of these contacts have already been touched upon. In particular, the example *aṙowoyt* (§ 2.1) is best analyzed by assuming that at the time of borrowing, Pre-Armenian had already seen the rise of the phoneme /ṙ/ from **rs*, **sr*, and **rH.* Perhaps, it had also undergone the first stage of the lenition of intervocalic labial stops **p* and **bʰ*. Otherwise, all examples must predate the Armenian sound shift, the metathesis of the clusters **dr* and **dʰr* (cf. *artoyt* < **droud-*), and perhaps the secondary palatalization (*kᶜowpič* < **koubigi̯V-*). Concerning the chronology of the diphthongs themselves, it is difficult to conclude much based on the material. Because the change of the diphthong **ou* (and perhaps **eu*) into *oy* most likely went through an intermediate stage such as **øy* (cf. Pedersen 1905: 324), it remains conceivable that the ostensible **ou* really reflects an adaptation of a foreign phoneme /ø/ *vel sim.*

19 The final *-n* and the corresponding *n*-stem declension pattern is not necessarily original. In some cases, it likely reflects the generalization of the accusative singular (thus *otn* 'foot' < **podm*), but in other cases the *-n* may be entirely secondary in origin (Weitenberg 1985).

A highly relevant morphological feature of the form *aṙowoyt* (§ 2.1) is the presence of a suffix that is most likely related to Greek -νθ- and its variants (referred to by the cover symbol *-*ṼD*-, cf. Šorgo 2020: 428). This suffix is usually associated with (a component of) the Pre-Greek substratum (Kuiper 1956: 216–219; Katičič 1976: 42–43; Kroonen, this volume). As shown by such examples as PGm. **arw-īt-* 'pea' against Gr. ἐρέβ-ινθος; PGm. **albut-/-et-* 'swan' against PSl. **olbǫdь/*lebedь*; and perhaps PGm. **samda-* 'sand' against Gr. ἄμαθος (where -αθ- could reflect **-n̥dʰ-*; Kroonen, this volume),[20] occasionally this suffix is also a feature of substrate words found in the Indo-European languages of Northern Europe. The most economic explanation for this observation seems to be that Greek, as well as several other Indo-European languages, were in contact with *genetically related* non-Indo-European substrate languages. These substrate languages may have been part of the lost linguistic landscape formed by farming societies that had spread from Southeast Europe during the Early and Middle Neolithic (Childe 1926: 83, Kallio 2003, Iversen & Kroonen 2017, Šorgo 2020). In contrast, Kuiper (1956: 218–219) assumes that forms like PGm. **arwīt-* "wandered from the Mediterranean to northern Europe." That is, they show the result of a secondary spread, not direct contact with the same substrate as Greek. Later, Kuiper (1995) operates with three different substrate layers in Europe. Yet, the reality of two of these layers is doubtful. In particular, the layer A2, or "the language of the geminates" (cf. Schrijver 2003: 220–224), has as its main diagnostic the alternation of geminate and singleton consonants in Germanic, but since this variation almost never occurs outside *n*-stems, it is better explained by the operation of Kluge's Law (Kroonen 2011: 127–131). Kuiper's A3, corresponding to Hans Krahe's "Alteuropäisch", is almost exclusively identified on the basis of toponyms. Therefore, we cannot conclude that it was ever in direct contact with the Indo-European languages. Moreover, it should be noted that the suffix **-ṼD-* is not the only recurring morphological feature of substrate words in both Greek and other European languages. Most notable of these is the prefix **a-*, associated with vowel reduction in roots (Schrijver 1997: 307–312, Iversen & Kroonen 2017: 518). Building upon the existing evidence for the influence of this European substrate upon the Armenian lexicon (cf. Martirosyan 2010: 805–807), we now have an exam-

20 Note, however, that this analysis becomes more difficult if we compare Arm. *awaz* (gen. *awazoy*) 'sand', as **sabʰn̥dʰ-* would yield ***awand*. Instead, it is traditionally assumed that *awaz* reflects **sabʰadʰ-o-* (HAB I: 351, Olsen 1999: 24). Alternatively, if the change **VdʰV* > *VzV* is rejected (e.g. Martzloff 2016: 129–35), we might instead posit a root noun **sabʰadʰ-s* > **awaj* with subsequent thematicization and intervocalic lenition **awaj-o-* > **awaz-o-*. Martirosyan (2010: 149–150) rejects the comparison altogether and considers *awaz* an Iranian loanword, citing P *āwāze* 'swamp', but this is semantically problematic.

ple of a recurrent phonemic alternation and a morphological substrate feature in the form of the suffix *-*oud*- > -*oyt*.[21]

References

Adonc, Nikolaj G. 1915. *Дионисий Фракийский и армянские толкователи* [Dionysius Thrax and the Armenian commentators]. Petrograd: Tipografija Imperatorskoj Akademii Nauk.

Bedevian, Armenag K. 1936. *Illustrated Polyglottic Dictionary of Plant Names*. Cairo: Argus & Papazian.

Beekes, Robert S. P. 1996. Ancient European Loanwords. *Historische Sprachforschung* 109. 215–236.

Beekes, Robert S. P. 2000. European substratum words in Greek. In Michaela Ofitsch & Christian Zinko (eds.), *125 Jahre Indogermanistik in Graz*, 21–31. Graz: Leykam.

Beekes, Robert S. P. 2010. *Etymological Dictionary of Greek*. 2 vols. Leiden/Boston: Brill.

Boutkan, Dirk. 1998. On the form of North European substratum words in Germanic. *Historische Sprachforschung* 111. 102–133.

Chantraine, Pierre. 1999. *Dictionnaire étymologique de la langue grecque: Histoire des mots*. 2nd edn. Paris: Klincksieck.

Childe, Gordon V. 1926. *The Aryans: A Study of Indo-European Origins*. London: Kegan Paul, Trench, Trubner & Co./New York: Alfred A. Knopf, Inc.

Clackson, James. 2017. The lexicon of Armenian. In Jared Klein, Brian Joseph & Mathias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 2, 1115–1132. Berlin/Boston: De Gruyter Mouton.

Clackson, James. 2020. The development of **o* in Armenian: The evidence of Greek loanwords. In Claire Le Feuvre & Daniel Petit (eds.), Ὀνόματων ἴστωρ: *Mélanges offerts à Charles de Lamberterie*, 33–44. Leuven/Paris: Peeters.

Derksen, Rick. 2000. Old Icelandic *jarpi* 'hazel-grouse', *rjúpa* 'ptarmigan' and their Germanic and Balto-Slavic cognates. *Amsterdamer Beiträge zur älteren Germanistik* 54. 75–84.

Dervischjan, P. Seraphin. 1877. *Armeniaca I: Das Altarmenische. Ein Beitrag zur Indo-Europæischen Lautlehre*. Vienna: Mechitharisten.

Ernout, Alfred & Antoine Meillet. 1951. *Dictionnaire etymologique de la langue latine: Histoire des mots*. 3rd edn. Paris: Klincksieck

ÈSSJa = Oleg N. Trubačev. 1974–. *Этимологический словарь славянских языков* [Etymological dictionary of the Slavic languages]. Moscow: Nauka.

FEW = Walther von Wartburg. 1922–. *Französisches etymologisches Wörterbuch: Eine Darstellung des galloromanischen Sprachschatzes*. Bonn: Fritz Klopp/Berlin: Teubner/Basel: Zbinden. Online access: https://lecteur-few.atilf.fr/index.php/.

Frings, Theodor. 1966. *Germania Romana*. 2nd edn. Halle: Niemeyer.

Furnée, Edzard J. 1972. *Die wichtigsten konsonantischen Erscheinungen des Vorgriechischen*. The Hague/Paris: Mouton.

21 Two other potential examples of the suffix variant *-*oud*-, both found in plant names, have no root comparanda, but I will cite them here for future reference: *xtowt* 'spleenwort, *Asplenium trichomanes*' (HAB II: 429, Bedevian 1936: 91), *kałnowt* 'restharrow, *Ononis repens*' (HAB II: 496, Bedevian 1936: 424). Both are attested very late and might show a levelling of the suffix *-oyt* → *-owt* based on the oblique cases, cf. also *aṙwowt, aṙowt* as spelling variants of *aṙowoyt*.

Gibbs, Kevin & Peter Jordan. 2013. Bridging the Boreal Forest: Siberian Archaeology and the Emergence of Pottery among Prehistoric Hunter-Gatherers of Northern Eurasia. *Sibirica* 12(1). 1–38.

Grammont, Maurice. 1918. Notes de phonétique générale. *Mémoires de la société de linguistique de Paris* 20. 213–259.

Greppin, John A. C. 1975. *Classical Armenian Nominal Suffixes: A Historical Study*. Vienna: Mechitharisten.

Greppin, John A.C. 1978. *Classical and Middle Armenian Bird Names*: *A Linguistic, Taxonomic, and Mythological Study*. Delmar, NY: Caravan Books.

Greppin, John A.C. 1983. An Etymological Dictionary of the Indo-European Components of Armenian (Cont. 1). *Bazmavēp: Revue d'études arméniennes* 1983. 235–323.

Greppin, John A.C. 1985. *Barkᶜ Gałianosi: The Greek-Armenian Dictionary to Galen*. Delmar, NY: Caravan Books.

Greppin, John A.C. 1992. Mediterranean Botanical Loanwords in Classical Armenian. In J.A.C. Greppin (ed.), *Proceedings of the Fourth International Conference on Armenian Linguistics*, *Cleveland State University*, *Cleveland, Ohio*, *September 14–18 1991*, 61–75. Delmar, NY: Caravan Books.

Greppin, John A.C. 1996. A Medieval Arabic-Armenian Botanical Dictionary. *Handēs Amsōreay* 110. 380–457.

HAB I-IV = Hračᶜeay Ačaṙyan. *Հայերէնի արմատական բառարան* [Armenian root dictionary]. 4 vols. 2nd edn. Yerevan: Erevani Hamalsarani Hratarakčᶜowtᶜyown.

Hamp, Eric P. 1981. Refining Indo-European Lexical Entries. *Zeitschrift für vergleichende Sprachforschung* 95(1). 81–83.

Hubschmid, Johannes. 1955. *Schläuche und Fässer: Wort- und sachgeschichtliche Untersuchungen mit besonderer Berücksichtigung des romanischen Sprachgutes in und außerhalb der Romania sowie der türkisch-europäischen und türkisch-kaukasisch-persischen Lehnbeziehungen*. Bern: Francke.

Hübschmann, Heinrich. 1877. Ueber die stellung des armenischen im kreise der indogermanischen sprachen. *Zeitschrift für vergleichende Sprachforschung* 23. 5–49.

IEW = Julius Pokorny. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern: Francke.

Isaksson, Sven, Kevin Gibbs & Peter Jordan. 2018. Maritime Nomads of the Baltic Sea: Ceramic Traditions, Collective Identities and Prehistoric Cuisine. In Peter Jordan & Kevin Gibbs (eds.), *Ceramics in Circumpolar History: Technology, Lifeways and Cuisine*, 63–80. Cambridge: Cambridge University Press.

Iversen, Rune & Guus Kroonen. 2017. Talking Neolithic: Linguistic and Archaeological Perspectives on How Indo-European Was Implemented in Southern Scandinavia. *American Journal of Archaeology* 121. 511–525.

J̌ahowkyan, Gevorg B. 1967. *Очерки по истории дописьменного периода армянского языка* [Essays on the history of the preliterary period of the Armenian language]. Yerevan: Academy Press.

J̌ahowkyan, Gevorg B. 1987. *Հայոց լեզվի պատմություն: նախագրային ժամանակաշրջան* [History of the Armenian language: Preliterary period]. Yerevan: Academy Press.

J̌ahowkyan, Gevorg B. 2010. *Հայերէնի ստուգաբանական բառարան* [Armenian etymological dictionary]. Yerevan: Asołik.

Jakob, Anthony. 2023. *A History of East Baltic through Language Contact*. Leiden/Boston: Brill.

Kallio, Petri. 2003. Languages in the Prehistoric Baltic Sea Region. In Alfred Bammesberger & Theo Vennemann (eds.), *Languages in Prehistoric Europe*, 227–244. Heidelberg: Winter.

Katičič, Radoslav. 1976. *Ancient Languages of the Balkans. Part One*. The Hague/Paris: Mouton.

Kortlandt, Frederik. 1983. On the relative chronology of Armenian sound changes. In John A. C. Greppin (ed.), *First International Conference on Armenian Linguistics: Proceedings*, 97–106. Delmar, NY: Caravan Books.

Kroonen, Guus. 2011. *The Proto-Germanic n-stems: A study in diachronic morphophonology*. Amsterdam/New York: Rodopi.

Kroonen, Guus. 2012. Non-Indo-European root nouns in Germanic: evidence in support of the Agricultural Substrate Hypothesis. In Riho Grünthal & Petri Kallio (eds.), *A Linguistic Map of Prehistoric Northern Europe*, 239–260. Helsinki: Suomalais-Ugrilainen Seura.

Kroonen, Guus. 2013. *Etymological Dictionary of Proto-Germanic*. Leiden/Boston: Brill.

Kuiper, Franciscus B. J. 1956. The etymology of ἄνθρωπος. In Heinz Kronasser (ed.), *ΜΝΗΜΗΣ ΧΑΡΙΝ: Gedenkschrift Paul Kretschmer I*, 211–226. Wien: Hollinek.

Kuiper, Franciscus B. J. 1995. Gothic *bagms* and Old Icelandic *ylgr*. *NOWELE* 25. 63–88.

Lamberterie, Charles de. 1982. Review of Rüdiger Schmitt, *Grammatik des Klassisch-Armenischen mit sprachvergleichenden Erläuterungen*, Innsbruck, 1981. *Bulletin de la Société de Linguistique de Paris* 77. 80–86.

Łapʿancʿyan, Grigor A. 1961. *Հայոց լեզվի պատմություն: հին շրջան* [History of the Armenian language: ancient period]. Yerevan: Haykakan SSH GA Hratarakšowtʿyown.

LIV^2 = Helmut Rix, Martin Kümmel, Thomas Zehnder, Reiner Lipp & Brigitte Schirmer (eds.). 2001. *Lexicon der indogermanischen Verben*. 2nd edn. Wiesbaden: Reichert

Macak, Martin. 2017. The phonology of Classical Armenian. In Jared Klein, Brian Joseph & Mathias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*. Vol 2, 1037–1079. Berlin/Boston: De Gruyter Mouton.

Martirosyan, Hrach K. 2010. *Etymological Dictionary of the Armenian Inherited Lexicon*. Leiden/Boston: Brill.

Martirosyan, Hrach K. 2013. The place of Armenian in the Indo-European language family: the relationship with Greek and Indo-Iranian. *Journal of Language Relationship* 10. 85–137.

Martzloff, Vincent. 2016. Le designation des «captifs» en albanais du Caucase (albanien), le verbe arménien *gerem* «emmener en captivité» et le traitement phonétique de $^*d^h$ intervocalique en arménien. *Wékʷos: revue d'études indo-européennes* 2. 109–177.

Matasović, Ranko. 2009. *Etymological Dictionary of Proto-Celtic*. Leiden/Boston: Brill.

Matasović, Ranko. 2020. Language of the bird names and the pre-Indo-European substratum. In Romain Garnier (ed.), *Loanwords and Substrata: Proceedings of the Colloquium Held in Limoges (5th to 7th June, 2018)*, 331–344. Innsbruck: Institut für Sprachwissenschaft.

Meillet, Antoine. 1936. *Esquisse d'une grammaire comparée d'arménien classique*. 2nd edn. Vienna: Mekhitharistes.

Meyer-Lübke, Wilhelm. 1911. *Romanisches etymologisches Wörterbuch*. Heidelberg: Carl Winter.

Olsen, Birgit A. 1999. The noun in Biblical Armenian: origin and word-formation. Berlin/New York: Mouton de Gruyter.

Olsen, Birgit A. 2020. An Armenian sound law revisited : *-eu̯- > *-iw-? In Claire Le Feuvre & Daniel Petit (eds.), *Ὀνομάτων ἴστωρ: Mélanges offerts à Charles de Lamberterie*, 115–124. Leuven: Peeters.

Pedersen, Holger. 1900. Wie viel laute gab es im Indogermanischen? *Zeitschrift für vergleichende Sprachforschung* 36(1). 74–110.

Pedersen, Holger. 1905. Die nasalpräsentia und der Slavische akzent. *Zeitschrift für vergleichende Sprachforschung* 38(3). 297–421.

Petersson, Herbert. 1916. Beiträge zur armenischen Wortkunde. *Zeitschrift für vergleichende Sprachforschung* 47. 240–291.

Polomé, Edgar C. 1986. The Non-Indo-European Component of the Germanic Lexicon. In Annemarie Etter (ed.), *o-o-pe-ro-si*: *Festschrift für Ernst Risch zum 75. Geburtstag*, 661–672. Berlin/New York: De Gruyter.

Ravnæs, Erling. 2005. The Relative Chronology of the Sound Changes from Proto-Indo-European to Classical Armenian. In Dag Haug & Eirik Welo (eds.), *Haptačahaptāitiš: Festschrift for Fridrik Thordarson on the Occasion of his 77th Birthday*. Oslo: Novus forlag/Instituttet for sammenlignende kulturforskning.

Salmons, Joe. 1992. Northwest Indo-European Vocabulary and Substrate Phonology. In Roger Pearson (ed.), *Perspectives on Indo-European Language, Culture and Religion: Studies in Honor of Edgar C. Polomé*, vol. 2, 265–279. McLean, VA: Institute for the Study of Man.

Schmitt, Rüdiger. 1981. *Grammatik des Klassisch-Armenischen mit sprachvergleichenden Erläuterungen*. Innsbruck: Institut für Sprachwissenschaft.

Schrijver, Peter. 1997. Animal, vegetable and mineral: some Western European substratum words. In Alexander Lubotsky (ed.), *Sound law and analogy: Papers in honor of Robert S. P. Beekes on the occasion of his 60th birthday*, 293–316. Amsterdam/Atlanta: Rodopi.

Schrijver, Peter. 2003. Early Developments of the Vowel Systems of North-West Germanic and Saami. In Alfred Bammesberger & Theo Vennemann (eds.), *Languages in Prehistoric Europe*, 195–226. Heidelberg: Winter.

Sen, Ranjan. 2015. *Syllable and Segment in Latin*. Oxford: Oxford University Press.

Smoczyński, Wojciech. 2018. *Lithuanian Etymological Dictionary*. Berlin: Peter Lang.

Solta, Georg R. 1960. *Die Stellung des Armenischen im Kreise der indogermanischen Sprachen*. Vienna: Mechitharisten.

Šorgo, Aljoša. 2020. Characteristics of lexemes of a substratum origin in Proto-Germanic. In Romain Garnier (ed.), *Loanwords and Substrata: Proceedings of the Colloquium Held in Limoges (5th to 7th June, 2018)*, 427–472. Innsbruck: Institut für Sprachwissenschaft.

Suolahti, Hugo. 1909. *Die deutschen Vogelnamen: Eine wortgeschichtliche Untersuchung*. Strassbourg: Trübner.

Thorsø, Rasmus. 2020. Two Balkan Indo-European Loanwords. In Matilde Serangeli & Thomas Olander (eds.), *Dispersals and Diversification: Linguistic and Archaeological Perspectives on the Early Stages of Indo-European*, 251–262. Leiden/Boston: Brill.

Vasmer, Max. 1955. *Russisches etymologisches Wörterbuch*. 2 vols. Heidelberg: Winter.

de Vaan, Michiel. 2008. *Etymological Dictionary of Latin and the other Italic Languages*. Leiden/Boston: Brill.

de Vries, Jan. 2000. *Altnordisches etymologisches Wörterbuch*. 2nd edn. Leiden/Boston/Cologne: Brill.

Weitenberg, Jos J. S. 1985. Additional -n in Armenian. *Annual of Armenian Linguistics* 6. 101–106.

Witczak, Krzysztof T. 2003. *Indoeuropejskie nazwy zbóż* [Indo-European cereal names]. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

WH I-II = Alois Walde & Johann B. Hofmann. 1938–1954. *Lateinisches etymologisches Wörterbuch*. 2 vols. Heidelberg: Winter.

Zhivlov, Mikhail. 2014. Studies in Uralic vocalism III. *Journal of Language Relationship* 14. 113–148.

Zsolt Simon

12 Indo-European substrates: The problem of the Anatolian evidence

The following study addresses the problem of whether the Anatolian branch of the Indo-European language family shares a substrate with any other Indo-European branch or language. The first section briefly outlines the methodology employed in this study, the second section contrasts this methodology with the limitations imposed by the philology of the Anatolian languages, and the third section provides the relevant etymological material. In the fourth section, I summarize the results.

1 The theory: Criteria of lexical substrates

The substrate(s) of a language (group) can be identified on phonological, morphological, syntactical, and lexical grounds as well as with the help of the toponymy. There are basically two possible approaches: identification with or without comparison to another language. While identification without comparison is not impossible, it is obviously more complicated, and the proposals will comparatively be weaker. If there is some kind of comparison, this can both be to an attested language and to traces of an unattested language in a third language (which in the end may turn out to be the traces of an attested language after all). Attributing grammatical changes to a substrate language is always more complicated and less falsifiable than identifying loanwords from a substrate. The toponym research also has its known constraints. Therefore, an approach will be employed here that is in accordance with the other studies of this volume: identifying substrate(s) on lexical grounds with comparison. Due to their geography, in the case of the Anatolian languages, a comparison can be made both with numerous non-Indo-European languages in their wider geographical horizon and other Indo-European languages. In accordance with the goal of this volume, this study investigates the latter: in other words, whether the Anatolian languages and any other Indo-European language or branch can be shown to share a common substrate.

Acknowledgments: I am grateful to H. Craig Melchert for sharing his new Cuneiform Luwian dictionary with me ahead of publication.

https://doi.org/10.1515/9783111337920-012

Any attempt to identify common substrates on lexical grounds faces the same key problem: how do we prove that the words proposed to represent a substrate show real cognacy with the words of the language being compared, in order to avoid incidental similarities and pre-scientific look-alikes (the *deus* / *theos* fallacy). My starting point is the frequently cited treatment by Schrijver (1997: 294–297), who conveniently summarized the arguments and presented counter-arguments (see the Introduction chapter, this volume). Here, I translate them into more general formulations with different and general criticism (substrate words and loanwords will be used synonymously in this section, see below on how to distinguish them):

1) We are dealing with words without etymology. This criterion is, unfortunately, subjective. In Section 3, we will see several cases which have an impeccable Indo-European etymology according to some researchers but are claimed to have no etymology whatsoever by other researchers. Furthermore, having no etymology does not necessarily imply anything (except the limits of our efforts), as such words could easily be inherited. For instance, it is instructive to have a look at the high number of words claimed to be inherited from Proto-Indo-European on formal grounds without a cognate outside Hittite or Anatolian in the etymological dictionary of the Hittite *inherited* lexicon (Kloekhorst 2008).
2) Our candidates have limited geographical distribution. This again does not necessarily imply that we are dealing with substrate origin. These words could simply be archaisms (see the case of the Sardinian vocabulary) or common, areal innovations.
3) The meanings of our words are conducive to borrowing. This criterion is partly subjective, but even assuming its validity, it again does not necessarily imply substrate words.
4) The proposed words show (recurring) phonological anomalies. This criterion comprises two subtypes: first, the occurrence of different, internally unmotivated variations of the same word within the same language is usually taken as a sign of (recent) borrowing; second, sets of non-inherited words in at least two languages that are semantically identical and formally almost identical, except a recurring difference – one of the languages consistently shows phoneme X where another language consistently shows phoneme Y. The classic example is a set of words showing **ai* in Proto-Germanic but **a* in the Proto-British cognates (Schrijver 1997: 299–307).
5) The candidates show recurring morphological anomalies. Within one or more language(s), several words have the same, foreign suffix. Unfortunately, this does not guarantee a borrowing since such suffixes may become productive and appear on inherited words (cf. the type G *halb-ieren*).

6) We find hints of non-Indo-European morphology or phonology in our words. On the one hand, this is a problematic argument from the point of view of phonology, since any borrowing has to be adapted phonologically.[1] If we are confronted with alleged cognates in different languages, their sound correspondences may reveal such irregular phonology – if we are in the lucky situation that they are recurring features, and then we are back to Point 4. On the other hand, unanalysable morphology may indeed point to a loanword. However, "unanalysable" does not necessarily mean the same for all scholars (see, e.g., the case no. 1.1 below).

The above considerations can be summarized as follows. Most of these arguments are supportive only, i.e., they do not prove anything, not even in combination (*contra* Schrijver 1997: 296) and they can offer a hint at best. Thus, we are probably dealing with loanwords / substrate words if:

1) They show phonological anomalies and/or non-transparent morphology, if only attested in one language. These apply also in the case of more than one language, but then these words must (otherwise) correspond to each other in a phonologically regular way.
2) They show a recurring phonological anomaly or multiple anomalies in the case of more than one language.

With these rules we will be able to make falsifiable claims for substrate words. However, two more details should be addressed:

1) Having a list of candidates, we still must determine whether we are dealing with a substrate or with (recent) loanwords. A key question is how the two can be separated. It is admittedly difficult to draw a line (and "substratum" is sometimes used in a more general sense, without sociolinguistic definition, see, e.g., Lubotsky 2001: 302), but the assumption of a substrate (and in general, any adstrate) implies a common society, i.e., cohabitation of different speech communities in the past. This is different from the adoption of loanwords from neighbours or from distant people, not to mention the special case of the *Wanderwörter*, as in their case it is frequently difficult to prove the necessary or alleged route. The problem is that the question of the societal background can be judged only with the help of historical sources and these are almost completely missing in the period we need to investigate. In

1 Note that frequently-used arguments such as "non-Indo-European" phonology (e.g. **/a/*) depends on the school of Indo-European studies, and irregular phonotactics in the sense of deviation from the root structure is in fact a *circulus vitiosus*.

other words, all cases must be evaluated individually and we can only hope that there will be enough evidence for a decision.

2) Assuming that we are dealing with a substrate word rather than a simple loan, we need a threshold that allows us to hypothesize the existence of a substrate on a lexical basis. If we have (a) specific reason(s) for excluding plain loanwords, then even one word could suffice. However, this is often not the case, as indicated in the previous point. The cases in which our candidates exhibit recurring phonological and morphological features are more straightforward, since they already imply that we are dealing with a securely connected set of words, making our claim falsifiable.

Unfortunately, reality is more complex. As I will show, many potential substrate cases consist of words of unknown etymology that are formally and semantically close to each other in at least two languages, and which are demonstrably not inherited from a common ancestor, but at the same time do not show any recurring anomalies, nor do they correspond to one another in a phonologically regular way. While one may refer to the criteria discussed above to provide supporting arguments, they do not in themselves suffice to demonstrate a substrate origin. Therefore, without further formal evidence (as per above), we are obliged to assume that they are not substrate words but rather chance resemblances, despite their closeness. This is admittedly a strict application of the formal criteria, probably stricter than most substrate researchers would prefer. This application also definitely excludes words that later might turn out to be substrate words, after all. On the positive side, only in this way can we obtain falsifiable claims about a substrate.

One of the reasons for this disturbing group is, of course, to be sought in the philological situation of the Anatolian languages, the relevance of which will be described in the next section.

2 The praxis: Perspectives and limits in the case of the Anatolian branch

In order to properly assess the question of the substrates of the Anatolian branch, I will first clarify the limits of the research as well as the potential substrates.

2.1 The limits of the research

When searching for substrates in the Anatolian languages, the following circumstances must always be kept in mind when evaluating the results or lack thereof:

a) The Anatolian languages are corpus languages, with variable levels of attestation: scarcely attested (Pisidian and Sidetic), of limited attestation (Carian, Lycian B, Lydian, Palaic), relatively well attested (Luwian and Lycian A), or fairly well attested (Hittite).
b) The phonetic forms of words are not necessarily clear due to undeciphered signs (Carian, Hieroglyphic Luwian, Sidetic, Pisidian) and due to the very nature of the writing systems (cuneiform and hieroglyphic).
c) Inscriptions of languages in alphabetic transmission frequently employ *scriptio continua* and accordingly, many words are not known at all or not known precisely, even when they are attested (Carian, Sidetic, and Pisidian).
d) A significant amount of words in all of the Anatolian languages have unknown or only vaguely known meanings (the etymological material in Section 3 will provide several examples).
e) None of these languages have an up-to-date full-scale dictionary (in the case of Hittite, only concise dictionaries are completed, but none of the thesauri CHD and HW^2 are; regarding the remaining languages, the eDiAna project planned to provide this type of dictionary, but the project was terminated prior to completion).
f) Only Hittite has a full-scale etymological dictionary (HEG). For the remaining languages, the eDiAna project would have provided this type of dictionary, but see above.

As a consequence, if demonstrating an Indo-European substrate proves impossible for the Anatolian branch, this does not necessarily imply that such a substrate did not exist, since due to the situation of the sources and the research we cannot evaluate whether this lack reflects a reality or it is incidental only. The usual maxim 'the absence of evidence is not evidence of absence' applies here in the most literal sense.

2.2 The possible substrates

There are two possible types of substrates: an "isolated" one affecting only the Anatolian branch (or any of its languages) and a "shared" one affecting the Anatolian branch (or any of its languages) *and* at least one other Indo-European language or branch. In accordance with the goal of this volume, this study is devoted to

the second type. The shared substrates, in turn, can again be divided into two subtypes, i.e. non-areal and areal:

Non-areal substrates are substrates shared with any Indo-European language that does not neighbour the Anatolian languages. The Anatolian languages were the first to branch off and did not, according to our current knowledge, interact with any other branch (except their neighbours). It would therefore be especially interesting to find such a common substrate (as we shall see, there have indeed been such proposals).

Areal substrates are substrates shared with any Indo-European language in the neighbourhood of the Anatolian languages, i.e., Greek, Phrygian, Thracian, and Armenian. This case requires some elaboration. First, Phrygian and Thracian will be omitted in the following, due to the simple reason that our knowledge of the vocabularies of these languages is exceedingly limited. Consequently, and unsurprisingly, I have found no proposals for a common substrate involving these groups. The remaining two languages imply three types of substrate, but we actually find proposals for four types in the literature (for references see Section 3):

1) A Greek-Anatolian-Armenian substrate. This hypothesis implies a substrate either stretching from the Aegean to East Anatolia or covering the Southern Balkans, the Aegean, and (Western) Anatolia. However, all such proposals that have been formulated also include additional Indo-European branches. For this reason, I have treated them in the "non-areal" and "Mediterranean" categories.
2) A Greek-Anatolian substrate. The interesting facet of this possibility is the known existence of Pre-Greek languages, even if we cannot identify them. Unfortunately, it is exactly those Anatolian languages that were spoken along the Aegean coast that are poorly known (at least Lydian and Carian),[2] unless the substrate extended into the Inner Anatolian regions. Note that the alleged Anatolian substrate of Greek is not considered here since this is not the object of the present study.
3) An Armenian-Anatolian substrate. Similar to the Greek-Anatolian substrate, the number of already known potential candidates makes this substrate interesting:
    a) unidentified local languages known, e.g., only from personal names such as those from Tušḫan (MacGinnis 2012);

2 Despite the frequent assertion, there is no evidence for Luwian as a spoken language in Western Anatolia, see first of all the discussion in Yakubovich (2010: 75–160), cf. also Simon (2018: 381, with refs.) for the subsequent discussion. At best, we can speculate on a third, Luwic language in Northwestern Anatolia (Simon 2017).

b) local non-Indo-European languages known only from onomastics (e.g., Kaškaean);
c) local non-Indo-European corpus languages (Hattian, Hurrian, Urartean);
d) local non-Indo-European languages only attested much later (the three Caucasian language families[3]).

It is of course a different question as to whether any of these could be connected to a substrate due to the obvious philological problems (in the first three cases) and the lack of generally accepted reconstructions (in the case of the Northeastern Caucasian languages). Nevertheless, this remains a future concern: by sifting through the evidence, I have not found any exclusively Anatolian-Armenian proposals,[4] and accordingly, I will not treat them separately.

4) The so-called Mediterranean substrate. In the secondary literature, several cases have been proposed as Mediterranean substrate in Anatolian words, which usually involved Latin cognates besides Greek ones.

3 The etymologies

This section provides a critical evaluation of the proposed substrate words in the Anatolian languages. It consists of Anatolian words of unknown, unclear, or disputed origin for which an Indo-European substrate connection has previously been proposed, either explicitly or by connecting them to substrate words in one or several other Indo-European languages. In other words, Anatolian words with established Indo-European or otherwise clear etymologies have not been taken into account, even if they have been attributed to substrates or labelled as *Wanderwörter* in the past (see the Appendix for a list). Two remarks are in order. First, this section is based on my sifting of the etymological literature, which inevitably means that some proposals have been overlooked despite my best intention. Second, next to "loanwords" and "substrate words", other labels have additionally been used in the secondary literature, notably "*Wanderwort*" and "culture word", sometimes within the same sentence, e.g., "a "Mediterranean" culture word (. . .) variously borrowed from a substrate" (HED M: 167). In other words, there was no

3 Despite the frequent assertions, none of these language families are demonstrably related to any of the above-mentioned Ancient Near Eastern languages: on Hattian see the critical overview in Simon (2012: 219–263), on Hurro-Urartean see, e.g., Smeets (1989).

4 There are known Anatolian-Armenian borrowings from common Ancient Near Eastern sources, but these are recent loanwords, not substrate words (cf. above).

clear distinction between these terms, and for this reason I have taken them all into account here[5] (see the concluding sections on each of the possible substrates regarding their separation).[6]

3.1 Non-areal substrates

Anatolian words have rarely been connected explicitly with previously proposed substrate words. However, there are several cases in which Anatolian words have been connected to words that should indeed be classified as substrate words, despite the previous interpretations of the relevant scholars.

3.1.1 Hitt. *alanza(n)-* '(a tree and its wood)': of unclear etymology (HED A: 30 and HEG A-K: 15, both with refs.; not included in EDHIL, implying a loanword)

PROP: According to Puhvel (1977: 598 and HED A: 30), the connection with Lat. *alnus*, Lith. *alk̃snis*, East Lith. *alìksnis* 'alder' (cf. also further Baltic forms [Lith. dial. *elk̃snis*]; Slavic **a/elisaH* [> Ru. *ol'xá* 'alder', dial. *ëlxa*, *elxá*, Bulg. *elxá* 'alder, spruce']; PGm. **aluz-* [> ON *ǫlr*, OE *alor*] and **alis/zō-* [> OHG *elira*, Du. *els*, Go. **alisa* > Sp. *aliso*] 'alder', see e.g., Friedrich 1970: 70–72 with refs. and possible Ancient Macedonian and Celtic forms) is the "most probable" one and "Hitt. has metathesized **al(i)sno-* to **alŋso-*" (alternatively, the Hittite word goes back to **aleno-* [> Lat. *alnus*]).

DISC: The Baltic, Germanic, and Latin words are nowadays held to be non-IE loanwords due to their *e*/*a*-variation in *Anlaut* and the different suffixes *-s-* and *-is-*.[7] Accordingly, Hitt. *alanza(n)-* would also reflect this substrate word. However, neither of the stems **alisn°* or **alsn°* fit the Hittite word and the assumed metathesis is *ad hoc* (cf. Melchert 2003a: 134). The alternative, to separate the Latin word and infer a Latin-Anatolian inherited isogloss is technically possible, but not compelling, especially in view of the meaning of the Hittite word not being fully known. Finally, the etymology proposed by Poetto (1973: 29) for the Hittite word, connecting it to Gr. ἐλάτη 'pine, fir', can now be classified as formally acceptable via **(h₁)elŋ̥tyo-* (on PIE **e* > Hitt. *a*/_RCC see EDHIL: 95, on **-Vntyo-* > Hitt. *-Vnza-* see Melchert 2003a:

5 Those cases in which the Anatolian languages were the source or both the Anatolian languages and the other Indo-European language(s) independently borrowed from an identified, non-substratal source (such as the above-mentioned Armenian and Anatolian borrowings from Ancient Near Eastern languages) are not included.

6 In the following etymological discussions, I do not intend to present the entire history of research (this can be found in the cited handbooks), but depart from the state of the art as reflected in the etymological handbooks.

7 EDSIL (371), EDBIL (50–51), EDLI (35); cf. also EIEC (11), Mallory – Adams (2006: 158), EDPG (22).

134–135). That said, while this form can be seen as a suffixed form by some scholars (e.g., by the present author), for other scholars it may imply a substrate origin based on its "unanalysable" morphology (see the methodological discussion in Section 1).

3.1.2 Car. γίσσα 'stone': of unknown etymology (Simon 2021 with refs.)

PROP: Georgiev 1966: 239 connected this word with G *Kies* (PGm. **kisa-*, cf. also **kisila-* 'gravel' as continued by OE *ciosol*, *ceosol*, *cisel* 'pebble', MDu. *kesel* 'id.', OHG *kisel* 'pebble') and Lith. *žizdras*, *žìgždras*, *žieždrà*, *žiezdrà*, *ziezdra* 'sand', OPr. *sixdo* 'sand' (EDPG: 289 and ALEW s.v. žieždrà).

DISC: These words are supposed to be loanwords in Germanic and Baltic due to their limited geography (cf. EDPG: 289, calling attention to similar Uralic and Kartvelian words), but the Carian word cannot formally be connected with them due to its initial voiced stop (for the details see Simon 2021).

3.1.3 Hitt. *kanen(iye/a)-* 'to bow down, to crouch, to squat': of disputed etymology (HEG A-K: 480–481, HED K: 42, and EDHIL: 434, all with refs.)

PROP: Puhvel (1981b: 352 and HED K: 42 ["most probably"]) identified it as yet another derivative of the PIE "root" **kneḭ-* extracted from **kneḭ-g^{wh}-* (Lat. *cō-nīveō* 'to close (the eyes)', Go. *hneiwan*, OE *hnīgan*, OHG *nīgan* 'to bend down, to bow') and **kneḭ-b-* (ON *hnípa* 'to be downcast', Lith. *knìbti* 'to collapse'), i.e., **kn(e)ḭ-n-* (for more precise forms see LIV^2: 365, 366: **Kneiǵ$^{\mu h}$-* '(sich) neigen' and ?**kneḭb^h-* 'hängen lassen, sinken lassen').

DISC: Some scholars identify the non-Anatolian verbs as substrate words (cautiously EDLI: 130 [about the first group due to its limited distribution and **T-D^H* structure] and EDHIL: 434 [the latter even arguing that this is exactly why "a connection with the Hittite verb is rather improbable"]), which would imply that Hittite shares this substrate. Nevertheless, any connection is possible only if the "suffixes" "*-g$^{\mu h}$-*", "*-b$^{(h)}$-*", and "*-n-*" can independently be confirmed, but this is not the case. Finally, EDHIL: 434 provides a plausible etymology, i.e. **ĝ-né-n-ti* from PIE **ĝen-* 'to bend' (whence **ĝen-u-* 'knee').

3.1.4 Hitt. and CLuw. *šapiyā(i)-* 'to cleanse': of unknown etymology (HEG L-S: 844–845 with refs. and HED Sa: 137; not included in EDHIL, implying a loanword[8])

PROP: According to Puhvel (HED Sa: 137), this verb is a cognate of OE *sāpe* 'soap', Lat. *sēbum* 'tallow', *sāpō* 'soap' (from Germanic) from PIE **sab-* (note that his

8 Rieken (2020) explains these words from Hitt. *ša/epiya-* and CLuw. *šapiya-* 'washbowl' (resp.), which, in turn, are explained by Sasseville and Rieken (2020) from Proto-Anatolian **seP-/*sP-ié/ó-* 'to scrub, to cleanse' (with Palaic *šapawina-* 'to scrub, to cleanse'), but no comparandum outside the Anatolian languages is provided.

Palaic "*sap*(*a*)-" "with inferential nuances of cleaning" is not connected: *šapāma*- means 'meal *vel sim.*' and *šapana*- is '(a substantive)', see Sasseville 2019b, 2019c).

DISC: Setting aside *sēbum* of unknown origin (EDLI: 550 with refs.), Latin *sāpō* cannot be a loan from Germanic for phonological reasons (cf. OHG *seifa*, *seipfa*, *seiffa*, MLG *sēpe*, MDu. *sepe*, OFri. *sēpe*, and OE *sāpe* 'soap' < **saip*°) and the Germanic and the Latin words can be explained in a phonologically regular way only if they originate in the local substrate identified by Schrijver, in which the correspondence Celtic *-a-* ~ Proto-Germanic *-ai-* is a recurring phonological anomaly (Simon 2020, cf. above). Whatever the case may be, neither the reconstructed substrate form (**saippwōn-*) nor the vocalism of the Latin and Germanic words in general can be reconciled with the Hittite word.

3.1.5 Hitt. *šumānzan*- '(bul)rush' (and perhaps CLuw. *šummanti*[. . .] of unknown meaning): of unknown etymology (EDHIL: 781, *contra* HEG L-S: 1151–1152 and HED SeSiSu: 158 [rejecting the meaning without arguments and thus, it is irrelevant])

PROP: G. Kroonen (pers. comm.) suggested a connection with PGm. **sem-eþa*/*ō*- 'rush' (> OS *semith*, MLG *sem*(*e*)*de*, OHG *semida*, etc.) and PC **sem-ino*- (?) (> OIr. *simin*, *sibin*(*n*)), on these forms see EDPG (432).

DISC: The Hittite word is formally ambiguous, allowing for the reconstruction of both **-m*- and **-w*-, but the Cuneiform Luwian word could decide the question, if it is indeed cognate (Melchert forthc. s.v.). Unfortunately, it is attested only fragmentarily (KBo 22.254 ii 3.11 and KBo 29.5, 6, Melchert forthc.) and although the proposed meaning is possible, the fragmentary contexts (which are, in fact, variations of a single one) do not allow any semantic assignment. In the case of **-m*-, we have a Hittite stem **sum*-, since *-anzan*- can regularly be explained as a suffix (Melchert 2003a). The main issue is whether this can be connected with **sem*-: while some scholars assume PIE **sm*- > Hitt. *šum*(*m*)-,[9] Kloekhorst (EDHIL: 783–785) provided convincing counterevidence and accordingly, these groups do not seem related. Moreover, it should be noted that for comparanda with a CVC structure the chances of chance resemblance are quite high (see Ringe 1992, 1999). Note that if the development of PIE **sm*- > Hitt. *šum*(*m*)- turns out to be correct, then we are dealing with a group of inherited words.

3.1.6 Hitt. *tašku*(*i*)- 'thigh-bone': of unknown etymology (HEG T,D: 255–256 and EDHIL: 852–853)

9 E.g., Kimball (1999: 199), Rieken (2000), Katz (2007), and Melchert (2016: 188–189).

PROP: This word has been compared to Gaul. *Tasco-* '(name element)', MIr. *Tadg* 'name of a king whose totem was a badger' (< **tazgo-*), PGm. **þahsan-* 'badger' (→ **þahsu-* [> dial. Nw. *toks*] and → **þahsa-* [> MHG *dahs*, etc.]; borrowed as VLat. *taxo*; EDPG: 531) as a *Wanderwort* under the old assumption that it means 'testicles' (Katz 1998, despite the phonological problems, EDPC: 372 assumed an inherited word).

DISC: As already Kloekhorst (2005: 37 with fn. 9, cf. also EDHIL: 852–853) pointed out (but overlooked in EDPC), Hitt. *tašku*(*i*)- means a body part between the pelvis and the shin-bone paired with feet, and thus, not testicles. Accordingly, this comparison must be given up (EDPG: 531).

We can conclude that the Anatolian words with alleged cognates in non-areal Indo-European substrate languages are either inherited or of unknown origin and in the latter case, the substrate connection cannot be maintained on phonological grounds.

3.2 Greek/Anatolian substrate

The numerous proposals are treated here in two groups. For the first group, I argue for recurring phonological features. The second group consists of the remaining cases, listed in alphabetical order.

3.2.1 Words with recurring phonological anomalies

(a) **Hitt. *-š-* : Gr. *-d-***

3.2.1.1 Hitt. *karšani*(*ya*)- 'sodaplant, soapwort': of unknown etymology (HEG A-K: 521 and HED K: 106–107, both with refs.; not included in EDHIL, implying a loanword)

PROP: Furnée (1972: 64 fn. 269) compared it with Gr. κάρδαμον / καρδάνη 'nose-smart / watercress', itself of unknown etymology (HEG A-K: 521: "vielleicht", not mentioned in HED; on the Greek word see GEW/I: 786–787, DELG: 497, and EDG: 643, all with refs.).

DISC: One of the Greek "endings" must be secondary and °*amon* could have easily been transformed by analogy to other plant names in °*amon* as in σήσαμον, κίννᾰμον, etc. (cf. Chantraine 1933: 133 and Schwyzer 1939: 494). This results in a phonologically compatible form, but the meaning is admittedly not identical. However, such semantic changes, i.e. the transfer of names among physically not necessarily very similar plants, are cross-linguistically ubiquitous (see e.g. the association of lentil and duckweed below, case no. 3.3.2).

3.2.1.2 Hitt. *paršna-* 'leopard': explained either from PIE **prs-no-* 'dappled, having spots' (Oettinger 1986: 22, followed by Melchert 1994: 175, EDHIL: 644, and HED Pa: 173 ["might be"]) or as a "*Wanderwort*" (compared to Hatt. *ḫa-prašš-un*, Turkmen *bars*, P *pārs*, Gr. πάρδαλις, πόρδαλις 'leopard, panther', etc., cf. HEG A-K: 478–479 with refs., HED Pa: 173, and Schrijver 2011: 247), a common Hittite-Greek substrate was suggested already by Gusmani (1968: 85).

DISC: The PIE etymology is formally simpler, but the loanword etymology is semantically more plausible. Needless to say, only the Greek and Hattian forms can be taken into account on geographical and chronological grounds. Greek and Hittite might have borrowed the word from a common source (**p(a)rC-*; perhaps even Hattian, too, which like the Greek variants would rather point to **prC-*).

In other words, one can argue that both cases originate from a third, common source with a phoneme unknown to both Hittite and Greek (e.g., [θ] or [δ]), which was substituted with [s] and [d], respectively.

(b) **Hitt. *nasal* + *labial stop* : Gr. *labial stop only***

3.2.1.3 Hitt. *impa-* (also *aimpa-*) 'weight, burden': of unknown etymology (HEG A-K: 6 and HED A: 15; not included in EDHIL, implying a loanword)

PROP: Furnée (1972: 271) compared the Hittite word to Gr. ἶπος 'weight, press' (cf. HED A: 15: "plausibly", "presumably (. . .) borrowed from some common source", but this suggestion was rejected by HEG A-K: 6, without a single argument: "ganz unwahrscheinlich").

3.2.1.4 Hitt. *kurimpa-* 'dregs, sediment': of unknown etymology (HEG A-K: 647 and HED K: 265; not included in EDHIL, implying a loanword)

PROP: Furnée (1972: 271) compared this word to Gr. κυρήβια 'husks, bran' (cf. HEG A-K: 647: "vielleicht"; HED K: 265: "possible"; but according to EDG: 806: "there seems little rea[s]on for this").

DISC: One can argue that both cases are borrowings from a third, common source with a phoneme unknown to both Hittite and Greek, presumably a nasalized vowel or a prenasalized consonant. At this juncture it is interesting to note that there is a language attested in the region between Greek and Hittite which seems to have prenasalized consonants, i.e. the Carian language. It would, however, be premature to identify the source with Carian. Setting aside the fact that these words are obviously not (yet) attested in Carian, there is considerable debate over the question of whether Carian actually had prenasalized consonants (Kloekhorst 2008: 138–139) or whether we are only dealing with an orthographical practice (Adiego 2019: 105). Besides, Carian prenasalized consonants were voiced, which is incompatible with the voiceless stop of Gr. ἶπος, although one could

argue that this word might have been borrowed in a phase preceding the voicing of the stops.

3.2.2 Proposals without recurring features

3.2.2.1 Hitt. *aršī-* 'plantation': of unknown etymology (HEG A-K: 68 and HED A: 174 with refs.; not included in EDHIL, implying a loanword)

PROP: The similarity with Gr. ἄρσεα 'meadows' (also of unknown etymology) has long been observed, but the Greek word has usually been interpreted as a Hittite loan (see the overview in Simon 2018: 384–385 with refs.). Simon (2018: 385) entertained the possibility of a common, third source.

DISC: The assumption of a common source would fit formally and would be understandable considering the geographical distance and the typology of Hittite loanwords in Greek (Simon 2018: 380–381). Nevertheless, it remains hard to prove. In fact, one may even wonder whether we might be dealing with inherited words here, i.e., by assuming that the Greek word is a non-Attic form. However, under this hypothesis, the geographically limited distribution is suspicious and a chance resemblance (note the different meanings) cannot be excluded.

3.2.2.2 Hitt. *ḫuḫupal-* 'lute or flapper, drum': of unknown etymology (HEG A-K: 273 and HED H: 359; not included in EDHIL implying a loanword)

PROP: According to HED (H: 359), "there is a definite culture-word relationship to (. . .) Gk. κύπελλον 'goblet', κύμβαλον 'cymbal' (. . .); cf. also Gk. κύμβη 'bowl'" (as well as to the formally similar Skt. *kumbhá-* m. 'a type of vessel' and Av. *xumba-* 'id.').

DISC: However, as discussed in Simon 2018: 396, 398 with refs., any connection must be excluded on obvious phonological and semantic grounds. The same applies to the proposals of HED H: 359, 387 to include Hitt. *ḫupallaš-* 'skull or scalp' and *kukupalla-* 'a vessel'.

3.2.2.3 Hitt. *iškiš-* 'back, backside, rear': of unknown etymology (HEG A-K: 401–402, HED E/I: 425, and EDHIL: 403, all with refs.)

PROP: EDHIL (403) claimed that *if* the usual comparanda, Gr. ἰσχίον 'hips', (Hes.) ἴσχι 'loins' (of unknown etymology, DELG: 472, GEW/I: 741, and EDG: 602–603), are related, then they are loans from a common, third source.

DISC: Since a borrowing is not possible in either direction on formal grounds, a borrowing from a common third source is indeed the only phonologically regular solution (the borrowing would then have to be a pre-Mycenaean one in the case of the Greek words due to the required sound change *-VsV- > -VhV- (> -VV-)).

Once more, however, the problem is that this borrowing is hard to prove.[10] Also in this case one may ask whether we are dealing with inherited words, although their geographically limited distribution is suspicious.

3.2.2.4 Hitt. *kurša-* 'skin bag': of unknown etymology (HEG A-K: 655–657 and HED K: 274–275, both with refs.; not included in EDHIL, implying a loanword)

PROP: The Hittite word is traditionally compared to Gr. βύρσα 'skin, hide', itself of unknown origin (GEW/I:278, DELG: 202, and EDG: 249), as a "culture word".

DISC: The phonological difficulties, including the correspondence of Hitt. *ku-* to Gr. β, can be solved only by the assumption of a Greek loanword from a Western Anatolian language that preserved voiced stops in initial position (Simon 2017, esp. 247 with refs.). Without this assumption, no phonologically regular connection of these words can be established.

3.2.2.5 Hitt. *laḫpa-* 'ivory': of unknown etymology (HEG L-S: 14–15 and HED L: 12–13, both with refs.; not included in EDHIL, implying a loanword)

PROP: This Hittite word is traditionally compared to Gr. ἐλέφᾱς, -αντ- 'ivory' as a *Wanderwort* (cf. Mallory & Adams 2006: 141, even though their alternative reconstruction, "'**lebh-* 'ivory'", fails to produce the Hittite word for obvious formal reasons).

DISC: Any connection between the proposed words is unacceptable due to the enormous and unexplained formal differences (cf. Simon 2018: 390). If Hitt. *laḫma-*, itself of unknown meaning, is indeed the same word (cf. CHD L-N: 12, supported by HED, but HEG remained sceptical), the spelling variation *-m-*/*-p-* in postconsonantal position would point to a Hattian origin (i.e., of the type *Ḫakp*/*miš* 'a settlement in the Hattian speaking region').

3.2.2.6 Lyc. B *laKra-* '(a ritual object or installation)': of unknown etymology (Melchert 2004: 119, Neumann 2007: 184, and Sasseville 2021b, all with refs.)

PROP: The Hittite word has long been compared to Gr. λάβρυς '(ritual) axe' (of unknown etymology, DELG 610–611, GEW/II: 67, and EDG: 819) as an Aegean substrate word (Neumann 2007: 184 with refs.).

DISC: The disputed phonetic value of ‹K› and the unclear meaning of the word prevent any solid etymological proposal, in spite of the numerous inner-Anatolian attempts (see the references given above). Note that if λάβρυς has anything to do with λαβύρινθος (an old, but heavily debated assumption, see Kroonen, this volume), then it goes back to a form with initial **d-* (cf. Valério 2017), which would be incompatible with the Lycian word.

10 This sound change involves Phrygian as well, which suggests that the substrate should be dated rather early. However, such an early date is difficult to falsify.

3.2.2.7 Hitt. *šulāi-*/*šūliya-* 'lead': of unknown etymology (HEG L-S: 1143–1144 and HED SeSiSu: 155–156, both with refs. [HED's (old) etymological connection with PIE **sle*/*oi̯H-* 'plum-coloured, blueish' is formally impossible, see already the refs. in HEG, ignored by HED], not included in EDHIL, implying a loanword)

PROP: It was suggested long ago that this Hittite word and Gr. σόλος 'iron mass' are common borrowings or substrate words (Gusmani 1968: 84; Egetmeyer 2010: 301–302).

DISC: Considering that a borrowing is not possible in either direction (Simon 2018: 404 *contra* HED SeSiSu: 156, accepting a Hittite loan in Greek, ignoring Simon's arguments), borrowing from a third source is definitely a possibility. However, this assumption does not provide a solution to the problem of the different first vowels: the [u] in the Hittite word is assured due to its *plene* spelling with ‹ú› (see the refs. in Simon 2018: 404). Potential solutions include the assumption of a common source language possessing a phoneme "between" [o] and [u] (interestingly enough, Hattian seems to be such a language, Simon 2012: 80–81) and the assumption of a borrowing mediated by a language without [u] or [o] (on typological grounds, the latter scenario seems more probable than the former; again, Hattian does not seem to have [o], Simon 2012: 80–81). Needless to say, however, no final decision can be made without further evidence.

All in all, setting aside the superficial comparisons, in most of the cases a common inheritance (or an Anatolian loan in Greek) cannot be excluded, except in the case of *šulāi-*/*šūliya-* 'lead' ~ σόλος 'iron mass', but no phonologically regular explanation can be given here that would prove a common origin.

3.3 The "Mediterranean" substrate

3.3.1 Hitt. *alel-*/*alil-* 'flower, bloom': of unknown etymology (HEG A-K: 16–17 and HED A: 33, both with refs.; not included in EDHIL, implying a loanword)[11]

PROP: According to a widespread view, we are dealing with a migratory culture word also seen in Brb. *alili* 'oleander', Bsq. *lili*, Alb. *lule* 'flower', Eg. *ḫrr.t* (Cpt. *hrēre*, *hlēli*), Gr. λείριον, Lat. *līlium* 'lily', etc. (cf. HEG and HED). The initial vowel of the Hittite form is explained as "prothetic *a*-" or "*a*-mobile" (Kronasser 1966: 324, followed by HEG). This view is rejected by HW^2: 59, albeit without argumentation. Schrijver (2017: 362) cautiously sees yet another exam-

11 The attempt of Sereni 1964: 20 to connect Hitt. *ulili-* 'greenery, vegetation; field' of unknown etymology to this word (and to its Mediterranean origins) is formally impossible, see already HEG U-Z: 39 with refs.

ple of a European substrate word with *a*-prefixation and subsequent syncope. (note that he reconstructs **a-leil-*, which does not fit his own theory, although, Hitt. *alel-/alil-* would exactly show the syncopated form that is expected within his framework).

DISC: The "prothetic *a-*" / "*a*-mobile" is an outdated concept: there is no such phenomenon in Hittite. Furthermore, it also implies that the Hittite form is secondary and the words with initial *l-* represent the primary form. This is, however, more than doubtful: the Greek and Latin words can alternatively be, and have indeed been, explained as borrowings from Egyptian (LEW/I: 801 with ref., *contra* Holton Pierce 1971: 105, no "obvious phonetic difficulties" exist, neither the semantics pose any problem) with loss of the foreign initial consonant in the initial consonant cluster. The Latin form, moreover, could have been the source of the other words, including the Berber word (REW: 409 with ref.). The Albanian word remains problematic in any of the theories due to its vocalism. As a result, Schrijver's analysis can be excluded as well. In other words, there is basically no comparandum for Hittite *alel-/alil-*.

3.3.2 Hitt. *ḫalenzu-* 'surface growth of stationary water (algae, duckweed, foliage, etc.)': of unknown etymology (HEG A-K: 128 and HED H: 20, both with refs.; not included in EDHIL, implying a loanword)

PROP: Pisani (1967: 403) connected this Hittite word to OHG *linsa*, Lat. *lēns* 'lentil', as well as OCS *lęšta*, as a culture word of unknown provenance (called an "unwarranted association" by HED H: 20).

DISC: Some scholars assume a Hattian origin for the Hittite word (Güterbock 1974: 309–310; HW^2 Ḫ: 27), which is formally convincing, since *ḫa=* is a typical nominal prefix in Hattian and *=u* can regularly be analysed as a Hattian morpheme (Soysal 2004: 217, 259). Furthermore, there is no semantic problem, since duckweed is cross-linguistically associated with lentils (e.g., G *Wasserlinse*, Hung. *békalencse* [lit. 'frog-lentil']). If the Hattian derivation of the Hittite word is correct, then the similarity of the European words with Hattian **lenz* would be remarkable (cf. §4 below).

3.3.3 Hitt. *ḫaš(š)ik(ka)-* '(a tree and its fruit, perhaps 'nut')': of unknown etymology (HEG A-K: 200–201, HED H: 230, HW^2 Ḫ: 423, cf. also EDHIL: 325)

PROP: Hoffner (1967: 43 fn. 58) connected *ḫa-š(š)ik(ka)-* 'a kind of fig' (as well as *maršigga-* and *šiggašigga-*) with Gr. σῦκον, τῦκον, and see also Lat. *fīcus*, Arm. *tᶜowz* 'fig' (which was classified by HED H: 230 as "at best conjectural").

DISC: The value of the Hittite comparanda is limited, since the meaning of *šiggašigga-* is unknown: it could be both a noun or an adjective (CHD Š: 359 with refs.) and while *maršikka-* is indeed a tree or its fruits (CHD L-N: 200), nothing more spe-

cific can be said about it. As a consequence, their etymologies remain unknown (HED M: 89 and HEG L-S: 147, 1035–1036, both with refs., not included in EDHIL). Moreover, there is an old Indo-European etymology for *ḫaš(š)ik(ka)-*:[12] a direct comparison with PGm. **aska-* 'ash' < **Hh₃esko-* and Alb. *ah* 'beech', Arm. *hacʿi* 'ash-tree' < **Hh₃esk-*. However, this proposal can be rejected, since **-sk-* does not acquire an epenthetic vowel in Hittite in this position (EDHIL: 60–61). A comparison with Baltic (**Heh₃-s-io-* > Lith. *úosis* 'ash-tree', etc.), Slavic (**Heh₃-s-en-* > SCr. *jȁsēn*, etc. 'ash-tree'), Latin (*ornus* 'a kind of ash-tree'), and Celtic (**Hh₃-es-no-* > MIr. *onn* 'pine tree', etc.) forms might be considered, but the derivation remains unparalleled and, more importantly, the spelling variations remain unexplained.

The variation in the spelling of the sibilant and the stop points to a loanword. The same variation of the stop is typical for adapted Hattian toponyms (of the type *Nerik(ka)*) and *ḫa=* is a typical nominal prefix in Hattian (Soysal 2004: 217). In conclusion, we are probably dealing with a Hattian loan in Hittite. Again, the similarity of Hatt. **sik* to Gr. σῦκον, τῦκον is striking, but we may be victims of the statistical probability of chance resemblance of words with a CVC-structure (cf. case no. 3.1.5).

3.3.4 Hitt. *kikri-* 'byword of BA.BA.ZA 'mash': of unknown etymology (HEG A-K: 570 and HED K: 175; not included in EDHIL, implying a loanword)

PROP: Neumann *apud* HEG compared this word to Lat. *cicer*, Arm. *siserṅ* 'chick-pea' and Mac. κίκερροι 'pale as the name of a sallow-coloured leguminous plant, Cyprus-vetch'.

DISC: No common proto-form can explain all these words, neither through a substrate nor as common inheritance, on phonological grounds, not to mention that we do not actually know the meaning of the Hittite word. Formally speaking it looks like an internal deverbal derivation in *-ri-*, but the derivational base remains obscure.

3.3.5 Hitt. *kištu-* '(support-)stand, rack, tray, shelf': of unknown etymology (HEG A-K: 594 and HED K: 200, both with refs.; not included in EDHIL, implying a loanword)

PROP: This Hittite word is frequently compared to Gr. κίστη 'basket, hamper', to which HED (K: 200) added Lat. *cista* 'chest, box', with the remark that "the semantics are only approximate (. . .) but perhaps tolerably close for an international culture word".

12 See, e.g., the ref. in HEG (A-K: 201), EDAIL (398), cf. also EIEC (32), Mallory and Adams (2006: 158–159).

DISC: The Latin word is generally explained as a borrowing from Greek,[13] although this is not necessary, and the Greek word is connected to Proto-Celtic **kistā* '(woven) basket' (MIr. *cess* 'basket, causeway of wickerwork, beehive' and OW *cest* 'fiscina').[14] The Greek-Celtic isogloss was, in turn, explained either as inherited (implicitly LEW/I: 223) or as a European substrate word (EDG: 705 and EDPC: 204–205; the Greek word is perhaps a borrowing according to DELG: 536). However, no European or Mediterranean substrate is needed: both the Hittite word and the Greco-Celtic forms can be explained either as internal derivations from PIE **ḱeis-* 'übrig lassen' (LIV^2: 321; in fact a back-projection of a Vedic verb), in which case the Hittite word would be a verbal abstract in *-tu-* with a zero-grade root (type Lat. *portus* 'harbour, refuge; door' from PIE **per-* 'hindurchkommen, durchqueeren' [LIV^2: 427–473]), or as derivatives from PIE **keis-*, a back-projection of Lith. *kišti* 'stecken' (ALEW s.v.), which was proposed for the Latin word already by Fick (1878: 266) (but rejected by LEW/I: 223 [without arguments] and Fraenkel 1962–1965: 260 [following an obsolete interpretation of the Baltic *ruki*-rule]).

3.3.6 Hitt. *mit(t)a-*, *miti-* 'red': of unknown origin (HED M: 167, HEG L-S 218–219, and EDHIL: 583, all with refs.)

PROP: According to Puhvel (1981a: 238, HED M: 167), we are "perhaps" dealing with a "Mediterranean" culture word" (cf. Gr. μίλτος 'ruddle, red earth, red color, etc.', Myc. *mi-to-we-sa* and Lat. *minium* 'red ochre'), "with *d* : *l* : *n* variation", "variously borrowed from a substrate, with discrepant phonetics (dental : lateral variation, stop and nasal confusion as in Etruscan)".

DISC: Formally speaking, these words have nothing in common with each other except the first two phonemes. The purported *d*/*l*/*n*- and dental/lateral-variations as well as the stop and nasal confusion do not exist in the quoted languages and they do not explain the forms.

3.3.7 Hitt. *:pulpuli*[(-). . .] 'beam or log?': of unknown etymology (HEG L-S: 646 and HED PePu: 118–119 ["phonesthetic reduplicates", whatever this should mean]; not included in EDHIL, implying a loanword)

PROP: HEG cautiously compared it with Lat. *pōpulus* 'poplar' ("Wanderwortcharakter nicht ausgeschlossen").

DISC: Setting aside that neither the form nor the meaning of the Hittite word are fully ascertained, these words do not correspond phonologically.

---

13 E.g., DELL: 220, LEW/I: 223, GEW/I: 860, and EDG: 705; EDLI does not even include it.

14 See e.g., LEW/I: 223, GEW/I: 860 (cautiously), EDG: 705, and EDPC: 204–205.

3.3.8 Hitt. ^{TU7}*ša*(*m*/*n*)*pukki*- '(a type of stew or soup)': of unknown origin (HEG L-S: 802 and HED Sa: 106–107; not included in EDHIL implying a loanword)

PROP: Following Knobloch (1955: 8–9), Puhvel (HED Sa: 106–107) connected it as "a pot-dish (ingredient)" and "typical culinary culture word" to Lat. *sa*(*m*)*būcus* 'elder-tree' and Gr. σάμψ(ο)υχον 'marjoram' (Knobloch and HEG L-S: 802 add Dacian σέβα and Gaul. σκοβιή 'elder(-tree/berry)' and prefer the Greek connection from a culinary point of view).

DISC: The motivation behind the name of the soup is unknown, but an ingredient is of course possible, although it must be noted that there is no hint of the soup's ingredients. The Greek, Dacian, and Gaulish words do not correspond formally, but the almost perfect identity with the Latin word is remarkable and cannot be attributed to chance due to their length. It might be debated whether elderberries are (or can be) used for soups, but they are definitely used for refreshing or alcoholic drinks, and since we are dealing with ritual texts, the word may refer to a potion. All in all, a connection is quite probable.

3.3.9 Luw. **tapa*(*ra*)- *vel sim.* 'hare': of unknown etymology

PROP: Katz (2001: 215–221, cf. already Kronasser 1966: 63–64) argued that this Luwian word together with Lat. *lepus* 'hare' and Arm. *napastak* 'hare' represent a *Wanderwort* **TaPVs*- 'hare' (not mentioned in EDLI: 335)

DISC: First, the Armenian word must be separated from this group on obvious formal grounds (for etymological proposals see Katz 2001: 217 with refs.). Second, the form of the Luwian word is unclear: it has long been reconstructed based on the observation that the Luwian word /tabariya-/ 'authority' and related words (all originating in a stem *tapar*-) can be written with the logogram LEPUS (a sign showing a hare, see the refs. already in Laroche 1960: 73–74). While this would indeed allow a reading **tapar(a)*-, the sign LEPUS is always followed by (at least) ‹+ra/i› (and further phonetic signs). This shows that without its presence the rebus spelling would not have been clear and thus, ‹+ra/i› is not the part of the stem. Accordingly, a reconstruction */taba-/ is more probable (cf. Katz 2001: 215 with fn. 28, see in general Katz 2001 for an attempt to identify this word in Hitt. *tapakaliya*-). This suffices to separate the Luwian word from the Latin (as well as from the Armenian). In fact, it can be separated already on the grounds that the Latin word has Western Mediterranean comparanda, cf. Siculian λέποριν (acc.), Turdetanian (Tartessian) and Massiliote λεβηρίς 'rabbit'. Even if their precise relationship is unclear, they cannot be combined phonologically with */taba(ra)-/.

Thus, one can conclude that most of the words suggested to belong to a "Mediterranean" substrate in Anatolian turn out to be superficial look-alikes (if they are look-alikes at all) or even just inherited words. The only possible exception is the

case of Hitt. ^{TU7}*ša*(*m*/*n*)*pukki*- '(a type of stew or soup)' and Lat. *sa*(*m*)*būcus* 'elder-tree' (if the vaguely known semantics do not mislead us). A common substrate including Italy and Central Anatolia is hardly possible and inferring such a phenomenon on the basis of a single comparison is definitely not advisable. An indirect loanword is more probable, with different possible routes: mediation by Greek is certainly required and historically entirely possible. The source of the Greek word could have been the Hittite word, directly (borrowing in the second millennium) or indirectly, via a Luwian (in the first millennium) or Western Anatolian borrowing (in both millennia). Finally, a common source, perhaps an Ancient Near Eastern language, cannot be excluded either.

4 Instead of conclusions

Although there is no lack of proposals, the number for convincing candidates for a substrate shared by the Anatolian languages with any other branch of the Indo-European languages is comparatively small. This is hardly surprising in view of the state of the art of Anatolian linguistics and of the circumstance that the Anatolian languages were the first branch to separate from the other Indo-European languages, allowing for no contact, according to our current knowledge, except with the local languages of Anatolia. The distribution of these candidates follows the expected patterns, but in addition, some remarkable observations can be made.

1. Non-areal substrates cannot be demonstrated at the moment. This fits with our knowledge of prehistory, but as discussed in the introduction, this observation cannot be used as an argument in Indo-European prehistoric research on methodological grounds.
2. An exclusive Armenian-Anatolian substrate does not seem to exist, either. Although this can be explained by historical and geographical factors, I personally find it slightly surprising in view of the high number of non-Indo-European languages, i.e. potential sources, in this area. It also seems surprising in view of the famously high number of Armenian words from unidentified substrate(s).[15]

15 The Armenian-speaking territory was outside of the territory of the Anatolian languages and any of the known Ancient Near Eastern contacts of Armenian are with languages and language families that are late intruders in Anatolia. Thus, no common substrate is expected. Hurrian, Urartian, and Akkadian definitely belong to this group, while it is unknown where the forerunners of the Kartvelian languages should be located in the 3rd and 2nd millennia B.C.: this may have been both within and outside of Anatolia. Nevertheless, it is exactly this unknown substrate layer of Armenian that militates against a simplified, historical explanation of the lack of evidence for this common substrate.

I would rather attribute this to the limits of the research outlined in the introduction, which may change in the future as research advances.

3. The Greek-Anatolian substrate is unique from that point of view that some cases may be proven. This requires further research, although the limited attestation of the Western Anatolian languages will seriously impede any such endeavour unless the situation changes dramatically.
4. The "Mediterranean substrate" can also provide at least one isolated case. However, here we have the problem of the geographical and chronological distance that would rather argue for a migratory loanword instead of a substrate.
5. Finally, some side-effects must be mentioned: while new arguments could have been presented for the Indo-European origin of some Hittite, Greek, and Latin words and for the Hattian origin of some Hittite words, two of these Hattian words show an unexpected but remarkable closeness with European substrate words. Only further research can decide whether this is a mere *fata morgana* or one more piece of evidence for the pre-Indo-European substrate languages in Europe: the formal differences and the short form in the case of **sik* rather argue for the former, but **lenz* provides a formally fitting parallel.[16]

Appendix: List of words excluded from present investigation

(1) Hitt. *amiyara-* 'ditch, canal, channel' is traditionally compared to Gr. ἀμάρα 'trench, channel' as a "Kultur- und Wanderwort" and common substrate words, but an Indo-European etymology is entirely possible, see the discussion in Simon (2018: 383–384 with detailed refs).

(2) Lyc. B *arm̃pa-* 'pillar' was compared to Lat. *rumpus* 'a horizontally trained vine-shoot' (Sereni 1964: 196; in his unfinished work Neumann 2007: 20 called it a "bloßer Anklang" without arguments), but it is probably a loan from Hitt. ^{GIŠ}*arimpa-* 'support, column, pillar', itself an Ancient Near Eastern loanword (see most recently Sasseville 2019a with refs.).

16 Hattian was proposed to be a substrate language in Europe (Schrijver 2011, 2014, 2017), but not convincingly (see the criticism in Simon 2012: 260–262), at least until now. Cf. also the Introduction (this volume), §3.1 with refs.

(3) Hitt. *ateš-* (secondary *atešša-*) 'adze, axe, hatchet': HED (A: 228) treats it as an international culture word (also EDHIL implies a loanword), purely due to its meaning (!) and because its nominative was not "accommodated" to "normal *s*-stem neuters like *nepis*" (!). However, it regularly corresponds to PGm. **adesan-* 'adze' (hence PIE **Hod^h-es-*, EDPG: 2, cf. also HEG A-K: 94, both with refs.).

(4) Lyc. A *epri-*$^{(ti)}$ of disputed meaning (plausibly 'to come (as backup) *vel sim.*' with Sasseville 2021a: 156–157) was compared to Gr. πρύτανις 'title of a leading official' as well as Etr. *purθne* of unknown meaning and *eprθnevc* 'title of an official or an institution' (on the semantics of the Etruscan words see Steinbauer 1999: 417, 456), which was classified by Neumann in his unfinished work as "kaum zutreffend" without arguments (2007: 64 with refs.). No explanation was provided for the formal and semantic differences and if Sasseville's interpretation is correct, it is an internal derivation from Lyc. A *epre*/*i-* 'back' (Sasseville 2021a: 156).

(5) Hitt. *kakkapa-* 'a small animal, object of hunting' is variously treated as the source of Gr. κακκάβη 'partridge' or both as traces of a common substrate. However, we are obviously dealing with an onomatopoetic word, not to mention that the comparison is hampered by our lacking understanding of the meaning of the Hittite word (see in detail Simon 2018: 393 with refs.).

(6) Hitt. *kant-* 'wheat' can regularly be connected with ToB *kanti* '±bread' and Gr. χόνδρος 'grain, seed, barley-grain' from PIE **g^{h}ond-* (Adams 2013: 146). Thus, there is no need for Puhvel's interpretation as a "culture term" (HED K: 56, note that *contra* his claim, a "Luwian *kant-*" does not survive in Lycian toponyms, not only because Lycian is a sister-language of Luwian but also for phonological reasons and because those toponyms have a different etymology, see now Pisaniello 2021).

(7) Hitt. *kunkumāti-* (a garden plant) and Luw. *kunkumā-* have previously been compared to Lat. *cucumis* 'cucumber' and Gr. (Hes.) κύκυον 'id.' (HED K: 251 ["probably culture word"], cf. also HEG A-K: 634–635 ["sicherlich (. . .) Wanderwort"], both with refs.; not included in EDHIL, implying a loanword). However, Rizza (2012) convincingly argued that we are dealing with words meaning 'crocus, saffron', which are rather connected with Vedic *kuṅkuma-* 'id.'.

(8) Hitt. *šēli-* 'grain pile, grain storage', falsely compared to several words in Vedic, Armenian, Greek, Romance, Old Saxon (HED SeSiSu: 15), has a clear Indo-European etymology (EDHIL: 744 with refs.). The absence of the agricultural sense in the cognate Hittite verb is not a problem (*contra* HED, see EDHIL: 695, ignored by HED).

(9) Hitt. *tabarna-* 'royal title' and the numerous derivatives of Luw. ***/tabar-/ 'powerful' have received a semantically and formally fitting etymology (see,

e.g., Melchert 2003b: 18–19, despite repeated protests, see HEG T, D: 118 and EDHIL: 830–831). Thus, there is no need for the repeated and phonologically arbitrary connection to λαβύρινθος (most recently Yakubovich 2002 with refs.).

(10) Luw. *targaš*(*ša*)*na-* 'horse (or donkey)' has been erroneously compared to Lat. *asinus* (see the critical discussion with refs. in Simon 2017 [2019] 328 fn. 58). For plausible PIE etymologies see Opfermann (2017), who starts from PIE **d*^{h}*reg*^{h}- 'to drag, haul'.

(11) Hitt. *tēta*(*n*)- 'breast, teat', CLuw. *tītan-* 'id.', *titaimma*/*i-* 'suckling', Lyc. *tideime*/*i-* 'son, child', *tidere*/*i-* 'collacteus' has been derived from the "Mediterranean substrate" (Gr. τίθη, Lat. *titta* 'breast', see the refs. in HEG T: 345; for PIE origin see EDHIL: 875–877, they might simply be nursery words).

(12) Hitt. ^{DUG}*urā-* '(a vessel)', which may simply mean 'big (sc. vessel)' (cf. $^{(DUG)}$ GAL) as Luwianism (see the refs. in HEG U-Z: 86, HEG's counterarguments are not compelling). The "Mediterranean" connection (Gr. ὕρχη, Lat. *orca*, *urceus*, *urna*, and now Car. *ork* 'a type of vessel, phiale, bowl, cup') should be excluded, since all these words go back to the Greek form (on the Carian word see Simon 2019: 302–304), and the Latin words probably to Greek via Etruscan (cf. LEW/II: 838–839). Whatever their origin is, all are formally incompatible with the Hittite word.

(13) Hitt. *wiyan-* 'wine', HLuw. *wiyan*(*i*)-, *win*(*i*)- 'vine', CLuw. *winiya-* 'of wine' is often taken from the "Mediterranean" substrate (see the refs. in HEG U-Z: 561–562, on the PIE origin see EDHIL: 1012 with refs.).

(14) Finally, Hitt. **wiya-* 'woman, wife', also attributed to the "Mediterranean" substrate (see the refs. in HEG U-Z: 555), should be mentioned here since it in fact does not exist (see most recently Yakubovich 2013). Similarly, the comparison of Hitt. **kikluba-* / Luw. **kikliba-* 'iron' with Gr. χάλυβος 'hardened iron, steel' in HED K: 175 can be rejected, since the latter rather originates in an ethnonym.

References

Adams, Douglas Q. 2013. *A dictionary of Tocharian B. Revised and Greatly Enlarged*. Amsterdam/New York: Rodopi.

Adiego, Ignasi-Xavier. 2019. Consonant Clusters, Defective Notation of Vowels and Syllable Structure in Caromemphite. In Ronald I. Kim, Jana Mynářová & Peter Pavúk (eds.), *Hrozný and Hittite: The First Hundred Years. Proceedings of the International Conference Held at Charles University, Prague, 11–14 November 2015*, 95–119. (Culture and History of the Ancien Near East 107). Leiden/Boston: Brill.

ALEW = Hock, Wolfgang (ed.). 2020–2023. *Altlitauisches etymologisches Wörterbuch*. Version 2.0. https://alew.hu-berlin.de/ (last accessed 18 September 2023)
Chantraine, Pierre. 1933. *La formation des noms en grec ancien*. Paris: Champion.
CHD = Güterbock, Hans G.,Harry A. Hoffner, Theo P. J. van den Hout & Petra Goedegebuure (eds.). 1980–. *The Hittite Dictionary of the Oriental Institute of the University of Chicago*. Chicago: The Oriental Institute of the University of Chicago.
DELG = Chantraine, Pierre. 1968–1980. *Dictionnaire étymologique de la langue grec. Histoire des mots*. Paris: Klincksieck.
DELL = Ernout, Alfred &Antoine Meillet. 1951. *Dictionnaire étymologique de la langue latine*. *Histoire des mots*, 3rd edn. Paris: Klincksieck.
EDAIL = Martirosyan, Hrach K. 2010. *Etymological Dictionary of the Armenian Inherited Lexicon*. Leiden/Boston: Brill.
EDBIL = Derksen, Rick. 2015. *Etymological dictionary of the Baltic inherited lexicon*. Leiden/Boston: Brill.
EDG = Beekes, Robert. 2010. *Etymological dictionary of Greek*. Leiden/Boston: Brill.
EDHIL = Kloekhorst, Alwin. 2008. *Etymological dictionary of the Hittite inherited lexicon*. Leiden/Boston: Brill.
eDiAna = Hackstein, Olav,Jared L. Miller, Elisabeth Rieken & Ilya Yakubovich (eds.). 2016–. *Digital Philological-Etymological Dictionary of the Minor Ancient Anatolian Corpus Languages*. München/Marburg. https://www.ediana.gwi.uni-muenchen.de/ (last accessed 18 September 2023).
EDLI = de Vaan, Michiel. 2008. *Etymological Dictionary of Latin and the Other Italic Languages*. Leiden/Boston: Brill.
EDPC = Matasović, Ranko. 2009. *Etymological Dictionary of Proto-Celtic*. Leiden/Boston: Brill.
EDPG = Kroonen, Guus. 2013. *Etymological Dictionary of Proto-Germanic*. Leiden/Boston: Brill.
EDSIL = Derksen, Rick. 2008. *Etymological Dictionary of the Slavic Inherited Lexicon*. Leiden/Boston : Brill.
Egetmeyer, Markus. 2010. *Le dialecte grec ancien de Chypre*. Berlin/New York: De Gruyter.
EIEC = Mallory, James P. & Douglas Q. Adams (eds.). 1997. *Encyclopedia of Indo-European culture*. London/Chicago: Fitzroy Dearborn.
Fick, August. 1878. Allerlei. *Beiträge zur Kunde der indogermanischen Sprachen* 2. 264–268.
Fraenkel, Ernst. 1962–1965. *Litauisches etymologisches Wörterbuch*. Heidelberg/Göttingen: Winter/Vandenhoeck & Ruprecht.
Friedrich, Paul. 1970. *Proto-Indo-European Trees. The Arboreal System of a Prehistoric People*. Chicago/London: University of Chicago.
Furnée, Edzard J. 1972. *Die wichtigsten konsonantischen Erscheinungen des Vorgriechischen. Mit einem Appendix über den Vokalismus*. The Hague/Paris: Mouton.
Georgiev, Vladimir I. 1966. *Introduzione alla storia delle lingue indeuropee*. (Incunabula Graeca 9). 2nd edn. Roma: Ateneo.
GEW = Frisk, Hjalmar. 1960–1972. *Griechisches etymologisches Wörterbuch 1–3*. Heidelberg: Winter.
Gusmani, Roberto. 1968. *Il lessico ittito*. Napoli: Libreria Scientifica.
Güterbock, Hans Gustav. 1974. The Hittite Palace. In Paul Garelli (ed.). *Le palais et la royauté* (*Archéologie et civilisation*). *XIX^e Rencontre Assyriologique Internationale*. 305–321. Paris: Geuthner.
HED = Puhvel, Jaan. 1984–. *Hittite Etymological Dictionary*. Berlin/New York: De Gruyter Mouton.
HEG = Tischler, Johann. 1977–2016. *Hethitisches etymologisches Glossar I-IV*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
Hoffner, Harry A. 1967. An English-Hittite Glossary. *Revue hittite et asianique* 25. 5–99.
Holton Pierce, Richard. 1971. Egyptian loan-words in Ancient Greek. *Symbolae Osloenses* 46. 96–107.

HW^2 = Friedrich, Johannes, Annelies Kammenhuber, Inge Hoffmann, Joost Hazenbos, Albertine Hagenbuchner-Dresel, Paola Cotticelli-Kurras & Federico Giusfredi (eds.). 1975–. *Hethitisches Wörterbuch*. Zweite, völlig neubearbeitete Auflage auf der Grundlage der edierten hethitischen Texte. Heidelberg: Winter.

Katz, Joshua T. 1998. Hittite *tašku-* and the Indo-European Word for 'Badger'. *Historische Sprachforschung* 111. 61–82.

Katz, Joshua T. 2001. Hittite *ta-pa-ka-li-ya-‹aš›*. In Onofrio Carruba & Wolfgang Meid (eds.). *Anatolisch und Indogermanisch. Anatolico e indoeuropeo. Akten des Kolloquiums der Indogermanischen Gesellschaft*, *Pavia*, *22–25. September 1998*, 205–237. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Katz, Joshua T. 2007. The Development of Proto-Indo-European **sm* in Hittite. In Alan J. Nussbaum (ed.). *Verba Docenti. Studies in historical and Indo-European linguistics presented to Jay H. Jasanoff by students*, *colleagues and friends*, 169–184. Ann Arbor/New York: Beech Stave.

Kimball, Sara E. 1999. *Hittite Historical Phonology*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Kloekhorst, Alwin. 2005. Hittite *ḫāpūša*(*šš*)- (formerly known as *ḫapuš-* 'penis'). *The Journal of Indo-European Studies* 33. 27–39.

Kloekhorst, Alwin. 2008. Studies in Lycian and Carian Phonology and Morphology. *Kadmos* 47. 117–146.

Knobloch, Johann. 1955. Holunder im Hethitischen? *PHMA* 1. 5–10.

Kronasser, Heinz. 1966: *Etymologie der hethitischen Sprache 1*. Wiesbaden: Harrassowitz.

LEW = Walde, Alois. 1938–1954. *Lateinisches etymologisches Wörterbuch*. 3. neubearbeitete Auflage von J. B. Hofmann. Heidelberg: Winter.

Laroche, Emmanuel. 1960. *Les hiéroglyphes hittites I*. *L'écriture*. Paris: Klincksieck.

LIV^2 = Rix, Helmuth (ed.). 2001. *Lexikon der indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen*. 2nd edn. Wiesbaden: Reichert.

Lubotsky, Alexander. 2001. The Indo-Iranian Substratum. In Christian Carpelan, Asko Parpola & Petteri Koskikallio (eds.). *Early Contacts between Uralic and Indo-European. Linguistic and Archaeological Considerations. Papers presented at an international symposium held at the Tvärminne Research Station of the University of Helsinki. 8–10 January*, *1999*, 301–317. Helsinki: Suomalais-Ugrilainen Seura.

MacGinnis, John. 2012. Evidence for a Peripheral Language in a Neo-Assyrian Tablet from the Governor's Palace in Tušḫan. *Journal of Near Eastern Studies* 71. 13–20.

Mallory, James P. & Douglas Q. Adams. 2006. *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*. Oxford: Oxford University Press.

Melchert, H. Craig. 1994. *Anatolian Historical Phonology*. Amsterdam/Atlanta: Rodopi.

Melchert, H. Craig. 2003a. Hittite Nominal Stems in *-anzan-*. In Eva Tichy, Dagmar S. Wodtko & Britta Irslinger (eds.). *Indogermanisches Nomen. Derivation*, *Flexion und Ablaut. Akten der Arbeitstagung der Indogermanischen Gesellschaft. Freiburg*, *19. bis 22. September 2001*, 129–139. Bremen: Hempel.

Melchert, H. Craig. 2003b. Prehistory. In id. (ed.). *The Luwians*, 8–26. Leiden/Boston: Brill.

Melchert, H. Craig. 2004. *A Dictionary of the Lycian Language*. Ann Arbor/New York: Beech Stave.

Melchert, H. Craig. 2016. Initial **sp-* in Hittite and *šip*(*p*)*and-* 'to libate'. *Journal of Linguistic Relationship* 14. 187–196.

Melchert, H. Craig. forthcoming. *A Dictionary of Cuneiform Luvian*. Ann Arbor/New York: Beech Stave.

Neumann, Günter. 2007. *Glossar des Lykischen*. *Überarbeitet und zum Druck gebracht von Johann Tischler*. Wiesbaden: Harrassowitz.

Oettinger, Norbert. 1986. "*Indo-Hittite*"*-Hypothese und Wortbildung*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Opfermann, Andreas (2017): Common Luwian /targasna-/. Proto-Indo-European Etymology. In Hackstein, Olav, Jared L. Miller, Elisabeth Rieken & Ilya Yakubovich (eds.). *Digital Philological-Etymological Dictionary of the Minor Ancient Anatolian Corpus Languages*. München/Marburg. http://www.ediana.gwi.uni-muenchen.de/dictionary.php?lemma=319

Pisani, Vittore. 1967. Review of *HW. 3. Ergänzungsheft*. *Paideia* 22. 402–404.

Pisaniello, Valerio. 2021. The Lycian toponym Κάδρεμα and the Anatolian wheat. In Federico Giusfredi & Zsolt Simon (eds.). *Studies in the languages and language contact in Pre-Hellenistic Anatolia*, 67–80. Barcelona: Universitat de Barcelona.

Poetto, Massimo. 1973. Su alcuni termini botanici etei. *Rendiconti dell'Istituto Lombardo. Accademia di Scienze e Lettere. Classe di Lettere e Scienze Morali e Storiche* 107. 25–32.

Puhvel, Jaan. 1977. Review of J. Friedrich & A. Kammenhuber: *Hethitisches Wörterbuch. Zweite Auflage. Lieferung 1. Journal of the American Oriental Society* 97. 596–599.

Puhvel, Jaan. 1981a. "Spider" and "Mole" in Hittite. In Yoël L. Arbeitman & Allan R. Bomhard (eds.). *Bono homini donum. Essays in Historical Linguistics in Memory of J. Alexander Kerns*, 237–242. Amsterdam/Philadelphia: Benjamins.

Puhvel, Jaan. 1981b. Review of *HEG. Lieferung 3*. *Bibliotheca Orientalis* 38. cols. 349–354.

REW = Meyer-Lübke, Wilhelm. 1935. *Romanisches etymologisches Wörterbuch*. 3rd edn. Heidelberg: Winter.

Rieken, Elisabeth. 2000. Hethitisch *šumumaḫḫ-*. *Historische Sprachforschung* 113. 171–175.

Rieken, Elisabeth. 2020. Cuneiform Luwian *šapiyā(i)-*$^{(di)}$ 'to scrub, rub, cleanse'. In Hackstein, Olav, Jared L. Miller, Elisabeth Rieken & Ilya Yakubovich (eds.). *Digital Philological-Etymological Dictionary of the Minor Ancient Anatolian Corpus Languages*. München/Marburg. http://www.ediana.gwi.uni-muenchen.de/dictionary.php?lemma=1288

Ringe, Donald A. 1992. On Calculating the Factor of Chance in Language Comparison. *Transactions of the American Philosophical Society* 82. 1–110.

Ringe, Donald A. 1999. How hard is it to match CVC-roots? *Transactions of the Philological Society* 97. 213–244.

Rizza, Alfredo. 2012. Polveri di parole – polveri di spezie. Sulle tracce dello zaffrano nell'area indo-mediterranea antica. In Paola Cotticelli Kurras, Mauro Giorgieri, Clelia Mora & id. (eds.). *Interferenze linguistiche e contatti culturali in Anatolia tra II e I millennio A.C. Studi in onore di Onofrio Carruba in occasione del suo 80° compleanno*, 231–265 (Studia Mediterranea 24). Genova: Italian University Press.

Sasseville, David. 2019a. Lycian B *arm̃pa-* 'pillar'. In Hackstein, Olav, Jared L. Miller, Elisabeth Rieken & Ilya Yakubovich (eds.). *Digital Philological-Etymological Dictionary of the Minor Ancient Anatolian Corpus Languages*. München/Marburg. http://www.ediana.gwi.uni-muenchen.de/dictionary.php?lemma=499

Sasseville, David. 2019b. Palaic *šapāma-* 'meal (*vel sim*.)'. In Hackstein, Olav, Jared L. Miller, Elisabeth Rieken & Ilya Yakubovich (eds.). *Digital Philological-Etymological Dictionary of the Minor Ancient Anatolian Corpus Languages*. München/Marburg. http://www.ediana.gwi.uni-muenchen.de/dictionary.php?lemma=646

Sasseville, David. 2019c. Palaic *šapana-* '(unknown)'. In Hackstein, Olav, Jared L. Miller, Elisabeth Rieken & Ilya Yakubovich (eds.). *Digital Philological-Etymological Dictionary of the Minor Ancient Anatolian Corpus Languages*. München/Marburg. http://www.ediana.gwi.uni-muenchen.de/dictionary.php?lemma=451

Sasseville, David. 2021a. *Anatolian Verbal Stem Formation*. *Luwian*, *Lycian and Lydian*. Leiden/Boston: Brill.

Sasseville, David. 2021b. Lycian B *lacra-* '(unknown)'. In Hackstein, Olav, Jared L. Miller, Elisabeth Rieken & Ilya Yakubovich (eds.). 2016–. *Digital Philological-Etymological Dictionary of the Minor Ancient Anatolian Corpus Languages*. München/Marburg. https://www.ediana.gwi.uni-muenchen.de/dictionary.php?lemma=1493

Sasseville, David & Elisabeth Rieken. 2020. Proto-Anatolian **seP-/*sP-yé/ó-* 'to scrub, to cleanse'. In Hackstein, Olav, Jared L. Miller, Elisabeth Rieken & Ilya Yakubovich (eds.). 2016–. *Digital Philological-Etymological Dictionary of the Minor Ancient Anatolian Corpus Languages*. München/Marburg. http://www.ediana.gwi.uni-muenchen.de/dictionary.php?lemma=1290

Schrijver, Peter. 1997. Animal, vegetable and mineral: some Western European substratum words. In Alexander Lubotsky (ed.). *Sound law and analogy. Papers in honor of Robert S. P. Beekes on the occasion of his 60th birthday*, 293–316. Amsterdam/Atlanta: Rodopi.

Schrijver, Peter. 2011. La langue hattique et sa pertinence possible pour les contacts linguistiques préhistoriques en Europe occidentale. In Coline Ruiz Darasse & Eugenio R. Luján (eds.). *Contacts linguistiques dans l'Occident méditerranéen antique*, 241–255. Madrid: Velasquez.

Schrijver, Peter. 2017. Talking Neolithic: The Case for Hatto-Minoan and its relation to Sumerian. In Guus Kroonen, James P. Mallory & Bernard Comrie (eds.). *Talking Neolithic: Proceedings of the workshop on Indo-European origins held at the Max Planck Institute for Evolutionary Anthropology*, *Leipzig*, *December 2–3*, *2013*, 336–374. Washington: Institute for the Study of Man.

Schwyzer, Eduard. 1939. *Griechische Grammatik auf der Grundlage von Karl Brugmanns Griechischer Grammatik I*. München: Beck.

Sereni, Emilio. 1964. Per la storia delle più antiche tecniche e della nomenclatura della vite e del vino in Italia. *Atti e Memorie dell'Accademia Toscana di Scienze e Lettere* 29. 75–204.

Simon, Zsolt. 2012. *Untersuchungen zur hattischen Grammatik*. *Phonologie*, *Morphologie und Syntax*. Budapest: Eötvös Loránd University dissertation.

Simon, Zsolt. 2017. δέπας und die anderen: Spuren eines verschollenen luw(o)iden Dialekts? *Wék^{w}os. Revue d'études indoeuropéennes* 3. 245–259.

Simon, Zsolt. 2017 [2019]. What did Taita import from Egypt? Hieroglyphic Luwian ASINUS$_{2(A)}$ Reconsidered. In Bori Németh (ed.). *Now Behold My Spacious Kingdom. Studies Presented to Zoltán Imre Fábián On the Occasion of His 63rd Birthday*, 317–330. Budapest: L'Harmattan.

Simon, Zsolt. 2018. Anatolian influences on Greek. In: Łukasz Niesiołowski-Spanò & Marek Węcowski (eds.). *Change*, *Continuity*, *and Connectivity*. *North-eastern Mediterranean at the turn of the Bronze Age and in the early Iron Age*, 376–418. Wiesbaden: Harrassowitz.

Simon, Zsolt. 2019. Zu den karisch-griechischen Lehnbeziehungen. *Glotta* 95. 295–308.

Simon, Zsolt. 2020. Latin *sāpō* 'hair-dye, soap', the Germanic Words for Soap, and the Common Substrate of Celtic and Germanic. *Amsterdamer Beiträge zur älteren Germanistik* 80. 391–400.

Simon, Zsolt. 2021. Carian γίσσα 'stone'. In Hackstein, Olav, Jared L. Miller, Elisabeth Rieken & Ilya Yakubovich (eds.). 2016–. *Digital Philological-Etymological Dictionary of the Minor Ancient Anatolian Corpus Languages*. München/Marburg. http://www.ediana.gwi.uni-muenchen.de/dictionary.php?lemma=1638

Smeets, Rieks. 1989. On Hurro-Urartian as an Eastern Caucasian Language. *Bibliotheca Orientalis* 46. cols. 259–279.

Soysal, Oğuz 2004. *Hattischer Wortschatz in hethitischer Textüberlieferung*. (Handbuch der Orientalistik I/74). Leiden/Boston: Brill.

Steinbauer, Dieter H. 1999. *Neues Handbuch des Etruskischen*. St. Katharinen: Scripta Mercaturae.
Yakubovich, Ilya. 2002. Labyrinth for Tyrants. *Studia Linguarum* 3. 93–116.
Yakubovich, Ilya. 2010. *Sociolinguistics of the Luvian Language*. Leiden/Boston: Brill.
Yakubovich, Ilya. 2013. Anatolian Names in *-wiya* and the Structure of Empire Luwian Onomastics. In Alice Mouton, Ian Rutherford & Ilya Yakubovich (eds.). *Luwian Identities. Culture, Language and Religion Between Anatolia and the Aegean*, 87–123. Leiden/Boston: Brill.

Peter Schrijver

13 East Caucasian perspectives on the origin of the word ‘camel’ and some notes on European substrate lexemes

1 A brief history of *camel*

The word family to which English *camel* belongs is nowadays widespread in all Germanic and Romance languages. Its immediate origin is Latin *camēlus*, which in turn was borrowed from Greek κάμηλος. The latter is attested since the early fifth c. BCE (Aeschylus, Herodotus; Chantraine 1968: 489). The Greek term is generally assumed to have been borrowed from West Semitic (Masson 1967: 66): compare e.g. Biblical Hebrew *gāmāl*, Syriac *gamlā*, Sabean *gml*, Arabic *ǯam(a)l* (Militarev & Kogan 2005: 116–118).

2 Formal problems facing the idea that Greek was borrowed directly from Semitic

Borrowing is entirely as expected of the name of an animal that is not native to Europe nor traditionally used there as a domesticated animal. In this particular case, the Greek form lies at the base of all related forms in European languages. The hypothesis that the Greek form was itself borrowed from Semitic requires two formal steps: the *-ā-* of the second syllable in Semitic (Hebrew) *gāmāl* was adopted into Ionian Greek, where it joined inherited **-ā-* in becoming *-ē-* (Kretschmer 1892: 287). Apparently the first *-ā-* was not so treated, however, for reasons that remain unclear. Furthermore, the initial *g-* of Semitic was replaced by Greek *k-*. This presents a serious problem (Heide 2011: 363; Militarev & Kogan 2005: 118): Greek possesses in its phonemic inventory a voiced *g-* like Semitic does, so there appears to be no reason to regard the substitution of *g-* by *k-* as an adaptation to Greek phonology. The *g* does make an isolated appearance in the compound name Γαυγάμηλα = καμήλου οἶκος ‘camel house’ (Strabo 16, 1, 3), but apart from this single instance in what is clearly a non-Greek name, Greek consistently shows *k-*. Therefore, whereas there can be no doubt that Greek κάμηλος was borrowed and that it is related to the Semitic word family, the attested Semitic forms may not be the direct source from which it was borrowed. The word for camel is not attested in Phoenician (apart from the word’s use as the name of the third letter of the alphabet, underlying Greek *gamma*), which

https://doi.org/10.1515/9783111337920-013

would be a good candidate for being the immediate source for historical reasons, but if the Phoenician word had initial *k-* rather than expected *g-*, that would have been irregular and in need of an explanation itself. Benjamin Suchard and Maarten Kossmann point out to me that Phoenician is an unlikely source of the Greek form for another reason as well: since long *ā* regularly became long *ō* in Phoenician, the Greek word for 'camel' should have had *ō* rather than *ā* > *ē* in the second syllable (cf. the letter name *yād* > Phoenician *yōd* > Greek *iota*).

According to Yakubovich (2016: 87 with footnote 30),[1] Greek *k-* may be explained if the Semitic word was borrowed into Greek via the intermediary of an Anatolian language, where word-initial voiced plosives do not occur and are replaced by voiceless plosives. He suggests that Hieroglyphic Luwian *ka-mara/i-* may mean 'camel' (which is a guess, cf. Bauer 2020). A borrowing via Anatolian may indeed account for the Greek *k-*, but even if Hieroglyphic Luwian *ka-mara/i-* means 'camel', it is unlikely to have been the immediate source of the Greek word because it accounts for neither the Greek *-ē-* nor the Greek *-l-*.

In this contribution, an alternative will be presented, viz. that the Greek term may have been borrowed from a language that is cognate with East Caucasian (see section 5).

3 Bactrian camels, dromedaries and their domestication

The Semitic world is an obvious area to which one might look for the origin of a word for 'camel', but it is not the only conceivable area. The Greek term (Chantraine 1968: 489), like its Semitic counterparts (Heide 2011: 363), is a generic word for 'camel', which denotes both the two-humped Bactrian camel (*Camelus bactrianus*) and the single-humped dromedary (*Camelus dromedarius*). Nowadays, the Bactrian camel is a rare native of the cold deserts of Central Asia, but its original territory probably included wider stretches of the Central Asian steppe. The earliest known traces of domesticated Bactrian camels are from Turkmenistan and date from the middle of the 3rd millennium BCE. In Namazga IV levels at the urbanized site of Altyn Depe, a clay model of a Bactrian camel drawing a cart was discovered which dates from ca. 2400 BCE (see recently Kirtcho 2009; Mallory & Adams 1997: 389, also 135–136). This early Bronze Age culture, which probably separated Indo-European pastoralists living on the steppe from more southerly

1 I am indebted to Zsolt Simon for drawing my attention to this idea and to Yakubovich's article.

regions that correspond to present-day Iran, Pakistan and India, probably spoke a non-Indo-European language. Slightly later, between ca. 2250 and 1700 BCE, it was replaced by the Bactria Margiana Archaeological Complex (BMAC), which plays an important part in the migration of Indo-Iranian speakers from the steppe to the south (Mallory in Mallory & Adams 1997: 72–74). Camelid skeletal remains in quantity appear around 3000 BCE in southern Central Asia and in eastern Iran (Kohl 2007: 142–143). By the third millennium BCE, the domesticated camel probably opened up access from the east towards the western steppe north of the Caucasus because it made regular travel across the desert steppes east of the Caspian possible (Kohl 2007: 62, 200, 250). Camel remains have been found in western steppe Yamnaya sites (3600-2200 BCE; Mallory and Adams 1997: 651).

Indirect evidence for a much earlier date of the domestication of the Bactrian camel derives from the Xinglongwa culture of Inner Mongolia, which flourished ca. 6100-5300 BCE: its desert sites lie so far apart across inhospitable terrain that people must have used camels to travel between them (Potts 2004: 148).

Very soon after the earliest evidence for domesticated camels in Turkmenistan they make an appearance in Mesopotamia. The earliest record of the Bactrian camel is in an animal list from the mid-third millennium BCE: Sumerian **am.si.ḫar.an**, literally 'elephant of the road/caravan'. A different Sumerogram is **am.si.kur.ra**, literally 'elephant of the land/mountains'. Both are transparent neologisms (Heide 2011: 354–60).

Wild populations of the dromedary do not exist anymore. It is presumed to have been native to the southeast of the Arabian Peninsula. Archaeological data indicate that the dromedary was domesticated in the south of the peninsula in the second half of the second millennium BCE (decrease in bone size, increase of bones in settlement contexts, first undoubted figurines of dromedaries; Uerpmann & Uerpmann 2012; see also Heide 2011: 339–343, Almathen *et alii* 2016).[2] It is around the same time that the dromedary makes its first appearance in cuneiform texts: Nippur, Middle Babylonian, 14th-13th c. BCE **anše.a.ab.ba**, Ugarit ca. 1200 BCE **[anše.a.a]b.ba** (Heide 2011: 352, 368). The Sumerogram translates literally as 'donkey of the sea', a neologism which presumably indicates that it arrived in Mesopotamia by trade ship, hence in domesticated form.

2 Potts (2004: 151 and 155) mentions dromedary remains at Tepe Ghabristan, Iran, during its Period 4 (3700-3000 BCE), which would be exceptionally early, as well as dromedary remains at Harappan sites of the later third and early second millennium BCE. It remains to be established whether the identifications and dates are reliable and if so, how they relate to data and ideas about the much later domestication of the dromedary in its native area in the south of the Arabian Peninsula.

In short, the Bactrian camel of the Central Asian steppes was domesticated perhaps as early as the late seventh millennium BCE in Inner Mongolia. The first definite traces of domesticated camels in southern Central Asia date from around the middle of the third millennium BCE, by which time their use spread rapidly towards the western steppe north of the Caucasus and via Iran towards the Middle East. This is at least a millennium earlier than the earliest secure evidence for the domestication of the dromedary, which was native to the southeast of the Arabian Peninsula. The literary evidence indicates that the Bactrian camel reached Mesopotamia by the middle of the third millennium BCE at the latest, while the dromedary made its first attested appearance over a millennium later.

4 Lexemes travelling with their denotata

This information is relevant to the theoretical places of origin of words for 'camel' and the routes that they may have travelled. Roughly two scenarios present themselves, one starting from the dromedary and the other from the Bactrian camel.

Scenario 1. The Semitic languages of the Arabian Peninsula may have had a very old word for the native dromedary that people encountered in the wild and no doubt hunted. This word, which under scenario 1 was the ancestor of the family of Hebrew *gāmāl*, spread over a wider area when dromedaries were domesticated and traded after approximately 1500 BCE. The word was subsequently also applied to the Bactrian camel, which had been known in southern Central Asia, in Mesopotamia and further south as a domesticate from the middle of the third millennium onwards. During the first millennium BCE the word was borrowed from Semitic into Greek, perhaps via an Anatolian language, and thence into Latin. This is approximately the scenario that agrees with the traditional narrative about the origin and spread of the word 'camel'.

This scenario may perhaps be both supported and complicated by the fact that North African Berber and Tuareg possess a word for 'camel' that is very similar to **gam(a)l*- but with all consonants in a different position: this is Berber *alɣ(^{w})əm*, Niger Tuareg *aɣlam*, which is discussed extensively by Kossmann 2005: 32–41 (he reconstructs **aa-lVqum*, where **aa*- is the obligatory nominal prefix).[3]

3 I am indebted to Maarten Kossmann for pointing me to the Berber forms and for discussion.

Scenario 2. An unknown language that was spoken somewhere between the eastern Central Asian steppes, where wild Bactrian camels lived and early domestication probably occurred for the first time, and Turkmenistan and eastern Iran, where domesticated camels appear in the third millennium BCE, had a word related with 'camel' (through inheritance or borrowing). This became widespread when domestication and long-distance transport made the Bactrian camel an economically important commodity. The word ended up in the Semitic world together with the animal itself, perhaps as early as the later third millennium BC, where it spread to become the word for 'camel' as well as 'dromedary' in the form of Hebrew *gāmāl* and its ilk. In that case *gāmāl* etc. replaced an earlier native word for (wild) 'dromedary'.[4] Since no attestation of Semitic *gāmāl* etc. is earlier than the first millennium BC (Militarev & Kogan 2005: 116–118), so more than 1500 years after the first attested appearance of domesticated Bactrian camels in Mesopotamia, this provides enough time for the scenario to unfold, rendering scenario 2 a theoretically conceivable scenario. Whether it is also a plausible scenario is a different matter: until cognates of the word family 'camel' surface on the Asian steppe and in the area north of Mesopotamia scenario 2 remains purely hypothetical.

5 'Camel' in Caucasian languages

This is where the evidence of the Caucasian languages comes in. The Caucasian languages comprise three language families: West Caucasian, with the languages Abkhaz, Circassian and extinct Ubykh; South Caucasian, or Kartvelian, to which four languages belong, of which Georgian is the most widespread and has the largest number of speakers; and East Caucasian, which consists of around 30 languages in two major branches: the western branch is Nakh, which comprises Ingush, Chechen and Batsbi; all other languages belong to the eastern branch, Daghestanian, of which Avar-Andi-Tsezic forms a major group. It is in the Nakh and Tsezic languages of East Caucasian and in Georgian that we find hitherto insufficiently acknowledged relatives (by borrowing or inheritance) of the 'camel' word family.

4 A candidate for being an old word for 'dromedary' is the group of Sabean *ʔbl* (7th c. BCE, Arabian Peninsula) 'dromedary', Arabic *ʔibil* 'dromedary, Bactrian camel', which is common in Semitic languages except for the Canaanite group (cf. Militarev and Kogan 2005: 4, Heide 2011: 346, 345 fn. 26).

In Caucasian studies, the existence of a word for 'camel' that is related to Semitic and Greek is recognized, for instance in Klimov & Xalilov 2003: 224. They regard Greek κάμηλος, which they state was itself borrowed from Semitic, as the direct source of Georgian *aklemi* 'camel', and Georgian *aklemi* as the source from which Chechen *emkal* and similar Nakh forms were borrowed. While it is reasonable to posit a relationship between those terms, given the overall similarity, the assumed route of borrowing gives rise to so many formal questions that a detailed investigation is called for. Let us first turn to the word for 'camel' in the Nakh languages, which has the most complicated history.

5.1 'Camel' in Nakh

Standard Chechen, which is based on the dialects of the plains and foothills of the central North Caucasus, has nominative singular *emkal*, an oblique stem (from which all cases except the nominative are formed) *emkal(a)-*, and a nominative plural *emkal-š*.[5] Similar forms are attested in the eastern Chechen Akki dialect (*emkal*, nominative plural *emkalš*) and in the southern Chechen Kist dialect (*enkal*, nominative plural *enkališ*, which shows assimilation of *mk* to *nk* /ŋk/).[6] In closely related Ingush, the word for 'camel' is *inkal*, oblique stem *inkal(a)-*, nominative plural *inkalaž*, which also shows assimilation of *mk* to *nk*. Batsbi has a word *aklam*, which will be addressed in section 5.3.

The history of the consonants in these words is relatively simple. Rather than the consonantal sequence *k-m-l*, which we encounter in the Semitic and Greek words for 'camel' that were discussed earlier, Chechen and Ingush show a sequence *m-k-l*. The consonants can be reconstructed as *m-k-l* for any stage between attested Nakh itself and Proto-East Caucasian (if the word existed here as early as that, which is of course not known, but see section 7; nor is the approximate date of Proto-East Caucasian known).

Reconstructing the vocalism is a different and more complicated story. First of all because Chechen and Ingush have undergone many vowel changes since the Proto-Nakh stage which come under the heading of umlaut (i.e. partial or total regressive vowel assimilation). Section 5.1.1. is devoted to the effects of umlaut on the word for 'camel' and to the reconstruction of Proto-Nakh. Deeper in time beyond that stage, the reconstruction of the vocalism becomes murkier but it needs to be addressed because the vowel changes that are known to have af-

5 Nichols and Vagapov 2004: 295.
6 Aliroev 1975: 89.

fected Pre-Proto-Nakh are relevant to an assessment of the earliest shape of the word for 'camel'. Those very early vowel changes are the subject of section 5.1.2.

5.1.1 Umlaut in Chechen and Ingush and the reconstruction of the Proto-Nakh word for 'camel'

The rules of palatal and labial umlaut in Chechen and to a lesser extent Ingush have been worked out in the unsurpassed study by Imnajshvili 1977: 51–125. A number of precisions, especially those pertaining to the effects that the close vowels **i, *u* had as opposed to the effects of the mid vowels **e* and **o*, were made by Schrijver 2021.

With few exceptions, Chechen and Ingush carry the stress on the first syllable. In many dialects, among those the dialects on which standard Chechen and Ingush are based, have lost all oppositions between old short vowels in unstressed, non-initial syllables, replacing them by *-a-* [ʌ]. This is the *-a-* in Chechen *emkal* and Ingush *inkal*, which consequently may go back to any of the Proto-Nakh short vowels **i, *e, *a, *o, *u*. We know that Proto-Nakh possessed those five vowels in non-initial syllables because the particularly archaic Cheberloj dialect of Chechen still preserves the opposition of all five short vowels, and so does the phonologically archaic Batsbi language.

The Chechen form *emkal* contains a clue to the original quality of *-a-*. In order to get a handle on this it is necessary to dwell on the phenomenon of palatal umlaut. The vowel *e*[7] in initial syllables in Chechen always resulted from palatal umlaut of an old short **a*, which was caused by either **e* or **i* in the second syllable. So Chechen *emkal* reflects Proto-Nakh **amkel* or **amkil*. Which of the two is the correct reconstruction can normally be decided on the basis of the Ingush cognate because in Ingush **i* does (**a* > *e*) and **e* does not cause palatal umlaut of **a*.[8] Like

7 This is *e* as opposed to *ie*, which in standard Chechen is never the result of umlaut. In standard Cyrillic orthography, *e* and *ie* are both spelled as <e>, which obscures this important opposition. The dictionary by Nichols and Vagapov (2004), the phonological description by Nichols 1997, the grammatical description by Nichols 1994, the Chechen morphology by Jakovlev (1960) and some other studies do distinguish *e* from *ie*. A similar situation obtains for Ingush, where *e* and *ie* are spelled in standard Cyrillic as <e>. In spoken Ingush, short **ie* (which usually but not exclusively occurs in closed syllables) merged with the *e* that arose as a result of umlaut (Nichols 2011: 25), which probably is a very recent phenomenon (Imnajshvili 1977 still distinguishes the two).

8 The different umlaut effects of **e* and **i* on **a* in Ingush are not specifically identified by Imnajshvili (1977), but they can be distilled from the Chechen material he provides if one compares Ingush cognates, e.g. Proto-Nakh **bali* 'shoulder(s)' > Chechen *bela-š*, Ingush *belaž*, both with umlaut (Cheberloj *bališ* and Batsbi *bali* preserve the original vocalism in first and second syllable); Proto-Nakh

standard Chechen, Ingush lost the opposition between the old short vowels in non-initial syllables, replacing them by *a* [ʌ]. Unfortunately, however, Ingush is not so obliging in this instance: its word for 'camel' is *inkal*, oblique stem *inkal(a)-*, nominative plural *inkalaž*, which shows *i-* rather than expected **e* (if from **amkil*) or **a* (if from **amkel*). The correspondence Ingush *i* ~ Chechen *e* is irregular and cannot be explained on the basis of known sound changes involving umlaut. For the time being, there is no other option than to reconstruct Ingush *inkal* as Proto-Nakh **imkVl*, with old **i* in the first syllable and with **V* in the second syllable representing any one of the Proto-Nakh vowels **a*, **e*, **o* or **i*.[9] We shall return to Ingush in section 5.1.2. So we cannot decide whether Chechen *emkal* reflects **amkel* or **amkil*.

Beside **amke/il* (Chechen) and **imkVl* (Ingush), a third form, **amkal*, must have existed, as is revealed by Cheberloj Chechen *ankal* and Itumkali Chechen *ankal* (Aliroev 1975: 89). As stated earlier, Cheberloj preserves Proto-Nakh vowel quality oppositions in non-initial syllables, so the second *-a-* in *ankal* represents original *-a-*. Itumkali, on the other hand, lost non-initial short vowel oppositions and underwent palatal umlaut much like standard Chechen did. The fact that Itumkali *ankal* does not show the effect of umlaut indicates that underlying it is the same form as is attested in Cheberloj, viz. Proto-Nakh **amkal*.

So at the end of this discussion, we arrive at the necessity to reconstruct three different forms of the word for 'camel' for Proto-Nakh:

1. **amkel* or **amkil* (Plains Chechen)
2. **amkal* (Cheberloj and Itumkali Chechen)
3. **imkVl* (Ingush)

5.1.2 Pre-Nakh origins of the Nakh vowel alternation **i* ~ **a*

In the Proto-Nakh reconstructions we observe an alternation in the first (stressed) syllable between **i* and **a*. In the second syllable, **e* or **i* alternates with **a*. Neither alternation can be explained on the basis of umlaut, but they do have an equivalent in the nominal morphology of the Nakh languages. Consider the fol-

**malin* 'warm' > Chechen *mela*^{N}, Ingush *mela* (Cheberloj *mali*^{N}, Batsbi *mali*^{N}); contrast Proto-Nakh **maqe* 'harrow' > Chechen *meqa* (umlaut), Ingush *maqa* (no umlaut; Cheberloj *maqe*); Proto-Nakh **laqen* 'high' > Chechen *leqa*^{N}, Ingush *laqa* (Cheberloj *laqe*^{N}, Batsbi *laqe*^{N}). See Schrijver 2021: 95–100. Nikolayev and Starostin (1994: 99) record the difference but they do not provide supporting material.

9 Not **u* because this would have caused labial umlaut resulting in Ingush **unkal*, cf. *=ug* 'leads' < **=ik'-u* of the Proto-Nakh verbal root **=ik'*, which is still represented as such in Batsbi *=ik'-*).

lowing examples (abbreviations: E = ergative case, G = genitive case, Loc. = one of the local cases, O = oblique case stem):

Proto-Nakh	Batsbi	Chechen	Ingush	meaning
1. **dikʾ* O **dakʾor-*	*dikʾ* pl. *dakʾvrĩ* beside *dikʾujr*	*dig* G *dagaraᴺ* E *dagaruo*	*dig* E *dogaruo*	‘axe’
The Batsbi plural forms are those of Desheriev 1953: 68 (Kadagidze & Kadagidze 1984 only provides the nominative singular); the nominative plural *dakʾvrĩ* probably represents **dakʾujrĩ* < **dakʾor-i*, with regular umlaut according to Batsbi rules (cf. Imnajshvili 1977: 120); *dikʾujr* reflects the generalization of the first-syllable vocalism of the nominative singular *dikʾ*. The Chechen and Ingush forms are those of Nichols & Vagapov 2004 and Nichols 2004, respectively. The Proto-Nakh reconstructions take the effects of umlaut in Chechen and Ingush into account. For the etymology see Giginejshvili 1977: 82 (on Lak and Lezgian cognates), Nikolayev and Starostin 1994: 944, Nichols 2003: 258.				
2. **niq* O **naqorV-*	-	*niq* G *naqaraᴺ* E *naqaruo*	*niq* E *noqaruo*	‘beehive’
The Chechen and Ingush forms are those of Nichols & Vagapov 2004 and Nichols 2004, respectively. See Nikolayev & Starostin 1994: 868 for the etymology.				

In both examples, **i* is found in the first syllable of the nominative singular, while the oblique stem, from which cases like the genitive, ergative and dative are formed, and sometimes also the plural, instead has **a*. The **a* is preserved without change in Batsbi and in Chechen, but Ingush has turned it into *o* as a result of labial umlaut caused by original **o* in the second syllable of the oblique stem (in standard Chechen as well as in the Plains dialect, **o* does not cause labial umlaut of **a* in the preceding syllable, cf. Imnajshvili 1977: 65–68, Schrijver 2021: 108–110). Subsequently, the unstressed **o* regularly became *a* [ʌ] in Ingush and Chechen.

This alternation of **i* in the nominative singular stem and **a* in the oblique stem is relatively rare in nominal paradigms, but it has counterparts which are more frequent and in which the nominative singular contains **o* or **u* and the oblique stem is characterized by **a*, e.g. [10]

10 Batsbi forms are taken from Kadagidze and Kadagidze 1984, Gagua 1961: 85, Holisky-Gagua 1994: 161, 167. Chechen forms from Nichols and Vagapov 2004 and Ingush forms from Nichols 2004.

Proto-Nakh	Batsbi	Chechen	Ingush[11]	meaning
3. **borc* O **barci-*	*borc* G *barciᴺ* E *barcav*	*buorc* G *bercaᴺ* E *bercuo*	*buorc* E *bercuo*	'millet'
4. **motˀtˀ* G **matˀtˀiᴺ* O **matˀtˀa-*	*motˀtˀ* G *matˀtˀiᴺ* E *matˀtˀav*	*muott* G *mettaᴺ* E *mattuo*	*muott* E *mettuo* [a]	'tongue'
5. **buc* G **ba:ciᴺ* O **ba:ca-*	*buc* G *baciᴺ* Loc. *bac-ma-k*	*buc* G *be:caᴺ* E *ba:cuo*	*buc* E *beacuo*[a]	'grass'
6. **butt* G **battiᴺ* O **batta-*	*butt* G *battiᴺ* E *battav*	*butt* G *bettaᴺ* E *battuo*	*butt* E *bettuo*[a]	'moon'

For reasons that remain to be worked out, the *a* in such paradigms appears as long *a:* in some etyma, e.g. in **buc* oblique **ba:c-* 'grass'. As argued by Schrijver 2021: 131–146, the Nakh vowel alternation of **i, *u, *o* in the first syllable of the nominative singular with **a* (or **a:*) in the first syllable of the oblique cases and the nominative plural has deep roots, which go back all the way to Proto-East-Caucasian, the common ancestor of Nakh and the Daghestanian languages. Nakh has a particularly close counterpart in the Tsezic languages of Daghestanian.

Proto-Nakh	Batsbi	Chechen	Ingush	Tsezic
7. **mott* O **matt-* 'bed, place'	*mott* Adessive *matteħ* pl. *mattiš*	*muott* G *mettaᴺ* E *mattuo*	*muott* E *mettuo*	**mɔčːˀe* O **mɨčːˀɔ-* > **mičːˀɔ-* in Tsez *moči* O *mečo-*, *moči-*, Hunzib *mɔče* O *mičo-* 'place, plot'
Etymology: Nikolayev & Starostin 1994: 803.				
8. **duqˀ* pl. **daqˀw-* 'yoke; mountain crest'	*duqˀ* 'yoke'	*duqˀ* G *duqˀaᴺ* E *duqˀuo* pl. *daqˀqˀaš*	*duqˀ* E *duqˀuo* pl. *douqˀaž*	**ruƛˀu* > Bezhta *ruƛˀo*; **riƛˀwe-* in Tsez *raʕƛˀu*, *roƛˀi*, Hunzib *riƛˀu* 'yoke'
Etymology: Giginejshvili 1977: 109, Nikolayev & Starostin 1994: 220, 954, Nichols 2003: 260.				

Where Nakh has an alternation nominative **u, *o* ~ oblique stem **a*, the Tsezic languages show an alternation **u, *o, *ɔ* ~ **i* in cognate etyma. On the basis of a close comparison of the Tsezic data with those from Nakh it is possible to establish that Proto-East Caucasian had the following features (Schrijver 2018a on Avar-Andi-Tsezic, Schrijver 2021 on Nakh and Proto-East Caucasian):

11 Palatal umlaut in the Ingush ergative is probably due to the generalization of the stem-final vocalism of the genitive.

(1) it possessed phonological stress, much like present-day Avar;
(2) nominal paradigms existed in which the nominative singular had the stress on the first syllable while the oblique stem had stress on a non-initial syllable (this is a type commonly attested in Avar, e.g. *moc:ʾ*, genitive singular *moc:ʾról* ‘moon’);
(3) the rounded vowels **u,* **o,* **ɔ* were regularly unrounded to **ɨ* in pretonic position in Proto-East Caucasian; hence the Proto-East Caucasian nominative (= absolutive) preserved the rounded vowel because it was stressed, while the oblique stem turned it into **ɨ* because it was pretonic.

With this original state of affairs in mind, the vowel alternation in Nakh receives a relatively simple explanation on the basis of the following regular vowel correspondences and reconstructions (Schrijver 2021: 138–149):

Proto-East Caucasian		***u**	***o**	***ɔ**	***ɨ**
Proto-Nakh	stressed and posttonic	**u* > **u*	**o* > **u*	**ɔ* > **o*	**ɨ* > **i*
	pretonic	**u* > **ɨ* > **a*	**o* > **ɨ* > **a*	**ɔ* > **ɨ* > *a*	**ɨ* > **a*

It is the last column in this diagram that is relevant to understanding the Proto-Nakh paradigms of **dikʾ*, oblique **dakʾor-* ‘axe’ and **niq,* **naqor-*, which resulted from Pre-Proto-Nakh *ˈ*dɨkʾ*, oblique **dɨˈkʾor-* and *ˈ*nɨq*, obique **nɨˈqor-*, depending on the original position of the stress.

In a few examples, the vowel alternation **i* ~ **a* is not only attested in the first syllable but also in the second, which suggests a more complex form of accentual mobility according to which the stress moved between the first, second and third syllables and consequently the second syllable was posttonic in the nominative, stressed in some oblique forms, and pretonic in others:

Proto-Nakh	**Batsbi**	**Chechen**	**Ingush**	**Pre-Proto-Nakh**
9. **jiš* O **a:šir-*, **a:šar-* ‘voice’	*iš* E *aširv*	*jiš* G *e:šaraᴺ* E *e:šaruo*	*jiš* E *a:šaruo*	*ˈ*ɨš* O **ɨˈširV-*, **išɨˈrV-*

Batsbi shows *i*-vocalism in the second syllable of the oblique stem, which agrees with palatal umlaut in the Chechen oblique stem. Ingush lacks palatal umlaut, which prohibits the reconstruction of **i* in the second syllable and necessitates the reconstruction of **a* (this is the only vowel which in Ingush does not cause umlaut of **a:* in the preceding syllable, cf. Schrijver 2021: 118*)*. In accordance with the stress-dependent behaviour of Proto-East Caucasian **ɨ* in Nakh, the deeper reconstruction of **jiš,* O **a:šir-*, **a:šar-* was probably *ˈ*ɨš* O **ɨˈširV-*, **išɨˈrV-*, respectively, which yielded Proto-Nakh **iš* O **a:širV-*, **a:šarV-*.

(continued)

Proto-Nakh	Batsbi	Chechen	Ingush	Pre-Proto-Nakh
10. **jis* O **aːsir-*, **aːsar-* 'hoarfrost'	–	*jis* G *eːsaraᴺ* E *eːsaruo*	*jis* E *aːsaruo*	**ˈɨs* O **ɨˈsɨrV-*, **ɨsɨˈrV-*
The situation is similar to that of 'voice' except that *i*-vocalism of the second syllable is not directly attested in Batsbi. For the etymology outside Nakh see Nikolayev & Starostin 1994: 675.				
11. **ʕi* O **ʕanar-*, **ʕanir-* 'steam'	*ʕa*, Lative *ʕanar-ʁ*	*ʕa* G *ʕänaraᴺ*	*ʕi* E *ʕanaruo*	**ˈʕɨ* O **ʕɨˈnɨrV-*, **ʕɨnɨˈrV-*
The Nakh oblique stem **ʕanar-* is required for Batsbi, while **ʕanir-* is presupposed by Chechen, which shows palatal umlaut (*ä* instead of *e* is regular in pharyngeal environment, see Imnajshvili 1977: 60, Schrijver 2021: 95–100). The absence of umlaut in Ingush *ʕanar-* suggests that it generalized the stem **ʕanar-*. Ingush is the only language to preserve the root vowel *i* in the nominative. See Nikolayev & Starostin 1994: 485 for possible etymological connection with Daghestanian.				

Consequently, what these paradigms show is the stress-related behaviour of Proto-East Caucasian **ɨ* in Nakh: it became Proto-Nakh **i* under the (Proto-East Caucasian) stress and **a* in pretonic position.

As a final twist, it seems that we are not justified in automatically reconstructing Proto-East Caucasian **ɨ* wherever the Nakh alternation **i* ~ **a* occurs, as I have done in the diagrams. The reason for that is that in the case of two etymologies involving Nakh **i* ~ **a* the evidence from Avar-Andi-Tsezic suggests the reconstruction of Proto-East Caucasian **a* rather than **ɨ*:

– Proto-Nakh **jis* O **aːsir-*, **aːsar-* 'hoarfrost' is related to Andi *asor*, Northern Axwax *aša* < Proto-Andic **asor*. Andi *a* can only reflect earlier **a*. The same vocalism is found in further cognates: Avar *ʕansí* 'snow drift', Tsez *az-qʾa* (with unclear *-qʾa*), Hunzib *aᴺza* < **ansV*. See Nikolayev & Starostin 1994: 675 for the etymology.
– Proto-Nakh **ʕi* O **ʕanar-*, **ʕanir-* 'steam' is probably related to Andi *hal* 'steam', which has **a*. A further relative is the verbal stem **hal-* (Avar *hal-*, *hwal-* 'boil'), **hel-* (Bezhta and Hunzib *hel-* 'to boil'). See Nikolayev and Starostin 1994: 485.

So in as far as can be judged on the basis of just two etymologies, it seems that nouns that had original **a* in the first syllable could join the alternation of nominative **i* ~ oblique **a* in Nakh. It is unclear whether this was a result of regular sound change (i.e. Proto-East Caucasian stressed **a* becoming **ɨ* > **i* in Nakh) or rather of analogy (i.e. the morphological pattern nominative **i* ~ oblique **a* (< **ɨ*) becoming productive at the expense of the pattern nominative **a* ~ oblique **a* (< **a*)).

On the basis of this information, it is possible to propose that the different reconstructions of the word for 'camel' in Nakh that we found in section 5.1.1 orig-

inally formed a single paradigm that looked like the paradigms of the words for 'voice', 'steam' and 'hoarfrost':

– **imkVl* (Ingush) can be reconstructed as the Proto-Nakh nominative, **'imkil*
– **amkel* or **amkil* (Plains Chechen) can be reconstructed as one of the two oblique stems, **am'kil-*
– **amkal* (Cheberloj and Itumkali Chechen) can be reconstructed as the other oblique stem, **amka'lV-*

The original paradigm **imkil*, oblique **amkil-* beside **amkalV-* was apparently so irregular that the various dialects of Checheno-Ingush generalized a single stem throughout the paradigm, but they did so by selecting a different stem.

5.1.3 A formal parallel to 'camel' in East Caucasian

There exists another etymon which in the Nakh languages shows an alternation between **i-* in Ingush and **a-* in Chechen, thus offering a close formal parallel to 'camel'. I am referring to the word for 'glass, porcelain':[12]

– Chechen *angali* (Matsiev 1961), *āngli, ä:ngli* (Nichols-Vagapov 2004) 'glass (shard), porcelain'. The Chechen dialectal forms Kist *engli*, Cheberloj *angali* and Itumkali *ängle* are provided by Aliroev 1975: 41. These forms reflect **a(:)ngalaj*.
– Ingush *ingalii* (Nichols 2004) 'porcelain', from **ingalaj*

The interesting aspect of this etymon is that it has Daghestanian cognates which directly point to a paradigmatic alternation between Proto-East Caucasian **i* and **ɔ* and hence provide evidence for the idea that the Nakh alternation **i* ~ **a* indeed stems from Proto-East Caucasian **i*: for Avar-Andi-Tsezic it is possible to reconstruct a paradigm absolutive **ħɔ́ngV(r)*, oblique stem **ħɨngwV́r(V)-*.[13]

– The vocalism of the first syllable of the absolutive was generalized in the Tsezic languages: Tsez and Hinuq *aki*, oblique stem *aki-* 'window'; Xwarshi *aᴺka* (Imnajshvili 1963: 16), Inxoqwar *aᴺka, aka*, oblique stem *aᴺka-* 'window, opening, hole' (Khalilova 2009: 89, 90; Kibrik & Kodzasov 1990, item 375). These forms reflect Proto-Tsezic **ankV(r)* from earlier **ħɔngV(r)*, with **a* < **ɔ* which is regular in word-initial position before nasal + plosive (/#_NT).[14]

12 I am indebted to an anonymous referee for pointing me to this etymon.

13 The reconstruction of **ħ-* rather than another pharyngeal or laryngeal consonant is arbitrary.

14 Direct evidence for a Tsezic rule **ɔ* > **a* /#_NT is provided by (1) Tsez *aqju*, Inxoqwar *aᴺqqu*, Hunzib *aqqu* 'female (animal)' < **anq-ju*, Hinuq *aqili*, Xwarshi *aqu*, Bezhta *aqo*, Hunzib *aqe* 'woman' < **anqə*; cf. Andi *horč:'i*, Botlix *hark:'i* 'wife' < **hɔrk:'i*; (2) Tsez, Hinuq *ata*, Xwarshi

– The Andic languages generalized the vocalism of the first syllable of the oblique stem: Andic *ingur* (Salimov 2010: 300), Northern Axwax *íᴺgo*, Southern Axwax (Tsegob, Tljanub) *ingo*, Southern Axwax (Ratlub) *ungor* 'window'; Chamalal *ìnna* 'opening, hole, window', Chamalal (Gigatl) *hingár* 'window', Tindi *hìᴺgwar* 'window, kind of ring-shaped bread roll', Karata *hingwár*, Karata (Tokita) *hingur* 'window; kind of pasty', Botlix *híngur*, Bagwalal (Xuštadi) *húᴺgar*, Bagwalal (Tlissib) *húᴺgwa* 'window', Godoberi *hingúr* 'window, pasty' (e.g. Kibrik & Kodzasov 1990, item 375, and the many dictionaries that appeared since the 1990s).

Further Daghestanian cognates: Lezgi *hamga* 'crystal', Tabassaran *hamg*, Aghul *ʕagug* 'glass', Kryz *ʕag* 'niche in wall, roof orifice', Budux *ħag* 'roof orifice' (cf. Nikolayev & Starostin 1994: 536).

The etymon does face the problem that the Batsbi cognate, *ank'raʔǒ* 'glass' (Kadagidze & Kadagidze 1984: 40), has *k'* rather than *g*, which would indicate that Chechen and Ingush *g* reflects Proto-Nakh **k'* rather than **g*, thus providing an obstacle to the Nakh-Daghestanian equation. However, the Batsbi word also means both 'pure, clear' and 'grass-snake'. The similarity of these Batsbi forms to Georgian *ank'ara* 'clear, pure (of water, voice)' and 'grass-snake' must be more than accidental, but the exact nature of the relationship is unclear: perhaps Georgian donated the item to Batsbi, or the native Nakh etymon for 'glass', which had **g*, was formally influenced by its Georgian look-alike and replaced **g* by **k'*.

5.2 'Camel' in Georgian

Before delving deeper into the history of the Nakh word for 'camel' it is necessary to study the other Caucasian languages that have cognates.

The Georgian word for 'camel' is *aklemi*. This has the same initial vowel *a-* as two of the Nakh forms but a different vowel in the second syllable (*e* rather than Nakh **a* and **i*). It also differs from Nakh by showing a different ordering of its consonants: *k-l-m* rather than Nakh *m-k-l*, an issue to which we shall return in section 5.5. *Aklemi* has an inflectional stem *aklem-*, the final *-i* being the productive suffix that was added to the Nsg. of all consonantal stems.

---

aᴺta, Inxoqwar *aˤta, aᴺta*, Bezhta *äᴺdä*, Hunzib *aᴺda* 'brain' < **aˤnda*, cf. Andi *honu* 'brain', Botlix *handa* 'forehead' < **hɔn(d)V*. Indirect evidence is provided by numerous instances in which Tsezic **a* /#_NT corresponds to Andic **i* (with **i* < Proto-East Caucasian unrounded pretonic **ɔ*), e.g. Tsez, Hinuq *ac*, Xwarshi, Inxoqwar, Bezhta *aᴺc* 'door', Andi, Botlix *hinc:'u* 'id.'; Tsez, Hinuq *aᴺɬu*, Inxoqwar *aᴺɬe*, Bezhta *aᴺɬo* 'armful', Andi *inɬir*, Botlix *intar* 'bosom'.

Closely related Mingrelian has *arkemi* 'camel', which Klimov & Xalilov 2003: 224 regard as a borrowing from Georgian.

5.3 'Camel' in Bezhta and Batsbi

Georgian *aklemi* was probably borrowed in two East Caucasian languages. One is Bezhta, a Daghestanian language that belongs to the Tsezic group of the Avar-Andi-Tsezic subfamily. This has *aklamo* 'camel' (Xalilov 1995: 26). Speakers of Bezhta are in frequent contact with Georgian speakers and the language contains many Georgian loanwords. It is therefore reasonable to suppose, with Komri, Xalilov & Xalilova (2015: 77), that Bezhta borrowed the word for camel from Georgian. They provide more examples in which Georgian *e* was borrowed as Bezhta *a*, e.g. Georgian *sabeli* 'rope' > Bezhta *sabali*; all instances have an *a* in the first and *e* in the second syllable, which is assimilated to *a - a* in Bezhta. The final *-o* is an innovation of Bezhta, the motivation for which is unclear.

The other language that probably borrowed from Georgian is the Nakh language Batsbi, which has *aklam* 'camel' (Kadagidze & Kadagidze 1984: 50). Speakers of Batsbi are all fluent in Georgian and the language abounds with Georgian loanwords. Such loanwords regularly lose the final *-i* of the nominative of Georgian consonantal stems, but there is no apparent reason why *aklem-* should have become Batsbi *aklam*. In general, Batsbi has many instances of the sequence *a - e*, e.g. *bader* 'child', *pχaner* 'shoulder', *sakʾer* 'neck'. It is perhaps conceivable that *aklam* represents a cross between borrowed *aklem-* and the native Nakh stem *amkal-*.

5.4 'Camel' in Tsez and Hinuq

The final representatives of 'camel' in Caucasian languages are found in two other Tsezic languages: Tsez *ʕumukúli*, the Asax dialect of Tsez *ʕomokuli*, and Hinuq *ʕomókilu* 'camel'. These forms share the consonantal sequence *m-k-l* with the Nakh languages and the fact that Tsez and Hinuq are spoken in the southwestern part of Daghestan that is close to Chechnya suggests that they may have been borrowed from Chechen at some prehistoric stage. The quadrisyllabic stem structure is highly unusual in Tsezic languages (one or two syllables are the norm). The similarity of the initial syllables to the West-Tsezic word for 'donkey', Tsez *ʕomoj*, Hinuq *ʕomoqʾi*, Xwarshi *umuqʾe*, Inxoqwar *omʿoqʾʷe* < Proto-Tsezic **ʕɔmɔqʾe* suggests that the loanword for 'camel' was remodeled by popular etymology on the native word for 'donkey' (for a semantic parallel for the association of camel and donkey, cf. Middle Babylonian **anše.a.ab.ba** 'dromedary', literally 'donkey of

the sea', see section 3). The final syllables in Tsez, *-kuli*, do not match those of Hinuq, *-kilu*. One approach to this difference is to assume metathesis of the vowels in one of the languages, but this has the drawback of being *ad hoc.* Alternatively, a common Tsez-Hinuq stage **-kwil-* may be reconstructed, which regularly became *-kul-* in Tsez and *-kwil-* in Hinuq, as in the following native etyma:[15]

Tsez	Hinuq	Proto-Tsez-Hinuq	Proto-Tsezic	meaning
ʁutʾ-	*ʁwitʾ-*	**ʁwitʾ-*	**ʁwetʾ-*	'to perish, die'
=iħu-	*=iχwi-*	**=iχʿwi-*	**=eχːʿwe-*	'to collect (in a stable), sit (on eggs)'
=aʿqu-	*=aqwi-*	**=aqʿwi-*	*=aqːʿwe-*	'to sew'
qʾuʿl	*qʾwil*	**qʷwil*	**qʷwel*	'rind, bark'
kud	*kwid, kud*	**kwid*	**kwed*	'basket'

The problem facing this reconstruction is that Hinuq and the Asax dialect of Tsez normally retain **wi* unchanged (Imnajshvili 1963: 28), but it may well be that Asax *ʕomokuli* was borrowed from another Tsez dialect. The difference in the word-final vowels between Tsez (*-i*) and Hinuq (*-u*) may reflect a different incorporation of this foreign lexeme into native inflexional morphology.

5.5 Conclusions on the attested Caucasian forms

It is possible to identify two related groups of words for 'camel' in Caucasian languages. The first has a consonantal sequence *k-l-m* and centres on Georgian *aklemi*. This, or a form very much like it, was probably borrowed into the East Caucasian languages Bezhta (*aklamo*) and Batsbi (*aklam*). The other group has a consonantal sequence *m-k-l* and centres on Checheno-Ingush. There we find a number of representatives whose vowel alternations are reminiscent of the vowel alternation in paradigms of native words such as Proto-Nakh **iš* O **aːširV-*, **aːšarV-*, allowing the reconstruction of a Proto-Nakh paradigm **imkil*, oblique **amkilV-*, **amkalV-*. This means that if the word for 'camel' is a loanword in Nakh itself, it was well integrated into the morphology of the receiving language. The Nakh form **amkil-* may be the source for the borrowing into Tsez and Hinuq, which can be tentatively reconstructed as **ʕomokwil-*. Formal differences between **amkil-* and **ʕomokwil-* can at least partly be explained if we assume that the Tsez-Hinuq form was influenced by the word for 'donkey', **ʕomoqʾe*.

15 Nikolayev and Starostin 1994 mention three of the etymologies (pp. 638, 701, 931). All items were checked against Xalilov 1999 and Xalilov and Isakov 2005.

The relationship between Georgian *aklemi* and Nakh **imkil*, oblique **amkil-*, **amkalV-* is more difficult to establish. It makes little sense that Georgian borrowed from Nakh or Nakh borrowed from Georgian: there are no phonotactic reasons in either language why the **m* should move from the left-hand side to the right-hand side, or vice versa, skipping across the medial *-l-*. It would seem that the only reason for assuming that the direction of borrowing was from Georgian to Nakh (Klimov-Xalilov 2003: 224) is the assumption that the word originated in Semitic and Georgian geographically lies between Semitic and Nakh. But it is not self-evident that the word originated in Semitic (see section 4), nor how the formal transformations between Semitic *gamal* and the like on the one hand and Georgian *aklemi*, Nakh **imkil*, oblique **amkil-*, **amkalV-* on the other should be explained.

What is plausible is that the missing link between Georgian *aklemi* and Nakh *amkil-* etc. is a form with the *-m-* between *-k-* and *-l-*, so exactly where it is in Semitic *gamal* and Greek κάμηλος. It stands to reason that if a borrowing of the shape **akmil-* or the like entered Nakh, it would be adapted as **amkil-*: consonant clusters of the type plosive + resonant are not tolerated (unless there is a morpheme boundary in between), while clusters of the type resonant + plosive are very common (see Nichols 2011: 89–98 on such consonant clusters in Ingush; she does not mention clusters of nasal resonant + plosive, but they are not uncommon, e.g. *q̓oandz* 'vinegar', *c̓andar* 'clean', *čʿonkar* 'mace', *enži:* 'coarse', *engar* 'co-wife', *onk* 'good quality, *sonka* 'mountain crest', all from Nichols 2004). It is less clear, however, why in Georgian a borrowing of the shape **akmel-* should become **aklem-* except by a sporadic metathesis. However that may be, as far as the consonants are concerned, an originally medial position of the *m* offers the best opportunity to account for its position in Georgian *and* Nakh.

How about the vocalism? Is it possible to propose a form that is capable of bringing together Georgian *aklemi* and Nakh **imkil*, oblique **amkil-*, **amkalV-*? Recall that if the Nakh item were native, its pre-Proto-Nakh reconstruction would be **imkɨl-*, with the development of **ɨ* to Proto-Nakh **i* or **a* depending on stress, which was mobile in this paradigm. Also recall that nouns with an original **a* rather than **ɨ* in the first syllable probably joined this pattern of noun inflection, so that beside a reconstruction **ɨmkɨl* also **amkɨl* is possible. It is either the latter, Pre-Proto-Nakh **amkɨl-*, or the later allomorph, Proto-Nakh **amkil-*, that comes closest to Georgian *aklem-* in as far as the vowels are concerned.

In conclusion, the original form of the word for 'camel' that is best able to account for the reflexes in Nakh and in Georgian is either **akmɨl-* or **akmil-*, with *-m-* as the middle consonant and *a* and *i/ɨ* as the vowels of the first and second syllable respectively.

6 The word ‘camel’ and a European lexical substrate

The question remains in which relationship Caucasian **akmɨl-*, which I shall henceforth use as a shorthand summarizing the conclusion of section 5 on Caucasian, stands to Semitic **gam(a)l-* and Greek κάμηλος. This brings us to the heart of the reason why this study on the history of the word ‘camel’ forms part of a collection of essays on the languages of pre-Indo-European Europe. For the formal relationship of **akmɨl-* to **gam(a)l-* and κάμηλος is strongly reminiscent of a set of etymologies that are believed to be of pre-Indo-European origin and that are widespread throughout Europe (Iversen & Kroonen 2017, Schrijver 2018b: 362–363, with references). I am referring here to a specific feature of some of those etymologies, viz. the presence or absence of an initial **a-*, which, if present, usually produces vowel changes or vowel loss in the body of the word. Here are a few relevant examples:

with *a-	**without *a-**
1. Old High German *amsala, amusla* ‘blackbird’ < **amsl-*	Latin *merula*, Welsh *mwyalch* < **mesal-* ‘blackbird’
2. Greek ἀστραπή ‘lightning’ < **astrp-*	Cypriote Greek στροπά ‘id.’, Old Irish *sraif* ‘sulphur’ < **str(a)P-*
3. Old High German *aruz* ‘ore’ < **arud-*	Latin *raudus* ‘ore’ < **raud-*
4. Welsh *erfin* ‘turnip’ < **arb^h-*	Old High German *ruoba*, Latin *rāpa*, Greek ῥάφανος ‘turnip’ < **rāp-*, **rab^h-*
5. Greek ἀχράς, -άδ- ‘wild pear’ < **ag^{h}rd-*	Albanian *dardhë* ‘pear’ < **ǵ^{h}a/ord-*
6. Hittite *alel* ‘flower’ < **aleil-*, Old Irish *ailestar* ‘flag-iris’ < **alil-stro-*	Greek λείριον, Latin *līlium* ‘lily’ < **leil-*

To this we can now add:

7. Caucasian **akmɨl-* ‘camel’	Greek κάμηλος, Semitic **gam(a)l-* ‘camel’

For some of those etyma Near Eastern counterparts have already been identified. The word for ‘ore’ (3) closely resembles Sumerian **urudu** /aruta/ (Schrijver 2018b: 363). The Hittite word for ‘flower’ (6) is usually connected to Middle Egyptian *ḥrr-t*, Coptic *hrēri, hlēli* ‘flower’, perhaps Berber *alili* ‘oleander’ (e.g. Tischler 1977:16, Puhvel 1984: 32–33). Kroonen 2012 drew attention to Greek γέλγις, ἄγλις ‘garlic’, possibly from **gedl-*, **a-gdl-* respectively and its relationship to Old Babylonian

gidlu ‘plaited string of onions and garlic’. So the word for ‘camel’ is not unique in this respect.

The linguistic stratum to which this group of originally non-Indo-European words belongs has been tentatively connected with the language or languages of the first farmers who entered Europe from Asia Minor from the seventh millennium BCE onwards (Iversen-Kroonen 2017, Schrijver 2018b). The fact that related lexemes are widespread in the Near East chimes well with this idea. The word for ‘camel’ obviously would not belong to this ancient stratum: on present evidence, the presence of domesticated Bactrian camels in Turkmenistan and subsequently on the western steppe and in the Near East cannot be pushed back beyond the third millennium BCE. But given its strikingly similar phonological pattern, it is possible that ‘camel’ originated in or, if it was a loanword, was morphophonologically adapted to the same language family to which the early European agriculturalists belonged and which may consequently be situated somewhere in the larger Near East (i.e. including Turkey, Iran, the Caucasus, Mesopotamia and the Levant).

Another interesting aspect of the word for ‘camel’ is that Caucasian languages are involved. It is quite possible that the word was borrowed there, but on the other hand it is striking how well the pattern found in the European substratal words with and without **a-* fits in with the morphophonology of reconstructed East Caucasian. Although the systematic reconstruction of that family is still in its infancy compared to the reconstructions of Indo-European, Uralic or Semitic, it is possible to outline features that were typical of the inflection of nouns. Of particular relevance is the behaviour of nouns of the original structure *(C)V_1CV_2C. Those nouns lose either the entire initial syllable (especially if that syllable has a weak consonantal onset, i.e. a laryngeal or pharyngeal), or they lose the vowel of the second syllable. The conditions under which either occurs remain to be worked out, but it is probable that some case forms lost the first and others the second syllable, no doubt in connection with accentual mobility. Consider the following examples of etyma which show both developments:

*(c)vcvc-	*(c)vcv/c-	*(c)vcv/c-	Nikolayev & Starostin 1994
1. **vmvrg-* ‘bracelet’	Tsez *aka*, Hunzib *aᴺga* < **amrg-*	Avar *mergó*, Godoberi *mingwa*, Axwax *magwa* < **mvrg-*	NS 234
2. **ħvnvcˀ-* ‘door’	Tsez *ac*, Bezhta *aᴺc* < **hancˀ-*; Andi *hincˀu*, Axwax *iᴺcˀo* < **hinčˀ-*	Avar *nucˀá*, Lak *nuz* < **nvcˀ-*	NS 218

(continued)

*(c)vcvc-	*(c)vcv/c-	*(c)vcv/c-	Nikolayev & Starostin 1994
3. *ħvmvs(:)- ‘grouse’	Avar ʃansá, Inxoqwar *eʷzo* < *ʃvms-	Chamalal *musːija*, Archi *musːal* < *mvsː-	NS 225, 530
4. *ħvnvcː’- ‘honey’	Avar *hocː’ó*, Andi *huncː’i* < *ħvncː’-	Tsez *nuci*, Hunzib *nucu* < *nucː’ə	NS 824
5. *ħvrvqː’- ‘house’	Inxoqwar *aʷq*, Andi *haqː’u*, Botlix *hanqː’u* < *hanqː’- < *harqː’-	Avar *ruqː’*	NS 522
6. *ħvƛ’vn- ‘bottom’	Andi *hilmu, hinƛ’u*, Bezhta *oʷƛ’o* < *hvnƛ’- < *hvƛ’n-	Avar *tʼínu*, Lak *čʼan* < *ƛ’vn-	NS 590
7. *bvƛ’vr- Avar *betʼér* ‘head’	Hinuq *miƛ’u*, Bezhta *möƛ’ö* ‘beak’ < *bvnƛ’v < *bvrƛ’v- < *bvƛ’rv-; Batsbi *marɬo* ‘nose’	Aghul *kʼil*, Tabassaran *kʼul* ‘head’, Archi *ƛil-li-ƛ’* ‘under the head’ < *ƛ’vl-	NS 1041
8. *ħvkːʼvl- ‘mouse’: Lak *uʿkʼulu*	Bezhta *aʷqʼo*, Avar ʃunkː’, Andi *hinkːʼu* < *ħvnkːʼv- < *ħvlkːʼv- < *ħvkːʼlv-	Lak *kʼulu*, Aghul, Tabassaran *qʿul* < *kːʼvl	NS 523, 935

Paradigms that show this alternation are not preserved in attested East Caucasian languages: they are reconstructed on the basis of a comparison of lexical cognates. These complex paradigms were almost completely replaced by more regular paradigms in the individual languages, by the generalization of one of the two stems, as the examples show. Only very rarely do we find fossils such as Lak *barz,* genitive *zurul* ‘moon’ or Asax Tsez *uži*, ergative *žojä* ‘boy’ (cf. Imnajshvili 1963: 73).

As examples 6, 7 and 8 indicate, if syncope of the second syllable resulted in a sequence of obstruent + resonant, metathesis occurred to the phonotactically admissible sequence resonant + obstruent. Recall that metathesis was also assumed to have affected the word for ‘camel’ in Nakh, where **akmɨl-* became Proto-Nakh **amkɨl-*. This is important because it shows how well the word for ‘camel’ was integrated in prehistoric East Caucasian morphophonology to all intents and purposes, it behaves like a native word. Suppose that it was exactly that: Proto-Nakh **amkɨl* < **akmɨl* may in that case reflect a Proto-East-Caucasian oblique **ɨkmíl-*, which was the oblique stem (with pretonic unrounding of **ɔ* to **ɨ*) to an absolutive **ɔ́kmɨl*, just like the Nakh word for ‘glass’, **ingal- ~ *angal-* reflects Proto-East-Caucasian oblique **ħɨngVl/r-* to an absolutive **ħɔngV(l/r)* (section 5.1.3).

Against this background it may be more than accidental that example 3, *ʕvmvs(:)-, whose representatives mean ‘grouse, black grouse, *ular* (Caucasian turkey)’ looks like the European substratum word *amsl-, *mesal- ‘blackbird’.

7 Conclusions and outlook

We have seen that Caucasian languages possess a word for ‘camel’ that is so similar to Semitic *gam(a)l- and Greek κάμηλος that it is probably related. A detailed reconstruction of this word in the Nakh group of East Caucasian reveals a Proto-Nakh paradigm *imkil, oblique *amkil-, *amkalV, which goes back to pre-Proto-Nakh *ɨmkɨl or *amkɨl, in a paradigm with accentual mobility. In order to be able to connect this with Georgian *aklemi* and with the Semitic and Greek forms, the original position of the *m was probably after rather than before *k: *ɨkmɨl or *akmɨl (section 5.5). The latter relates to Semitic *gam(a)l- and Greek κάμηλος according to a pattern that is familiar from a series of European substratum words, such as *amsl- ~ *mesal- ‘blackbird’ and *arud- ~ *raud- ‘ore’ (section 6). While the latter substrate is implicated in the initial spread of agriculturalists from Asia Minor into Europe from the seventh millennium BCE onwards, the word for ‘camel’ is unlikely to be that old: the domesticated camel reached Eastern Europe and the Near East probably no earlier than the third millennium BC.

The morphophonological pattern observed in *amsl- ~ *mesal- and *akmɨl- ~ *gam(a)l-, *kameːl- has a close counterpart in the morphophonology of the native East Caucasian lexicon, which may indicate that the language that donated the European substrate lexemes was of East Caucasian stock (section 6). The word for ‘camel’ is either a native East Caucasian lexeme, which found its way into Greek and Semitic together with the domesticated Bactrian camel (scenario 2 of section 4); or it is of Semitic origin (and in that case probably originally denoted the dromedary) and spread into East Caucasian so early that it was completely nativized according to the rules of East Caucasian morphology (scenario 1 of section 4). The latter scenario is faced with a serious problem, however. In East Caucasian, the initial syllable that now appears and then disappears in examples like Tsez. *ac* ‘door’ < *ancːʾ- < *ħanVcːʾ-, Avar. *nucːʾá* ‘door’ < *nVcːʾ- (see the examples in the table in section 7) is not a prefix but the initial syllable of the root of the lexemes in question. Hence, if the connection of East Caucasian *ħVmVs(:)- ‘grouse’ with European substrate *amsl- ~ *mesal- ‘blackbird’ is correct, the *a- in the latter is not a prefix either. Nor is the *a- in the word for ‘camel’, *akmɨl- ~ *gam(a)l-, *kameːl-. If that is so, and if ‘camel’ is a loanword from Semitic into East Caucasian (scenario 1 of section 4), how, then, would it have acquired the initial *a-,

which is absent from Semitic? The converse is more likely: an East Caucasian alternating **akmɨl-* ~ **kVmVl-* was borrowed into Semitic and Greek on the basis of the latter root form. Hence it is necessary to come out in favour of scenario 2.

If 'camel' and/or the European substrate lexicon referred to above are indeed of East Caucasian origin, this has an important consequence. The East Caucasian family is attested in the northeastern Caucasus and is not known to have ever resided elsewhere. However, if the language of the European substrate lexicon is East Caucasian *and* the language of the first European farmers, that must mean that East Caucasian was more widely distributed in the Near East by the early Neolithic. Similarly, if 'camel' is a native East Caucasian lexeme, the fact that it became widespread in Semitic and ended up in Greek implies that East Caucasian could be found somewhere outside the Caucasus area, and more precisely even as far east as the area between Turkmenistan and eastern Turkey, by the third millennium BCE. And the fact that East Caucasian is nowadays spoken only in the Caucasus would put it on a par with Ossetic and Armenian, as a representative of a once widespread language family that managed to hold on to life only in the Caucasus while it was replaced by other languages families in the lowlands to the north and south. These are far-reaching conclusions, which require a stronger foundation than can be provided on the basis of our present state of knowledge. Whether the ideas expressed here will ever surpass the status of ideas will depend on progress on the reconstruction of East Caucasian, but a tantalizing piece of material may support the idea, however tentatively: Tocharian, an Indo-European language that was spoken in oases along the Silk Road in present-day Xinjang, possesses a word for 'elephant' that is formally similar to Nakh **akmɨl*, viz. Tocharian B **onkolmo*, Tocharian A *onkaläm* < Proto-Tocharian **enkelmo*.[16] For the plausibility of a semantic transition from 'camel' to 'elephant', compare the Sumerian neologism for 'Bactrian camel', **am.si.ḫar.an**, which literally means 'elephant of the road/caravan' (see section 3); also note the plausible etymological connection between Gothic *ulbandus* 'camel' and Greek ἐλέφᾱς, -αντ- 'elephant' (e.g. Lehmann 1986: 375).

Finally, to return to the beginning of this study, pre-Proto-Nakh **akmɨl-* and its presumed co-existence with a form without **a-*, **kamɨl-*, may be relevant to explaining why Greek κάμηλος has an initial *k-* and a front vowel in the second syllable, issues that are both left unexplained by the idea that the Greek word was borrowed from Semitic **gam(a)l-*.

16 I am indebted to Guus Kroonen for pointing me to this form. See Adams 1999: 113 for attempts to provide the word with an Indo-European etymology. The Tocharian word for 'camel' is hiding behind the Tocharian B adjective *partāktaññe* 'pertaining to a camel', which lacks a convincing etymology (Adams 1999: 358–59).

References

Adams, Douglas Q. 1999. *A Dictionary of Tocharian B*. Amsterdam/Atlanta: Rodopi.

Aliroev, Ibragim Junusovich. 1975. *Sravnitel'no-sopostavitel'nyj slovar'otraslevoj leksiki čečenskogo i ingušskogo jazykov i dialektov* [*Comparative Dictionary of the Word-Field Lexicon of the Chechen and Ingush Languages and Dialects*]. Maxačkala: Čečeno-Ingušskoe Knižnoe Izdatel'stvo.

Almathen, Faisal et al. 2016. Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proceedings of the National Academy of Sciences*, June 14, 2016. 6707–6712.

Bauer, Anna. 2020. Hieroglyphic Luwian */kamara?/*, *ka-mara/i-* '(a commodity)'. In *EDiAna* (*Digital Philological-Etymological Dictionary of the Minor Ancient Anatolian Corpus Languages*), Beta-version 0.2, consulted 1-1-2022 Available at https://www.ediana.gwi.uni-muenchen.de/dictionary.php?lemma=988.

Chantraine, Pierre. 1968. *Dictionnaire étymologique de la langue grecque*. Paris: Klincksieck.

Desheriev, Junus D. 1960. *Sovremennyj čečenskij literaturnyj jazyk, čast' I, Fonetika* [*The Modern Chechen Literary Language, Part I, Phonetics*]. Groznyj: Čečeno-Ingušskoe Knižnoe Izdatel'stvo.

Gagua, R.R. 1961. Sklonenie suščestvitel'nyx v Batsbijskom jazyke [The inflection of substantive nouns in the Batsbi language]. In E.A. Bokarev (ed.), *Voprosy izučenija iberijsko-kavkazskix jazykov* [*Questions of the Study of the Ibero-Caucasian Languages*], 72–88. Moscow: Izdatel'stvo Akademii Nauk.

Giginejshvili, Bakar K. 1977. *Sravnitel'naja fonetika dagestanskix jazykov* [*Comparative Phonetics of the Dagestan Languages*]. Tbilisi: Izdatel'stvo Tbilisskogo Universiteta.

Heide, Martin 2011. The domestication of the camel: Biological, archaeological and inscriptional evidence from Mesopotamia, Egypt, Israel and Arabia, and literary evidence from the Hebrew bible. *Ugarit-Forschungen* 42. 331–384.

Holisky, Dee Ann & Rusudan Gagua 1994. Tsova-Tush (Batsbi). In Rieks Smeets (ed.), *The Indigenous Languages of the Caucasus*, Vol. 4: *The North East Caucasian Languages*, Part 2, 147–212. Delmar/New York: Caravan Books.

Imnajshvili, David S. 1963. *Didojskij jazyk v sravnenii s ginuxskim i xvaršijskim jazykami* [*The Dido Language in Comparison with the Ginukh and Khwarshi Languages*]. Tbilisi: Izdatel'stvo Akademii Nauk Gruzinskoj SSR.

Imnajshvili, David S. 1977. *Istoriko-sravnitel'nyj analiz fonetiki naxskix jazykov* [*Historical-Comparative Analysis of the Phonetics of the Nakh Languages*]. Tbilisi: Mecniereba.

Iversen, Rune & Guus Kroonen. 2017. Talking Neolithic: linguistic and archaeological perspectives on how Indo-European was implemented in Southern Scandinavia. *American Journal of Archaeology* 121(4). 511–525.

Jakovlev, Nikolaj F. 1960. *Morfologija čečenskogo jazyka* [*Morphology of the Chechen Language*]. Groznyj: Čečeno-Ingušskoe Knižnoe Izdatel'stvo.

Kadagidze, Niko & David Kadagidze. 1984. *C̓ova-tušur-kartul-rusuli lek̓sikoni* [*Tsova-Tush-Georgian-Russian Lexicon*]. Tbilisi: Mecniereba.

Khalilova, Zaira. 2009. *A Grammar of Khwarshi*. Utrecht: LOT.

Kibrik, Aleksandr E. & Sandro V. Kodzasov 1990. *Sopostavitel'noe izučenie dagestanskix jazykov* [*Comparative Study of the Dagestan Languages*]. Moscow: Izdatel'stvo Moskovskogo Universiteta.

Kirtcho, Lyubov B. 2009. The earliest wheeled transport in southwestern Central Asia: new finds from Altyn Depe. *Archaeology, Ethnology & Anthropology of Eurasia* 37(1). 25–33.

Klimov, Georgij A. & Madzhid Sharipovich Xalilov. 2003. *Slovar' kavkazskix jazykov*. *Sopostavlenie osnovnoj leksiki* [*Dictionary of the Caucasian Languages. Comparison of the Basic Lexicon*]. Moscow: Vostočnaja Literatura.

Kohl, Philip L. 2007. *The Making of Bronze Age Eurasia*. Cambridge: Cambridge University Press.

Komri, Bernard & Madzhid Sharipovich Xalilov, Zaira Xalilova. 2015. *Grammatika beztinskogo jazyka* [*Grammar of the Bezhta Language*]. Leipzig: Max Planck Institut for Evolutionary Anthropology, and Maxačkala: Institut Jazyka, Literatury i Iskusstva im. G. Cadasy, Dagestanskij Naučnyj Centr Rossijskoj Akademii Nauk.

Kossmann, Maarten. 2005. *Berber Loanwords in Hausa*. Cologne: Rüdiger Köppe.

Kretschmer, Paul. 1892. Zum ionisch-attischen Wandel von ᾱ in η. *Zeitschrift für vergleichende Sprachforschung* 31. 285–296.

Kroonen, Guus. 2012. On the etymology of Greek ἄγλις and γέλγις 'garlic'. *Journal of Indo-European Studies* 40. 289–299.

Lehmann, Winfred P. 1986. *A Gothic Etymological Dictionary*. Leiden: Brill.

Mallory, James P. & Douglas Q. Adams (eds.). 1997. *Encyclopedia of Indo-European Culture*. London/Chicago: Fitzroy Dearborn Publishsers.

Masson, Émilia. 1967. *Recherches sur les plus aciens emprunts sémitiques en grec*. Paris: Klincksiek.

Matsiev, Axmet G. 1961. *Čečensko-russkij slovar'* [*Chechen-Russian Dictionary*]. Moscow: Gosudarstvennoe izdatel'stvo inostrannyx i nacional'nyx slovarej.

Militarev, Alexander & Leonid Kogan. 2005. *Semitic Etymological Dictionary. Volume II Animal Names*. Münster: Ugarit-Verlag.

Nichols, Johanna. 1994. Chechen. In Rieks Smeets (ed.), *The Indigenous Languages of the Caucasus*, Vol. 4*: The North East Caucasian Languages*, Part 2, 1–77. Delmar/New York: Caravan Books.

Nichols, Johanna. 1997. Chechen phonology. In A. S. Kaye (ed.), *Phonologies of Asia and Africa (including the Caucasus)*, 941–971. Winona Lake, IN: Eisenbrauns.

Nichols, Johanna 2003. The Nakh-Daghestanian consonant correspondences. In Dee Ann Holisky & Kevin Tuite (eds.), *Current Trends in Caucasian, East European and Inner Asian Linguistics. Papers in Honor of Howard I. Aronson*, 207–264. Amsterdam/Philadelphia: John Benjamins.

Nichols, Johanna. 2004. *Ingush-English and English-Ingush Dictionary*. London/New York: Routledge.

Nichols, Johanna. 2011. *Ingush Grammar*. Berkeley/Los Angeles/London: University of California Press.

Nichols, Johanna & Arbi Vagapov. 2004. *Chechen-English and English-Chechen Dictionary*. London/New York: Routledge.

Nikolayev, Sergei L. & Sergei A. Starostin. 1994. *A North Caucasian Etymological Dictionary*. Moscow: Asterisk Publishers.

Potts, Daniel T. 2004. Camel hybridization and the role of *Camelus bactrianus* in the ancient Near East. *Journal of the Economic and Social History of the Orient* 47. 143–165.

Puhvel, Jaan. 1984. *Hittite Etymological Dictionary.* Vol. 1: *Words beginning with A*, Vol. 2: *Words beginning with E and I*. Berlin/New York/Amsterdam: Mouton Publishers.

Salimov, Xangerej S. 2010. *Gagatlinskij govor andijskogo jazyka* [*The Gagatl Dialect of the Andi Language*]. Maxačkala: IJaLI.

Schrijver, Peter. 2018a. The origin of vowel alternation in Avar-Andi-Dido (North-East Caucasian), with special reference to the Dido languages. *International Journal of Diachronic Linguistics* 15. 199–223.

Schrijver, Peter. 2018b. Talking Neolithic: the case for Hatto-Minoan and its relationship to Sumerian. In Guus Kroonen, James P. Mallory & Bernard Comrie (eds.), *Talking Neolithic* (Proceedings of the workshop on Indo-European origins held at the Max Planck Institute for Evolutionary

Anthropology, Leipzig, December 2–3, 2013), 336–374. Washington: Institute for the Study of Man.
Schrijver, Peter. 2021. A history of the vowel systems of the Nakh languages (East Caucasian), with special reference to umlaut in Chechen and Ingush. *Languages of the Caucasus* 5. 84–151.
Tischler, Johannes. 1977. *Hethitisches etymologisches Glossar*, Lieferung 1. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
Uerpmann, Hans-Peter & Margarethe Uerpmann. 2002. The appearance of the domestic camel in southeast Arabia. *Journal of Oman Studies* 12. 235–260.
Xalilov, Madzhid Sharipovich 1995. Bezhtinsko-russkij slovar’ [Bezhta-Russian Dictionary]. Maxachkala: Institut Jazyka, Literatury i Iskusstva im. G. Cadasy.
Xalilov, Madzhid Sharipovich 1999. *Cezsko-russkij slovar'* [*Tsez-Russian Dictionary*]. Maxachkala: Institut Jazyka, Literatury i Iskusstva im. G. Cadasy.
Xalilov, Madzhid Sharipovich & Isak Abdulvaxidovich Isakov. 2005. *Ginuxsko-russkij slovar'* [*Ginukh-Russian Dictionary*]. Makhachkala: Institut Jazyka, Literatury i Iskusstva im. G. Cadasy.
Yakubovich, Ilya. 2016. Some transitive motion verbs and related lexemes in Late Luwian. *Indogermanische Forschungen* 121. 69–92.

List of contributors

Anthony Jakob was born in London in 1992. After completing a dual honors BA in Russian and Music at the University of Sheffield, he moved to Leiden. Here, he completed an MA and subsequently a PhD at the Leiden University Center for Linguistics in 2023 as part of the EUROLITHIC project under the supervision of Guus Kroonen. His dissertation discussed the East Baltic languages in the context of language contact, on which basis he published a book *A History of East Baltic through Language Contact* (Brill) in 2024. His research interests also include broader topics of Indo-European and Uralic linguistics.

Anders Richardt Jørgensen was born in Copenhagen in 1977. He obtained his PhD in Indo-European linguistics from the University of Copenhagen in 2007. His research focuses on comparative Indo-European and Celtic linguistics, Middle Breton philology and Modern Breton dialectology. He is currently a researcher at Uppsala University in the project *Languages and Myths in Prehistory* (LAMP).

Guus Kroonen is associate professor at the Leiden University Center for Linguistics and part-time research professor at the Department of Nordic Studies and Linguistics of the University of Copenhagen. He studied Scandinavian languages and cultures at the University of Amsterdam as well as comparative Indo-European linguistics at Leiden University, where he later received his PhD. His research interests include the historical linguistics of the Nordic, Germanic and Indo-European languages, the contact linguistics of the Indo-European subgroups and the interdisciplinary study of the Indo-European language dispersal. In addition to his papers, key research contributions are his *Etymological Dictionary of Proto-Germanic* (Brill, 2013) and *The Indo-European Puzzle Revisited* (Cambridge University Press, 2023). In 2016, he became the principal investigator of the ERC-funded research project on *The Linguistic Roots of Europe's Agricultural Transition* (EUROLITHIC), which allowed him to prepare the present volume.

Ranko Matasović was born in Zagreb in 1968. In 1995 he received his PhD from the University of Zagreb, where he is full professor and Head of the Chair of Comparative Indo-European Linguistics. In 1997 he was a Fulbright fellow in the University of Wisconsin (Madison) and in 2003 he became a Humboldt fellow in Bonn University. He was also a visiting scholar or guest professor in the universities of Oxford, Vienna, Leiden, Heidelberg, Athens (Georgia) and Leipzig. In 2012 he became a regular member of Croatian Academy of Sciences and Arts. His research is focused on Indo-European, Celtic and Slavic languages, as well as on language typology. Ranko Matasović was awarded the Croatian Academy's Prize for a lasting contribution to science (in 2002) and the Croatian Philological Association's Bulcsú László Prize in 2019. He published more than 150 articles and thirteen books, including Gender in Indo-European (Heidelberg, 2005), Etymological Dictionary of Proto-Celtic (Leiden, 2008) and An Areal Typology of Agreement Systems (Cambridge, 2018).

Lotte Meester is currently employed as a free-lance editor. She completed an MA in Comparative Indo-European Linguistics (Leiden University, 2020). During her studies, she co-developed and marketed an *etymological tear-off calendar*, i.e. the Dutch *Etymologische Scheurkalender* (2019, 2020), published by Brill Publishers. For this, she and her co-developers won the annual *Popularization Award* of the *Netherlands National Graduate School of Linguistics*. In the period 2019–2020, Lotte

https://doi.org/10.1515/9783111337920-014

Meester was employed as an intern at the ERC-funded research project on *The Linguistic Roots of Europe's Agricultural Transition*.

Peter Schrijver is professor of Celtic languages and culture at Utrecht University. He publishes on the history of languages of Ancient and Early Medieval Western Europe, with a focus on language reconstruction and on the effects of ancient language contact. In recent years he has also been working on the reconstruction of the East Caucasian language family. Recent publications comprise *Language Contact and the Origins of the Germanic Languages* (Routledge 2014); 'Talking Neolithic: The Case for Hatto-Minoan and its relation to Sumerian', in G. Kroonen et alii (eds.), *Talking Neolithic: Proceedings of the workshop on Indo-European origins held at the Max Planck Institute for Evolutionary Anthropology, Leipzig, December 2–3, 2013* (Washington: Institute for the Study of Man 2018), 336–374; 'Pruners and trainers of the Celtic family tree: the rise and development of Celtic in the light of language contact', in L. Breatnach et alii (eds.), *Proceedings of the XIV International Congress of Celtic Studies* (Dublin: School of Celtic Studies, Dublin Institute for Advanced Studies 2015), 191–219.

Zsolt Simon (Budapest, 1981; PhD 2013 Eötvös Loránd University Budapest) is a research fellow of the Hungarian Research Centre for Linguistics. Previously he was junior research fellow at the Research Center for Anatolian Civilizations of the Koç University (Istanbul), a TÜBITAK research fellow at the same university, and a research associate of the Institut für Assyriologie und Hethitologie of the Ludwig-Maximilians-Universität München. Next to numerous papers focusing especially on Anatolian languages, the Neo-Hittite world, and the loan contacts of the ancient Indo-European languages, he is a co-author of the *Digital Philological-Etymological Dictionary of the Minor Ancient Anatolian Corpus Languages* (München – Marburg, 2016-) and a co-editor of the *Hungarian Assyriological Review* (2020). Since 2024, he is a Ramón y Cajal researcher at the Institut del Pròxim Orient Antic of the University of Barcelona.

Paulus van Sluis completed a BA and MA in Celtic studies at Utrecht University, graduating with a thesis on lenition following verbs in Middle Welsh. He then joined the Fryske Akademy, where he produced a report on the sociolinguistic vitality of Bildts, a minority lect spoken in the north of the Netherlands, as well as an article on the distribution of Dutch and Frisian grammar in Bildts. He then wrote his PhD thesis at the University of Wales Centre for Advanced Welsh and Celtic Studies on the phonological and orthographical development of lenited voiceless stops in British Celtic and Middle Welsh. In his role as postdoctoral researcher in the EUROLITHIC project at Leiden University, Paulus has researched various aspects of Celtic historical linguistics and linguistic paleontology, with a personal focus on prehistoric Celtic-Germanic language contact, the linguistic paleontology of beekeeping in Celtic, and the substrate languages found in the Celtic lexicon. Paulus additionally takes keen interest in visualization of linguistic and geographical data.

David Stifter is professor of Old and Middle Irish at Maynooth University in Ireland. He is founder and editor of the interdisciplinary Celtic Studies journal *Keltische Forschungen* and founding member of *Societas Celtologica Europaea* (European Association of Celtic Studies scholars). His research interests are language variation and change in Old Irish, the epigraphy and philology of the Ancient Celtic languages (Lepontic, Celtiberian, Gaulish, Ogam Irish), comparative Celtic linguistics and the Indo-European context of Celtic. His introductory handbook *Sengoídelc: Old Irish for Beginners* (Syracuse University Press 2006) has been adopted for teaching Old Irish in universities worldwide. *Cis-alpine Celtic* and *Ogam* (Prensas de la Universidad de Zaragoza 2020 and 2022) are introductory handbooks to those epigraphic traditions. Research projects include a dictionary of the Old Irish

glosses in the Milan Codex Ambrosianus C 301 inf., *Lexicon Leponticum*, the ERC-funded *Chronologicon Hibernicum*, the IRC-funded *DiAgnostic*, and several projects for the digitisation of early medieval Irish epigraphy in the Ogam alphabet and the Latin script.

Cid Swanenvleugel completed his BA and MA in comparative Indo-European linguistics in Leiden, where he graduated with an MA thesis on shared innovations of Italic and Celtic. He is currently a PhD candidate at Leiden University on the ERC-funded EUROLITHIC project. His research focuses on the pre-Roman elements found in the Sardinian language and their implications for our understanding of Sardinian and Mediterranean linguistic prehistory. His main interests are historical Romance linguistics and language contact in the Mediterranean. Additionally, he is interested in language conservation and revitalization.

Rasmus Thorsø obtained a BA and MA degree in Indo-European Linguistics from the University of Copenhagen. His MA thesis, concerning the shared lexicon of Armenian and Greek, was awarded the prize for best master's work by the Indogermanische Gesellschaft in 2017. In 2018, Thorsø moved to Leiden University as a PhD candidate under the EUROLITHIC project, and in 2023, he was awarded the PhD degree with the thesis *Prehistoric loanwords in Armenian,* which examines three of the oldest strata of loanwords in the Armenian language. Thorsø has published mainly on Indo-European and the linguistic history of Armenian, but his research interests also encompass the languages of the Caucasus, in particular Kartvelian and East Caucasian.

Andrew Wigman completed his undergraduate education at the Pennsylvania State University, earning BScs in Archaeological Science and World Language Education as well as a BA in Classics & Ancient Mediterranean Studies. His honors thesis, submitted to fulfill the graduation requirements of the Schreyer Honors College, documented the status of language revitalization and education being undertaken by five Native American groups. With a scholarship through the *Deutscher Akademischer Austauschdienst*, he earned a Master's degree in Comparative Indo-European Linguistics at the Ludwig Maximilian University in Munich. His Master's thesis "Linguistic and Archaeological Insights on the Migration of the Proto-Tocharians" discussed the details and timing of the split of Tocharian from Proto-Indo-European and its arrival in the Tarim Basin using linguistic, genetic, and archaeological lines of evidence. He earned his PhD from Leiden University as part of the ERC-funded EUROLITHIC project. His dissertation *Unde venisti? The Prehistory of Italic through its Loanword Lexicon* investigates the layers of non-Indo-European vocabulary within the Italic languages to shed light on how and when the family entered the Italian peninsula.

Index of cited forms

https://doi.org/10.1515/9783111337920-015