

On the comparative method, internal reconstruction, and other analytical tools for the reconstruction of the evolution of the Basque language: An assessment

José Ignacio Hualde*

University of Illinois at Urbana-Champaign

ABSTRACT: This paper is an attempt to present the state of the art in Basque historical phonology. The accomplishments and limitations of different methodologies are evaluated. These methodologies include the application of the comparative method to Basque dialects, the analysis of old borrowings in Michelena's work, internal reconstruction, and Lakarra's canonical root hypothesis. I also discuss the possibilities afforded by internal reconstruction and root theory for discovering genetic relationships between Basque and other languages, focusing on recent proposals.

KEYWORDS: Comparative method, internal reconstruction, Proto-Basque, Old Common Basque.

1. Introduction

In this paper I would like to evaluate the results of the different techniques that can and have been used to reconstruct different stages in the evolution of the Basque language over the last few millennia. I believe it may be useful at this point in the development in the field of Basque Historical Linguistics to take stock of what are securely-established achievements in the reconstruction of the (pre-)history of the Basque language, focusing on the methods that made them possible; and assessing

* **Corresponding author:** José Ignacio Hualde. Dept. of Spanish and Portuguese & Dept. of Linguistics (University of Illinois at Urbana-Champaign) (4080 FLB - Urbana, IL 61801, USA) – jihualde@illinois.edu – <https://orcid.org/0000-0003-4969-2995>

How to cite: Hualde, José Ignacio (2020 [2022]). «On the comparative method, internal reconstruction, and other analytical tools for the reconstruction of the evolution of the Basque language: An assessment», *ASJU*, 54 (1-2), 19-52. (<https://doi.org/10.1387/asju.23021>).

Received: 2021-01-12; Accepted: 2021-03-26. Published online: 2021-08-09.

ISSN 0582-6152 - eISSN 2444-2992 / © 2022 Instituto-Seminario de Filología Vasca «Julio de Urquijo» (UPV/EHU)



This work is licensed under a
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

whether significant further progress has actually been made by the application of other methodologies that have been used in more recent work. Necessarily, this review of the state of the art will reflect the theoretical commitments of the writer. Other people may disagree.

First, in section 2, I define the terms that I will be using in this paper for the different stages under discussion. In section 3, I consider the reconstruction of Old Common Basque, the last common ancestor of all living and historically attested Basque dialects, through the comparative method. Then, in section 4, I consider Michelena's masterful application of essentially the same methodology to the comparison of early loanwords with their Latin and Romance sources in order to obtain information about an older stage than Old Common Basque. In section 5, I discuss other techniques, including internal reconstruction and Lakarra's canonical root-based methodology. Recent extensions of Lakarra's methodology with comparative goals in mind, including Blasco Ferrer's Proto-Basque-Paleo-Sardinian hypothesis and Blevins's Proto-Indo-European-Euskarian hypothesis, are discussed in section 6. Finally, section 7 offers a summary.¹

2. On the concept of Proto-Basque

A «proto-language» can be defined as the last common ancestor of a group of related languages or language varieties. My preference would, thus, be to reserve the term *Proto-Basque* to refer to the last common ancestor of all living and historically attested Basque varieties, which Michelena (1981) dates to the early Middle Ages. However, the term Old Common Basque (in Basque, *euskara batu zaharra*) is the one that has become established in the last few decades, and that is the name I will use here for this proto-language, abbreviated OCB. Instead of the three-word name *Old Common Basque* to refer to Proto-Basque, in the sense defined in the first sentence of this paragraph, it would seem that the shorter expression *Common Basque* should be enough. However, its Basque equivalent, *euskara batua*, has the drawback that this is precisely the name that has been given to modern Standard Basque, so that the qualification *zaharra* 'old' becomes necessary. The term *euskara batu zaharra* appears to be originally due to Lakarra (1997: 516, fn. 104, 2011b [2014]).

Michelena (1981), in fact, makes a distinction in some places between Common Basque (*vasco común*) and Proto-Basque (*protovasco*), where a given feature is Common Basque if it is found in all Basque dialects, independently of whether it was present before dialectal diversification started or is a later common innovation (see also Lakarra 1995). This is a conceptually important distinction, since we do find many, even very recent, innovations that have spread throughout the entire Basque-speaking territory. These include, to mention a few important ones, with different chronologies, the grammaticalization of free demonstratives as bound articles, the creation of a new second person plural pronoun with corresponding

¹ A post-scriptum has been added where I discuss Blevins & Sproat (2021), which was published after this paper had already been accepted and formatted for publication.

verbal forms, and a radical restructuring of the time/aspect/mood system (Mounole 2014; Mounole & Lakarra 2018).

For the first centuries of our era, and thus before the medieval OCB language, we have inscriptions containing names in one language or group of closely related dialects that is traditionally known as Aquitanian, since the first inscriptions of this type were discovered in the territory of Roman Aquitania, although more recently epigraphic evidence for the same language has also been found south of the Pyrenees, in Navarre and even Castile. Gorrochategui (2020a) refers to the language of this whole area as Aquitanian-Vasconic. The Aquitanian(-Vasconic) names show an evident relation to Basque, but what the exact nature of this relation is remains uncertain. The language of the Aquitanian names may be either the direct ancestor of OCB or a close relative (see Gorrochategui 2015; Campbell 2017). Since the Aquitanian-speaking area appears to have been rather large (as Aquitanian-type names have been found as far south as Soria, in Castile), it is also likely that this language was dialectally diversified, in which case OCB would directly continue one of the Aquitanian dialects. Some Latin borrowings into Basque may also predate OCB. It is thus convenient to have a label such as Proto-Basque-Aquitanian for the reconstructed common ancestor of OCB and other dialects of the Aquitanian (or Aquitanian-Vasconic) group.

Any comparison between Basque and Aquitanian has to be made with the caveat that, whereas we know what the Basque words mean, for Aquitanian we only have proper names. It seems almost evident that, for instance, the Aquitanian woman's name ANDERE must be identified with Bq *andere*, *andre* 'woman, lady', used as a proper name in the Middle Ages (Gorrochategui 1984, with the identification already in Luchaire 1877). Similarly the equations² Aq VMME = Bq *ume* 'child', Aq SAHAR = Bq *zahar* 'old' (VMME SAHAR in Lerga, Navarre, Michelena 1961), Aq SESENCO (in Soria, Gorrochategui 2009, 2011) = Bq *zezenko* 'little bull', Aq SENICCO, SENIPONNIS, etc = Bq *sein*, *sehi* (pre-OCB **seni*) 'child' and quite a few others seem eminently reasonable, but the fact is that, lacking, as we are, the semantic component, we cannot have the same certainty as when we equate, say, Sp *hoja* 'leaf' with Fr *feuille* 'leaf' or Lat *equus* 'horse' with Gr *hippos* 'horse' and Sansk *aśva* 'horse'.

Before Proto-Basque-Aquitanian, one may define as many Pre-Proto-Basque-Aquitanian stages as one finds expedient. The dating of these stages, however, becomes impossible to establish.

3. The application of the comparative method in the reconstruction of OCB

As historical linguists, the most reliable tool for the reconstructions of earlier stages of a language that we have in our kit is the comparative method: the comparison of cognates across language varieties in the search for systematic correspondences

² The following abbreviations are used in this paper for languages and dialects: Aq = Aquitanian, Ar = Araban Basque, Ast = Asturian, Bq = Basque, Bizk = Bizkaian Basque, Eng = English, Fr = English, Gip = Gipuzkoan Basque, Gr = Greek, Lap = Lapurdian or Labourdin Basque, LN = Low Navarrese Basque, Lat = Latin, NBq = Northern Basque, OCB = Old Common Basque, PIE = Proto-Indo-European, Rom = Romance, Sansk = Sanskrit, Zub = Zuberoan or Souletin Basque, Sp = Spanish.

among phonemes. In the case of Basque, this method is, unfortunately, of limited application, given, on the one hand, how few and generally superficial the differences among Basque dialects are and, on the other, the fact that no other language is immediately identifiable as genetically related to Basque, in the sense that Sanskrit, Greek and Latin are readily identifiable as having sprung from a common source from a quick look at morphological paradigms (see Michelena 1963).

When we compare Basque dialects, there are few significant differences in segmental phonology where the application of the comparative method would lead us to postulate proto-forms for the last common stage that are significantly different from those found in any historically attested variety or where questions arise regarding which of the attested variants is the oldest one. Things are only different in a few domains, such as prosody, where we do find a great diversity of accentual systems, and, to some extent, also in inflectional morphology.

OCB, as a proto-language, can be reconstructed by the application of the comparative method to its descendants, the dialects of the Basque language. In this section, I briefly review some of the main inter-dialectal differences in phonology and the proposals that have been made for the proto-language. I try to separate the results that we obtain through the application of the comparative method (OCB) from the additional information for somewhat earlier stages that we can gather through internal reconstruction (pre-OCB).

3.1. Nasalized vowels in OCB and pre-OCB intervocalic /-n-/

Michelena (1963, 1977) reconstructs **ardano* ‘wine’ by comparing the dialectal variants Bizk *ardau*, Gip *ardo*, Lap *arno*, Zub *ardū*, in addition to *ardan-* found everywhere in derived forms; **arrani* from *arrai*, *arrain*, *arran-* ‘fish’; and **seni* ‘child’ by comparing the attested dialectal forms Bizk *sein*, Ar *segi*, RS /s̺i/³ ‘child’, Lap-LN *sehi* ‘servant’, among quite a few other similar examples.

Notice that, since the weakening of intervocalic /-n-/, that is, VnV > ṼṼ, is common to all Basque dialects, strictly speaking the last common stage (OCB) had nasalized vowels in these words, e.g. OCB **ardāō* ‘wine’.⁴ It is by comparing this free form with the bound form *ardan-* (in *ardantz*a ‘vineyard’, etc.), that is, by internal reconstruction, that we can reconstruct a pre-OCB form **ardano* (pre-OCB **ardano* > OCB **ardāō*, *ardan-*).

Regarding words like pre-OCB **seni*, where the weakened /-n-/ was on the onset of the second syllable, the presence of (nasalized) aspiration in Northern Bq⁵ *sehi* may be ascribed to the common stage (-n- > -h̃- before a stressed vowel), which is

³ In forms within phonemic slanted lines I use /s̺/ for the voiceless apico-alveolar fricative represented as <s> in Basque orthography (and the roughly identical sound in Old Spanish) and /s/ for the voiceless lamino-alveolar or lamino-dental fricative written <z> and the similar or identical fricative that in Old Spanish and other medieval Romance languages resulted from the deaffrication of older dental affricates.

⁴ Julen Manterola (p.c.) points out to me that the form *ardāō*, with a diacritic over the last two vowels, is actually documented in a manuscript by the 16th century writer Esteban de Garibay («Endechas de doña Emilia de Lastur», photographically reproduced in Guerra 1924: 276).

⁵ Here I use the name «Northern Basque» to refer to Basque varieties spoken north of the Pyrenees.

the standard view (Michelena 1977: 411; Igartua 2015), or may reflect a later, stress-related development only in northern dialects, since the treatments of borrowings shows that in those dialects aspiration developed in that position after other segments as well, e.g. Gallo-Rom /oneʃt/ > NBq *onhest* ‘honest’, Rom /ʃola(t)s/ > NBq *solhas* ‘conversation’, etc. (Hualde 2018).

Accepting Aq SENI- as a cognate of **seni* ‘child’, as Michelena does, we conclude that the pre-OCB stage when intervocalic /n/-weakening took place postdates Proto-Basque-Aquitanian. If, as seems reasonable, we accept that this sound change did not take place independently from the identical sound change in the neighboring Gascon language (Lakarra 2011a [2014]), we obtain a more precise dating for it (with its inception in the 5th c. and intermediate stages before the complete deletion of the nasal stop, Chambon & Greub 2002; see also Hualde 2015 for the chronological ordering of this phenomenon with respect to the palatalization of velars in Romance).

3.2. Aspirated vs. unaspirated voiceless stops

Another dialectal difference of some importance is found in the plosive consonants. Since the first texts entirely written in Basque (from the 16th century), we find a three-way distinction north of the Pyrenees, voiced /b d g/ vs voiceless unaspirated /p t k/ vs voiceless aspirated /ph th kh/, but only a two-way contrast, voiced vs voiceless, in the rest of the country.

Given the fact that, in northern dialects, the aspirated stops are restricted to the onset of one of the two first syllables of the word—and, in native words, even more precisely to the onset of the second syllable—it seems reasonable to conclude, with Michelena (1977: 216, 2011: VI, 178), that the /p t k/ vs /ph th kh/ contrast in those dialects is historically secondary and has arisen through phonemic split, where the position of the stress-accent may have conditioned allophonic distribution at a somewhat earlier time; e.g. *ekhárr* ‘to bring’ vs **eákin* > *jákin* ‘to know’; with aspiration becoming contrastive after the shift of the accent from the pen-initial to the penultimate, e.g. **akhér* > *ákher* ‘billy goat’, now contrasting with *jákin* (Michelena 1977: 424, 2011: VI, 344-345). Notice that the shift of the stress-accent leaving one of its correlates behind requires the prior phonologization of stop-aspiration, which may have resulted from a Grassmann’s Law-like dissimilation in forms such as **harthu* > **hartú* ‘to take’, with an unaspirated stop on the onset of the stressed syllable, vs. **sarthú* (> Zub *hártü*, *sárhü*, after the shift) (Hualde 2018, and in press).

3.3. Post-nasal and post-lateral voicing

In all dialects except for the easternmost varieties, Roncalese (Erronkari) and Zuberoan, stops have become voiced after nasals and laterals. Lakarra (2011b [2014]) considers this to be one of the very oldest isoglosses in Basque, separating Zuberoan-Roncalese from the rest of the dialects. The examples in (1) illustrate this dialectal difference (for Roncalese, see Estornés Lasa 1997; for Zuberoan, see Larrasquet 1939):

(1)	Central	Roncalese	Zuberoan	OCB	
	<i>ongi</i>	<i>onki</i>	<i>hunki</i>	*/(h)on-ki/	‘well’
	<i>igande</i>	<i>igante</i>	<i>igante</i>	*/igan-te/	‘Sunday’
	<i>alde</i>	<i>alte</i>	<i>althe</i>	*/al-te/	‘side’
vs.	<i>aldi</i>	<i>aldi</i>	<i>aldi</i>	*/aldi/	‘time (Fr foix, Sp vez)’
	<i>beldur,</i>	<i>beldur</i>	<i>beldür</i>	*/beldur/	‘fear’
	<i>bildur</i>				

The comparison of loanwords confirms that the Zuberoan-Roncalese dialect is conservative in this respect (i.e. that it has not undergone devoicing in this position, instead), since it has preserved the voicing feature of the etyma in this context, whether voiced or voiceless:

(2)	Lat	Central-Western	Roncalese	Zuberoan	
	<i>angelum</i>	<i>aingeru</i>	<i>aing(u)ru</i>	<i>aingürü</i>	‘angel’
	<i>in gyrum</i>	<i>inguru</i>	<i>unguru</i>	<i>ingürü</i>	‘surroundings’
	<i>incudem</i>	<i>ingude</i>		<i>ünkhüde</i>	‘anvil’
	<i>plantarium</i>	<i>landare</i>	<i>lantare</i>	<i>lanthari</i>	‘(young) plant’
Celt	<i>landa</i>	<i>landa</i>	<i>landa</i>		‘field; outside’

The sonorization of post-lateral and post-nasal stops, however, does not appear to have ever been a fully regular sound change, since we have exceptions in all dialects, including, for instance, *malko* ‘tear’ and *anker* ‘cruel’, and some suffixes such as diminutive /-ko/ do not seem to undergo voicing anywhere, e.g. *zezenko* ‘little bull’ (cf. Aq SESENCO, in Soria). Since, on the other hand, morpheme-internally, sequences agreeing in voice are more common than disagreeing ones in all dialects, including Zuberoan-Roncalese, it is possible that, as Michelena (1977: 355, 2011: VI, 291) suggests, a tendency to prefer voiced stops after nasals and laterals was already present in OCB, and, in a non-Neogrammarian fashion, was spreading throughout the lexicon. The dialectal difference would then be due to Zuberoan-Roncalese having abandoned this tendency to voice stops in this context at an earlier time.

3.4. Other segmental differences

Other interdialectal correspondences do not offer any doubt regarding which varieties have innovated and how. Thus, for instance, the correspondence /y/ (noted *ü*) in Zuberoan and neighboring varieties vs. /u/ elsewhere (e.g. Zub *sü* ‘fire’, elsewhere *su*) is no doubt due to the fronting of /u/ in Zuberoan. Similarly, the systematic sound correspondence that can be illustrated with /jan/ - /jan/ - /zan/ - /jan/ - /χan/ ‘to eat’ shows an evolution of word-initial palatal glides along the same path that we find in the history of Spanish. By internal reconstruction, the form /jan/ obtained through the comparative method can be further determined to derive from earlier */e-an/, with the same prefix as we find in, e.g., *e-torri* ‘to come’ and other old verbs (Michelena 1977: 408). Such hiatus-initial participles are not necessarily very old, since the earlier

hiatus has left remnants detectable in the distribution of aspiration in northern dialects (e.g. NBq *ekharri* ‘to bring’ vs **eakin* > *jakin* ‘to know’), as Michelena pointed out (see section 3.2 above), as well as in the position of the accent itself up to the present-day in Goizueta (in northwestern Navarre). In the local variety of Goizueta, the stress-accent normally falls on the second syllable, e.g. *artú* ‘to take’, (*i*)*kúsi* ‘to see’, but with yod-initial verbs we find initial accent instead, e.g. *yákin* ‘to know’, *yárrri* ‘to put’, etc. (Hualde 2018). Given these facts, presumably we could reconstruct a word-initial hiatus from interdialectal comparison as well, although in a more indirect way than through internal reconstruction.

3.5. Accentuation

Whereas segmental differences among Basque dialects are relatively small, we find, on the other hand, a great diversity of prosodic systems (see, Hualde 1997, 1999; Gaminde 1998). In Hualde (2012, and in press) a rather complex accentual system is reconstructed for OCB. This prosodic system would include a contrast between lexically accented and unaccented words, which nowadays we find in some Bizkaian regions, and, within the set of lexically accented words, a contrast between words with a high (or rising) accent and words with a low (or falling) accent. This contrast in tonal melodies is nowadays found in the speech of Goizueta and until very recently was also found in a large area of western Navarre, as well as, across the provincial boundary, in Oiartzun, Gipuzkoa. In addition, at the phrasal level, there would have been an initial tonal rise, from the first to the second syllable, as in modern northern Bizkaian varieties. The geographically widespread pattern of accent on the pen-initial syllable may have arisen through the reinterpretation and lexicalization of initial phrasal-boundaries as the locus of prominence at very different points in time in different areas (Hualde 2003).

This is what the comparative method leads us to conclude. Arguably no other hypothesis can account for the accentual correspondences that we can establish among Basque varieties (see also Egurtzegi & Elordieta 2013 and in press). However, since lexically-accented words turn out to be essentially borrowings and morphologically complex words, we can arrive at a much simpler prosodic system for a pre-OCB stage, where there were no accentual contrasts among words and, at the phrasal level, there was only a tonal rise at the beginning and a fall at the end of the phrase (already in Hualde 1993).

Michelena’s reconstruction of pen-initial accent for an old stage of Basque (which was made at a time when many important facts of dialectal variation had not yet been described) is not based on the comparative method, but on internal reconstruction. It is deduced from the distribution of /h/ and aspirated stops in eastern dialects. The conclusion that was reached in Hualde (1995) is that the system reconstructed by Michelena should be properly understood as being ancestral to Northern Basque, but that it cannot be accepted for OCB, since it cannot account for certain accentual facts in Western dialects. The strongest evidence is the following. In those Bizkaian varieties where monomorphemic native words are lexically unaccented, among the words in the accented class we find old borrowings like those in (3), which may be accented either on the first or the second syllable.

This is incompatible with the assumption that at the time of the borrowing the language had either initial or pen-initial accent:

- (3) Getxo: lexically accented borrowings (Hualde 1995):
- | | | | |
|----|--------------------------------|---|--|
| a. | <i>dénpora</i> ‘time’ | < | <i>témpora</i> |
| | <i>ántzar</i> ‘goose’ | < | <i>ānser</i> |
| | <i>léku</i> ‘place’ | < | * <i>luécu</i> (<i>lucco</i>) < <i>lócum</i> |
| | <i>sékula</i> ‘never’ | < | <i>sáecula</i> |
| b. | <i>autóno</i> ‘September’ | < | <i>autúmnum</i> |
| | <i>doméka</i> ‘Sunday’ | < | <i>domínica</i> |
| | <i>eskóla</i> ‘school’ | < | <i>schóla</i> |
| | <i>puxíka</i> ‘bladder’ | < | <i>vēsíca</i> |
| | <i>kinpúla</i> ‘onion’ | < | <i>caepúlla</i> |
| | <i>makilla</i> ‘walking stick’ | < | <i>baccilla</i> ⁶ |

If pen-initial accent had been the regular pattern at the time of the borrowing, loanwords accented on the second syllable would have joined the majority class and would be unaccented nowadays. Given the old dating of some of these borrowings, it seems unavoidable that for OCB we must reconstruct (phrase-)final accent, as we still have in Getxo and other Bizkaian varieties in the unmarked case. The high vs low accent contrast in Goizueta must also be reflected in some way in our reconstruction of OCB prosody.

3.6. Conclusions on the application of the comparative method for the reconstruction of OCB

A clear limitation for the application of the comparative method to the Basque dialectal data is that the differences that we find among dialects are relatively small. This is so except in a few specific domains such as the prosodic system, which we have just examined, and aspects of inflectional morphology that we are not considering here (see Gómez & Sainz 1995 on auxiliary verbs and Manterola 2015 on the *-ak* vs *-ek* dialectal alternation in the ergative plural).

Michelena (1963) made the entirely valid point that the comparative method does not require for the languages being compared to be genetically related. The only requirement is that the *items* being compared should be cognates. That is, the method can be applied to borrowings. In this way, by analyzing the adaptation of Latin and early Romance loanwords, he was able to reconstruct earlier stages of Basque than OCB and provide evidence for important sound changes. Since the same adaptations are found in all Basque varieties, the application of the comparative method to these varieties would not be effective. It is by comparing the Basque words with their Latin etyma that we can make diachronic progress. I consider this topic in the next section.

⁶ Julien Manterola (2019: 1187-1188) notes that for the source of this borrowing a plural Latin form *baccilla* is unexpected, and suggests instead that the source must have been a verbal derivative, which is actually attested in Occitan.

4. The comparative method applied to early borrowings

As just mentioned, for the comparative method to be applicable, the items compared need to be genetically related, but the languages do not have to have a common source. Thus, Michelena (1963) points out that one could find a good number of cognates (of Latin origin) when comparing Basque and Italian or even Albanian. Michelena in fact showed that we can learn much about the evolution of Basque in the first centuries of our era by observing how borrowings were adapted.

In this comparison, there is no issue of reconstructing proto-forms. We know that Basque borrowed the relevant words from Latin/Romance, but the comparison offers us information about the evolution of the Basque language some centuries before the OCB stage. Notice the time limitation: the analysis of changes in loanwords may allow us to go further back in time than OCB, which, as mentioned, developed in the early Middle Ages, but we can only go as far back as the time when the loanwords were borrowed.

To give a sense of the procedure, let us review some of the sound changes for which the evidence comes only or primarily from the adaptation of borrowings, all of them identified in Michelena's work (although the way they are formulated here does not always coincide with Michelena's formulation). Since this is just an illustration of the method, we will be concerned with only some of the most important sound changes and adaptations.

4.1. Voicing of word-initial stops

The facts are not in dispute: in the earliest borrowings from Latin, word-initial voiceless stops become voiced, whereas the voiced vs voiceless distinction is preserved in word-internal intervocalic position (Michelena 1977, examples taken from Hualde 2020):

(4) Adaptation of Latin stops

a. Word-initial stops

/b-/ & /p-/	>	/b-/	<i>benedica-</i> > <i>bedeinka</i> 'bless'
			<i>pacem</i> > <i>bake</i> 'peace'
/d-/ & /t-/	>	/d-/	<i>digitalem</i> > <i>ditare</i> 'thimble'
			<i>taratrum</i> > <i>daraturu</i> 'drill'
/g-/ & /k-/	>	/g-/	<i>gulam</i> > <i>gura</i> 'desire'
			<i>cellam</i> > <i>gela</i> 'room'

b. Word-medial stops

/-b-/	>	/-b-/	<i>habēre</i> > <i>abere</i> 'domestic animal'
/-p-/	>	/-p-/	<i>sapōrem</i> > <i>zapore</i> 'flavor'
/-d-/	>	/-d-/	<i>auditum</i> > <i>aditu</i> 'to listen'
/-t-/	>	/-t-/	<i>anātem</i> > <i>ahate</i> 'duck'
/-g-/	>	/-g-/	<i>rēgem</i> > <i>errege</i> 'king'
/-k-/	>	/-k-/	<i>lacum</i> > <i>laku</i> 'lake'

Regarding native words, what we find is consistent with the way borrowings have been adapted: word-initial stops are almost always voiced.⁷ In fact, with exceedingly few exceptions, the only initial stops in native Basque words are /b-/ and /g-/ (/d-/ only in inflected verbs). In word-internal intervocalic position, on the other hand, the voiced vs voiceless contrast is robust (also postconsonantly, e.g. *arte* ‘living oak’, *ardi* ‘sheep’, except for the facts noted in 3.3):

(5) Distribution of stops in the native vocabulary

	p	t	k	b	d	g
Initial	—	—	(<i>ke</i>) ‘smoke’	<i>beso</i> ‘arm’	(<i>daki</i>) ‘knows’	<i>gorri</i> ‘red’
Medial	<i>apar</i> ‘foam’	<i>ate</i> ‘door’	<i>aker</i> ‘b. goat’	<i>abar</i> ‘branch’	<i>eder</i> ‘beautiful’	<i>ager</i> ‘appear’

In (5), parentheses around an example indicate that the pattern is exceptional. Only a small handful of words have initial /k-/ and, at least for some of these, there is evidency that they have undergone aphaeresis. This is the case for *ke* ‘smoke’ in (5), which has a dialectal variant *eke*, and *kendu* - *ekendu* - *gentu* ‘remove’. Word-initial /d-/ is also exceptional. Leaving borrowings aside, it is only found in inflected verbal forms. Michelena (1977: 257-258) explains this gap by lateralization of word-initial dental stops; /d-/ > /l-/.

What is disputed and has been the object of much discussion is how exactly this situation came about. This is an interesting situation because neutralization in favor of the voiced series in word-initial position is unexpected from what we know about sound change and, furthermore, goes against what we find in the neighboring Western Romance languages, where the Latin opposition is maintained word-initially but there is systematic intervocalic voicing of /p t k/ (although, notably, with exceptions in Pyrenean Gascon and Aragonese varieties, perhaps due to their Basque-Aquitainian substrate, see Saroihandy 1913, Rohlfs 1977, Allières 1992). For different proposals, see Martinet (1950), Trask (1985), Hualde (1999).

Michelena (1977) assumes a phonological system with word-initial neutralization of the two stops series for the Basque of the Roman Empire, whereas for Martinet (1950) this state of affairs would have been a later development, caused by the loss of strongly aspirated word-initial stops. Since in the Aquitainian names we find examples of word-initial /t/ and perhaps /k/, we may conclude that the neutralization of word-initial stops in favor of the voiced series postdates Proto-Basque-Aquitainian.⁸

⁷ Exceptions to the generalization that word-initial stops are never voiceless in old borrowings and native words have been explained in various ways, including assimilation in voicing/voicelessness (*galte* > *kalte* ‘harm’), dissimilation (*gibel* > *kibel* ‘back’), the influence of Romance cognates in the case of loanwords, phonosymbolism and four-part analogy (*bekatu* - *pekatu* ‘sin’ : *bidaiia* - x ‘trip’) (Gavel 1920; Martinet 1950: 225; Michelena 1977: 240-244 [2011: VI, 197-200]; Trask 1997: 129-130; Lakarra 2009).

⁸ It is hard to know what to think of apparent alternations like TALSCONIS and HALSCONIS in the Aquitainian onomastics. We may be dealing with two variants of the same name, like *Eddy* and *Teddy*, or with unrelated names, like *Annie* and *Danny*.

4.2. Deaffrication of word-initial affricates

In Basque, the contrast between the voiceless fricatives orthographically noted *s*, *z* (apical and laminal, respectively) and the corresponding affricates *ts* and *tz* is robust word-medially, but word initially we only find the fricatives. Since Latin lacked affricates, the treatment of affricates in loanwords can only be observed from the moment these segments arose in Romance. What we see is that they are systematically simplified, e.g. Rom /tselu/ > Zub *zeliu* ‘sky, heaven’, other Bq *zeru*; Rom /tsentena/ > Bq *zendea* ‘county, district’. Michelena (1977) unifies the voicing of initial stops and the deaffrication of initial affricates within his proposal that, in the Basque spoken in Roman times, what are now a voiceless stop/voiced stop contrast and an affricate/fricative contrast were part of a single more general contrast between *fortis* and *lenis* consonants, with this contrast being neutralized in favor of the *lenis* series in word-initial position. Notice again that since Classical Latin did not have affricates, we actually cannot have any evidence for or against the existence of a constraint against word-initial affricates in Basque from the treatment of loanwords until these develop in Latin at a later stage, from the palatalization of velars.

As far as I am aware, there are no clear morphophonemic alternations in the native lexicon that would allow us to suspect that some native words that start with a fricative used to begin with an affricate at an earlier time. It would appear that we do not have any clear evidence to determine whether word-initial affricates were already absent in Roman times, or the situation was brought about by a later, but pre-OCB, development. If in Aq XEMBVS (Gorrochategui 2011), relatable to Bq *seme* ‘son’, the letter X represents an affricate /tʃ/, as seems likely, that could be taken as an indication that in Roman times at least some close relatives of the Basque language had word-initial affricates and that the change /tʃ-/ > /ʃ-/ may be more recent.⁹

4.3. Rhotacism of intervocalic /l/

This is another sound change in pre-OCB for which the evidence comes primarily from the treatment of borrowings; e.g. Lat *gulam* > Bizk *gura* ‘desire’, Lat *voluntātem* > Bq *borondate* ‘will’, etc. This sound change necessarily preceded the simplification of geminate /l.l/, mentioned immediately below.

Evidence for this sound change could also be gathered from morphophonological alternations in native words such as *gari* ‘wheat’, but in compounds *gal-*, for which

⁹ Gorrochategui (2011, 2015), however, suggests that X- may have represented a prepalatal fricative. This seems less likely to me. Notice that the use of <x> to represent /ʃ/ in Old Spanish seems to be directly related to the particular evolution of the group /ks/ in this language, e.g. Lat *dixi* /diksi/ > OSp *dixe* /dife/ ‘I said’. Incidentally, in Northern Basque, the letter <x> was traditionally used mostly to represent the affricate now noted <ts>. The presence of SEMBVS (Gorrochategui 1984) next to XEMBVS may be indicative of dialectal variants or may instead be due to the difficulties that the carvers of the inscriptions experienced in writing a sound /tʃ/ for which the Roman alphabet did not have a letter (cf. the great variety of spellings for the same affricate phoneme found in *RS*). This is not, however, a topic about which I am able to formulate a well-founded opinion, since I do not know enough regarding how the affricates found in names from other languages spoken in the Roman Empire were represented in inscriptions in the Roman alphabet, or what other values, besides /ks/, the letter <x> could have in Latin when writing non-Latin names.

we could reconstruct **gali*. Nevertheless, the existence of examples like *abere*, *abel-* ‘domestic animal’, from a descendant of Latin *habere* (Old Sp /aβeres/ ‘possessions’) forces us to be prudent in our reconstruction (see *EHHE*, s.v. *gari*).

4.4. Simplification of geminates

In loanwords, all Latin geminate consonants except for *-rr-* are simplified, e.g. Lat *cattum* > Bq *katu* ‘cat’, Lat *saccum* > Bq *zaku* ‘bag’, Lat *castellum* > Bq *gaztelu* ‘castle’, Lat *annōna* > Bq *anoa* ‘provisions’. In the neighboring Romance languages, the same degemination process took place, except for /l.l/ and /n.n/, which were kept for a longer time and had a different evolution from other geminates.

In the case of geminates other than alveolar sonorants, it is often hard or impossible to know whether the geminate was already simplified in the source of the borrowing. For instance, Bq *katu* may derive from Lat *cattum* /kat.tu/ or it may come from a later form /katu/. Although the results in Basque for all Latin geminates except for /r.r/ is the same, degemination, it is likely that the simplification of geminate /l.l/ and /n.n/ took place later than other degeminations; again, like in neighboring Romance languages, where these two geminates have a distinct evolution. In Castilian Spanish, the change /l.l/ > /k/, as in *castellum* > /kasti.ko/, necessarily postdates the delateralization of /k/ from other sources, as in Proto-Western Romance **/fi.kal/* > OSp /hi.za/ ‘daughter’ (< Lat *filiam*), PWR **/o.lu/* > OSp /o.ʒo/ ‘eye’ (< Lat *oculum*). Similarly, a relatively late age for the degemination of /-l.l-/ in Basque must be assumed. The change /-l.l-/ > /-l-/ must have postdated /-l-/ > /-r-/. Otherwise we would have, for instance, **gazteru* instead of *gaztelu* ‘castle’. As for the degemination of the alveolar nasal geminate, it must have taken place after at least the initial stages in the weakening (and ultimate deletion) of /-n.n-/ (< /-n-/) > /-ŋ-/, which, for Gascon, Chambon & Greub (2002) place by the end of the 5th century at the latest.

To understand the power of Michelena’s approach, it may be useful to consider an example. The comparison of old and new Basque dialects does not allow us to reconstruct for OCB anything very different from, say, *aingeru* ‘angel’ and *gaztelu* ‘castle’. Comparison with their Latin sources *angelum* and *castellum*, on the other hand, reveals something more interesting. Furthermore, it forces us to consider that a native word like (*h*)*iri* - *uri* ‘town’ may originally have had intervocalic /-l-/ and *zulo* ‘hole’ may have had a geminate, or as Michelena prefers, fortis lateral. Similarly, the different treatment of Latin singleton and geminate alveolar nasals (cf. *annōna* > *anoa* ‘provisions’), immediately suggests that the reason for the maintenance of the intervocalic nasal in words like *ene* ‘my’, *anaia* ‘brother’ and *arrano* ‘eagle’ is that these words had a geminate (or fortis) nasal at the relevant time (although that is not the only possible explanation).

4.5. Simplification of stop+ll clusters

At the time of the oldest borrowings from Latin/Romance, Basque did not tolerate muta-cum-liquida clusters. These clusters are completely lacking in the native lexicon. In many cases, those clusters, when present in borrowings, were

adapted in the expected way, by vowel epenthesis, whether they were word-initial or word-medial: (Lat *crucem* >) Rom /krutse/ > Bq *gurutze* ‘cross’, Lat *granum* > Bq *garau(n)* ‘grain’, Lat *librum* > Bq *liburu*, (Lat *regulam* >) Rom *regla* > Bq *erregela* ‘rule’, Rom *cleta* > Bq *kereta* ‘gate’, Lat *frontem* > Bq *boronde* ‘forehead’, etc. Nevertheless, in clusters with /l/, an alternative solution that we find in some words is the deletion of the stop: Lat *planum* > Bq *lau(n)* ‘flat’, Lat *plumam* > *luma* ‘feather’, (Lat *ecclesiam*) > Rom **egleisa* > Bq *eleiza* - *eliza*, Lat *clarum* > Bizk *laru* ‘pale yellow’, Lat *gloriam* > Bq *loria* ‘glory’, Lat *florem* > Bq *lore* ‘flower’, Rom *cloca* (Sp *clueca*) > *koloka* - *loka* ‘broody hen’, among others.

Blevins & Egurtzegi (2017) make the observation that, from a cross-linguistic point of view, the adaptation of stop+/l/ clusters in loanwords by deletion of the stop is uncommon. They suggest that the deletion of the stop may have taken place in the Romance source, instead. That is, to give an example, Bq *lau(n)* ‘flat, plain’ would have had as its source not /planu/ but rather /lanu/, cf. Sp *llano*, Ast *llanu*, with depalatalization of the palatal lateral (/ʎ / > /l/) taking place either in Basque or also in the Romance source. For at least some of the words in this group, the same idea regarding the source of the Basque word appears to have been expressed before by other authors (see Agud & Tovar 1988-1994, s.v. *laun*), but not in such a well-reasoned way. In the case of *lore* ‘flower’ (cf. Sp *flor*), an alternative source could have been pre-Gascon **hlore* (see Agud & Tovar 1988-1994, s.v.). If this is so, there was no pre-OCB stop+/l/ > /l/ sound change.

Blevins & Egurtzegi’s (2017) suggestion may be correct, although it requires us to postulate a sizable number of hypothetical sources with initial /ʎ-/ or /l-/ in Romance dialects in contact with Basque. It is not impossible that Spanish words like *pluma* ‘feather’, *gloria* ‘glory’, *claro* ‘clear’, etc. are learned words that have replaced earlier variants showing palatalization of the cluster. It is just that the evidence does not seem to be there for many of these words. This is a point that Blevins & Egurtzegi (2017) explicitly address. Nevertheless, an adaptation of muta-cum-liquida clusters in Basque with deletion of the stop, as is traditionally assumed, would not lack well-known parallels (e.g. in the treatment of Germanic loanwords in Finnish), even if it is not the cross-linguistically most common adaptation, as Blevins & Egurtzegi (2017) also acknowledge. In any event, Blevins & Egurtzegi’s (2017) work forces us to entertain the hypothesis that pre-OCB may not have had a stop+/l/ > /l/ sound adaptation in some words, the relevant cluster simplification having taken place in the Romance source instead. (But see, e.g., Lat *placet* > *laket* ‘please’).

This is as far as the comparative method (and a very shallow application of internal reconstruction) can take us. Michelena expanded the narrow horizons offered by interdialectal comparison by application of the method to the analysis of the borrowed vocabulary. The question naturally arises of whether the same method could not also be profitably applied to the Aquitanian names of Roman times, which would be roughly contemporary with Michelena’s reconstructed stage, since they belong to a language or set of dialects clearly related to Basque. Besides the relatively small number of Aquitanian elements that we have for the comparison, the reconstruction of Proto-Basque-Aquitanian has the two added difficulties that we actually do not know the meaning of the potential components of Aquitanian names and that we do not know which Aquitanian names may belong to a direct

ancestor of Basque and which may belong, instead, to a close relative. We can use the evidence provided by Aquitanian as proof that Basque at that time already had words like *andere* ‘woman, lady’, *neskato* ‘girl’, *zahar* ‘old’, *gizon* ‘man’ and *zezenko* ‘little bull’. The Aquitanian evidence can also in some cases be used as potential confirmation of our reconstructions, as with pre-OCB **seni* and Aq *SENI-*. But in other cases, like with Bq *aker* ‘billy goat’ and Aq *AHER*, it is less evident that we are dealing with cognates and, if so, whether we are dealing with two diachronic stages of the same dialect or with forms that belong to two sister languages, which may lead us to different conclusions for the reconstruction.

In the next section I consider other methodologies that have been applied to the reconstruction of older stages than OCB and Proto-Basque-Aquitanian.

5. Beyond the comparative method: Internal reconstruction and Lakarra’s canonical root hypothesis

Besides the comparative method, the other main tool that we have at our disposal is internal reconstruction. To have some certainty in applying this method we need to ascertain that the words that we are comparing all contain the same morpheme. Generally, thus, the method can be safely and fruitfully applied to inflectional paradigms.

In Basque, by combining the comparative method and internal reconstruction, we can reconstruct verbal forms that are quite different from those historically attested, and in this reconstruction we can have a great deal of certainty; e.g. from *dut - det - dot* ‘I have’, *du, dau* ‘s/he has’ we can obtain the partial paradigm **daut* ‘I have’, *dau* ‘s/he has’. Further comparison with *dudala* ‘that I have’ and similar forms (as well as Roncalese *dud - dur* ‘I have’) leads us to postulate even earlier **dauda* ‘I have’, and from comparison with other forms, we finally obtain a reconstructed paradigm **daduda* ‘I have’, **dadu* ‘s/he has’, **dadugu* ‘we have’, etc. (Gómez & Sainz 1995).

Nevertheless, what we can learn from applying internal reconstruction to Basque inflectional morphology is somewhat limited, since only a handful of verbs are conjugated, and nominal inflection takes the form of suffixes attaching at the end of noun phrases, without modification of nominal bases (except for some dialectal changes in vowel sequences). Although some problems remain to be solved in the domain of inflectional morphology, the fact is that we do not have anything like the different noun declensions, grammatical genders, and verb conjugations of the Indo-European languages.

In semantically transparent compounds and transparently derived words, we have another safe foothold. Here, Basque does offer some interesting phonological alternations, e.g. *betazal* ‘eyelid’ < *begi* ‘eye’ + *azal* ‘skin’, *arkazte* ‘young sheep’ < *ardi* ‘sheep’ + *gazte* ‘young’, *satitsu* ‘shrew’ (with a dialectal variant *saguitsu*) < *sagu* ‘mouse’ + *itsu* ‘blind’, etc.¹⁰ Having identified the relevant phonological alternations

¹⁰ For the sound changes, see Hualde (2006), where an intermorphemic glottal stop is postulated as the trigger of changes like /-gi/ > /-t/ at the end of the first member of a compound.

in semantically transparent compounds, like the examples just given, we can be either very certain or reasonably certain about the etymology of more opaque compounds, such as *bekaitz* ‘envy’ < *begi* ‘eye’ + *gaitz* ‘bad’ (see Agud & Tovar 1988-1994, s.v.), *ekaitz* ‘storm’ < *egu(n)* ‘day’ + *gaitz* ‘bad’ (Michelena 1977: 310), etc.

We are only on very slightly less safe ground when we identify a root **bel* ‘black’ as being present in both *beltz* ‘black’ and *bele* ‘raven’, as well as in rather transparent compounds like *goibel* ‘dark skies’ (< *goi* ‘top’ + **bel*), *harbel* ‘slate’ (< *harri* ‘stone’ + **bel*), *orbel* ‘dry leaves’ (< *orri* ‘leaf’ + **bel*), etc. (see Agud & Tovar 1988-1994 for these etymologies).

Beyond this «mild» internal reconstruction, Michelena provided some bolder etymologies, such as his identification of the verbal form **din* ‘to become’ (found in modern subjunctive forms like *dadin*, e.g. *etor dadin* ‘may it come’, etc), as a suffix in words like *berdin* ‘same’, which he analyzes as *ber* ‘self’ + *din* ‘become’, *urdin* ‘blue’ < *ur* ‘water’, *gordin* ‘raw’ < *gorri* ‘red’, and several other words with the same ending (Michelena 1970). In this case, Michelena’s crucial insight was the identification of *-din* as a verbal form. A long tradition of scholarship had already agreed on the identity of the first component of many of these items.¹¹

All of this still leaves much of the native lexicon unexplained. In an attempt to expand the horizons of Basque etymological research, Lakarra (1995 and later work) proposes to consider the canonical structure of Basque roots, following a methodology that has been successfully applied to other language families. Using this as a heuristic procedure, any word whose phonological structure does not comply with the canonical pattern may be suspect of being either a borrowing or morphologically complex.

To test the potential value of the hypothesis, it may be useful to engage in a sort of *Gedankenexperiment*. Suppose we do not know anything at all about the history of the Spanish language. We have no old documents and we have no knowledge of Latin or any other Romance language either. If we examine the basic Spanish lexicon, we will immediately notice that the majority of words are disyllabic, although some are longer and a few are shorter. Trying to name the objects around me in Spanish, I find disyllabic *mesa*, *silla*, *puerta*, *cuarto*, *sala*, *techo*, *suelo*, *pared*, *papel* and longer *cocina*, *ventana*, *visillo*, *persiana*, *alfombra*, *lámpara*, *bolígrafo*. Adjectives reveal a similar pattern: *rojo*, *verde*, *azul*, *blanco*, *negro*, *marrón*, *alto*, *bajo*, *grueso*, *grande* vs *amarillo*, *naranja*, *delgado*, *pequeño*. We would find the same distribution within basic lexical domains such as parts of the body, etc. In the case of words ending in an unstressed vowel, it is also clear that the last vowel is an inflectional suffix, since it is deleted in derived words, e.g. *mesa*, *mesilla*; *silla*, *sillón*. We may then hypothesize that words whose root does not conform to the monosyllabic pattern illustrated by *mes-*, *sill-*, *puert-*, *cuart-*, etc., are likely to originally be something other than monomorphemic native words. This hypothesis would prove useful, since it

¹¹ Michelena was at times more speculative. For instance, in proposing an etymological connection between the suffix *-doi*, ‘grove, orchard, forest, etc’ and the noun *lohi* ‘body’, with a change **d- > l-* word-initially (Michelena 1977: 258, 2009, vol. 6: 212), One of his most celebrated etymologies is **abari-bagako-egun* ‘day without supper’ for Bizkaian *bari(a)ku* ‘Friday’ (Michelena 1971, English translation 2020).

would lead us to discover that *ventana* is indeed to be analyzed as *vent-an-a* and similarly with *pers-ian-a*, *coc-in-a*, *vis-ill-o*, *amar-ill-o* (although *amar-* would remain problematic). We would also correctly conclude that *boli-graf-o* is a compound and we might suspect that *alfombr-a* and *naranj-a* are borrowings, since they are longer than canonical words but are not easily amenable to decomposition within the language. On the other hand, it might be less clear to know what to do with words like *lámpar-a*, *pared*, *papel*, *azul* and *marrón*. The results of our little thought experiment speak in favor of the methodology, but not without some caveats.

In the case of Spanish, we know what the answers to our etymological questions are. All the Spanish words listed above have well-known, uncontroversial etymologies. For Basque, on the other hand, most of the time we do not know the correct answer, but the method should in principle work in a similar way. It may allow us to discover the hidden etymology of many words, but it may also produce spurious results, where the strict application of our canonicity criterion had led us astray. In general, we will not know when we are right and when we are mistaken.

As a first step, Lakarra (1995) proceeds in a similar way as we have illustrated with Spanish. Since most native Basque words are mono- or disyllabic, longer words like *arrano* ‘eagle’, *anaia* ‘brother’, *al(h)aba*¹² ‘daughter’, *iturri* ‘fountain, spring’, *arraultze* ‘egg’, **ardano* ‘wine’, **gaztane* ‘cheese’, **arrani* ‘fish’, etc., should be scrutinized.

Then Lakarra takes a second, bolder and more controversial step. He notices that quite a few native disyllabic words may contain suffixes (e.g. participial *-i* in *gorri* ‘red; bare, naked’, *hori* ‘yellow’, etc.). A few other words show a reduplicative pattern, as in *gogor* ‘hard’, relatable to *gor* ‘deaf’ and *zezen* ‘bull’ (< *zen?). An additional observation is that in disyllabic VCVC words starting with a vowel, the two vowels tend to match, e.g. *adar* ‘horn, branch’, *azal* ‘skin’, *eder* ‘beautiful’, *odol* ‘blood’, etc. (although not exceptionlessly: *aker* ‘billy goat’, *oker* ‘twisted’, etc.). He thus proposes to analyze these and other vowel-initial disyllabic words as containing a CVC root: *a-dar*, *e-der*, *o-dol*, and suggests that the initial vowel has resulted from reduplication, with eventual deletion of a word-initial consonant. Presumably, another possibility would be to assume an ancient vowel prefix that underwent vowel harmony.

By noticing a number of distributional gaps in CVC «roots», Lakarra then proceeds to constraint the possible shapes of what he takes to be the only type of root at the Pre-Proto-Basque-Aquitania stage that he reconstructs. His last step is to identify these roots throughout the lexicon. His methodology allows Lakarra to find, for instance, the root of *e-dan* ‘to drink’ in Pre-OCB **ardano* ‘wine’, which I find eminently plausible and could not have been found using more traditional methodologies. The admirably well-documented *Historical and Etymological Dictionary of Basque* (EHHE, Lakarra, Manterola & Seguro 2019) offers a demonstration of the value of root-based etymological research. At the same time, following this bold approach, Lakarra has put forward many other etymologies

¹² In the heterosyllabic group /l.h/ we can be certain that the aspiration is ancient when we find confirmation in medieval Southern texts (e.g. *olba* ‘hut; forge’, *elborri* ‘thorn bush’). In the absence of this Southern evidence, it is less clear whether this feature is old, since aspiration after /l/ preceding the nucleus of the second syllable was generalized at some later point in Northern Basque, even to borrowings (cf. *solhas* ‘conversation’, etc.).

that other scholars may not be ready to accept without further proof (as in, to give just one example, the identification of a common root **dar* in *adar* ‘horn, branch’ and *indar* ‘strength’). The etymologies in *EHHE* signalled by the authors with an exclamation mark—and a few that lack such a mark—fall within this category of highly speculative proposals (for my taste).

Perhaps it needs to be stated explicitly, since this point appears to be lost sometimes, that the Pre-Proto-Basque-Aquitainian stage that Lakarra reconstructs is just as far as a specific methodology will take us when we apply it consistently to the Basque data. It is a hypothetical, undatable stage resulting from the consistent application of a certain hypothesis. There is no possibility of getting at the most remote origins of Basque or any other oral language, which would be the moment when our ancestors developed spoken language. If, within a given linguistic tradition, there was a monosyllabic stage at some point, at an even earlier point in time the language may have been polysyllabic, in a circle that for all practical purposes may be considered endless.

6. Recent extensions of root theory with comparative goals

In the last decade, several authors have extended Lakarra’s methodology in different directions, proposing genetic connections between Basque and other languages on the basis of such reconstructions. Here I will discuss a couple of these attempts.

6.1. The Proto-Basque-Paleo-Sardinian hypothesis

Blasco Ferrer (2016 and other work) has interpreted a number of Sardinian toponyms as compounds containing CVC morphemes that for the most part coincide with the roots identified in Lakarra’s work. Lakarra (2013b) himself has expressed skepticism about the Proto-Basque-Paleo-Sardinian hypothesis. There are indeed reasons to remain skeptical regarding both the meaning and the form of the Paleo-Sardinian *comparanda* that Blasco Ferrer has proposed. The meaning given to the hypothetical Paleo-Sardinian morphs is based exclusively on the interpretation of toponyms, which is pretty much the only evidence we have for this language. Regarding form, as Lakarra (2013b) remarks, a problem is precisely that the Paleo-Sardinian roots that Blasco Ferrer identifies are so similar—in fact, most of the time identical—to Lakarra’s reconstructed Pre-Proto-Basque-Aquitainian roots.

Recent paleo-genetic research has shown that the westward expansion of agriculture from Anatolia starting about 10,000 years ago was accompanied by massive human genetic replacement (for the Iberian Peninsula, see Olalde *et al.* 2019). As a hypothesis, it seems reasonable to speculate that both Paleo-Sardinian and Basque may have descended from one of the languages spoken by the first European agriculturalists. What seems less reasonable is to think that this language would have remained essentially unchanged for millennia in both areas. Although Lakarra’s reconstructed CVC roots are undated and quite possibly undatable, it is highly unlikely that they are over 8,000 years old (which is when the Anatolian agriculturalists reached Iberia). For comparison, the time distance between Sanskrit

and Latin and their common ancestor is generally assumed to be much shorter; yet, an interpretation of, say, modern Portuguese or Italian toponyms based on pre-Sanskrit roots strikes me as a hopeless endeavor.

6.2. Blevins's Proto-Basque and Proto-Indo-European-Euskarian

Blevins (2018) also adopts Lakarra's root-based approach, but without accepting the premise that all ancient roots had the same phonological shape. Using, then, her own, insightful and interesting, but problematic, internal reconstruction, she identifies a number of roots in what she calls «Proto-Basque». Some of the roots she identifies coincide with Lakarra's, but many are novel. Her hypothesis («the Proto-Indo-European-Euskarian Hypothesis») is not that Basque is an Indo-European language, as several other authors have attempted to demonstrate (e.g. Forni 2013; see Gorrochategui & Lakarra 2013; Kassian 2013 for reviews), but, rather, that «[...] Proto-Basque and Proto-Indo-European should be seen as sister languages, or daughters of sister languages, descending from the same mother tongue» (Blevins 2018: 134).

If Basque is not an IE language, what is exactly the claim, then? One of Blevins's suggestions appears to be that Basque may be related to PIE in the same way as Anatolian is related to the rest of the IE branches. The other possibility she allows is that the relationship may be even more distant («daughters of sister languages»). It is not clear to what extent postulating this type of very distant relationship is really different, in practical terms, from making the proposal that all human languages are ultimately related, which, of course, may very well be true. Saying that two or more languages are related implies making a claim that at some point in time and in some specific place there was a community of speakers of their common ancestor, but this is not explored in Blevins's book.

The fact that Basque inflectional morphology does not look anything like that of the Indo-European languages does not speak in favor of a relatively recent genetic link; but, of course, in the case of a very remote relation of the type that Blevins envisions anything is possible. In any event, one would have to assume either that the hypothetical common ancestor of Basque and PIE was spoken at a time before any of the IE inflectional morphology developed or that Basque underwent massive sound change and lost all of this morphology. This second hypothesis does not seem compatible with the great conservatism during millennia that one needs to assume in order to be able to compare Basque roots obtained through internal reconstruction and PIE roots.

My evaluation of the book largely coincides with Gorrochategui's (2020b: 307-308) assessment. One example might be enough to illustrate the methodological problems that I see. I will consider only one of the roots from the first list of Basque roots in the book, which the author provides on p. 27 (Table 2.1). Whereas I am able to recognize several of the words from which Blevins may have extracted the Proto-Basque roots that she provides in this table, several others are completely new to me. I will choose just one of those, although any of the others would also serve the purpose: **dol* 'go along'.

In the text above the table, Blevins mentions the Bq word *odol* 'blood', which in agreement with Lakarra, she analyzes as containing the root **dol*. But, where does

she get the meaning ‘go along’ from a word that means ‘blood’? If we go to the appendix, we find this (p. 252):

***dol** ‘long; extending; run through’,
 B *odol* < *ha-dol (*ha -NMZ) ‘that which runs through’.

The reason for the gloss is still not clear. If we are to be allowed maximum freedom in assigning meaning to reconstructed forms, why not interpret ‘blood’ as ‘red’ or ‘life’ or ‘spirit’ or any other possible meaning vaguely related to ‘blood’, instead of ‘that which runs through’? The extended meaning that the author provides for *odol* ‘blood’ allows her to compare it with the PIE verbal roots that she lists right below:

PIE **delh*₁- ‘to extend, lengthen’ [LIV3]¹³
 PIE **del*¹ ‘long’ (CW: 16)

But broadening the meaning of *odol* ‘blood’ in order to facilitate this comparison would be methodologically unsound. Is there any other evidence for this interpretation? Under the entry for **dol*, the author tells the reader to also see **delh* ‘long; along; extending; run through’. If one goes there, one sees that Blevins finds evidence for this Proto-Basque root in *adel*- ‘arrangement, going along, extension’ (which she analyses as deriving from **an-del*, **an* ‘on the side’). This is a word that I don’t know, which, in itself, would not be particularly surprising. It turns out that the hypothetical *adel*- has been extracted from the words *adelatu* ‘arrange, prepare, launch; disperse (of smoke)’ and *adelu* ‘preparation, arrangement; adornment, trim; release, launch’. To me, and I believe, almost any specialist in the language, these do not look like native Basque words at all and appear to be, instead, transparent Romance borrowings, cf. Old Sp *adeliñar* ‘arrange, prepare, fix’ (Modern Sp *aliñar*), from Latin *ad delineāre*.

Blevins finds further corroborating evidence for a Proto-Basque root **del(h)* ‘long, etc.’ in *andel* ‘deposit, streak; extension, arrangement’. First of all, what the Basque dictionary that she is consulting (*OEH*) provides as definition for this «word» is only ‘depósito’. The other meanings that Blevins adds do not seem to be found in the dictionary entry. What the *OEH* does add is that this word has been taken from Azkue’s (1918) *Diccionario de bolsillo* (‘Pocket dictionary’). As the *OEH* further points out, Azkue (1905-1906) tells the reader in his larger dictionary, that *-andel* is not a free word, but in fact only a suffix, whose meaning he has deduced from compounds where the first member is clear. If we consult Agud & Tovar (1988-1994), we see that other authors have disputed Azkue’s reasoning.

In addition to all of this, Blevins, who postulates a rule of initial *d*-deletion, adds to the set *elbar* ‘invalid’ and *elkarri*, which the *OEH* translates as ‘perpend stone’ (‘perpiaño, piedra que traba en la pared’), as further evidence for Proto-Basque **delh* ‘long, etc.’ About *elbar*, leaving aside the rather tenuous semantic relation between ‘invalid’ and ‘long’, other proposals for its etymology mentioned in Agud & Tovar

¹³ I have not been able to consult LIV3 (= 3rd edn. of *Lexicon der Indogermanischen Verben*, Rix et al. 2014, which Blevins cites as «electronic file»). The 2nd, 2011 edition has **delh*₁- ‘behauen, spalten’ (= ‘cut, split’).

(1988-1994) may certainly be preferable. As for the word *elkarri* ‘perpend stone’, the authors of the *OEH* tell us that they have taken this word from Larramendi’s (1745) dictionary.¹⁴ It looks like a transparent compound of *elkar-* ‘to join’ and *(h)arri* ‘stone’, whether Larramendi had heard it somewhere or he created it himself as a possible Basque translation equivalent for the Spanish word *perpiaño*. Haplogy, *elkarrarri* > *elkarri*, is common in Basque with this sequence, cf. *sagar* ‘apple’+*ardo* ‘wine’> *sagardo* ‘cider’, and Larramendi would certainly have been able to notice this phenomenon and apply it to his neologism, if this is indeed what this word is.¹⁵

The connection with the meaning ‘long’ also seems unclear in the case of two other words given as evidence for **del(h)* in this entry: *adala* ‘scouring stone; dump’, [actually, Sp ‘piedra de fregar, vertedero’]¹⁶ and *sedel* ‘pus coming out of a boil’ (cf. Sp *cedal*, *OEH* s.v. *sendel*), both documented by Azkue in local Bizkaian varieties.

The semantic match is somewhat better in an additional word mentioned by Blevins in this entry, *eskudel* ‘handrail of a stairway’ [Sp ‘pasamano(s), barandilla’], which is clearly a compound of *esku* ‘hand’ and was also collected by Azkue in an area of Bizkaia, together with *eskuzal* for the same meaning; but this could hardly be an ancient formation. In the same dialectal region we also find *eskamel* ~ *eskabel* ‘bellows handles’ [Sp ‘brazos de los fuelles’]. Most likely these different variants, *eskudel* ~ *eskuzal* ~ *eskamel* ~ *eskabel* all have the same origin, with the meaning of ‘handhold’, and may be based on Sp *manivela* ‘crank, handle, handgrip’ with replacement of Sp *mani-* ‘hand’ with *esku*, *eska-*.

In summary, after all of this, what is the real evidence for a pair of Proto-Basque roots **dol* ‘long; extending; run through’, **del(h)* ‘long, along; extending; run through’, to be related to PIE **del(h)* ‘long’? I cannot find any unproblematic evidence. The author is nevertheless to be commended for being very explicit about the evidence that she has used for her reconstructed Proto-Basque roots. It is only because of this careful documentation in the book that other scholars can evaluate the evidence for each etymological proposal and reach their own conclusions regarding its strength.

¹⁴ In creating the *OEH* and making it available online, the Basque Academy has rendered a huge service to scholarship on the Basque language. But using it to search for ancient Basque words without a deep understanding of the Basque lexicographical tradition is as risky a business as mushroom hunting without proper training.

¹⁵ In the Royal Spanish Academy’s *Diccionario de Autoridades* (1726-1739), which Larramendi used to make his *Trilingual Dictionary*, the Spanish word *perpiaño* is defined as «la piedra que traba en la pared» and *trabar* is defined as «juntar, ò unir una cosa con otra para mayor fuerza, ò resistencia». Larramendi’s *elkar* + *-arri* = *elkarri* «joining stone» thus provides a perfect semantic match for the definition of *perpiaño* in the Spanish Academy’s dictionary. This is consistent with one of Larramendi’s strategies for creating neologisms (Urgell 2004).

¹⁶ Incidentally, in the case of words of limited currency, coming from Azkue’s dictionary, an additional difficulty can be understanding his definitions in Spanish. In the case of this dictionary entry, most likely ‘piedra de fregar’ and ‘vertedero’ are to be interpreted as synonyms, with the meaning of ‘washing trough’ (‘fregadero’). Neither with this meaning, nor in Blevins’s interpretation, is it clear to me that the word *adala* recorded by Azkue offers any support for an ancient root **del(h)* ‘long’. Notice also that in this case we have *-dal-*. It thus appears that any word containing *-dol-*, *-del-*, *-dal-*, *el-* could be used as potential evidence for the proposed Proto-Basque root (or pair of roots), as long as the author can imagine some way to connect the meaning of the word to the notion of ‘long’.

For Blevins, a virtue of her reconstruction of Proto-Basque roots is that, unlike Lakarra, she always assigns meaning to her reconstructed roots. Instead, I see this as problematic, since it is open to a very high degree of subjectivity.

My conviction is the same as the one Meillet (1925) expressed almost a century ago: for the comparative method to work properly, researchers should allow themselves no latitude in interpreting the meaning of potential *comparanda*. Bq *odol* means ‘blood’. That’s it. It does not mean ‘that which runs through’. Bq *odol* can be compared only with a word in a different language that means ‘blood’. It cannot be compared with a word that means ‘long’. If we give ourselves such freedom to modify and stretch meanings, we are vitiating the method and the whole enterprise becomes vacuous. It is only after a genetic relationship has been solidly established and the sound correspondences have been worked out that one can fruitfully compare, for instance, English *bone* with German *Bein* ‘leg’ (which is still a trivial case of semantic shift).

For comparison, Blasco Ferrer (2016) prefers, instead, to interpret **dol* as ‘red’, for which he finds evidence in Sardinian toponyms. Lakarra (2013b) relates **dol* to the concept of ‘fall’ (based on his identification of the root in other words like *e-ror-i* ‘fallen’, and his principle of no homonymy). In *EHHE*, a possible relation between *odol* ‘blood’, *jori* ‘abundant’ and *erori* ‘fall’ is also suggested, albeit tentatively. Under a different hypothesis, one could decide, for instance, to give the root **dol* the meaning of ‘liquid’, which also seems sensible. Then one could find this root perhaps in *hordi* ‘drunk’ < **dol-di*, *uholde* ‘flood’ < **ur-dol-te* and even *dolare* ‘cider press’ < **dol-are* (if we also leave aside more plausible etymologies, see below on borrowings), etc.; and it would not be too hard to find additional examples to justify the necessary sound changes. Or suppose that, instead, one prefers the meaning ‘vigor, vital force’ for **dol*, then confirmation could be found in *adore* - *adole* ‘vigor, energy’ < **a-dol-e* (again ignoring more sensible etymologies). If we sought support for the Proto-Basque-Japanese hypothesis, we might relate Bq *odol* ‘blood’ to Jp *odori* ‘to dance’, etc. I don’t see any way to constrain this procedure. Among all these proposals and pseudo-proposals, a semantic connection between ‘blood’ and ‘long’ strikes me as particularly implausible.

I do not know what method Blevins has used to extract the semantics of her reconstructed roots. If interpretations such as ‘blood’ = ‘that which runs through’ are the result of a constrained method, this would need to be made very clear to readers.

In my opinion, in this kind of reconstructive and comparative work, confirmation bias is a serious problem that can make the results rather unreliable and unconvincing for other scholars who are not already committed to finding evidence for a particular genetic connection. One could argue that what I am calling a «bias» is what other people would call a «hypothesis», and that without a hypothesis we are unlikely to make progress. I would not disagree.

Let us consider the issue of bias/hypothesis in linguistic comparison. I remain fully convinced that someone else using the same internal-reconstruction techniques as Blevins, but with a different hypothesis (say, that Basque is related to Uralic, North Caucasian or the Afro-Asiatic family) would produce a Pre-Proto-Basque-Aquitanian that looks significantly different. In fact, such comparisons continue to be produced at a steady pace and any or all of these putative genetic relationships

could be true (after all, human languages have been evolving for a really long time), but, once one allows oneself as many degrees of freedom as proponents of these relationships typically do, there cannot be any satisfactory demonstration. Proponents of one of these genetic relationships are unlikely to be persuaded by the evidence presented by defenders of a different one.

In order to avoid potential bias from preordaining the results, one could think of charging a committee of independent judges (say, twenty graduate students in a class on Historical Linguistics whose only motivation is to get a good grade) with comparing the basic vocabulary (not reconstructed roots!) of Basque to the lexicon of several languages in the Indo-European, Uralic, North Caucasian, Afro-Asiatic and perhaps other families, in an exploratory, hypothesis-free manner, and then calculating agreement rates among judges. Bengtson (2017) reports on a version of this experiment, in which he himself was the only participant, and concludes that the evidence favors the Basque-Caucasian Hypothesis. But, again, it is impossible to know whether the author was uncommitted to any hypothesis before running the experiment.¹⁷

Comparison of Basque with Indo-European, in particular, has its own inherent difficulties. Since Basque has been in intense contact with the Italic branch of Indo-European for two millennia, one needs to be extra-careful with cognates due to borrowing. The minimal precaution we can take is to accept that, from the time of the Roman Empire, borrowing has been essentially unidirectional: from Latin/Romance to Basque. Clear counterexamples are exceedingly few. Besides this, given its geographical location, it is certain than before coming into contact with Latin, Basque-Aquitainian was in contact with Celtic and possibly other Indo-European branches, from which it may have also received perhaps a large number of undetected borrowings.

In this respect, I am troubled by the fact that Blevins rejects a number of well-established etymologies of Basque words for which scholars have come to a consensus that they are borrowings from Romance. Thus, she disagrees with those of us who think that Bq *leku* 'place' is a borrowing from Ibero-Romance **luecu* 'place' (in the Glosas Emilianenses, *lueco*) < Lat *locum*, that *harpoi* 'harpoon' and *eskalapoin* 'clog' are from Gallo-Romance, that Bq *asmatu* 'to imagine, invent, discover' is just Old Spanish *asmar*, that Bq *hanka* 'leg' and Eng *haunch* are cognates, that Bq *abia-n* 'going, underway' is from Sp *aviar* (< Lat *ad viam*), etc., preferring to see in every one of these cases a Proto-Basque root relatable to a reconstructed Proto-Indo-European root. She also defends, for instance, that *zapata* 'shoe' is a native Basque word, an etymological proposal that is not new (see Moguel 1881 [1802]: *diálogo tercero*), but seems unlikely, given the geographical diffusion of its cognates. The burden of the proof is clearly on Blevins's side. She either would have to argue convincingly that cognates of the Basque words in Romance (and sometimes Latin) are in fact simple look-alikes and not true cognates or she would have to provide a

¹⁷ In a test of Blevins's methodology, Bakker (2020), in an otherwise glowing review of Blevins's book, offers a brief comparison of Proto-Basque (PB) roots as reconstructed by Blevins (2018) with Proto-Austronesian-Ongan (PAO) roots, obtaining the result that «PB looks surprisingly close to PAO, perhaps closer to PAO than to PIE» (Bakker 2020: 23).

satisfactory account for how, when and why Spanish, French and other languages borrowed these words from Basque.

However, even if we could agree on the proper treatment of words that are traditionally thought not to belong to the native Basque lexicon and we were to run such a bias-free experiment, I am pessimistic that any clear picture would emerge regarding this or any of the other comparisons that have been proposed. The reason is that very little probative evidence is likely to have survived for a genetic relationship at the time depth that must be contemplated. Chances are that very little of what we can determine about Proto-Indo-European as was spoken about 5,500 years ago was also true of its Pre-Proto-Indo-European ancestor of 15,000 or 20,000 years ago, and the same is true of Basque or any other language. And it is in the nature of language change that very solid evidence for common ancestry often disappears within the span of a few centuries.

For the one etymological proposal that I have examined here, **dol* ‘go along’, I was not able to find any solid evidence. But I have not examined all of Blevins’s etymologies and perhaps there are some that are more convincing and appear to have good matches in Proto-Indo-European. Even if that were the case, however, it would be impossible to know whether that could be taken as a clear indication of language relatedness. The best matches could very well be ancient borrowings. Once more, I agree with Meillet and other «conservative» historical linguists in thinking that when a linguistic relationship is remote beyond a certain point in time, as we must assume between Basque-Aquitanian and any purported relative, common inheritance cannot be distinguished from the results of borrowing or chance.

7. Conclusions

In this paper, I have reviewed the different methodologies at our disposal for the reconstruction of earlier stages of the Basque language and have considered the degree of confidence that we can have in their results. The most reliable methodology is the comparative method, which, in the case of Basque, in principle, can only take us as far as the last stage before dialectal fragmentation, known as OCB. We can accept the reconstructed forms obtained through this method with a very high degree of confidence. Equally certain are pre-OCB forms like **ardano*, which can be obtained by internal reconstruction applied to OCB pairs like *ardãô/ardan-*.

Michelena offered a masterful demonstration that, through the application of what is essentially also the comparative method to old loanwords in Basque and their Latin sources, we can obtain valuable information about sound changes that likely took place a few centuries before OCB (since Basque dialects do not vary in the relevant respects). These sound changes are reflected in the adaptations undergone by such loanwords and sometimes their formulation leads us to discover older possible forms of native words.

Internal reconstruction produces the safest results when we can be absolutely certain that several forms contain the same old morpheme, as when we compare the different forms of an inflectional paradigm. In Basque, such paradigms are found only in a limited number. On the other hand, semantically transparent and

unquestionable compounds show evidence for a number of striking phonological processes, whose analysis can then be extended to more opaque forms, sometimes with a high level of confidence.

Lakarra (1995, 2011, 2013a, among many other publications) has proposed other methodologies in order to make further progress in our study of the past of the Basque language. By adopting, as a discovery procedure, a very strict view of what an ancient Basque root may look like, with no tolerance for deviations from that pattern or for homonymy, he has arrived at a large number of novel etymologies. The historical reality of etymologies discovered using this method, however, must remain in most cases highly hypothetical in the absence of such confirmatory evidence as comparison with other languages could provide, even if the establishment of lexical families, as is done in *EHHE*, can serve to strengthen some of these etymological hypotheses.

Finally, the (in my view) unconstrained reconstruction techniques in work like Blevins (2018) do little to advance our knowledge of the prehistory of the Basque language. My opinion is that the methodology is too unreliable for real progress to be made.

Through the careful application of standard methodologies to the reconstruction of pre-Proto-Basque-Aquitainian we cannot go very far back in time with the evidence at our disposal if we hope to obtain results that are convincing. My sense is that, unless there is some unexpected breakthrough in the available evidence, we will never be able to demonstrate a close genetic relationship between Basque-Aquitainian and any other language. This does not imply that all the hypotheses that have been advanced are wrong, only that no convincing demonstration can be provided based on the available data.

Post-scriptum

Some of the issues that I raised above in my review of Blevins (2018) are addressed in Blevins & Sproat (2021). Unfortunately, the arguments given there do not significantly alter my general evaluation of that work.

First, regarding the issue of conscious or unconscious bias, the explanation given in Blevins & Sproat (2021) does little to assuage my worries about the methodology. This is what the authors have to say:

The question of bias should also be addressed. Do the Proto-Basque reconstructions in Blevins (2018) resemble Proto-Indo-European reconstructions because they are influenced by knowledge of PIE forms? In some clear cases, marked here by **, the answer is yes. However, at a more global level, this question is difficult to answer and may be best dealt with by comparing competing alternatives. For example, where Blevins (2018) proposes initial PB **sth* clusters and Michelena (1977) posits **z*, as in **sthun* ‘support, carry weight’ [JB:371], analyses with **z* are unable to account for Basque doublets like B *astun* ‘heavy, pregnant’, B *azun* ‘full, loaded, pregnant’. The idea of initial **sth* clusters in particular Euskarian roots may have been influenced by Indo-European lexemes, but what ultimately matters is whether the Proto-Basque reconstructions are internally supported, based on adherence to principles of the comparative method and internal reconstruction, and offer a better

understanding of internal properties of the language, in this case, for example, the limited distribution of Common Basque /z/ [...]. (Blevins & Sproat 2021: 509)

Let us consider the one example that Blevins & Sproat offer here. The word *astun* is the basic word for ‘heavy’ in the central and western area. The word *azun* - *axun*, on the other hand, is of much more restricted usage and, according to the *OEH*, has the meaning ‘pregnant’ primarily applied to sows and sometimes to dogs, cats and other animals. The *OEH* also provides a second meaning ‘full, loaded, pregnant’ (Sp ‘lleno, cargado, preñado’), but making clear that, with this meaning, this word is only found in the poetic language of a couple of writers from the beginning of the 20th century, who in their poems speak, respectively, of «the pregnant cloud» (*odei azuna*) and the waves of the sea pregnant with amazing strength. Leaving this aside, however, *astun* ‘heavy’ and *azun* ‘pregnant’ could certainly still be related (cf. Port *grávida* ‘pregnant’), and a connection has indeed been proposed many times and has been explained in a number of ways, none of which involves proposing **sth* (see Agud & Tovar 1988-1994, s.v. *azun*). Among other possibilities, both *astun* ‘heavy’ and *azun* ‘pregnant (sow)’ could contain the common derivational suffix *-dun* ‘having, bearing’, but attached to different roots, with devoicing of /d/ after a sibilant in *astun*, for which *hats* ‘breath’ + *dun* is proposed in the *EHHE*, and with intervocalic /d/ deletion in *azun*, perhaps from *hazi-dun* ‘bearing seed’ (*azidun* > **aziun* > *axun*).¹⁸

For the other four «doublets» in Blevins (2018: 90) the connection is less than clear. It cannot be totally discarded that there could be a connection between the two members of some of these pairs, since most of the words are of uncertain etymology, but this is by no means self-evident. It could be that deriving *istil* ‘puddle; mud’ from Lat *distillāre* ‘to drip’, as Michelena does, is incorrect and this is instead a native Basque word related to the poorly documented *zildu* ‘to dirty, muddy’, and to *isil* ‘silent’ (for which Blevins prefers the minority form *izil*, which may have arisen from depalatalization of the expressive variant *ixil*), all containing a root *(s)*thil* ‘still; stagnant’, related to PIE **teHl* ‘still, silent’, so that Bq *istil* and Eng *still* would be cognates. As far I can see, however, the main argument in favor of accepting these etymologies over other etymologies that could and have been proposed would be to be able to relate the Basque words to the PIE roots, which would make the comparative procedure circular. Incidentally, or perhaps crucially, there is a high likelihood that *zildu*, with the meaning ‘to dirty, muddy’ is a mistake. It is the answer given to the question ‘ensuciar, enlodar’ to a worker gathering information for the Basque Academy’s *Triple Questionnaire* in the 1920s by a 90 year old informant in a village on the southern border of the Basque language. Nowhere else was this answer given by any informants (see Echaide 1964: 111, and for the identity of the informant, 45) and there is no other documentation for this word with this meaning. In the *OEH*, under *zildu*¹ we can see that this verb actually

¹⁸ For the palatalization of the sibilant with absorption of the following glide, other examples may be *axola* ‘worry’, if Michelena is correct in proposing that it comes from a derivative of Gascon *ànsia* ‘anxiety’ (see Arbelaz 1978, s.v.) and *azeri* - *axeri* - *axari* ‘fox’ < **azeari* < *azenari*, also following Michelena’s etymological proposal, if it has not resulted from affective palatalization.

means ‘to sprout’ (Sp ‘brotar’). In a bias-free analysis, there is no compelling reason that I can see to take any of the sets of words provided in Blevins (2018) as evidence for a Proto-Basque cluster **sth* (which is impossible in historical Basque, where stops are always unaspirated after a sibilant, Michelena 1977: 204).¹⁹

Blevins’ (2018) statistical arguments for considering that the contrast between *s* /s/ and *z* /z/ is not old in Basque, namely that «first, the two sounds do not contrast in syllable-final position; second, in word-initial position, *z* is heavily over-represented and may out-number *s* by 2:1» (Blevins 2018: 71), are interesting but not persuasive. First, as Gorrochategui (2020b) insists, in final position there are minimal pairs like *hitz* ‘word’ vs. *hits* ‘pale’, *hatz* ‘finger’ vs. *hats* ‘breath’ (Notice that in Michelena’s system only affricates, not fricatives, are expected word-finally). Secondly, regarding word-initial frequency, in English, to give an example, there appear to be about two and a half times as many words starting with /m/ as with /n/, which I, for one, don’t take as evidence that the /m/ - /n/ contrast in English is very recent.

Regarding the matter of semantic freedom, Blevins & Sproat (2021: 516) state that they have adopted stricter semantic criteria for their comparison with PIE than those allowed in Blevins (2018). They provide the example of Bq *malko* ‘tear(drop)’, which Blevins (2018) reconstructs as Proto-Basque **malkho*. They discuss the similarity of this word with words in several Indo-European languages meaning ‘milk’ and argue in favor of an etymological connection («milk comes in drops», p. 11), but, with good criteria, decide to exclude this word in order to constrain the semantics of the pairs of Proto-Basque and Proto-Indo-European roots being compared.

In the *EHHE*, *malko* ‘tear(drop)’ is analyzed as containing the diminutive suffix *-ko* and being related to *madari* ‘pear’ (which has a variant *udare* ‘pear’) and *maluta* ‘snow flake; tuft of wool’. There are very good phonological reasons for accepting this hypothesis and rejecting the hypothesis that *malko* ‘tear drop’ comes to us virtually unchanged from a Proto-Basque **malkho* of thousands of years ago. First of all, plosives are generally voiced after a lateral within morphemes in the native lexicon (and, except in the easternmost dialects, also in old borrowings, see section 3.3 above), but there is no voice assimilation with some suffixes, including diminutive *-ko* (cf., e.g., *zezen-ko* ‘little bull’). Secondly, this word has a marked accentual pattern. In Goizueta, it is *malko* with a low accent on its initial syllable (see Lujanbio 2013 [2018]), which is consistent with its having undergone contraction (cf., e.g., *pātu* ‘to place’ < *paratu*). To explain these facts, we need to postulate a derivation that cannot be very different from **madal-ko* > **maalko* > *malko* (cf. *EHHE*: 173). Finally, Michelena (1977: 267-276) noted that, although the phoneme /m/ is not infrequent in Basque, in a large number of words containing it, it clearly derives from other phonemes and phoneme sequences. This led him to make

¹⁹ There are, in fact, several somewhat similar alternations (although involving also a rhotic) in pairs or triplets of words where we can be certain that we are dealing with the same lexical item: *zurzai* - *sustrai* - *sustai* ‘root’, *mazurza* - *masustra* - *masusta* ‘strawberry’, *(h)este* - *(h)er(t)ze* ‘fence’, *beste* - *ber(t)ze* ‘other’, *bost* - *bortz* ‘eight’, etc., all of which have been analyzed as post-OCB evolutions, without any need for postulating that in Basque *z* /z/ is a recent phoneme that derives from older **sth* (Michelena 1977: 364-365; Manterola 2020).

the radical claim that, at some point in time, this phoneme was absent from the Basque inventory. To account for the fact that word-initial *ma-* is rather common, Michelena hypothesized that this is an expressive prefix, of no great antiquity, where the presence of the uncommon phoneme /m/ (introduced through loanwords and sound changes; e.g. /nb/ > /m/) reinforced its expressive character, accounting for its success in new formations. The doublet *udare* ‘pear’ and *madari* ‘pear’ (from whose diminutive, *màlko* ‘tear’ may have arisen) is one of several examples showing that *ma-* must have been originally a prefix. Blevins’s etymological hypothesis leaves all these facts unexplained.

The semantic constraints that Blevins & Sproat (2021) have now added to the Basque-PIE comparison in Blevins (2018) do not solve the essential problem with this enterprise, which is the method through which Proto-Basque roots and their meanings have been extracted. In their list of comparanda in Table 3, we still find a Proto-Basque root **del(h)* ‘long, extending through’, for which I was not able to find any good evidence.

To exemplify the limited amount of semantic latitude they have allowed themselves, the authors ask whether it would be licit to compare Proto-Basque **ben* ‘thought, feeling, mind’, which is a noun, with Proto-Indo-European **men-* ‘to think’, a verb. Their answer is affirmative. This does indeed seem very reasonable. The question is what the evidence is for Proto-Basque **ben* ‘thought, feeling, mind’. Here again, Blevins and Sproat are to be commended for laying out the evidence they have used, which is the following list of entries from the *OEH*:

- Proto-Basque **ben*² ‘thought, feeling, mind; mental power’
- B *amets, aments, amens, amentx* ‘dream, fantasy, illusion’ < **ha-men-Vs*
- B *ben, men* ‘serious, formal, grave’ (*ben* backformed from *-pen*)
- B *men*¹ ‘power, authority’
- B *men*³ in *men on* ‘respect’ (*on* ‘good’)
- B *men*⁴ ‘faculty, capacity’
- B *men*⁵ ‘great desire to do’
- B *-men, -pen* abstract noun-forming suffix, with central association to properties of mind, thoughts, expressions and feelings, as in: *gogamen* ‘thought, idea’ (*gogo* ‘thought, desire,... will, mind, intelligence’); *orhoipen, orhoitpen* (*oroipen*) ‘memory, recollection’ (*orhoit* ‘memory’); *erakuspen* ‘teaching, instruction’ (*erakutsi* ‘show, demonstrate’) *iziapen* ‘fright, fear, terror, horror’ (*izu, izi* ‘scared, frightened’); *mingapen* ‘pronunciation’ (*mingain* ‘tongue’)
- B *mendu* ‘character, propensity, inclination’
- B *omen* ‘opinion’

(Blevins & Sproat 2021: 517).

Let us start with the numbered dictionary entries for *men*. For *men*¹ ‘power, authority’, the *OEH* states that it is usually found declined in local cases: *menean* ‘in the power, locative’, *menera* ‘allative’, *menetik* ‘ablative’; e.g. *etsaiaren menetik* ‘from under the power of the enemy’ in the first book written in Basque, Etxepare (1545); *Jainkoaren menean* ‘under the power of God’, etc. As a full noun, not in a local case and not following another noun in the genitive case, *men* with the meaning ‘power,

authority' is only found in the writings of S. Mendiburu, one of Larramendi's Jesuit followers, and two later writers. Its exact relation to *mende* 'time; century; authority' is unclear. Or perhaps it may have arisen as a variant of *-pean*, *-bean* 'under' (e.g. *mahai pean* 'under the table'), from the locative case of *behe* 'low'. That is, *Jainkoaren behean* 'under God' > *Jainkoaren menean*, with historical nasalization of /b/ by the combined action of the two nasals of the genitive and the locative. A derivation from Latin *mens*, *mentem* for both *men* and *mende* has also been suggested, but this is rejected by Michelena for semantic reasons (Agud & Tovar 1988-1994; Arbelaiz 1978; see also Bastardas & Manterola 2020: 220-222). Notice that if, against Michelena's opinion, we accept a semantic link between 'under the power' and 'mind', we cannot discard that, rather than going back to Blevins' Proto-Basque, the word was borrowed from Latin.

For *men*³ 'humble, docile, simple', Blevins & Sproat are only interested in the sub-entry *men on* 'respect', an expression first found in a 1934 poem by N. Ormaetxea, who used the pen-name Orixé.

*Men*⁴ 'faculty, capacity' appears to be the creation of the lexicographer R.M. Azkue, who may have extracted it from *ikus-men* 'sight', *entzu-men* 'hearing', etc. It was used by several of his contemporaries, at the beginning of the 20th century, writing in the literary and cultural magazine *Euskal Esnalea*. The neologistic character of this word (or this meaning) is betrayed by the fact that these authors sometimes felt compelled to add a gloss to the word. For *men*⁵ the entire *OEH* entry reads «(L-côte), con grandes ganas de hacer' A». That is, the only source is Azkue's dictionary and no examples of it have been found in the corpus.

The rather productive derivational suffix *-men*, *-pen*, has often been related to Gascon *-men(t)* (< Lat *-mentu*), which may or may not be correct. Most of the examples of words with this suffix provided by Blevins & Sproat are neologisms of lexicographers and writers.

There are other meanings of *men*, such as 'moment; last moment' which Blevins & Sproat leave aside, but may be also crucial for establishing the original meaning or meanings of this word.

The genetic connection, if any, between *men*, in any of its meanings, and the other words that Blevins & Sproat have added to this list: *ame(n)ts* 'dream', *omen* 'reputation; apparently' and the rare western form *mendu* 'character, propensity, inclination' is totally unclear. Most likely these are just other words that happen to contain the sequence /men/. In the introduction of the *EHHE*, it is proposed for *omen* that its use as a particle ('apparently') is older than its use as a noun ('reputation') and that it ultimately derives from the verb *emon* 'to give'. In the same etymological dictionary, a couple of hypotheses for the origin of *amets* are also discussed.

Once again, I cannot find any uncontroversial evidence for a Proto-Basque root **ben* 'thought, feeling, mind'. The oldest documented form appears to be *men-*, with locative inflection and following another noun in the genitive case, meaning something like 'submission, being under someone's power'.

The etymologies that I have examined here are the ones that Blevins & Sproat (2021) decided to showcase. The inescapable conclusion is that pro-Indo-European bias—whether conscious or unconscious—is pervasive and has affected the

reconstruction of Proto-Basque to a much larger extent than the authors are willing to acknowledge or are aware of. This, in my view, invalidates the conclusions that we may be able to obtain from the statistical modelling presented in the paper. In many cases, this pro-IE bias may have been completely unconscious and almost impossible to avoid, since word associations arise spontaneously in our mind. If you are an English-speaker, **dol*, the Lakarrian root of *odol* ‘blood’, may bring the word *dolicocephalic* ‘long skull’ to mind;²⁰ *malko* ‘tear’ may immediately evoke *milk*, which has the same consonants; the Basque word *istil* ‘puddle, mud’ does sound like *still*, and the Basque word for ‘still, quiet’ is *isil*, etc. These spontaneously-arising mental word associations, which vary from person to person, may in fact be of great value when the languages that you are examining have already been shown to be related and may lead to true advances in etymology; but, if you are trying to establish a new genetic relation, they are bound to be a real problem. A more robust methodology for establishing Basque roots and their meanings is required.

Finally, regarding the treatment of Basque words with obvious cognates in Romance languages that I mentioned in 6.2 (*abiatu* = Sp *aviar*, *hanka* = Sp *anca*, *harpoi* = Sp *harpón*, *asmatu* = Old Sp *asmar*, *eskalapoin* = Gasc *esclopon*, etc.), Blevins & Sproat (2021) make the following statement, which I find hard to interpret:

Basque (and Aquitanian) data used as the basis for reconstruction of Proto-Basque in Blevins (2018) and in this study has been carefully screened to eliminate loanwords, especially loans from Indo-European languages with which Basque has been in contact. (Blevins & Sproat 2021: 513)

Examining all the etymologies proposed by Blevins (2018) would take many hours of work. Based on the cases that I have examined, my sense is that the likelihood—or the Bayesian priors, if one wishes—that a good number of the novel etymologies in that book would withstand this sort of scrutiny is rather low.

Gorrochategui (2020b), for similar reasons to those expressed here and some additional ones, ends his review of Blevins (2018) with these words:

Por todo ello, aun reconociendo el interés de ciertos análisis fonológicos y prosódicos y agradeciendo el esfuerzo por una disposición ordenada del material, su hipótesis a favor de una relación de parentesco entre su protovasco y el indoeuropeo no alcanza el nivel de evidencia comparativa suficiente. (Gorrochategui 2020b: 308)

I can only agree. Paraphrasing and updating Gorrochategui’s conclusion, Blevins’ (2018) book is admirably well organized and some of the phonological analyses that it contains are interesting, but it has serious methodological issues. The statistical modelling in Blevins & Sproat (2021) is not convincing, given the little confidence that we can have in the reconstructed Proto-Basque roots.

²⁰ This mental association may be particularly strong for those of us who were trained in Generative Phonology. A fact that arises from memory for me is Lightner’s (1983: 205) proposal that, in the synchronic phonology of modern English, the words *long*, *length* and *dolicocephalic* have an underlying representation containing the root /dl-/.

Abbreviations

- ASJU* = *Anuario del Seminario de Filología Vasca «Julio de Urquijo»*
EHHE = Lakarra, Manterola & Seguro 2019.
FLV = *Fontes Linguae Vasconum*
OEH = Euskaltzaindia 2017
RIEV = *Revista Internacional de los Estudios Vascos*
RS = Anonymous 1596

References

- Agud, Manuel & Antonio Tovar. 1988-1994. *Materiales para un diccionario etimológico de la lengua vasca I-XXI* (incomplete). Published in fascicles in *ASJU*. Available online as *Supplements of ASJU* 37. <https://ojs.ehu.eus/index.php/ASJU/article/view/8625>.
- Allières, Jacques. 1992. Gascón y euskera: afinidades e interrelaciones lingüísticas. *ASJU* 26 (3). 801-812.
- Anonymous. 1596. *Refranes y sentencias*. Pamplona: Pedro Porrallis de Amberes. (Critical edn. by Joseba A. Lakarra, 1996, Bilbao: Euskaltzaindia.)
- Arbelaz, Juan José. 1978. *Las etimologías vascas en la obra de Luis Michelena*. Tolosa: Kardaberatz.
- Azkue, Resurrección María de. 1904-1905. *Diccionario vasco-español-francés*. Bilbao. (Repr. 1969, Bilbao: La Gran Enciclopedia Vasca.)
- Bakker, Peter. 2020. Review. *Advances in Proto-Basque Reconstruction with Evidence for the Proto-Indo-European-Euskarian Hypothesis*. *FLV* 130. 563-592.
- Bastardas, Reina & Manterola, Julen. 2020. Pour un traitement digne du 21e siècle des emprunts au protoroman du basque dans le DÉRom. In Éva Buchi & Wolfgang Schweickard (eds.), *Dictionnaire Étymologique Roman (DÉRom) 3: Entre idioroman et Protoroman (Beihefte zur Zeitschrift für romanische Philologie 443)*, 217-242. Berlin-Boston: De Gruyter. <https://doi.org/10.1515/9783110654264-009>.
- Bengtson, John. 2017. *Basque and its closest relatives: A new paradigm. An updated study of the Euskaro-Caucasian (Vasco-Caucasian) Hypothesis*. Cambridge, Mass.: Association for the Study of Language in Prehistory.
- Blasco Ferrer, Eduardo. 2016. Protovascuence y paleosardo: reconstrucción y comparación / Palaeo-Basque and Palaeo Sardinian: Reconstruction and comparison. *Liburna* 9. 73-88.
- Blevins, Juliette. 2018. *Advances in Proto-Basque reconstruction with evidence for the Proto-Indo-European-Euskarian Hypothesis*. New York & London: Routledge.
- Blevins, Juliette & Ander Egurtzegi. 2017. Unexpected obstruent loss in initial obstruent-sonorant clusters: an apparent example from Basque. *Phonology* 34. 507-522.
- Blevins, Juliette & Richard Sproat. 2021. Statistical evidence for the Proto-Indo-European-Euskarian hypothesis: A word-list approach integrating phonotactics. *Diachronica* 38(4). 506-564. <https://doi.org/10.1075/dia.19014.ble>.
- Campbell, Lyle. 2017. Language isolates and their history. In Lyle Campbell (ed.), *Language isolates*, 1-18. Abingdon, Oxon & New York: Routledge.
- Chambon, Jean-Pierre & Greub, Yan. 2002. Note sur l'âge du protogascon. *Revue de linguistique romane* 66. 473-495.

- Echaide Itarte, Ana María [prepared by]. 1984. *Erizkizundi Irukoitza / Triple cuestionario / Triple questionnaire (= IKER 3)*. Bilbao: Euskaltzaindia.
- Egurtzegi, Ander & Gorka Elordieta. 2013. Euskal azentueren historiaz. In Ricardo Gómez, Joaquín Gorrochategui, Joseba A. Lakarra & Céline Mounole (eds.), *Koldo Mitxelena Katedraren III. Biltzarra*, 163-186. Bilbao: UPV/EHU.
- Egurtzegi, Ander & Gorka Elordieta. in press. A history of the Basque prosodic systems. *Diachronica*.
- Estornés Lasa, Bernardo. 1997. *Diccionario español-uskara roncalés/Erronkariko uskararen hiztegia*. Iruña/Pamplona: Nafarroako Gobernua/Gobierno de Navarra.
- Euskaltzaindia [Basque Academy]. 2017. *Diccionario General Vasco/Orotariko Euskal Hiztegia*. Online: <https://www.euskaltzaindia.eus/oe>.
- Forni, Gianfranco. 2013. Evidence for Basque as an Indo-European language. *Journal of Indo-European Studies* 41 (1/2). 2-142.
- Gaminde, Iñaki. 1998. *Euskaldunen azentuak*. Bilbao: Labayru.
- Gavel, Henri. 1920. *Éléments de phonétique basque*. Paris: Champion.
- Guerra, Juan Carlos de. 1924. *Viejos textos del idioma: Los cantares antiguos del euskera*. San Sebastián: Martín y Mena. http://www.liburuklik.euskadi.eus//applet/libros/JPG/diputacion/FSS_001450/FSS_001450.pdf.
- Gómez, Ricardo & Koldo Sainz. 1995. On the origin of the finite forms of the Basque verb. In José Ignacio Hualde, Joseba A. Lakarra & R. L. Trask (eds.), *Towards a history of the Basque language*, 235-273. Amsterdam: Benjamins.
- Gorrochategui, Joaquín. 1984. *Onomástica indígena de Aquitania*. Bilbao: UPV/EHU.
- Gorrochategui, Joaquín. 1995. The Basque language and its neighbors in antiquity. In José Ignacio Hualde, Joseba A. Lakarra & R. L. Trask (eds.), *Towards a history of the Basque language*, 31-64. Amsterdam: Benjamins.
- Gorrochategui, Joaquín. 2009. Vasco antiguo: Algunas cuestiones de geografía e historia lingüística. *Palaeohispanica* 9. 539-555.
- Gorrochategui, Joaquín. 2011. Euskara zaharra. In *IKER 26: Pirinioetako hizkuntzak: lebena eta oraina*, 695-713. Bilbao: Euskaltzaindia.
- Gorrochategui, Joaquín [Gorrotxategi, Joakin]. 2015. Akitaniera eta euskara. In Beatriz Fernández & Pello Salaburu (eds.), *Ibon Sarasola, gorazarre. Homenatge, homenaje*, 291-302. Bilbao: UPV/EHU. (Repr. in Gorrochategui, 2018, I, 445-454.)
- Gorrochategui, Joaquín. 2018. *Opera selecta I*, ed. by Blanca Urgell & José M. Vallejo. Vitoria-Gasteiz: UPV/EHU.
- Gorrochategui, Joaquín. 2020a. *Aquitanian-Vasconic: language, writing, epigraphy* (AELAW booklet 9). Zaragoza: Prensas de la Universidad de Zaragoza.
- Gorrochategui, Joaquín. 2020b. *Fontes Linguae Vasconum*: orígenes y documentos para una Historia del Euskara. In Ekaitz Santazilia, Dorota Krajewska, Eneko Zuloaga & Borja Ariztimuño (eds.), *Fontes Linguae Vasconum 50 urte*, 295-313. Iruña/Pamplona: Nafarroako Gobernua/Gobierno de Navarra. <https://doi.org/10.35462/fontes50urte.19>.
- Gorrochategui, Joaquín & Joseba A. Lakarra. 2013. Why Basque cannot be, unfortunately, an Indo-European language. *Journal of Indo-European Studies* 41(1/2). 203-237.
- Hualde, José Ignacio. 1993. On the historical origin of Basque accentuation. *Diachronica* 10(1). 13-50.
- Hualde, José Ignacio. 1997. *Euskararen azentuerak (Supplements of ASJU 42)*. Donostia & Bilbao: Gipuzkoako Foru Aldundia & UPV/EHU.

- Hualde, José Ignacio, 1999a. Basque accentuation. In Harry van der Hulst (ed.), *Word prosodic systems in the languages of Europe*, 947-993. Berlin: Mouton de Gruyter.
- Hualde, José Ignacio. 1999b. Pre-Basque plosives. In Jon Franco, Alazne Landa & Juan Martín (eds.), *Grammatical analyses in Basque and Romance linguistics*, 77-104. Amsterdam: Benjamins.
- Hualde, José Ignacio. 2003. From phrase-final to postinitial accent in Western Basque. In Paula Fikkert & Haike Jacobs (eds.), *Development in prosodic systems*, 249-281. Berlin: Mouton de Gruyter.
- Hualde, José Ignacio. 2006. Laringalak eta euskal azentuaren bilakaera. In Beatriz Fernández & Itziar Laka (eds.), *Andolin gogoan: Essays in honour of Professor Eguzkitza*, 497-511. Bilbao: UPV/EHU.
- Hualde, José Ignacio. 2008. Acentuación y cronología relativa en la lengua vasca. *Oihenart* 23. 199-217.
- Hualde, José Ignacio. 2012. Two Basque accentual systems and the notion of pitch accent language. *Lingua* 122. 1335-1351.
- Hualde, José Ignacio. 2015. Dialektologia dinamikoa. In Irantzu Epelde (ed.), *Euskal hizkerak eta dialektok gaur egun. Lapurdum*, Special Issue 3. 125-153.
- Hualde, José Ignacio. 2018. Aspiration in Basque. *Papers in Historical Phonology* 3. 1-27. <https://doi.org/10.2218/pihph.3.2018.2602>.
- Hualde, José Ignacio. 2020. Euskara Batu Zaharraren eta lehenagoko garaian berreraiketaren gainean: adibide bi. In Iñaki Camino, Xabier Artiagoitia, Irantzu Epelde & Koldo Ulibarri (eds.), *Eibartik Zuberoara euskalkietan barrena. Koldo Zuazori gorazarre* [Festschrift for Koldo Zuazo], 295-317. Bilbao: UPV/EHU.
- Hualde, José Ignacio. in press. Accent shift and the reconstruction of Old Common Basque accentuation. In Haruo Kubozono, Junko Ito & Armin Mester (eds.), *Prosody and prosodic interfaces*. Oxford: Oxford University Press.
- Igartua, Iván. 2015. Diachronic effects of rhinoglottophilia, symmetries in sound change and the curious case of Basque. *Studies in Language* 39. 635-663.
- Kassian, Alexei. 2013. On Forni's Basque-Indo-European Hypothesis. *Journal of Indo-European Studies* 41 (1/2). 181-201.
- Lakarra, Joseba A. 1985. Larramendiren hiztegitzaren inguruan. *ASJU* 19 (1). 1-50.
- Lakarra, Joseba A. 1995. Reconstructing the Pre-Proto-Basque root. In José Ignacio Hualde, Joseba A. Lakarra & R. L. Trask (eds.), *Towards a history of the Basque language*, 189-206. Amsterdam: Benjamins.
- Lakarra, Joseba A. 1997. Euskararen historia eta filologia: arazo zahar, bide berri. *ASJU* 31 (2). 447-535.
- Lakarra, Joseba A. 2009. Forma canónica y cambios en la forma canónica de la lengua vasca: hacia los orígenes del bisilabismo. *Palaeohispanica* 9. 557-609. (Acta Palaeohispanica, X.)
- Lakarra, Joseba A. 2011a. Erro monosilabikoaren teoria eta aitzineuskararen berreraiketa: zenbait alderdi eta ondorio. *FLV* 113. 5-114.
- Lakarra, Joseba A. 2011b [2014]. Gogoetak euskal dialektologia diakronikoaz: Euskara batu zaharra berreraiki beharraz eta haren banaketaren ikerketaz. In Irantzu Epelde (ed.), *Euskal dialektologia: Lehena eta oraina (Supplements of ASJU 69)*, 155-241. Bilbao: UPV/EHU.

- Lakarra, Joseba A. 2012. Mailegaketa eta berreraiketa euskararen historiaurrearen ikerketan. In Iván Igartua (ed.), *Euskara eta inguruko hizkuntzak historian zehar*, 17-74. Vitoria-Gasteiz: Eusko Jaurlaritza.
- Lakarra, Joseba A. 2013a. Aitzineuskararen berreraiketa sakonagorantz: forma kanonikoa, kronologia, gramatikalizazioa. In Ricardo Gómez, Joakin Gorrotxategi, Joseba A. Lakarra & Céline Mounole (eds.), *Koldo Mitxelena Katedraren III. Biltzarra*, 275-324. Vitoria-Gasteiz: UPV/EHU.
- Lakarra Joseba A. 2013b. Protovasco: comparación y reconstrucción... ¿para qué y cómo? (Por una vascológia autocentrada, no ensimismada). In Paolo Francalacci *et al.* (eds.), *Iberia e Sardegna. Legami linguistici, archeologici e genetici dal Mesolitico all'Età del Bronzo*, 127-150. Firenze: Le Monnier.
- Lakarra, Joseba A., Julen Manterola & Iñaki Segurola. 2019. *Euskal hiztegi historiko-etimologikoa (EHHE-200)*. Bilbao: Euskaltzaindia.
- Larramendi, Manuel de. 1745. *Diccionario trilingüe del castellano, vascuence y latin*, 2 vol. Donostia-San Sebastián: B. Riesgo y Montero.
- Larrasquet, Jean. 1939. *Le basque de la Basse-Soule Orientale (Collection Linguistique 46)*. Paris: Société de Linguistique de Paris.
- Lightner, Theodore. 1983. *Introduction to English derivational morphology*. Amsterdam: Benjamins.
- Lujanbio, Oihana. 2013 [2018]. Goizuetako azentu hiztegia. *ASJU* 47(2). 1-86.
- Luchaire, Achille. 1877. *Les origines linguistiques de l'Aquitaine*. Pau: Veronese.
- Manterola, Julen. 2015. *Euskararen morfologiaren historiarako: artikulua eta erakusleak/ Towards a history of Basque morphology: Articles and demonstratives*. Vitoria-Gasteiz: UPV/EHU. Doctoral dissertation.
- Manterola, Julen. 2019. Euskaltzaindiaren *Euskal Hiztegi Historiko-Etimologikoa* (EHHE-200): etorkizunari begirako aurkezpen bat. *Euskera* 64(2, 2). 1165-1212.
- Manterola, Julen. 2020. Erdi Aroko onomastika eta euskal lexikoaren berreraiketa: *Zuazoren* sustraiak. In Iñaki Camino, Xabier Artiagoitia & Koldo Ulibarri (eds.), *Eibartik Zuberoara euskalkietan barrena: Koldo Zuazori Gorazarre*, 511-534. Bilbao: UPV/EHU.
- Martinet, André. 1950. De la sonorisation des occlusives initiales en basque. *Word* 6. 224-236. (Repr. in André Martinet, 1964, *Économie des changements phonétiques*, 2nd edn., 370-388. Bern: A. Francke.)
- Meillet, Antoine. 1925. *La méthode comparative en linguistique historique*. Oslo: Instituttet for Sammenlignende Kulturforskning.
- Michelena, Luis. 1957. Las antiguas consonantes vascas. In *Miscelánea homenaje a A. Martinet*, vol. 1, 113-157. Universidad de La Laguna. (Repr. in Michelena 2011, vol. VII. 57-92. English translation by R. L. Trask, The ancient Basque consonants. In José Ignacio Hualde, Joseba A. Lakarra & R. L. Trask (eds.). 1995. *Towards a history of the Basque language*, 101-136. Amsterdam: Benjamins.)
- Michelena, Luis. 1963. *Lenguas y protolenguas*. Salamanca, Universidad de Salamanca (= *Acta Salmanticensia, Filosofía y Letras*, 17, no. 2.) (Repr. in Michelena 2011, vol. I. 1-98.)
- Michelena, Luis, 1961. Los nombres indígenas de la inscripción hispano-romana de Lerga (Navarra). *Príncipe de Viana* 22. 65-74. (Repr. in Michelena 2011, vol. V. 261-264.)
- Michelena, Luis. 1970. Nombre y verbo en la etimología vasca. *Fontes Linguae Vasconum* 2. 67-93. (Repr., in Michelena 2011, vol. VII. 235-267.)

- Michelena, Luis [Mitxelena, L.] 1971. Egunak eta egun izenak. *Munibe* 23. 583-591. (English translation by José Ignacio Hualde, Days and names of days, 2020, *FLV* 130. 595-616.)
- Michelena, Luis. 1977. *Fonética histórica vasca*, 2nd edn. Donostia-San Sebastián: Diputación Foral de Guipúzcoa. (Repr. in Michelena 2011, vol. VI.)
- Michelena, Luis. 1981. Lengua común y dialectos vascos. *ASJU* 15. 291-303. (Repr. in Michelena 2011, vol. VII. 517-544.)
- Michelena, Luis. 2011. *Obras completas*, 15 vols. (*Supplements of ASJU* 54-69), ed. by Joseba A. Lakarra & Iñigo Ruiz Arzalluz. Donostia-San Sebastián & Vitoria-Gasteiz: Gipuzkoako Foru Aldundia & UPV/EHU.
- Moguel, Juan Antonio. 1881 [1802]. *El doctor Peru Abarca, catedrático de la lengua vascongada en la Universidad de Basarte*. Bilbao: Elizalde. (Online: <https://klasikoak.armiarma.eus/idazlanak/M/MogelJAAbarka.htm>).
- Mounole, Céline. 2014 [2018]. *Le verbe basque ancien: étude philologique et diachronique*. Vitoria-Gasteiz: UPV/EHU (= *ASJU* 48). <https://ojs.ehu.eus/index.php/ASJU/issue/view/1701>.
- Mounole, Céline & Joseba A. Lakarra. 2018. Euskara arkaikoa. In Joakin Gorrotxategi, Iván Igartua & Joseba A. Lakarra (eds.), *Euskararen historia*, 352-468. Vitoria-Gasteiz: Eusko Jaurlaritzza/Gobierno Vasco.
- Olalde, Iñigo, et al. 2019. The genomic history of the Iberian Peninsula over the last 8,000 years. *Science* 363 (6432). 1230-1234. <https://doi.org/10.1126/science.aav4040>.
- Rix, Helmut, Martin Kümmel, Thomas Zehnder, Reiner Lipp & Brigitte Schirmer. 2011. *Lexicon der Indogermanischen Verben*, 2nd edn. Wiesbaden: Reichert.
- Rohlf, Gerhard. 1977. *Le gascon*, 3rd ed. (= Beihefte zur Zeitschrift für romanische Philologie, 85). Tübingen & Pau: Max Niemeyer Verlag & Mairimpouey Jeune.
- Saroihandy, Jean. 1913. Vestiges de phonétique ibérienne en territoire roman. *RIEV* 7 (4). 475-497.
- Trask, R. L. 1985. On the reconstruction of pre-Basque phonology. In José L. Melena (ed.), *Symbolae Ludovico Mitxelena septuagenario oblatae*, 885-891. Vitoria-Gasteiz: UPV/EHU.
- Trask, R. L. 1997. *The history of Basque*. London: Routledge.
- Urgell, Blanca. 2004. Etimologia eta neologia Larramendiren *Hiztegi Hirukoitz-ean* (1745). *Lapurdum* 9. 299-310. <https://doi.org/10.4000/lapurdum.882>.
- Watkins, Calvert. 2011. *The American heritage dictionary of Indo-European Roots*, 3rd edn. Boston & New York: Houghton Mifflin Harcourt.