

Mikel Martínez-Areta (ed.)

# Basque and Proto-Basque

Language-Internal and Typological Approaches  
to Linguistic Reconstruction

**Mikroglottika**

Minority Language Studies

Volume 5



PETER LANG  
EDITION

## Introduction

*Mikel Martínez-Areta*

*(University of the Basque Country, UPV/EHU)*

Much of the literature on the history of the Basque language —the only surviving language of non-Indo-European origin in western Europe— has dealt with its origin and possible relatives. A list of the many attempts to relate it to another language or language family would be exasperatingly long, and most probably not exhaustive. Leaving aside the obvious relationship of historical Basque to Aquitanian, Trask (1995a) refuted all of them, or more precisely, denied that any of the alleged relationships had been proven so far, and in Trask (*HB*: 358-429) he made a magnificent *reductio ad absurdum* of the methods which had been used in such attempts, arguing that such a methodology can lead us to prove whatever linguistic connection we wish (as he demonstrated with Basque and Hungarian). Since then, such attempts have not ceased, some of them unprecedented. The Basque language, however, continues to be a language isolate, and there is no indication that it will cease to remain as such in the near future.

Unfortunately, this ceaseless effort of historical research to establish unprovable theories has often diverted attention from paths which might have proven successful for making progress in the elucidation of its prehistorical aspects. As a result, Basque diachronic linguistics suffers from a historical delay in comparison with the equivalent discipline in many other languages or language families. In particular, progress made in diachronic cross-linguistic typology since the seventies has not been systematically applied to Basque until very recently, and then only in literature mostly written in Basque, Spanish or French. The purpose of this book is precisely to fill this gap, to provide an account of this situation, and to turn the spotlight on this possibility for research.

The starting point for this purpose, then, will be that Basque has as yet no proven relative, and hence no possibility for linguistic comparison beyond its geographical boundaries (for the Basque-Iberian question, see below). Our sources of knowledge are limited, and any honest attempt to explore the past of the language ought to start from this assumption. Written texts do not appear until roughly 1500, and then only in particular dialects and geographical areas. The attestations from the five centuries prior to this, mainly consisting of onomastic material in documents written and transmitted in medieval monasteries, do not provide any crucial data which can be of help for historical grammar. The same holds true for the Aquitanian inscriptions engraved on stone in the Roman

period (ca. 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> centuries AD), and for the place names provided by Greek and Latin authors from Antiquity.

With such a scarcity of sources, Bascologists have had to maximize the utility of the information provided by the linguistic contact with Latin and Romance. It cannot be a coincidence that the first scholar who tried to carry out systematic historical linguistic work on Basque in the modern sense of the word, Hugo Schuchardt (1842 – 1927), was a Romanist. Since Schuchardt's work,<sup>1</sup> Basque linguistic studies have made much progress, but, as I have said, it is hard to escape the impression that serious diachronic work has been delayed by misconceptions, uchronic conceptions and clichés, most of them derived from its status as a language isolate and from the fact that much research has focused on the search for relatives. Thus, it was not until the 1950s that the structuralist method was first employed to establish some basic game rules in historical reconstruction. This enabled Basque diachronic linguistics to catch up with the functionalist-structuralist methodology which flourished in the post-war period, but, indeed, in the diachronic research of other language families such as Romance, Germanic, Slavic or Indo-European linguistics, the Neo-Grammarians and their phonetic law had already provided a useful tool for reconstructing analysis as early as the end of the 19<sup>th</sup> century.<sup>2</sup>

These functionalist-structuralist approaches entered the Basque discipline thanks to André Martinet (1908 – 1999, who was not a Bascologist himself, but was the spearhead of the functionalist school), René Lafon (1899 – 1974), and especially Luis Michelena (1915 – 1987). Basque linguistics in the second half of the 20<sup>th</sup> century, even after Michelena's death and in fact until the present day, is dominated by this prominent figure. Before I move on to comment on his work in some detail, however, I would like to observe that, in my opinion, Basque historical grammar has still much to profit, in several respects, from a deeper reading of Lafon's work. Some crucial keys to understanding the formation of both the nominal and the verbal inflections, such as the lack of grammaticalized categories of number and diathesis, were already pointed out by him, were later ignored —or very superficially considered— for decades, and only in recent years are they being timidly recovered. The explaining potential of these views in the light of the progress made in recent typological research is, as I see it, enormous.

Luis Michelena's work, and particularly his *Fonética Histórica Vasca* (1961), is the sieve through which any reconstructing hypothesis must pass be-

---

1 For a history of the linguistic work on Basque, cf. Trask (*HB*: 49-70).

2 It is perhaps worth recalling that Schuchardt vigorously opposed the Neo-Grammarian views.

fore being subjected to ulterior tests. It is with Michelena that diachronic research on Basque reaches the maturity necessary to locate any posed problem or hypothesis within the coordinates of the scientifically possible and reasonable. As Trask (*HB*: 67) pointed out, Michelena worked on practically every linguistic area: historical phonology, the verbal system, place names, personal names, etymology, syntax, genetic relations, Romance influences, dialectology, literary criticism, the word-accent, etc. And for practically all of these areas, his studies have been the starting point for further research in recent decades.

After Michelena's death, Basque diachronic linguistics came to an impasse which lasted for some time. Indeed, the impression somehow caught on that his work on historical issues was insuperable, and that little or no contribution could ever be made to the field beyond his work.<sup>3</sup> In addition, the 70s and 80s brought the generative fever, and —perhaps somewhat unexpectedly in a discipline like Basque linguistics— scholars' interest drifted towards synchronic issues.

In spite of all that, Basque philologists continued their meticulous labor, which brought major lexicographic and editing achievements. The accentual diversity of Basque, a realm in which Michelena insisted that research was still in its infancy, was carefully studied by Hualde in the nineties, from both a synchronic and diachronic point of view. And since 1995, upon the solid basis of Michelena's reconstructing paradigm, Lakarra has developed another paradigm which attempts to go even further back in the history of the language, taking the study of the canonical root as the keystone. This diachronic enterprise was also strongly favored by foreign Bascologists such as Roger Larry Trask (1944 – 2004), Rudolf de Rijk (1937 – 2003) and Georges Rebuschi. The premature death of the first two was particularly unfortunate for the field, as one cannot avoid the impression that they left us just when a discussion largely generated and fostered by them was beginning to become interesting, and that their participation in it would have been of the utmost relevance.

As a consequence of that perseverance, Bascologists are more optimistic today than several years ago about the possibility of obtaining fruitful results in diachronic research, and in the last ten years there has been a renewed interest in the field among scholars and graduate students (the contributors in this volume are a good token of this). In comparison to giants like Lafon or Michelena —to use the well-known metaphor attributed to Bernard of Chartres—, we may be dwarfs, but we are dwarfs standing on their shoulders. In addition to having their work already completed, we have more sources and more typological information than they ever had access to. The situation in this respect is also more

---

3 With regard to historical phonology, Trask (*HB*: 6) stated: “[s]ince Michelena, little remains beyond tidying up the details”.

favorable than that of 16 or 18 years ago, when Trask's *The History of Basque* (Trask 1997) and *Towards a History of the Basque Language* (Hualde & Lakarra & Trask (eds.) 1995) were respectively published. Within the realm of Basque philology and linguistics, several texts have been discovered, Lazarraga's manuscript being the most important. Dialectal research has filled some gaps that we had about archaizing dialects, and a number of inscriptions from Antiquity containing Basque onomastic data have also been found since Gorrochategui's canonical work on the issue (Gorrochategui 1984). While all this has not drastically changed the overall picture and we still have to face the same, at times demoralizing, scarcity of sources, the sum of all of these elements has enlarged our knowledge of several important aspects.

As Romance and Indo-European diachronic linguistics show, apart from the number of records we have, a prerequisite for progress is often simply that enough time elapses for scholars to process and arrange the information at their disposal. A Bascologist of 2012 is lucky to have, after many years of laborious work, a *General Dictionary of Basque* (Euskaltzaindia 1987-2005), in 16 volumes, with copious records from the written tradition for each lexical item, and Yrizar's *Morfología del Verbo Auxiliar Vasco* (Yrizar 1991-2008), in 15 volumes, which is virtually an encyclopedia of the Basque auxiliary verb. Similarly, in addition to the discovery of the new texts mentioned above, other texts that we already had but which had either remained unedited, or only very unsatisfactorily so, have now been properly edited and more profoundly discussed.

In addition to the progress made within Basque philology and linguistics, cross-linguistic typological research at an international level has also considerably advanced in the last four decades. Since the 1970s, new linguistic universals have been set up, some that were already known have been better attested, some others refuted, and still others have given rise to endless controversies between opposing views. By the same token, frequent grammaticalization paths have been discovered and/or discussed. Trask (1977, 1979, i.a.) was a pioneer in the use of these resources for diachronic research on Basque, but the bibliographical flood has been particularly copious in the last two decades. As a result, today we have essential handbooks such as Corbett (2000), Siewierska (2004), Bybee & Perkins & Pagliuca (1994), Heine & Kuteva (2002), etc., as well as many more descriptive monographs on particular languages or language families. A path of research which is particularly promising in relation to Basque, and which has begun to be systematically exploited by Lakarra and some of his disciples, is the study of serial verb constructions and their grammaticalization, especially into different postpositions, which in turn may become suffixes. Here again, without the assistance of contributions like Lord (1993) and Aikhenvald & Dixon (eds., 2006), such an enterprise would prove much more difficult.

By contrast, a field which has not rendered any helpful results for Basque diachronic research, at least as far as grammar is concerned, is its comparison with the Iberian language. The theory that Basque and Iberian are originally the same, or at least relatives, lost ground in the course of the 20<sup>th</sup> century, first with the decipherment of the Iberian script by Gómez-Moreno in the second quarter of the century, and then with the work, critical of their relationship, by Tovar and Michelena in the 1950s. Some serious Iberists such as Orduña (2005) and Ferrer i Jané (2009) —and any reader who is acquainted with the literature which is often published in this realm will understand the relevance of the adjektivization of ‘Iberists’ here— have recently revived that line of research.<sup>4</sup> However, their analysis has mainly focused on numerals and some lexical items. Hence, even if we accepted their analysis, and/or even if we believed that Basque and Iberian were to some degree relatives, it would still continue to be true that Iberian has not yet been of any help in the elucidation of any aspect related to the historical grammar of Basque.<sup>5</sup> Consequently, Iberian data are barely mentioned throughout the entire work.

This book is an attempt to present a general picture, for both Bascologists and typologists —especially those interested in diachronic phenomena—, of the

- 
- 4 Against the possibility of such a comparison, see Lakarra (2010b). Orduña explained the similarities between Basque and Iberian numerals by assuming that the former borrowed them from the latter, whereas for Ferrer i Jané those numerals would be cognates descending from a common language. However, Orduña (2011) has recently adhered to Ferrer i Jané’s view on this particular point.
- 5 The Iberian attestations which have been related to Basque and which could theoretically provide some help for grammatical comparison are very few, and extremely dubious. Thus, an *ekiar* found in Caminreal (Teruel) and an *ekien* found in Mendigorriá (Navarre) have been related to western Basque *egian* ‘(s)he made’. More possible forms with this lexeme *eki*- ‘make’ are identified by Orduña (2010), who even suggests that an *eroki*-, allegedly a causative of *eki*-, might be the lexeme in two further inscriptions. Such a semantics certainly belongs to the narrow lexical scope which we expect to find in inscriptions, but the comparison with *egian* has, among others, the crucial problem that we reconstruct *\*(z)e-gin-a/e-n* for Basque, and hence we would expect the appearance of an intervocalic *-n-* —or at least *-h-*, which does not exist in Iberian— in *ekiar*, *ekien*, etc. Some other classical similarities with Iberian, such as the filiation suffix *-(t)ar* or the alternation *-n / -r / -Ø* are interesting and worth exploring, but still insufficient for helping us to better understand any grammatical aspect of Basque from a diachronic point of view. Another question which has not been settled is whether Iberian was a unitary language throughout the entire peninsular east —and if so to what degree—, or it only corresponded to a limited region within that area (perhaps that of the Contestani of Antiquity and some other tribes around them), and in other regions where Iberian inscriptions are found —such as those closest to the Basque-speaking territory— it was simply a *lingua franca*, as proposed by de Hoz (1993).

current situation of that research. Its structure, then, resembles —or at least is aimed to resemble— that of a historical grammar. Extra-linguistic issues such as population genetics or paleo-anthropologic research are left aside. While smatterings of Basque grammar, philology and culture can be useful for grasping some details, the work is intended to be comprehensible for as broad a public as possible. Hence, each chapter explains the linguistic features and facts in historical Basque, and the ample section of Abbreviations at the beginning contains a guide to the dialects and the main linguistic sources that appear throughout the work.

Chapter 1 (by Barreña & Ortega & Amorrortu) is a description of the Basque language today, in which an accurate account of the number of speakers and their percentages in each administrative region, and of the evolution of those rates during the last decades, is given. The socio-political background underlying all those data is likewise discussed, as well as the historical factors which have led to them. Finally, the sociolinguistic aspects of today's linguistic situation are described. Speakers of Basque are classified into different types, and an attempt is made to account for their subjective attitudes towards the language.

In Chapter 2 (by Martínez-Areta), the degree of intelligibility between Basque dialects is explained, they are enumerated one by one and the basic information about each one (its geographical situation, approximate number of speakers, degree of idiosyncrasy, the antiquity of the oldest sources, etc.) is provided. The issue of how many dialects there are is also briefly touched upon, but the diachronically most relevant section discusses at what chronological point a Common Basque can be postulated as the origin of all dialects, and how the branching process of those dialects may have occurred. Before entering the linguistic discussion proper, in Chapter 3 (by Ulibarri) a short external history of the Basque language is given. Since this also serves the purpose of describing the majority of the sources available to us for the historical research of Basque, this external history is complemented by two sections about dialectology and toponymy, which are two additional sources of linguistic evidence.

Chapters 4 and 5 deal with phonetics, phonology and root structure. In Chapter 4 (by Egurtzegi), the phonemic inventory of Basque in the historical period is described, and then the reconstructions proposed by Michelena for Proto-Basque, and by Lakarra for Pre-Proto-Basque, are set out. First, this is done for the vocalism, then for the consonantism. Finally, phonotactics and the old accent systems are also discussed. Chapter 5 (by Lakarra) is not only an explanation of the root structure in diachronic terms, but also of how its study has served as a starting point for the development of a new reconstructing paradigm at all levels. It also contains two appendices. The first consists of three tables in which all attested root structures are analyzed with respect to all their possible

theoretical combinations. The second is a sample of 50 proposed etymologies which exemplify most of the root structure processes defended throughout the chapter.

We move next to the morphological part of the volume, which comprehends chapters 6 to 9. Chapter 6 (by Santazilia) is a study of the Basque noun morphology, which sets out the noun paradigm of historical Basque and the structure of its NP, and then goes on to list the explanations proposed so far for every aspect concerning the formation of that paradigm: first the number/definiteness axis, and then the individual cases, classifying these into primary cases (grammatical and local), secondary cases (those built upon the allative and the genitive), and pseudo-flectional morphemes or non-cases. In Chapter 7 (by Martínez-Areta), demonstratives—in both their determinative and pronominal functions—and personal pronouns are discussed. The discussion about the origin of the former is considered in connection with the origin of the noun paradigm. As for the latter, of special relevance is the alternation between weak and strong (or even hyperstrong) forms, the syntactic/pragmatic conditioning of that alternation, and, in the case of the genitive forms, the origin, functioning period and decline of the Linschmann-Aresti Law.

Chapters 8 and 9 deal with the verbal morphology, splitting it into the study of the non-finite verb and the finite verb, as is usual in analysis, according to the structure of the Basque verb. In Chapter 8 (by Padilla-Moyano), first the morphological formations—by means of an ample array of allomorphs of diverse origin—of the participle, of the gerund (both built, at least originally, upon the verbal root), and of the prospective participle are explained. Secondly, the author discusses the origin of the different kinds of periphrases which resulted from the possible combinations of all those non-finite verbal forms with the different kinds of auxiliaries—built upon verbal lexical roots which over the course of time became auxiliary—, thus giving rise to a tense/aspect/mood system whose diachronic development from Common Basque to contemporary Basque is extremely interesting. In Chapter 9 (by Ariztimuño), in turn, the morphological structure and origin of the finite verbal forms—i.e. of both the synthetic verbs and auxiliary verbs of periphrastic constructions—are described and discussed. For that purpose, the different categories which intertwine in that structure, such as person, number, tense, aspect, mood and valency are dissected and analyzed both individually and in relation to the rest of the parameters. Then, one of the most characteristic features of the Basque finite verb, allocutive agreement, is covered, and finally the differences between auxiliary and synthetic verbs are discussed, both synchronically and diachronically. The chapter ends with a set of interesting conclusions and indications for future research.



In Chapter 10 (by Reguero-Ugarte), which closes the volume, the main features of Basque word order are summarized, first those corresponding to clause-level phenomena, which can be syntactically or pragmatically conditioned, and then those within the phrase. Some of the latter show diachronic and/or diatopic variation in the attested data, and hence a directionality with respect to the OV vs. VO dichotomy is worth analyzing, as it can provide us with important clues about the changes which have occurred in all the other parameters, including the phonological and morphological ones. Finally, the main hypotheses about the word order in Proto-Basque, which inevitably have to resort to internal reconstruction —focusing especially on the morphology of the finite verb—, are put forward.

For place names which have an equivalent in English, the form in English has been used. Obviously, it is not always easy to draw a line between those which have an English form and those which do not. Clear cases are e.g. Navarre and Biscay, but what about Cizain or Aescoan? In general, I have interpreted that the names of the provinces (*herrialde* in modern Standard Basque) and dialects can be given in English, and here this work differs from Trask (*HB*), in which e.g. the Souletin dialect is called *Zuberoan*. By contrast, town and village names are given in their official Basque forms (except in some cases such as Bilbao, where the Basque official form *Bilbo* sounds somewhat awkward). Elements from intermediate levels, such as subdialects, varieties or valleys are given in Basque (*Sakana*, *Deba Valley*, etc.), unless we have considered that they have an English equivalent. Obviously, this decision may be arbitrary in some cases (as is the case of *Roncalese*, *Salazarese*, *Aescoan*, *Cizain*, etc.). In the cases of town names in which the Basque and the Spanish/French official names are completely different (e.g. Basq. *Iruñea* / Sp. *Pamplona*, capital of Navarre) —not mere spelling (e.g. Basq. *Azkoitia* / Sp. *Azcoitia*) or pronunciation variants (e.g. Basq. *Sara* / Fr. *Sare*), or variants which are somewhat similar to each other due to their common origin (e.g. Basq. *Beskoitze* / Fr. *Briscons*)—, both variants are usually given.

As for the cited authors, the names of those corresponding to researchers mentioned in the references are spelled as they themselves signed their work. So are in general those corresponding to the linguistic sources, but this is not always done, and a justification for the election is not systematically provided. To give but one example, Lazarraga will be spelled with <z>, and its abbreviation will be Laz., even though he signed as *Laçarraga*, in order to strengthen the distinction from Lç. (corresponding to *Leiçarraga*), as they will both be frequently cited throughout some chapters of the text.

By contrast, the extracts of texts taken from the written tradition are given, for practical reasons, in the spelling system of Standard Basque, following the

procedure used in the *General Dictionary of Basque*. However, data from medieval documents, or even from later documents but which do not belong to the literary corpus proper —such as those found in archives, diplomatic collections, etc.—, are given in their original spelling.

Some of the abbreviations for dialects, authors and works have been inspired by the system used in Trask (*HB*) or in the *General Dictionary of Basque* (Euskaltzaindia 1987-2005), but the resulting whole is original, and is hence detailed in the Abbreviations. My final resolution comes across a couple of possible duplicities. One is B, which by itself stands for Biscayan, but in PB stands for Proto-Basque, in CB for Common Basque, and in SB for Standard Basque; the other is S, which stands for Souletin when referring to the dialect, but for Subject when referring to the clause constituent, and for Subject Intransitive when referring to the argument type. Since the context makes it absolutely clear which one is involved in each case, I have not considered it necessary to take any measure beyond warning against them at this point.

The chronological terminology which is used throughout the work contains two basic terms. One is Common Basque (= CB), which is the diachronic projection of all Basque historical dialects, and hence the old common language prior to the dialectal fragmentation, from which all dialects branched. Michelena (1987 [1981a]) dated it to around the 5<sup>th</sup>-6<sup>th</sup> cc., and while there is no conclusive proof, it is commonly accepted since then that this might be close to the reality. This does not exclude, however, the existence of multiple innovations which have occurred after that period but which have reached all dialects. The other basic term is Proto-Basque (= PB), which would be a stage of the language prior to CB, commonly associated with Michelena's classical reconstruction of the consonant system. Much more arbitrarily, it is dated to the last centuries before the Christian Era, its chronological position being sometimes defined as the stage of the language immediately prior to the beginning of its contact with Latin. Throughout the volume, however, some authors stratify PB into different phases, which they can specify by adjectivizing PB with Old, Classical, Pre-, etc. In particular, Lakarra's theory of the root is to be located in a phase previous to PB, which can therefore be labeled as Pre-Proto-Basque or Old Proto-Basque.

This volume is a monograph published within the series of studies *Mikroglottika*, an international journal of minority language philologies printed by the Peter Lang Publishing Group, and co-edited by Raúl Sánchez-Prieto, Daniel Veith, and myself. As the editor of this particular volume, I would like to thank my co-editors of *Mikroglottika* Raúl Sánchez-Prieto and Daniel Veith for their encouragement in the production of this volume, and the publisher Peter Lang for their agreement on suggested ideas and their assistance in carrying them out.

The impression and a linguistic supervision of the text have been funded by the Department of Culture of the Government of the Basque Autonomous Community, to which the editor and the contributors of this volume wish to thank for their generous help. I am also thankful to Ruth Yates, for her patient reading and linguistic supervision of the text, and to Alphax Studio for the design of the figures and maps of Chapters 1 and 2. Likewise, I would like to mention that I belong to a number of research projects, within the framework of which this work has been carried out:

- 1) The research project “Monumenta Linguae Vasconum III-IV: historia, crítica y edición de textos vascos” [= FFI 2008-04516, FFI2012-37696], led by Joseba Lakarra and funded by the Ministry of Education and Science of the Spanish Government.
- 2) The research project “Historia de la Lengua Vasca y Lingüística Histórico-Comparada” (HLMV-LHC), Ref. GIC 10/83, IT 486-10, led by Joaquín Gorrochategui and funded by the Government of the Basque Autonomous Community.
- 3) The Training and Research Unit “Hizkuntzalaritza Teorikoa eta Diakronikoa: Gramatika Unibertsala, Hizkuntza Indoeuroparrak eta Euskara” (UFI 11/14), funded by the University of the Basque Country (UPV/EHU).