ADVANCES IN PROTO-BASQUE RECONSTRUCTION WITH EVIDENCE FOR THE PROTO-INDO-EUROPEAN-EUSKARIAN HYPOTHESIS

Juliette Blevins

Advances in Proto-Basque Reconstruction with Evidence for the Proto-Indo-European-Euskarian Hypothesis

This book presents a new reconstruction of Proto-Basque, the mother language of modern Basque varieties, historical Basque, and Aquitanian, grounded in traditional methods of historical linguistics. Building on a long tradition of Basque scholarship, the comparative method and internal reconstruction, informed by the phonetic bases of sound change and phonological typology, are used to explain previously underappreciated alternations and asymmetries in Basque sound patterns, resulting in a radically new view of the proto-language. The comparative method is then used to compare this new Proto-Basque with Proto-Indo-European, revealing regular sound correspondences in basic vocabulary and grammatical formatives. Evaluation of these results supports a distant genetic relationship between Proto-Basque and Proto-Indo-European, and offers new insights into specific linguistic properties of these two ancient languages. This comprehensive volume, which includes a detailed appendix including Proto-Basque/Proto-Indo-European cognate sets, will be of general interest to linguists, archeologists, historians, and geneticists, and of particular interest to scholars in historical linguistics, phonetics and phonology, language change, and Basque and Indo-European studies.

Juliette Blevins is Full Professor in the Linguistics Program at the Graduate Center, the City University of New York, USA, where she is Director of the Endangered Language Initiative. Her main research interests are sound patterns and sound change, with a special focus on phonological typology, as detailed in her influential book *Evolutionary Phonology: The Emergence of Sound Patterns*.

Routledge Studies in Historical Linguistics

Claire Bowern,
Yale University, USA

The Diachrony of Verb Meaning
Aspect and Argument Structure
Elly van Gelderen

Advances in Proto-Basque Reconstruction with Evidence for
the Proto-Indo-European-Euskarian Hypothesis
Juliette Blevins

Advances in Proto-Basque Reconstruction with Evidence for the Proto-Indo-European-Euskarian Hypothesis

Juliette Blevins

For Molly, Lucy, and Rebecca,

eguzkia, ilargia, eta izarrak,
lurra, sua, eta ura,
hiruak zarete nire gogo-bihotza.

*   *   *

Contents

Foreword

This book presents a new reconstruction of Proto-Basque (or Proto-Euskarian) and evidence that Proto-Basque is related to Proto-Indo-European. Both of these proposals diverge radically from earlier work. The Proto-Basque sound inventory and phonotactics deviate from previous proposals. And the arguments that this new Proto-Basque is a sister or cousin of Proto-Indo-European goes against the prevailing view of Euskarian as a language isolate. The new detailed analysis presented here is based on classical techniques within the field of historical linguistics: internal reconstruction and the comparative method. At the same time, it is embedded within my own understanding of the phonetic basis of sound change, phonological typology, and linguistic typology more generally.

  The research for this book began about five years ago. In February 2013, Ander Egurtzegi, then a graduate student at Euskal Herriko Unibertsitatea (UPV/EHU, the University of the Basque Country) under the primary supervision of Professor Joseba Lakarra, came to the Graduate Center at CUNY to spend a semester working with me on Basque historical phonology. At the time, I was co-supervisor of Dr. Egurtzegi's dissertation *Towards a phonetically grounded diachronic phonology of Basque*, completed in 2014. To prepare for this supervision, I read the materials Ander had been kind enough to send in advance, including Luis Michelena's collected works, Gorrochategui's Aquitanian monograph, and Professor Lakarra's studies of Proto-Basque root structure. The semester was an intense immersion in Basque historical phonology, and while Ander considered sound changes from the Romance contact period forward, I began to explore alternatives to traditional views of the oldest reconstructable state of the language. The most serious difficulty in this area is figuring out what is native Basque and what is not. Romance loans have overwhelmed the language, but Celtic and Germanic loans have also been suggested. By the end of April 2013, I had the beginning of the system presented here, based on inherited Basque words as vetted by Ander. The most striking aspect of the Proto-Basque roots that I reconstructed was their similarity in form and meaning to Proto-Indo-European roots. While my focus was primarily on the Basque material, I was soon side-lined by what appeared to be identifiable regular sound correspondences between Proto-Basque, as I understood it, and Proto-Indo-European. An early version of this work was presented with Ander at the Twenty-First International Conference on Historical Linguistics in Oslo in August 2013 and feedback from that conference was helpful

in many ways. It became clear that there was no agreed upon standard for what constituted evidence for a long-distance genetic relationship of the kind I was suggesting. At the same time, I was reminded of the minimal requisites: regular sound correspondences with good semantic matches in basic vocabulary, and, if possible, some statistical evidence for the relationship.

For the last four years, I have devoted a great deal of research time to this project. I have tried to do this, for the most part, on my own, consulting with specialists as need be. One area where I have sought collaboration is in statistical arguments for genetic relatedness. Dr. Richard Sproat of Google Inc., a longtime friend and colleague, has been kind enough to offer his expertise and assist in this area. Chapter 8 of this book presents potential statistical evidence for a relationship between Proto-Basque and Proto-Indo-European, including reference to Blevins and Sproat (in progress).

As the literature in Basque historical linguistics and Indo-European historical linguistics is vast, I hope colleagues will forgive me for only citing in the text references that are encyclopedic, easily available or directly relevant to the points of discussion. The appendix, which contains proposed Proto-Basque etymologies and external comparisons, is more thorough in referencing, and, wherever possible, cites Proto-Indo-European reconstructions that are widely agreed upon, and easily accessed.

At the same time, there is a great degree of detail in the specific analyses presented. To help readers new to Basque historical linguistics and/or Indo-European historical linguistics, respectively, there are brief introductions to Parts I and II of the volume which provide general background references on these two language families and grammatical profiles, including the linguistic features under comparison.

Though I am a relative new-comer to the fields of Basque historical linguistics and Indo-European historical linguistics, three fortunate circumstances have allowed me to explore the Proto-Indo-European-Euskarian Hypothesis with some degree of rigor. The first is the digital revolution that has made the entire *Orotariko Euskal Hiztegia*, the majority of Aquitanian inscriptions, and medieval cartularies with Basque placenames, easily accessible and searchable. The second is the generosity of colleagues in offering comments and criticisms in their areas of expertise, and correcting the many mistakes and oversights in my earlier work. I am sure that many small errors, inconsistencies, and omissions remain, but the overall strength of the arguments presented owes a great deal to the expertise and critical eye of others, noted in the acknowledgments and throughout the text where relevant.

Finally, as a linguist of the present era, I have the superb history of scholarship in Indo-European historical linguistics and Basque historical linguistics behind me. Everything that follows attempts to build respectfully on those foundations, with hope that there is something in this volume that might intrigue and delight scholars of the past.

Juliette Blevins
Maplewood, New Jersey
March 16, 2018

Acknowledgments

Writing this book has been a challenging and fulfilling endeavor, and many people deserve thanks for their teaching, inspiration, guidance, criticism, patience, and general support.

First and foremost, I thank Ander Egurtzegi for introducing me to the field of Basque historical linguistics and for providing me with the preliminary tools, readings, and cultural experiences that allowed this book to germinate and grow. Ander's role in this work is briefly sketched in the foreword. What is not mentioned there is his vast knowledge of Basque historical linguistics, and his firm and insightful criticism over the past five years, that have played a central role in informing and limiting the hypotheses presented here. Eskerrik asko, Ander, hain ikasle, lankide eta lagun ona izateagati, eta eskerrak zuri, Beari, eta zure guraso maitiei, Nekane eta Martxel, hain harrera eskuzabalagatik, zuengandik ikasitako herri jakituria (eguzkiloretik kardora), Euskal Herriko hainbeste txoko bereziak, bertan izandako elkarrizketa biziak eta hainbat aldi alaiengatik. Mila esker ere lan hau arreta handiz irakurtzeagatik eta hainbeste hobekuntza eskaintzeagatik.

Equal thanks go to Jabier Elorrieta, a warm, inquisitive, and cheerful presence in my life for 30 years. Eskerrik asko zure zintzotasunagatik, leialtasunagatik eta argitasunagatik. Eskerrik asko /s/ vs. /z/-ri buruz emandako eskolengatik, zure pazientziagatik, entzuteagatik eta irakurtzeagatik, baita zalantzan egiteagatik eta arduratzeagatik ere. Eskerrik asko *m-n sinesteagatik eta *zapi*-ri buruz galdetzeagatik. Eskerrik asko kritika zorrotzak jasotzerakoan emandako animoengatik, Bilbon piperrak jaten gonbidatzeagatik, eta nesken eta lauron lagun bikaina izateagatik. Ezin izango nukeen hau zure etengabeko babesik eta gomendio zuhurrik gabe egin.

I am also extremely grateful to other Basque scholars who have read and commented on earlier versions of this work, including José Ignacio Hualde, Iván Igartua, Joseba Lakarra, and Joaquín Gorrochategui.

I am greatly appreciative of other linguists who took the time to read and comment on previous drafts, despite their rough form, or who were kind enough to answer specific queries in their areas of specialization, or who, knowingly or unknowingly, said the right thing at the right time, including James Clackson, Lyle Campbell, Bernard Comrie, Andrew Garrett, Martin Kümmel, Johanna Nichols, Bill Poser, Ed Vajda, Joe Salmons, and Richard Sproat. Richard, thank you, especially, for not only taking an interest but also being the most patient co-author I have ever worked with. Correspondence with Leah Saddy regarding archeological evidence opened my eyes to the potential relevance of the Ozieri culture, while historian David Peterson

was kind to offer detailed answers to philological questions concerning Basque material in the Becerro Galicano. Earlier versions of parts of this volume were also read by several anonymous reviewers. I thank all of you for your time, consideration, and very helpful criticism, and I am sorry not to be able to mention you by name.

I also am grateful to Ben Fortson for writing *Indo-European Language and Culture: An Introduction*, thereby making so much of Indo-European scholarship accessible to a wider audience and for his friendly replies to my unsolicited e-mails. And for long-term support and challenges concerning evidence for long-distance relationships, I remain appreciative of my long-time colleagues in Austronesian linguistics, Bob Blust and Andy Pawley.

I thank the organizers of the Twenty-First International Conference on Historical Linguistics in Oslo in August 2013 for inviting me to speak there, and for audiences there, and at UPV/EHU, Harvard University, CUNY, the University of Cambridge, UT Austin and Penn, where parts of this work have been presented over the past four years. Some of you who attended these talks have written to me asking for drafts of the book, or detailed comparison sets, as the work evolved. Please forgive me for not sending anything sooner, and thank you, again, for your patience. I greatly appreciate your continued interest and hope you will find this book worth waiting for.

The Graduate Center at the City University of New York has been my academic home while writing this book. I offer sincere thanks to former president Bill Kelly, President Chase Robinson, former provost Louise Lennihan, and Provost Joy Connolly for generously supporting the Endangered Language Initiative at the Graduate Center and for their general interest in promoting scholarship on minority languages in New York City and around the world. I also thank Gita Martohardjono, executive officer of the linguistics program, and Nishi Bissoondial, our program administrator, who, together, have made everything easier—from organizing details of Ander Egurtzegi's original visit, to arrangements for students during my sabbatical last year, which allowed me to make further progress on this manuscript. Merci aussi to Alexandre François and others at LaCiTO/CNRS for having me as an EFL International chair in Paris in Autumn 2016, where I returned to foundational issues of sound change typology touched upon here in Chapter 9. And rounds of applause to all of my wonderful students, especially those who took historical linguistics and Indo-European historical linguistics with me over the past four years and put up with forays into Proto-Basque phonology. Special thanks to Michela Cresci for teaching me about Camuno and showing me the mysterious petroglyphs of Val Camonica; to Ignacio Montoya for quiet, solid support; and to Daniel Barry for assistance with Iranian lexical materials. Two other colleagues at the Graduate Center have offered constant inspiration and good cheer: thanks to Andy Beveridge for saying, "I just *knew* you would do something like this!" and to Daniel Kaufman, at guerilla headquarters—*Oi nagusi, mila esker*!

With support from the Endangered Language Initiative of the Graduate Center at CUNY, I was able to spend four weeks in the Summer of 2016

Guide to the Reader

A. Abbreviations

i Languages

aB	Archaic Basque
Alb.	Albanian
An.	Anatolian
Arm.	Armenian
Aq	Aquitanian
Av.	Avestan
B	Basque, modern Basque
Cast.	Castillian
Cat.	Catalan
Class	Classical
Celt.	Celtic
cB	Common Basque
Eng.	English
Eu	Euskarian
Fr.	French
Gasc.	Gascon
Gaul.	Gaulish
Gk.	Greek
Gmc.	Germanic
Goth.	Gothic
Hit.	Hittite
I-I	Indo-Iranian
Ital.	Italic
It.	Italian
Kurd.	Kurdish
Lat.	Latin
Latv.	Latvian
Lith.	Lithuanian
LL	Late Latin
Luv.	Luvian
mB	Medieval Basque
ME	Middle English
Occ.	Occitan
OCS	Old Church Slavonic
OE	Old English
OFr.	Old French
OIr.	Old Irish
OHG	Old High German
ON	Old Norse
OPr.	Old Prussian
pAn.	Proto-Anatolian
PB	Proto-Basque
pCelt.	Proto-Celtic
Pers.	Persian
PIE	Proto-Indo-European
pI-I	Proto-Indo-Iranian
pInd.	Proto-Indic
Pol.	Polish
pSl.	Proto-Slavic
Rom.	Romance
Russ.	Russian
S-Cr.	Serbo-Croatian
Skt.	Sanskrit
Slav.	Slavic
Sp.	Spanish
Toch.	Tocharian
UB	Unified Basque (Euskara batua)
Ved.	Vedic

Basque dialects

A	Alavese
AE	Aezkoan

V	Bizkaian
G	Gipuzkoan
HN	High Navarrese
L	Lapurdian
LN	Low Navarrese
R	Roncalese
Z	Zuberoan
S	Salazarese

ii Sources, scholars

AK	Kloekhorst (2007)
Az1, Az2	Azkue (1905–06), vol. 1 and vol. 2
BF	Fortson (2010)
BG	Becerro Galicano of San Millán
CW	Watkins (2011)
EDB	Trask (2008)
FHV	Michelena (1977 [2011a])
JG	Gorrochategui (1984)
HB	Trask (1997)
IELex	Indo-European Lexicon
LIPP	Dunkel (2014), Band II
LIPP1	Dunkel (2014), Band I
LIV	Rix et al. (2001)
LIV3	Rix et al. (2014)
M	Michelena (collected works)
MW	Weiss (2009)
NIL	Wodtko et al. (2008)
OEH	Orotariko Euskal Hiztegia
Pk	Pokorny (1959)
RLT	Turner (1966–69)

iii Grammatical terms

ABL	ablative
ABS	absolutive
ADJ	adjective-forming
adj.	adjective
ADV	adverb-forming
adv.	adverb
AFF	affirmative
AGT	agentive
ALL	allative
ALT	alternative
ATTR	attributive
AUG	augmentative
AUX	auxiliary
BAB	nursery word
BFD	backformed
CAUS	causative
CHAR	characteristic
COMP	comparative
comp.	comparanda
COND	conditional
CONJ	conjunction
CM	combining form
DAT	dative
def.	definite
DER	derivational
DET	determiner
DIM	diminutive
DRG	derogatory
EMPH	emphatic
ERG	ergative
EXC	excessive
EXP	expressive palatalization
fam.	familiar
GEN	genitive
gr.	grade, wrt PIE ablaut
INFL	inflectional
INT	interrogative
IMP	imperative
INSTR	instrumental
INTJ	interjection
IRR	irrealis
KIN	kinship
lit.	literally
LOC	locative
n.	noun
NEG	negative
NF	non-finite verb marker
NMZ	nominalizer
NUM	numeral
OBL	oblique case
ONM	onomastic
PERF	perfective
PRT	partitive

PL, pl.	plural		sth.	something
PLUR	pluractional		so.	someone
PRES	present		SOC	sociative
PRO	pronoun		SUP	superlative
PTCL	particle		TOP	toponym
PTCP	participle		var.	variant
Q	question marker		vb.	verb
RECP	reciprocal		VBZ	verbalizer
RED	reduplicated		VOC	vocative
SBJ	subjunctive		1	first person
SG, sg.	singular		2	second person
SPAC	spacial		3	third person
SS	sound symbolic			

B. Phonological categories and symbols

C	consonant
V	vowel
´	stress or accent
N	a nasal consonant
R	a sonorant consonant
S	sibilant
H	{ph, th, kh, h} for Proto-Basque/Basque
H	*h_1, *h_2, *h_3 for Proto-Indo-European
(h)	a root-final *h in PB, continued only before vowels
G	velar stop
Gʲ	palatalized velar stop
Gʷ	labialized velar stop
T	voiceless stop
T–	In combining forms that occur in Euskarian compounds, this symbol indicates the devoicing or /t/-insertion effect of a histori-cal oral stop in the stem. For example, /arT+bin/ > *arpin* and /arT+ile/ > *artile*, where /arT/ is the combining form (CM) of *ardi* 'sheep' and T is a reflex of *d
D	voiced stop
*X	a PB consonant of unknown quality
x–y	two elements separated by morpheme boundary
x+y	two elements separated by compound boundary
x=y	two elements separated by clitic boundary
x.y	two elments separated by syllable boundary
ab/*cd*	only *ab* or *cd* is being compared to something else
ø	zero, as in zero-grade

Other

>	changes to, by regular sound change
<	has come from, by regular sound change
>>	changes to, but *not* regular sound change (used for loan phonology, semantic shift, and analogy, clarified by context)
<<	comes from, but *not* regular sound change (used for loan phonology, semantic shift, and analogy, clarified by context)
*	reconstructed proto-form
**	Proto-Basque reconstruction has phonological, morphological, or semantic features motivated by PIE comparison.
**	Proto-Indo-European construction is non-standard, as explained
!	Expected but unattested (synchronic or diachonic)
?	questionable (when preceding PB reconstruction, indicates possible loan; when preceding PIE reconstruction, taken from relevant source)
c.	century
IPA	International Phonetic Alphabet

C. Spelling and transliteration conventions

There are two kinds of words cited in this book: reconstructions, preceded by *, and attested words taken from a range of sources. Reconstructions are copied from the original source, unless they are my own for Proto-Basque, in which case, they follow the arguments in Part I of this book. This means that Proto-Indo-European constructions will not all use the same symbols since they are taken from different authors. See the notes at the beginning of the appendix for more details. Basque words cited are also copied from the original source. Forms written here may differ from OEH head entries in reflecting conservative dialect forms. When this is the case, OEH head entry forms are given in parentheses, *except* when the only difference is presence or absence of one or more <h>s.

Part I
Reconstruction of Proto-Basque

Introduction to Part I

The central argument of this book is that Proto-Basque, the ancestral language of modern Basque (*Euskara*), is related to Proto-Indo-European, the reconstructed ancestor of Anatolian, Indo-Iranian, Tocharian, and most European languages. The book has two parts that represent two steps of an argument for a long-distance relationship between Proto-Basque and Proto-Indo-European. Part I argues that Basque evolved from a reconstructed language, Proto-Basque, that has sounds and sound patterns that are somewhat different from modern Basque, and that are different from other linguists' conceptions of Proto-Basque. Part II argues that this new reconstruction of Proto-Basque yields roots and grammatical morphemes that show regular sound correspondences with Proto-Indo-European reconstructions across a wide range of semantic fields, including basic vocabulary. These regular sound correspondences strongly suggest a genetic relationship between Proto-Basque and Proto-Indo-European, as they are highly unlikely to have arisen by chance. At the same time, the nature of the sound correspondences leads one to conclude that Proto-Basque and Proto-Indo-European were not in a mother-daughter relation. Rather, these two ancient tongues were distinct languages themselves descending from a common ancestor. This is the Proto-Indo-European-Euskarian Hypothesis.

Since Basque is widely believed to be an isolate, unrelated to any other known living language, the central proposal of this book is radical. And since earlier proposals for long-distance relationships between Basque and other language families, including Indo-European, have been strongly criticized by expert Bascologists (e.g. Lakarra 1996, 1999, 2010; Trask 1997:358–429; Gorrochategui and Lakarra 2013), the basic premises of this study must differ in some fundamental way from those of previous scholars. This is indeed the case. The basic premise that is unique to this volume is a new reconstruction of Proto-Basque. Part I of this book spells out, in detail, the arguments for this new reconstruction.

The starting point for any hypotheses about earlier stages of the Basque language is a close analysis of the Basque language, its dialects, and its form at different historical periods. This is the focus of Chapter 1. In order to give readers with little knowledge of the Basque language the background

necessary to appreciate this, and the arguments about earlier stages of the language that follow, I offer a brief overview of Basque here, with a focus on phonology and morphology, followed by an overview of Part I to guide them through the steps in the argument. For a detailed reference grammar of Basque, see Hualde and Ortiz de Urbina (2003). For those looking for a longer introduction to the general structure of the modern language, see Chapter 2 of Trask (1997). A detailed phonological description of Basque can be found in Hualde (1991a), while Michelena's *Fonética histórica vasca* (1977) remains the most comprehensive and influential work on Basque historical phonology. For a brief up-to-date survey of Basque historical phonology, see Egurtzegi (2013a).

The Basque Language

The Basque language, called *Euskara* in Basque, is spoken by about 900,000 people in the Basque Country—a 100-mile-long strip at the western edge of the Pyrenees. The Basque Country is divided by the modern border of France and Spain, with most speakers of Euskara on the Spanish side and less than 100,000 on the French side. While Basque is not considered endangered, the language is no longer spoken in southern parts of the Basque Country, where the Alavese dialect is now extinct. The language has also greatly receded in the easternmost regions, where Roncalese and Salazarese are no longer spoken. The first written records of Basque are in Latin manuscripts of the 10th century from Navarre. These records include many placenames, personal names, nicknames, and even some short phrases (Michelena 1977[2011]:27ff).

  One of the best-known linguistic features of Basque is that it is an "isolate". This means that there is no known genetic relationship between Basque and any other known language or language family. Technically, this is not exactly right. An ancient language called Aquitanian is known to have been spoken on both sides of the western Pyrenees around 300 BCE, and there is good linguistic, archaeological, and historical evidence that Aquitanian was related to Basque. The linguistic evidence is of two kinds: placenames in Aquitaine (later known as Gascony) and Aquitanian names of people and gods that occur in Latin funerary and votive texts from about 0–300 CE (Michelena 1977[2011]:20ff; Gorrochategui 1984, 1995; Campbell 2012). Some common Basque words with identifiable cognates in Aquitanian proper names are B *bihotz* 'heart' as in Aq BIHOSSI and B *neska, neskato* 'girl' as in Aq NESCATO. Given this kind of evidence, there is general agreement among Bascologists that Aquitanian was either an ancestral form of Basque or a language closely related to an ancestor of Basque. The language family that includes Basque and Aquitanian is referred to as Euskarian (Martinet 1955; Gorrochategui 1995). However, because many linguists believe that Aquitanian is simply an older form of Basque, the reconstructed proto-language based on data

from modern Basque, historic Basque, and Aquitanian, is called Proto-Basque. I follow earlier scholars in their use of both of these terms.

General Typological Profile

What kind of language is Basque? One of the most striking aspects of the Basque language as a whole is its regularity. Sound patterns and word-structure are relatively transparent, with very little in the way of irregular or exceptional forms. This regularity extends from word order to expression of grammatical relations to morphological paradigms and sound patterns. The one highly irregular verb is *izan* 'be', and there are no irregular nouns.

Modern Basque tends to have Subject-Object-Verb word order, but the order of words is determined, to a great extent, by information structure. The focus, or most important information in the sentence, precedes the verb (or verb phrase); while the topic, or general theme of the discourse, is in sentence-initial position. Thematic roles in Basque are largely determined by case-marking. Case-marking and verbal agreement are mostly ergative-absolutive, with only small corners of verbal morphology (e.g. some past-tense verbs) showing nominative-accusative patterns. The ergative case pattern shows ergative case-marking on the subjects of transitives, while subjects of intransitives and objects of transitives have absolutive case-marking. Nominal case-markers are added to the last element of the noun phrase and follow articles (-*a* singular, -*ak* plural, which encode number and definiteness). The three nuclear cases, determining verbal agreement, are absolutive (zero), ergative (-*k*), and dative (-*i*). Additional local cases are inessive (-*n*), allative (-*ra*), ablative (-*tik*), and local genitive (-*ko*), while non-local cases include the possessive genitive (-*en*), instrumental (-*z*), comitative (-*ekin*), benefactive (-*entzat*), and causal (-*engatik*). Allomorphy of these case suffixes is regular and determined by the form of the base. This regularity and transparency is typical of the language as a whole: with the exception of a handful of inflecting verbs, morphology is highly agglutinative, with easily identifiable morpheme boundaries. The majority of verbs are nonfinite and co-occur with auxiliaries, which encode tense-aspect-mood and agreement (with object and indirect object). Within the verb phrase, auxiliaries are final, except when fronted under negation occurring after the particle *ez* 'no, not'. A small number of verbs, including the auxiliaries, are finite, and some, such as the verb *izan* 'be', show irregular (suppletive) forms. The language is almost wholly suffixing, with the exception of fossilized prefixes (e.g. *e-* a verbalizer, -*ra-* a causative) and neologisms, and compounding is a productive word-formation process. While gender is not a grammatical feature generally encoded in the inflectional system, Basque has the rare feature of allocutive agreement, where the gender of the addressee is marked when using the second-person familiar /hi/ pronoun. For example, if I am speaking to a close friend, I would say *Juliet naun*, "I am Juliette", where the final /-n/ shows I am speaking informally to a female,

or *Juliet nauk*, "I am Juliette", where the final /-k/ marks that I am speaking informally to a male.

Word-structure in Basque is relatively simple with productive compounding and suffixation, and clear morpheme boundaries allowing for root and stem identification. Most Basque roots are monosyllabic (*hats* 'breath', *lan* 'work', *on* 'good', *su* 'fire') or disyllabic (*aho* 'mouth', *argi* 'light', *begi* 'eye', *garbi* 'clean', *labur* 'short', *mihi* 'tongue'). Most modern Basque verbs do not occur as bare roots, but the root or stem can be discovered by subtracting the initial /e-/, /i-/, a historic verbalizer (e.g. -*gin* of *egin* 'do', -*dan* of *edan* 'drink'), and, where relevant, a final participle *-i (e.g. -*thor-* of *ethorri* 'come', -*sur-* of *isuri* 'pour'). Seeming roots of three or more syllables are usually analyzable as having fossilized suffixes. For example, *alaba* 'daughter', *osaba* 'uncle', *iloba* 'nephew' all appear to contain a kinship formative /-ba/, which is no longer productive. A wide range of productive derivational suffixes gives rise to nominal, verbal, adjectival, and adverbial stems/words: for example, from *on* 'good' (adjective); we have *ongarri* 'manure, fertilizer' (noun); *ontza* 'friendship' (noun); *ondu* 'improve, get better' (verb); *onen* 'best' (adjective); *ondo* 'well' (adverb); to *ongi* 'well' (adverb). Compounding is also highly productive, and compounds may be head-initial (*buru* 'head' + *beltz* 'black' = *burubeltz* 'black-headed'), head-final (*iturri* 'spring' + *buru* 'head' = *iturburu* 'fountainhead'), or exocentric (*buru* 'head' + *goi* 'high' = *burugoi* [R] 'haughty, arrogant'). At the same time, it is possible to find compounds that are frozen, reflecting earlier conservative phonological features in the form of root allomorphy. For example, Roncalese *burugoi* 'haughty' appears to be reformed from older *burgoi*, where *bur-*, the combining form of *buru* 'head' is in evidence. Old compounds are very important in the discovery of ancient Basque sound patterns since there is extensive evidence that word-initial lenition of many consonants has occurred. When a root was embedded as the second element of an old compound, its initial consonant was protected from lenition and can be recovered. For example, compare *hume (ume)* 'child; young, of animal' as a free form, with -*khume* 'offspring, child', second member of *arkume* 'lamb' (cf. *ardi* 'sheep') and *orakume* 'puppy' (cf. *hor*, *hora* (*or*) 'dog'). Here root-initial *kh of *khume is preserved in compounds, though, as argued in Chapter 3, together with *th and *ph, it was weakened to /h/ word-initially. Another arguably archaic feature visible in *orakume* 'puppy' is the final /a/ of *hora. In Chapter 3, I suggest that modern *ora-* 'dog' reflects *Hora based on the vowel in *orakume* 'puppy'. In this case, the stem-final vowel was reanalyzed as an instance of the B determiner /=a/. In the modern language, the definite determiners /=a/ (singular) and /=ak/ (plural) are suffixed to the last element of the noun phrase, and could be considered phrasal affixes. The phonological alternations that occur under suffixation of these elements vary across dialects and are known to many linguists from problem sets in phonology textbooks, as, for example, the set in Kenstowicz (1994:22) that includes *buru* 'head', *burue* (Baztan, and many other varieties), *buruwe* (Bizkaian) 'the head', where final vowel raising of /-a/ is triggered by the preceding high

vowel, and glide insertion occurs in Bizkaian (which may also show glide strengthening as in *buruba, burube*). While the occasional inflectional suffix has been reconstructed in Part I, the central focus of reconstruction are lexical roots, particles, and derivational suffixes, many of which are grammaticized instances of earlier roots and particles.

Of all of the typological features just noted, the ones that have played the biggest role in attempts to relate Basque to other languages are its ergative case-marking and direct and indirect object agreement patterns. These patterns are not found in any other European languages, but some are found in some Caucasian languages and, within Indo-European, in the Anatolian languages. However, typological comparisons of this kind are not valid methods for assessment of relations of descent in historical linguistics, where genetic relationships between languages are validated by the comparative method, including detailed demonstration of sound-meaning correspondences across basic lexical items and grammatical formatives. Since sound-meaning correspondences can only be assessed after a comprehensive and explanatory account of Proto-Basque phonology has been put forward, a proposal for a long-distance relationship must be grounded in a firm reconstruction of Proto-Basque sound patterns. This is the goal of Part I of this volume, with several general features of Basque phonology presented here.

Orthography

Basque words are written using the standard Basque orthography, a writing system that is phonemic in nature, representing all and only the phonological contrasts made in the standard language (though some of these are neutralized in various dialects), with only a few exceptions. Table 1.3 in Chapter 1 shows all orthographic symbols and their equivalents in the International Phonetic Alphabet. Cases where the standard orthographic symbol does *not* have the standard IPA value are listed below.

Basque orthography	Approximate IPA value	
<r>	/ɾ/	alveolar rhotic tap
<rr>	/r/	alveolar rhotic trill
<s>	/ş/	voiceless apical sibilant
<z>	/s̺/	voiceless laminal sibilant
<x>	/ʃ/	voiceless alveo-palatal fricative
<ts>	/tş/	voiceless apical affricate
<tz>	/ts̺/	voiceless laminal affricate
<tx>	/tʃ/	voiceless alveo-palatal fricative
<ll>, <il>	/ʎ/	voiced palatal lateral
<ñ>, <in>	/ɲ/	palatal nasal stop
<tt>	/c/	voiceless palatal stop
<dd>	/ɟ/	voiced palatal stop
<ph>, <th>, <kh>	/pʰ/, /tʰ/, /kʰ/	voiceless aspirated stops

Basque Orthographic Symbols with Non-IPA Values

As in Spanish, the Basque tap vs. trill distinction is represented by a single rhotic <r> vs. a double rhotic <rr>. And, like Spanish, in word-final position, this contrast is neutralized, where one finds only <r>. However, in Basque when a vowel-initial suffix, such as the definite marker /=a/ is added to an <r>-final word, the typical result is <rr> as in *adar* 'branch; horn', *adarra* 'the branch' or *sagar* 'apple', *sagarra* 'the apple'. A small number of exceptions to this pattern are discussed in Chapter 5, including *ur, hur* 'water' with definite *ura, hura* 'the water'.

Another important aspect of the orthography that readers unfamiliar with Basque should remember is that the <s> vs. <z> and <ts> vs. <tz> contrasts do not involve voicing. In all cases, <s> writes the apical voiceless sibilant while <z> writes the laminal voiceless sibilant: both are phonologically voiceless. Likewise, <ts> writes the voiceless apical affricate, while <tz> writes the voiceless laminal affricate. While voicing (or aspiration) is contrastive for oral stops in Basque, no other sounds in the inventory contrast for voicing or aspiration.

Finally, in citing Basque forms, I write <h> wherever aspirating dialects (or historic forms) show <h> or aspiration. This means that in some cases, the OEH dictionary form will not show an orthographic <h>, and in these cases, I write the OEH orthographic form after the *h*-ful form in parentheses in the text (but not in the appendix, for reasons of space). For example, I write *hon* (*on*) 'good' based on Old Lower Navarrese <hon> and Zuberoan *hun*, but the standard orthographic form of this word is *on*, and I write *aphal* (*apal*) 'low' based on the Lower Navarrese and Zuberoan pronunciation, though the OEH main entry is *apal*. The distribution of /h/ in aspirating varieties like Zuberoan is extremely important in this work. Initial /h/ contrasts with zero in these dialects as in *hagin* 'molar tooth' vs. *agindu* 'order, mandate'. As detailed in Chapter 3, this initial /h/ is argued to continue Proto-Basque *h or is a debuccalized continuation of *pʰ, *tʰ, or *kʰ. In Part II, Proto-Basque *h plays an important role in external comparison where it is argued to correspond with the Proto-Indo-European laryngeals.

Phonology

Most varieties of Basque have the five basic vowels /i u e o a/ and the falling diphthongs /ai, ei, oi, au, eu/. Contrastively, nasalized vowels as well as the front rounded vowel /y/ are found in northern dialects and argued to be due, in part, to contact with neighboring Romance languages (Egurtzegi 2017, to appear a). There are many vowel and /h/-initial words in Basque, due to historical weakening of initial consonants, detailed in Chapter 3. The basic consonant system includes voiceless stops /p t tt k/ (aspirated in some dialects); voiced stops /b d dd g/; voiceless affricates /ts, tx, tz/; voiceless fricatives /f s z x/; nasals /m n ñ/; laterals /l il/, rhotics /r rr/; a palatal glide /j/, which varies greatly in pronunciation across dialects; and an aspirate /h/, which is not pronounced in some dialects. While there is little disagreement that Proto-Basque should be reconstructed with the vowels *i, *u, *e, *o,

*a (see Chapter 2), the consonant system and phonotactics are less widely agreed upon. Where modern Basque has 24 contrasting consonants /p, t, k, b, d, g, m, n, l, r, rr, s, ts, z, tz, f, h, tt, dd, ñ, il, x, tx, j/, Michelena's (1977) reconstruction of Proto-Basque has only 16 consonants *t, *k, *b, *d, *g, *h, *ts/*s, *tz/*z, *N/*n, *L/*l, *R/*r, where the five final pairs constitute fortis/lenis contrasts, first proposed by Martinet (1955). Points of agreement between most Bascologists regarding the consonant system are (i) that the palatal series /tt,dd,ñ,il,x,tx/ is relatively new, deriving, in part, from expressive palatalization (there is also allophonic palatalization adjacent to /i/); (ii) native /j/ derives from initial *e(d), *i(d) (but it is also found in loans, e.g. *jaboi* 'soap' from a Romance reflex of Lat. *sāpōnem*); (iii) Basque initial /f/ is a good indicator of loanword status (as in *fede* 'faith' << Lat. *fedem*), since the only internal source of /f/ is an earlier /uh/ sequence (e.g. *afari* < *auhari* 'dinner'; Hualde 1997b:422f); and (iv) Basque initial /d/, /t/ and /k/ in uninflected content words (e.g. nouns, verbs, and adjectives) is a good indicator of loanword status since there is some evidence that these sounds were weakened or lost initially in the history of the language. Though a wide range of syllable types are found in Basque due to lexical borrowings (see below), native Basque words show very limited syllable types. Excluding the inflectional suffixes ending in /-k/, and a handful of words ending in /t/ (*bat* 'one', *bost* 'five', *bart* 'last night') the native vocabulary is primarily limited to syllables of the shape (C)V(V)(R)(S), where C is any consonant, V is a vowel, R is a sonorant consonant {r,l,n}, and S is a coronal fricative or affricate. Many words begin with vowels, and many words begin with /h/ (in *h*-ful varieties). Though consonant clusters are absent in initial position in native words due to this syllable template, they are common word-internally: *alde* 'side', *ardi* 'sheep', *argi* 'light', *asko* 'many', *astun* 'heavy', *esku* 'hand', *garbi* 'clean', *handi* 'big', *urte* 'year'. All varieties of Basque have words where one syllable is pronounced with more prominence than another, however, patterns of prominence vary greatly across dialects. In some dialects, particularly in the Western Basque Country, these systems are best described in terms of pitch accent, while in other dialects, the system is best described in terms of word stress. Chapter 5 presents an overview of these systems and a slightly revised reconstruction of the Proto-Basque accentual system.

Loanwords and Loanword Phonology

One of the most difficult problems facing the historical linguist attempting an internal reconstruction of the Basque language is how to distinguish native (inherited) vocabulary from loanwords. As discussed in Chapter 1, Euskarian has had a long history of contact with Indo-European languages, from Celtic tribes occupying the Iberian Peninsula as early as 600 BCE, to the Roman conquest, and later invasions by Germanic tribes. The general strategy taken here is a conservative one: if a Basque word *could* be a loan, it is generally disregarded as potential material for internal reconstruction.

If others have treated a particular word as a loan, and I do not, justification is provided in the notes section of the etymology in the appendix. In general, early loans reflect native Basque sound patterns. Since the great majority of identifiable early loans are from Latin (or later Romance languages), the clearest patterns are seen there. These can be summarized as follows: (i) syllable-initial CR clusters are broken up by an epenthetic vowel (B *garau* 'grain' << Lat. *grānum*; B *liburu* 'book' << Lat. *librum*); (ii) word-initial sC clusters are preceded by a prothetic vowel (B *eskola* 'school' << Lat. *schŏla*; B *estrata* 'road' << Lat. *strata*); (iii) a word-initial rhotic is preceded by a prothetic vowel (B *errege* 'king' << Lat. *rēgem*; B *erripa* 'slope, sloping' << Lat. *rīpam*); (iv) word-initial /b/, /d/, /g/, /l/, /m/, and /n/ are borrowed without change (B *balea* 'whale' << Lat. *ballaena*; B *dolu* 'pain, repentance' << Lat. *dŏlum*; B *garau* 'grain' << Lat. *grānum*; B *liburu* 'book' << Lat. *librum*; B *merkatu* 'market' << Lat. *mercātum*; B *nothatu* 'to mark, stain' << Lat. *notāre*); (v), in contrast, word-initial /p/, /t/, /k/ and /f/ are often borrowed as voiced stops (B *bake* 'peace' << Lat. *pācem*; B *denbora* 'time' << Lat. *tempŏra*; B *gorputz* 'body' << Lat. *corpus*; B *boronde* 'front' << Lat. *frontem*). Though it may appear that some initial clusters in Romance loans undergo consonant loss (e.g. B *lore* 'flower', cf. Lat. *flōrem*), a recent study argues that Basque words with this pattern are borrowed from Romance languages that have simplified these clusters under palatalization (Blevins and Egurtzegi 2017). A further finding of this same study is that variation in loanword phonology can be associated with the degree of exposure and familiarity one has with the contact language. That being said, the clear patterns outlined above all reflect native Basque phonotactics. Since syllable-initial clusters are illicit, they are resolved by epenthetic vowels as in patterns (i) and (ii). And since words cannot begin with rhotics, epenthetic vowels are inserted before a rhotic as in pattern (iii). The contrast between patterns (iv) and (v) is, perhaps, the most interesting. In Chapter 3, the tendency to borrow initial /p/, /t/, /k/ as voiced stops is associated with the historical loss of word-initial voiceless stops in the language. To date, no careful linguistic stratigraphy has been carried out on Basque in relation to its contact with Celtic, Romance, and Germanic, for the simple reason that loans from Celtic and Germanic are not easy to identify, and not widely agreed upon. However, some of the Proto-Basque derivational affixes reconstructed here allow for a preliminary cut. In particular, the nominal prefix *ha-, the collective *hi- and the nominalizing suffix *-s occur only with native roots, suggesting that their productivity pre-dates contact with Celtic, Italic, and Germanic.

Overview of What Follows

Part I of this book presents a new reconstruction of Proto-Basque, the hypothesized ancestor of Basque and Aquitanian, while Part II compares this new reconstruction of Proto-Basque with Proto-Indo-European, using the

comparative method. Chapter 1 provides background to the arguments which follow, including historical sources for Basque and Aquitanian, an introduction to Basque dialects, a brief history of language contact, and the standard orthography, as it is used throughout this work. Chapter 1 also introduces the sound system of Proto-Basque as conceived by earlier scholars, and the changes I suggest in the following chapters. Chapters 2–5 detail arguments for the revised Proto-Basque phonology as summarized in Chapter 6. Chapter 2 follows earlier work on the Proto-Basque vowel system, with some critique of the evidence for diphthongs. Chapter 3 presents evidence for the revised consonant system, proposing *m, a series of aspirated stops including *ph, and introduces the single-sibilant hypothesis, suggesting that Proto-Basque had only *s, with modern /z/ derived from *sC and *Cs clusters. Chapter 4 argues for revised word phonotactics and syllable structure, presenting a range of heuristics for root identification, and couples this revised morphological analysis with internal evidence for *sC and *Cs clusters. Chapter 5 reviews work on Basque pitch accent and stress, proposing a mixed system for Proto-Basque where lexical accent and quantity-sensitive stress assignment both play a role. A summary of the primary findings is presented in Chapter 6 and compared with earlier proposals.

The arguments presented in Chapters 2–5 build on earlier proposals. However, the reconstructions make use of two new kinds of evidence: language-internal alternates and doublets that had not been previously appreciated in the context of internal reconstruction; and language-internal asymmetries in phoneme distribution that had not been fully appreciated. For example, the extension of Martinet (1950) and Michelena's (1977) mention of possible initial debuccalization to a regular sound change *ph, *th, *kh > h is based on a thorough investigation of potential *ph/h*, *th/h* and *kh/h* alternates in the language, conditioned by medial vs. initial position of the stem, respectively. In contrast, the novel suggestion that initial *z* derives historically from *sT where T is *ph, *th, or *kh is supported by newly discovered doublets such as *astun* 'heavy', *azun* 'loaded, pregnant', and by distributional asymmetries of /s/ and /z/ in the modern language. Native speakers of Basque may notice that in some cases, Basque words cited here, such as *astun* above are very common, while other words, such as *azun*, are rare in common usage and may not be known to many native speakers. Rare words are used as evidence for Proto-Basque reconstructions, and are often words relating to traditional farm-life (agriculture, animal husbandry, dairying, etc.) and indigenous flora, fauna, and geography. A long history of technological innovation, including the industrial revolution, and heavy urbanization of parts of the Basque Country have resulted in many of these words falling out of use.

When examining the details of the arguments for a revised view of Proto-Basque sound patterns and morpheme structure in Chapters 2–5, the reader should keep in mind two important points. First, the techniques used to arrive at this reconstruction are almost as old as the field of modern historical

linguistics itself. These techniques include the comparative method and internal reconstruction. The comparative method, as its name suggests, is the comparison of distinct languages or dialects that show cognate forms. Once cognates are found, one attempts to find regular sound correspondences, and from there, to reconstruct pre-existing forms on the basis of proposed regular sound changes. For Basque, one can apply the comparative method to compare distinct dialects of the language, and to compare Basque to Aquitanian. The method of internal reconstruction, in contrast, involves the analysis of patterns within a single language (or single dialect of a language), to hypothesize earlier states of the language, with a focus on morpheme alternates, morphophonological alternations, and skewed sound distributions. So, although some aspects of the sound system and certain sound changes I propose for Proto-Basque are novel, the methods I have used to arrive at these are standard in the field of historical linguistics. A second point to keep in mind is that most of the specific revisions I propose for Proto-Basque phonology are conservative and would likely be considered (or have been considered), at least in passing, by any historical linguist familiar with the Basque native vocabulary. For example, given arguably inherited Basque words such as *madari* 'pear', *malgu* 'soft', and *musker* 'lizard', it is not unreasonable to reconstruct *m continued as /m/ in these words. Yet the standard reconstruction of Proto-Basque has no *m.

In putting forward new hypotheses, such as the existence of *m in Proto-Basque, I present new evidence, review previous proposals, and present the strengths and weaknesses of my own position. In this way, the reader can evaluate each point independently. If the reader decides that certain Proto-Basque reconstructions are suspect, for example, all those with *m, the same reader can ignore or put less weight on external comparisons in Part II that involve *m/*m correspondences. All of the proposals in Part I are supported by the Proto-Basque etymologies offered in the appendix. Unless specified otherwise, the sound changes that relate the Proto-Basque form to attested Basque forms are regular, with the full set of sound changes adopted summarized in Table 6.2.

A short introduction to Part II is included for readers with limited background in historical Indo-European linguistics. It provides an overview of the comparative method, the Proto-Indo-European reconstructions used in this volume, and evidence for long-distance relationships of the kind proposed here for Proto-Basque and Proto-Indo-European.

1 Basque and Proto-Basque

The Basque language, *Euskara*, is spoken by approximately 900,000 people (Gobierno Vasco 2012), primarily in the Basque Country, *Euskadi*, and has at least six widely recognized dialects (Map 1.1) and as many as a dozen or more. Some of the earliest linguistic work on Basque dialects is Bonaparte (1863, 1869 [1991]).

Since the 19th century, Bonaparte's general view has been maintained with slight revisions (Michelena 1977 [2011]),[1] resulting in some consensus of classification of the language into ten different dialects (*Euskalkiak*). The ten dialects and their approximate historical ranges in the 19th century are illustrated in Map 1.1, taken from Egurtzegi (2014:7).[2] In Table 1.1, the ten

Map 1.1 19th-Century Basque Dialects (Euskalkiak) from Egurtzegi (2014:7)

Table 1.1 Major Basque Dialects

Egurtzegi and This Volume	*Abbr.*	*Michelena*	*Martínez-Areta*	*Trask*
Aezkoan	AE	Aezcoano	Aescoan	Aezkoan
Alavese[†]	A	Meridional	Alavese	Southern
Bizkaian	V	Vizcaíno	Biscayan	Bizkaian
Gipuzkoan	G	Guipuzcoano	Guipuscoan	Gipuzkoan
High Navarrese	HN	Alto navarro	High Navarrese	High Navarrese
Lapurdian	L	Labortano	Labourdin	Lapurdian
Low Navarrese	LN	Bajo navarro	Low Navarrese	Low Navarrese
Roncalese[†]	R	Roncalés	Roncalese	Roncalese
Salazarese[†]	S	Salacenco	Salazarese	Salazarese
Zuberoan	Z	Suletino	Souletin	Zuberoan

[†] Dialects no longer spoken

dialect names used by Michelena (1977), Trask (1997), and Martínez-Areta (2013b) are compared with those used in this volume, which follows the names and abbreviations used in Egurtzegi (2014). Note that three dialects, Alavese, Roncalese, and Salazarese, are no longer spoken.[3]

Modern Basque dialects descend from historic Basque as recorded in documents dating back to the Middle Ages. Prior to this time, at the beginning of the Middle Ages, the Basque koiné spoken is generally referred to as 'Common Basque' (Michelena 1977, 1981).[4] A closely related language or Basque ancestor is Aquitanian (Gorrochategui 1984, 1995) and the language family that includes Basque, Aquitanian, and all potentially related languages is referred to as Euskarian (Gorrochategui 1995). Aquitanian is the earliest attested Euskarian language, known primarily from personal names and names of deities recorded in Latin texts from the beginning of the Common Era to about 300 CE. In this work, Aquitanian data is taken from Gorrochategui (1984), unless noted otherwise. The Euskarian language family, sometimes referred to as 'Vasconic' or simply as 'Basque' or 'Euskara', includes all present and historic varieties of Basque, as well as Aquitanian and the term 'Euskarian' in the title of this book, is used in this way to refer to the language family as a whole.

Euskarian languages across time are presented in Table 1.2 prior to the diversification of dialects shown in Table 1.1. Table 1.2 also includes abbreviations for these languages, as well as primary attestations or data supporting these chronological states. Between Aquitanian and well-attested stages of the language beginning in the 17th century are Medieval Basque and Archaic Basque, with primary data discussed in detail in Michelena (1964, 1977), Gárate and Knörr (1982) and Irigoien (1997). The oldest hypothesized state of the language, taking into account Aquitanian data, is referred to as "Proto-Basque" (see Note 4). One central goal of this work is to reconstruct the sound system and root lexicon of Proto-Basque, the mother language of all Euskarian languages, as far back in time as evidence permits.

Table 1.2 Euskarian Languages Over Time

Time Period	Euskarian Language	Abbr.	Attestations/ Primary Data
Before 200 BCE	Proto-Basque	PB	—(reconstructed in this volume)
Before 200 BCE	Pre-Proto-Basque	PB	—(reconstructed, Lakarra 1995)
? 200 BCE	Proto-Basque	PB	—(reconstructed, Michelena 1977)
0–300 CE	Aquitanian	Aq	in Latin inscriptions
400–600 CE	Common Basque	cB	—(reconstructed, Michelena 1981)
900–1400	Medieval Basque	mB	Emilian glosses, 10th century Cartulary de San Millán, 1025 (placenames) Other cartularies, 11th–12th centuries Picaud, ~1140 (word-list) Le Censier Gothique de Soule, 14th century (housenames) Von Harff, 15th century (word-list)
15th–16th c.	Archaic Basque	aB	*Orotariko Euskal Hiztegia* citations, including: Dechepare 1545 (Low Navarrese); Landucci 1562; Lazarraga 1567–1605; de Betolaza 1596 (Alavese); Garibay, ca. 1592, Anon. Refranes y Sentencias 1596 (Bizkaian); Leiçarraga, New Testament, 1571 (Lapurdian)
17th–19th c.	Literary Basque	lB, B	*Orotariko Euskal Hiztegia* citations. See also Ulibarri (2013)
20th–21st c.	Modern Basque, Unified Basque (*Euskara Batua*)	B UB	Numerous, including *Hiztegi Batua* 2010, OEH

Historical and synchronic work on Basque sound patterns is rich and varied, and it will be impossible to review all of it here. General works on the synchronic and diachronic phonology of Basque include Uhlenbeck (1903), Gavel (1920), Hualde (1991a), Hualde et al. (1995), Trask (1997), Hualde and Ortiz de Urbina (2003), and Martínez-Areta (2013a). Most relevant to the present work and the foundation of modern Basque historical phonology is Luis Michelena's *Fonética Histórica Vasca* (= *FHV*, Michelena 1977), originally published in 1961. Michelena's monumental work attempts to reconstruct Proto-Basque as spoken sometime between 300 BCE and 600 CE. A thorough overview of the field of Basque historical linguistics with detailed summaries of earlier work and modern developments is Martínez-Areta (2013a). The volume includes an overview of the Basque

language today, as well as chapters on Basque dialects (Martínez-Areta 2013b), sources for historical research (Ulibarri 2013), and grammatical structure and development of Proto-Basque, including historical phonetics and phonology (Egurtzegi 2013a), root structure (Lakarra 2013), nominal morphology (Santazilia 2013), demonstratives and personal pronouns (Martínez-Areta 2013c), non-finite verbal morphology (Padilla-Moyano 2013), finite verbal morphology (Ariztimuño 2013), and word order (Reguero-Ugarte 2013). Egurtzegi (2014) provides a current phonetically grounded diachronic phonology of Basque from the Middle Ages onward. Part I of this book can be seen as an attempt to push back in time from where Michelena (1977) and Egurtzegi (2013a) left off.

Within the field of historical linguistics, Basque, and Euskarian more generally, is viewed as an isolate, meaning that there is no genetic relationship between Euskarian languages and any other known languages, living or extinct. As a consequence, the comparative method of historical linguistics, which involves comparing distinct languages to reconstruct proto-languages, contributes to the reconstruction of Proto-Basque only at the level of dialect comparison, or comparison of Basque with Aquitanian. Given the commonly accepted view of Euskarian as an isolate, the majority of historical work on Proto-Basque combines this comparative work with internal reconstruction along with loanword analysis and typological comparison to discover earlier stages of the language (Martínez-Areta 2013a). The method of internal reconstruction, like the comparative method, is based on the Neogrammarian hypothesis that sound change is regular. In this chapter, the most widely accepted proposals for Proto-Basque phonology that make use of dialect comparison and internal reconstruction are reviewed, with focus on Michelena's work as detailed in *FHV* and Lakarra's (1995, 2013) study of root structure.

Chapters 2–5 present evidence for a new reconstruction of Proto-Basque that builds on these earlier proposals. The general methods are essentially the same. However, the new reconstruction makes use of two new kinds of evidence: language-internal alternations that had not been previously appreciated in the context of internal reconstruction and language-internal asymmetries in phoneme distribution that had not been fully appreciated. The entire sound system and its development are informed by an articulated theory of phonological typology and phonetically based sound change, as detailed in Blevins (2004, 2006, 2008, 2009, 2015).

Overall, the new view attempts to motivate phonetic and phonological features of Proto-Basque that differ from earlier proposals. Some of these features make Proto-Basque appear more commonplace than was previously thought. The Proto-Basque I reconstruct has some unremarkable segments and no striking gaps: it has *m; *pʰ; a single sibilant *s; a single rhotic *r; and a contrast between plain voiced stops and voiceless aspirated stops. On the other hand, Proto-Basque as reconstructed in the chapters that follow has some features that are remarkable in the context of the modern

language, and earlier reconstructions of Proto-Basque. For example, all instances of /z/ in native vocabulary are derived, either from *sC or *Cs sequences, though Basque generally does not tolerate initial *sC clusters.

The new Proto-Basque presented in Part I of this volume was arrived at by the comparative method internal to Euskarian and internal reconstruction. This reconstruction has led to a potential revision of Basque's status as an isolate, explored in Part II. However, it should be kept in mind that most of the discussion in Part I is Euskarian-internal. Proto-Basque forms preceded by single asterisks are those that are arrived at by examining Basque dialects and historical Basque and Aquitanian data, and applying the comparative method and methods of internal reconstruction to these.[5]

The Basque language has had extensive contact with a range of Indo-European languages from the earliest historic periods.[6] The Celts, speaking Celtic languages, are thought to have passed through Basque-speaking areas between 600 BCE and 100 CE, with evidence of their occupation in the form of Bronze Age artifacts and placenames (*Nervión, Deba, Ulzama*).[7] Pre-Roman Basque was surrounded by diverse languages, most of them Celtic: in the west was the Indo-European, presumably Celtic, language of the Cantabrians; to the south was Celtic Celtiberian; Iberian, an unclassified language, was spoken to the east; and Celtic Gaulish was spoken to the north, beyond the Garonne River (De Hoz 1981; Echenique 1987:47; Gorrochategui 1995:57). About 400 years later, the Roman invasion began, and from 197 BCE into the beginning of the Common Era, contact with Latin steadily increased (Gorrochategui 1995:35). This influence is visible in the earliest attested Euskarian language: Aquitanian (Gorrochategui 1984). Latin inscriptions include Aquitanian names of people and divinities, often with Latin endings and function words. Germanic tribes, including the Visigoths, are known to have reached the Basque Country in 407 CE, with Leovigildo, Visigoth King, founding Victoriacum on the Basque village of Gasteiz in 581.[8] During this period, Basque was in contact with Gothic, the earliest attested Germanic language. Viking raids began in 844 CE and continued intermittently for almost 50 years. During this time, older forms of Basque may have been in contact with Old Norse, another Germanic language. From the early Middle Ages to the present, influence from other Romance languages to the north and south has been continuous and includes intense contact with Navarro-Aragonese and Gascon, and, later, Spanish and French. Given this extensive history of contact, one of the most difficult tasks before anyone attempting to reconstruct Proto-Basque is to identify directly inherited vocabulary. Here I follow closely *FHV* (Michelena 1977), *Orotariko Euskal Hiztegia* (*OEH*, see the following) and Egurtzegi (2014, Appendix II) in the identification of loans—that is, words borrowed from other languages into Euskarian.[9] In any case where my position on native versus loan vocabulary differs from these earlier sources, it is noted in the note section of the lemmas in the appendix. Newly identified or proposed loans are mentioned in the text where relevant.

The primary source for lexical material in this work is *Orotariko Euskal Hiztegia* (OEH), the *General Basque Dictionary*, and Gorrochategui's (1984) volume on Aquitanian, as well as the older sources listed in Table 1.2. OEH includes citations from the earliest attested Aquitanian and Basque texts to the present when they are available. OEH head entries represent Standard or Unified Basque (*Euskara Batua*), as standardized in 1968, where they exist. When citing OEH forms in the text, only the most conservative (dialect) forms are listed; if these differ from the dictionary headword, the relevant headword is included in parentheses for easy reference. No reference source is given for the OEH, as this is the default reference work. For example, UB *lur* 'earth' is a head entry in OEH. However, under the analysis proposed here, the most conservative attested form of this word is the dialect form *luhur* [LN], further supported by *luur* (with /h/-loss) in other attestations, including varieties of Bizkaian. In cases like this, the comparison sets show: B *luhur* [LN] (*lur*), a modern Basque word, *luhur*, from the Low Navarrese dialect, whose OEH head entry is *lur*. All OEH forms are taken from the current on-line version and can be accessed by the headword provided. In cases where placenames or other terms not found in these sources are used, references are provided. Throughout this book, my own Proto-Basque reconstructions are presented in boldface (preceded by *) and are supported with relevant internal evidence in the chapters that follow, and in the appendix.

1.1 Classical Views of the Proto-Basque Sound System

The most widely accepted reconstruction of the older stages of the Basque language is the result of a combination of methods. The comparative method is applied to modern and historical varieties to arrive at Common Basque, while internal reconstruction is used to reconstruct pre-Latin and pre-Aquitanian Basque, usually called Proto-Basque. Proto-Basque is tentatively placed around the beginning of the Common Era, before most Latin loanwords entered the language. This date also precedes 500 CE, the approximate date at which Common Basque is thought to have split into the precursors of its modern varieties (cf. Michelena 1981).

All modern varieties of Basque have a contrast between a voiced series of stops /b d g/, and a voiceless series /p t k/, which, in some varieties, may be aspirated in certain contexts. The phonological oral stop inventory of Proto-Basque, as originally conceived of by Martinet (1950 [1970]) was distinct from that of the modern language. Instead of a contrast between voiced and voiceless or voiced and voiceless aspirated stops, Martinet suggested an opposition between fortis and lenis consonants. The fortis series of stops had a tense articulation while the lenis series were lax. Voicing and aspiration were both non-contrastive for stops. Martinet's system also included nasals *n and *m[b] (1950 [1974]:387–388).

The fortis/lenis opposition first proposed by Martinet (1950 [1970]) was extended by Michelena (1951, 1957 [1988]). Michelena added a fortis/lenis

series for sonorants and fricatives as well, viewing this contrast as a central organizing principle of all consonants in the language. Figure 1.1 shows the Proto-Basque consonant inventory as reconstructed by Michelena (1977 [2011]:305), where N, L, R represent the fortis sonorant series. In contrast to Martinet's suggestion of *p, in *FHV*, the labial fortis was missing (cf. Michelena 1977 [2011]:215) and, in contrast to Martinet's *m[b], Michelena (1977 [2011]:220) has no labial nasal. In addition to the oral stops, Michelena (1977 [2011]:34) also has a contrast between *s, an apico-alveolar fricative, and *ś a lamino-alveolar fricative, and each of these has a fortis affricated counterpart as well.

Under Michelena's analysis, the opposition between the fortis and lenis series was limited to word-medial position. Word-initially, consonants were neutralized to lenis, while word-finally they were neutralized to the fortis.[10] Rhotics were absent from initial position. The reconstructed inventory also includes *h. The *h was limited to pre-vocalic position, but did not take part in the fortis/lenis opposition.

In addition to the consonants in Figure 1.1, Michelena reconstructs the vowel system in Figure 1.2. Proto-Basque has five vowels (*a, *e, *i, *o, *u) and five descending diphthongs (*ai, *au, *ei, *eu, *oi).

Michelena (1979 [1985]) proposed that the syllabic structure of Proto-Basque was (C)V(W)(R)(S)(T), where any consonant could occupy the onset, W represents the second half of diphthongs with {i,u}, R represents sonorants, {r,l,n}, S represents the sibilant affricates and fricatives, and T stands for an oral stop. Although no Proto-Basque word shows the maximal realization of this template (Michelena 1957/58 [1988]), words with final WRS, WRT, and WST are reconstructed.

Michelena (1957/58 [1988]) suggests peninitial (second-syllable) stress in Proto-Basque, a system which may indeed characterize the post-Roman contact period, as discussed further in Chapter 5. My proposal, which attempts to reach back farther in time, makes use of intervocalic consonant distribution in CVCV roots to reconstruct a quantity-sensitive system of accent, with evidence for lexical accent as well.

fortis	—	*t	*k	* ć /t͡ʃ/	*c /t͡s/	*N	*L	*R /r/	
lenis	*b	*d	*g	* ś /ʃ/	*s /s̺/	*n	*l	*r /ɾ/	
neutral									*h

Figure 1.1 Proto-Basque Consonants after Michelena (1977)

*i		*u		*e		*o		*a
				*ei		*oi		*ai
				*eu				*au

Figure 1.2 Proto-Basque Vowels after Michelena (1977)

Michelena's system was widely adopted with little modification until the end of the 20th century. The most significant changes are those proposed by Lakarra (1995, 2005, 2013). These include changes in the consonant system and a more restrictive view of root/morpheme structure. Lakarra's system, shown in Figure 1.3, is usually referred to as Old Proto-Basque or pre-Proto-Basque.

Lakarra's inventory has fewer proto-consonants than Michelena's, with even more restricted phonotactics. Lakarra's proto-inventory includes five stops *t, *k, *b, *d, *g; three fricatives *ş,* ṣ, *h; and three sonorants *n, *l, *ɾ. Further, Lakarra's reconstruction of the proto-language incorporates a strict monosyllabic CVC root structure (1995, 2005, 2013). With only initial and final positions in CVC roots, Michelena's fortis/lenis opposition proposed for medial (intervocalic) position is no longer viable. Lakarra concludes that sonorants and sibilants never took part in this opposition, and suggests *t vs. *d and *g vs. *k as the only laryngeal/strength contrasts in the proto-language. Within *CVC roots, final position is limited to a sonorant or one of the two sibilants.

The proto-inventories are related to the modern language by a limited number of regular sound changes. Following Martinet (1950 [1970]), Michelena (1977 [2011]:200) proposes a shift of initial fortis oral stops to *h*, discussed in detail in § 3.3, and a (potentially systematic) word-initial *d- > l- (FHV:211–212) in order to account for the apparent absence of root-initial /d/. Lakarra makes a stronger claim: in his system, all inherited instances of word-initial /l/ continue older *d (2006a, 2013:197–198). The only sonorant that Lakarra (2013:198) reconstructs in onset in the monosyllabic period is /n/. As noted above there is no rhotic contrast in Lakarra's system. In order to account for the intervocalic flap /ɾ/ vs. trill /r/ contrast in Basque (e.g. *gori* 'burning, incandescent' vs. *gorri* 'red'), he proposes late developments as the source of the opposition (2013:200), mentioning medieval intervocalic l >/ɾ/ as one source of onset taps, and consonant clusters such as /nɾ/ giving rise to /r/, alongside a gradual word-final /ɾ/ > /r/. Under Lakarra's proposal, sibilants had lenes allophones word-initially and fortes allophones word-finally (Lakarra 2011a, 2011b, 2013:198).

Lakarra (2013) assumes no changes for the vowel system proposed by Michelena, except that there are no diphthongs in the proto-language (2013:201); this forces earlier *VV sequences to be treated either as bimorphemic, or as *VCV. Lakarra also adopts Michelena's peninitial stress (2006b). Within his model, *CVC roots can be preceded by *(C)V- prefixes: assuming root stress, the prefixed word will have peninitial stress (Lakarra

| Onset C$_1$ | *b | *t, *d | *k,*g | *ş | *ṣ | *n | — | — | *h |
| Coda C$_2$ | — | — | — | *ş | *ṣ | *n | *l | *ɾ | — |

Figure 1.3 Pre-Proto-Basque Consonants in C$_1$VC$_2$ Roots after Lakarra (2013)

2013:178). See Chapter 5 where I propose further evidence for peninitial stress in words of the shape CVCVC, as well as a more general weight-sensitive stress system for earlier stages of the language.

Michelena's proto-inventory, lexical reconstructions, and inventory of sound changes together with Lakarra's root theory offer great insight into earlier stages of Basque. I build on these in the chapters that follow. In §1.2, I briefly highlight the differences between Proto-Basque as I reconstruct it and these earlier proposals. Chapters 2–5 present sustained arguments for these differences.

1.2 A Revised View of Proto-Basque Sound Patterns

Chapter 2 presents a revised Proto-Basque vowel system. Though the vowel system adopted follows that proposed by Michelena (1977) shown in Figure 1.2, some gaps in vowel distribution are explained in terms of new or revised sound changes. In addition, the status of some Proto-Basque diphthongs is questioned.

The reconstruction of the Proto-Basque consonant system presented here differs in several ways from the classical proposals just reviewed. First, the Proto-Basque consonant inventory differs from those proposed by Michelena and Lakarra. The revised Proto-Basque consonant inventory is shown in Figure 1.4. (Here and throughout, phonemic representations of all voiceless aspirates, attested or reconstructed, are written <ph>, <th>, <kh>, following Basque orthographic conventions for aspirated stops.)

Like earlier proposals, including *FHV*, Proto-Basque is reconstructed with two series of oral stops and with *h, a glottal aspirate. Unlike earlier proposals, proto-phonemes *m and *ph (= IPA [pʰ]) are argued to be part of the Proto-Basque inventory. Also distinct from earlier proposals is the oral stop contrast: under the present account, the distinction is not between fortis and lenis stops, but between voiceless aspirated stops and unaspirated voiced stops. Another significant difference is in the sibilant/affricate inventory. I propose a single fricative *s, in contrast to earlier proposals where an apical-laminal *s vs. *z contrast was inherited from Proto-Basque. Apart from the addition of *m, the sonorant system is essentially the same as that proposed by Lakarra (1995, 2005, 2013). Chapter 3 reviews widely agreed upon features of the Proto-Basque consonant system and presents

Voiceless aspirate	*ph = [pʰ]	*th = [tʰ]	*kh = [kʰ]	
Voiced (unaspirated)	*b	*d	*g	
Fricative		*s		*h
Nasal stop	*m	*n		
Liquid		*l, *r		

Figure 1.4 Revised Proto-Basque Consonant Inventory

arguments for *m and *ph as proto-phonemes, as well as arguments for treating the voiceless series as a series of voiceless aspirated stops.

A second difference between Proto-Basque as detailed here and earlier proposals involves the structure of roots and syllables. Where Lakarra adheres to a strict CVC root structure, I argue for monosyllabic and disyllabic roots. And where all earlier researchers assume simple syllable onsets, I suggest a limited inventory of syllable-initial *sC clusters. Chapter 4 presents the general phonotactics of Proto-Basque, including evidence for initial *sC clusters, a single sibilant *s, and a regular change of *sC > z in tautosyllabic onset clusters. A final section discusses the status of disyllabic *CVhV(C) roots.

A third difference between this proposal and earlier ones is the proposal of sound changes that are sensitive to the position of Proto-Basque accent. Chapter 5 presents the basic accentual system and proposed sound changes that include reference to accent in their structural descriptions.

Chapter 6 presents a summary of the major differences between Proto-Basque phonology as reconstructed here and other proposals. When taken together, phonological, morphological, and lexical properties of Proto-Basque display superficial similarities with Proto-Indo-European. These similarities are explored in Part II, where evidence is shown to support a long-distance relationship between Proto-Basque as reconstructed here, and Proto-Indo-European, including regular sound correspondences between these two proto-languages.

In the chapters that follow, every effort has been made to make the data and arguments accessible to general linguists as well as Basque language specialists. For this reason, I have made use of standard Basque orthographic conventions, unless a form is enclosed in square brackets [. . .], in which case it is written in the International Phonetic Alphabet. The only difference between the orthography used here and that in the OEH concerns the representation of /h/ and aspiration in dialects that maintain /h/ and aspiration. In this volume, I write this aspiration wherever there is evidence for it, dialectally or historically, and put the OEH entry form in parentheses in the text. For example, I write *handi* 'big', as in the OEH, but I write *hon (on)* 'good', based on dialectal *hon, hun*, and I write *aphal (apal)* 'low, humble', based on dialectal *aphal*. Wherever a variant of the OEH main entry is cited instead of the main entry, the main entry is also in parentheses, as in, for example, *gahar (gar)* 'flame', *luhur (lur)* 'earth'. (For reasons of space, this practice is not followed in the appendix.)

The values of the orthographic symbols used throughout this volume for Basque and Proto-Basque are shown in Table 1.3.

As noted above, full lemmas for Proto-Basque reconstructions with Proto-Indo-European comparisons, where relevant, are in the appendix. I imagine Basque specialists will take a special interest in Part I of this book, while Indo-Europeanists, historical linguists, and general linguists will take a special interest in Part II, but the appendix, which includes relevant cultural

Table 1.3 Standard Basque Orthographic Symbols and Their IPA Equivalents

	Orthographic	*IPA Equivalent*
Voiceless stops	\<p\>	/p/
	\<t\>	/t/
	\<k\>	/k/
	\<tt\>	/c/
Voiceless aspirated stops	\<ph\>	/pʰ/
	\<th\>	/tʰ/
	\<kh\>	/kʰ/
	\<tth\>	/cʰ/
Voiced stops	\<b\>	/b/
	\<d\>	/d/
	\<g\>	/g/
	\<dd\>	/ɟ/
Voiceless fricatives	\<f\>	/f/
	\<s\>	/s̺/ (apical)
	\<z\>	/s̻/ (laminal)
	\<x\>	/ʃ/ (alveo-palatal)
	\<h\>	/h/[11]
	\<h̃\>	/h̃/[12]
Voiceless affricates	\<ts\>	/ts̺/ (apical)
	\<tz\>	/ts̻/ (laminal)
	\<tx\>	/tʃ/ (alveo-palatal)
Nasals	\<m\>	/m/
	\<n\>	/n/[13]
	\<in\>, \<ñ\>	/ɲ/
Liquids	\<l\>	/l/
	\<il\>, \<ll\>	/ʎ/
	\<r\>	/ɾ/
	\<rr\>	/r/
Glides[14]	\<j\>, \<i\>	/j/
	\<u\>	/w/
Vowels	\<i, e, a, o, u, ü[15]\>	/i, e, a, o, u, y/

information, should be appreciated by all readers, and basic aspects of the historical phonology are repeated there to facilitate accessibility. I have tried to write the volume so that it can be read as two separate studies, or as one, but given the nature of some of the arguments, cross-referencing is some-times unavoidable.

Notes

1. The 2011 re-publication of Michelena's (1977) *Fonética histórica vasca* will be referenced from hereon simply as Michelena (1977), abbreviated as FHV.
2. I am grateful to Ander Egurtzegi for allowing me to reproduce Map 1.1.

3. Zuazo (1998, 2008) distinguishes five modern dialects and one extinct variety: Western (Bizkaian and Alavese), Central (Gipuzkoan), High Navarrese, Lapurdian-Navarrese, Zuberoan, and Eastern Navarrese (Salazarese and Roncalese), which was no longer spoken by the end of the 20th century. Hualde and Ortiz de Urbina (2003) follow the classification of Michelena (1977) and Trask (1997). Martínez-Areta (2013b) distinguishes between northern and southern High Navarrese and Western and Eastern Low Navarrese. See Martínez-Areta (2013b) for further discussion.

   Note that 'V' is used as an abbreviation for the Bizkaian dialect (from Sp. *Vizcaíno*), not to be confused with 'B', the abbreviation used in this book for 'Basque' as a general term for the language.

4. Michelena (1981) makes a distinction between Proto-Basque, the oldest (hypothetical) form of the language, and 'Common Basque', the form of the language spoken at the beginning of the Middle Ages, before dialect differentiation. Note that what is referred to as 'Proto-Basque' in Michelena's work, and the tradition that follows, is not a classic proto-language, but a pre-language, since it is based primarily on methods of dialect comparison and internal reconstruction. If Aquitanian is a distinct language from pre-Basque, then use of Aquitanian data allows one to refer to the common ancestor, rightly, as Proto-Basque (or Proto-Euskarian, following Gorrochategui 1995), which I do throughout this volume.

5. Proto-Basque forms preceded by two asterisks ** are arrived at on the basis of external comparative data of the kind presented in Part II of this book. If any aspect of the reconstruction (including the semantics) relies on extra-Euskarian data, ** precedes the PB form, with an explanation of the reason for the double asterisk provided in the note section of the lemma in the appendix.

6. There has also been contact with Andalusian Arabic as spoken by the Moors since about 714, when they arrived at the Ebro River, the south border of Nafarroa. Arabic loans are suggested by Michelena (1961), Trask (1997), and others; some have come into the language through Romance. Other Medieval sources may include Aramaic: compare Basque *zohar* 'luminescence, brilliance (of the sky)' with Aramaic *zohar*, Biblical Hebrew *zohar* 'splendor, radiance'. The Book of Zohar, an Aramaic text, first appeared in Spain in the 13th century, authored by Moses de León (1240–1305).

7. *Nervión* may have the same root as *Nervii*, a Celtic-speaking Belgic tribe. *Deba*, a river name in Guipuzkoa, is from Celt. *deva* 'goddess', a common river name (cf. Belgica river names *Deve, Devere*, *Dieppe* < *Divisapa, etc.). *Ultzama* (< Utzama), the name of a valley in Navarre may be be from *uksama- < *upsama- 'the highest one' (Gorrochategi 2002:107). Possible Celtic loans into Basque from this early period that are not placenames include *mando* 'mule' and *gori* 'red hot'(Igartua and Zabaltza 2012:51). Many compare Basque *adar* 'horn; branch' with OIr. *adarc* 'horn'; however, Trask (1997:369) notes that the Old Irish term has no known Indo-European etymology and may well have been borrowed from Basque into Celtic. This is not a new idea: Buck (1949:209) suggests that Basque *adar* is the basis of Gallo-Latin ADARCA 'spongy growth on sedge', loaned into Celtic, with a Celtic suffix. If the form was borrowed from the B plural, *adarrak*, then the final velar is accounted for as well.

8. Ivan Igartua (personal communication, 2015) points out that there are some doubts regarding the indentification of Victoriacum with Gasteiz; the original settlement could have been Vitoriano, a village about 20km away, or even Veleia.

9. In the task of loan identification, I have been greatly aided by Ander Egurtzegi, who has spent many hours reading over proposed etymologies and highlighting

potential Romance loans. I am also grateful to Jabier Elorrieta, José Ignacio Hualde, and Ivan Igartua for assistance in potential loan identification.

10. Since word-initial position is typically strong, and word-final position weak, the proposed neutralization of initial consonants to lenis, and final consonants to fortis might be unexpected on typological grounds. However, see Chapter 5 for evidence of earlier initial iambic (weak-strong) prosody.

11. The standard orthography writes <h> in many words, even though /h/ is not pronounced in the standard language. Initially, /h/ may be voiceless, but intervocalically, it is predictably voiced.

12. This is not standard orthography. See Hualde (1993b) and Egurtzegi (2014) on nasalized aspirates in Zuberoan.

13. The nasal /n/ is homorganic with a following consonant. For example, *kanpo* 'outside' (<< Sp. *campo*) is [kampo].

14. Context determines when vowels <i> and <u> are realized as non-syllabic.

15. The phoneme /y/ <ü> is an innovation and occurs only in Zuberoan and Mixean Low Navarrese. See Egurtzegi (2017) on the evolution of the /y/ vs. /u/ contrast in these varieties of Basque.

2 The Proto-Basque Vowel System

As noted in Chapter 1, Michelena reconstructs the vowel system in Figure 1.2, repeated in Figure 2.1. Under his analysis, Proto-Basque has five simple vowels: *a, *e, *i, *o, *u as well as five descending diphthongs: *ai, *au, *ei, *eu, *oi. In analyzing these as diphthongs, the hypothesis is that they were tautosyllabic vowel sequences.

   In this chapter, I offer additional evidence for the basic five-vowel system proposed by Michelena. However, in agreement with Lakarra (2013:201), I find little evidence for true diphthongs in Proto-Basque, with no clear contrast between *VV and *VhV sequences. I also highlight distributional and lexical evidence for pre-rhotic front-vowel lowering, and stress-conditioned vowel reduction and loss. In §2.4, I introduce preliminary evidence for an ancient system of root-vowel alternations in Proto-Basque. This system is not discussed further in Part I, but since it involves root pairs with distinct vocalism, it merits mention in the discussion of the Proto-Basque vowel system.

2.1 Evidence for Five Vowels in CV(C) Roots

While Michelena (1977) reconstructs the Proto-Basque vowel system *a, *e, *i, *o, *u on the basis of internal reconstruction of native roots, stems, and affixes, it has yet to be clearly demonstrated that these five vowels contrast in all monosyllabic root types. In Chapter 4, Proto-Basque root, stem, and syllable phonotactics are detailed, including heuristics for root identification. Here, as summarized in Table 2.1, I illustrate vowel distribution and contrast with roots that, with only one exception, have *CV and *CVC

*i	*u	*e	*o	*a
		*ei	*oi	*ai
		*eu		*au

Figure 2.1 Proto-Basque Vowels after Michelena (1977)

structures, consistent with proposals in *FHV*. These roots are also conservative in following Michelena's syllable-structure constraints: in $*C_1VC_2$ roots, C_2 must be one of the consonants *r, *l, *n, or *s. Since *CV roots are relatively rare, I have filled in the cell for *e with a Proto-Basque *CVCV root, *behe 'below'. Disyllabic *CVCV roots are a departure from Lakarra's (1995, 2013) conceptions of Proto-Basque phonotactics, as they are disyllabic and cannot be reduced to CVC roots. Motivation of these roots is taken up in Chapter 4. So as not to diminish the legibility of Table 2.1, I have not included modern reflexes of the roots illustrated. Full lemmas are provided in the appendix. Representative Basque forms continuing roots in parentheses, include *madari* 'pear' (*ma); *behe* 'below' (*behe); *bi* 'two' (*bi[1]); *oso* 'whole' (*so); *su* 'fire' (*su); *gari* 'wheat' (*gar); *igeri* 'float, swim' (*ger); *okhor* 'slice' (*khor); *agur* 'greeting of respect' (*gur[1]); *galdu* 'lost' (*gal[2]); *geldi* 'still, stagnant' (*gel); *bil* 'turn; round' (*bil); *odol* 'blood' (*dol); *ahul* 'lacking' (*hul); *eman* 'give' (*man); *mendi* 'mountain' (*ben[1]); *mihi, min-* 'tongue' (*bin); *monho (muino)* 'hill' (*bon); *asun* 'stinging nettle' (*su-n); *ikhasi* 'learn' (*khas); *iñesi, ihesi* 'flee' (*nes); *hits* 'pale' (*this); *hots* 'sound' (*hos); and *huts* 'empty' (*thus).

   This exercise has proven useful in identifying potential gaps in the *CVC root inventory: although *e appears to have otherwise free distribution, *Ce and *Cir roots are unattested, and Cer roots, with coda *r are rare, with only one or two solid reconstructions, including *ger 'float, swim' and *kher 'twisted'. Interestingly, in examining Vr sequences across dialects, one finds instability of /e/ before /r/, with some doublets suggesting pre-rhotic vowel lowering of *e > a.[1] Since most developments in the vowel system after Common Basque show assimilatory (or dissimilatory) vowel raising, this lowering is notable. Some potential examples of pre-rhotic vowel lowering post-dating Common Basque are shown in

Table 2.1 Five Proto-Basque Vowels in CV and CVC Roots

	*CV	*CVr	*CVl	*CVn	*CVs
Vowel					
*a	*ma	*gar	*gal	*man	*khas
	'fruit'	'grain'	'loss'	'give'	'learn'
*e	(*behe)	*ger	*gel	*ben	*nes
	'below'	'float, swim'	'still'	'rise'	'flee'
*i	*bi	—	*bil	*bin	*this
	'two'		'turn; round'	'tongue'	'muted'
*o	*so	*khor	*dol	*bon	*hos
	'whole'	'cut'	'go along'	'hill'	'sound'
*u	*su	*gur[1]	*hul	*su-n	*thus
	'fire'	'esteemed'	'lacking'	'burning'	'empty'

(2.1a–c), while more interesting cases, suggesting an earlier sound change are shown in (2.1d,e).

(2.1)   Pre-rhotic Front-Vowel Lowering

    i   Dialectal é > a/ __ r

	UB	Bizkaian	
a	*aker*	*akar*	'billy goat'
b	*berri*	*barri*	'new'
c	*bertzun*	*bartzun*	'iron tool for gathering coals (cf. *bertz* 'cauldron')

    ii   Regular: *é > a/ h__ r.

d	*behe*	'bottom, below' < **behe**
	behar	'need, obligation' < **behe-r**
e	*mehe*	'thin, slender' < *ben-e (**ben³**)
	mehar	'tight, narrow' < *ben-e-r (**ben³**)

The sound change in (2.1ii) appears to have pre-dated Common Basque, as *!beher* and *!meher* are unattested in aspirating varieties, while variants such as those in (2.1a–c) suggest that later extensions to other contexts, not preceded by /h/, occurred after dialect diversification. Consistent with this are minimal pairs such as *bartz* 'nit' (< *bar-s) vs. *bertz* 'caldron' (< *ber-s).[2]

2.2 Re-evaluating PB *CVV Roots

While there is good evidence for simple vowels *i, *e, *u, *o, *a in Proto-Basque roots, evidence for Michelena's *ei, *oi, *ai, *eu, *au is less convincing. Either the VV sequence is suspect because it has a possible loan source, or it is suspect because of internal evidence for an earlier intervocalic consonant that has been lost. Surface VV clusters in Basque have multiple sources: intervocalic *n was weakened to /h̃/ and from that stage, to /h/ or zero, with nasalization sometimes reinstated as syllable-final /n/, as discussed in Chapter 3. PB *h also reduced to zero in Western dialects. A final source of VV sequences is sporadic dialectal *r*-loss, most widespread in Zuberoan (FHV). In Table 2.2, I list suspect PB $*V_1V_2$ diphthongs on the left and proposed $*V_1nV_2$ and $*V_1hV_2$ sequences with matching vowels on the right.

    The general problem is that, apart from the one *au sequence in *hauts* 'dust' it is difficult to motivate tautosyllabic *VV. Though I derive *sei* 'six' < **sahi,** given its similarity to IE words for 'six' it could be viewed as a potential loan. B *goi* 'high' may have the same root as *gora* 'up' (< *go-ra), and medieval placenames show <h> suggesting *gohi: cf. <Orengohin>, <Arbelgoihen> [BG]. B *aita* 'father' is thought by many to derive from earlier *atta* (cf. Aq ATTA-) with medial palatalization, later phonologized as a pre-consonantal diphthong (see OEH summary under *aita*). Michelena argues