The History of Basque

R. L. Trask

ROUTLEDGE
London and New York

First published 1997
by Routledge
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by Routledge
29 West 35th Street, New York, NY 10001

Typeset in Times by
RefineCatch Limited, Bungay, Suffolk
Printed and bound in Great Britain by
Mackays of Chatham, PLC, Chatham, Kent

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication Data
Trask, R. L. (Robert Lawrence), 1944–
The History of Basque / R. L. Trask.
Includes bibliographical references and index.
1. Basque language–History. I. Title.
PH5024.T73 1997
499'.92'09–dc20 96–15702

ISBN 0–415–13116–2 (hbk)

Estherri,
aspaldikoari

Chapter 3

Phonology

3.1 PRELIMINARIES

The phonological history of Basque during the last 2,000 years or so is very well understood. That understanding is due almost entirely to the efforts of one man, the great Basque linguist Luis Michelena. Before Michelena's work, virtually nothing of significance had been done on the historical phonology of Basque, apart from Schuchardt's (1887) demonstration that word-initial *p-* must be of recent origin.

Michelena summarized his conclusions about the history of the consonant system in his 1957 paper, but the entire body of his work is presented in his book *Fonética histórica vasca* (*FHV*), first published in 1961. A second and considerably expanded edition followed in 1977, and a third edition in 1985. Most of what follows in this chapter is explicated in much greater detail in that book, though I also include some further observations and suggestions from my own more recent work and from that of others. Here I cite *FHV* in its 1977 edition.

Many of the linguists who preceded Michelena were in fact deeply pessimistic about the possibility of recovering any useful information about the prehistory of Basque on the ground that, since Basque is a genetically isolated language, there are no other languages to compare it with, and hence no possibility of applying the comparative method, our principal instrument in historical phonology. Saussure himself remarks, in his celebrated *Cours* (Saussure 1916): 'On ne peut rien tirer du basque parce que, étant isolé, il ne se prête à aucune comparaison.' And the distinguished comparativist Meillet declares in his 1925 book:

> Entre l'état du basque au xvi[e] siècle et l'état du basque aujourd'hui, il y a des différences, mais les changements ne sont pas essentiels; en substance, la langue est restée la même. Si donc on ne trouvait pas le moyen de rapprocher le basque de telle ou telle autre langue, il n'y aurait aucun espoir d'en faire jamais l'histoire.

It was Michelena's achievement to show that such pessimism was completely unfounded. To begin with, he brought to the study of Basque an unprecedented command of linguistic theory. Almost all earlier investigators had operated with a deeply atomistic and prestructuralist view of phonology, in which absolute sounds were held to be the primitive units, and there was no conception of a phonological system. This shortcoming naturally induced a great deal of confusion. For example, the observation that Latin loans like PACE, TEMPORA and CASTELLU appear in Basque as *bake*, *denbora* and *gaztelu* had led to the positing of a 'voicing' process for initial plosives. In fact, as Michelena demonstrated, Pre-Basque simply had no word-initial voiceless plosives, and further had no voicing contrasts at all, and the attested forms merely reflect the best available rendering in terms of the phonological system of Basque in Roman times.

Furthermore, Michelena was able to take advantage of a great deal of evidence whose significance had not previously been recognized. This evidence includes

- the existence of systematic phonological differences among the various dialects of Basque;
- the presence in all dialects of a number of systematic alternations;
- the presence in Basque of a large number of loan words from Latin and from its Romance descendants;
- the presence of numerous Basque personal names and place names in medieval texts;
- the existence in medieval documents of a not insignificant number of individual Basque words, phrases, glosses, glossaries, verses and other items, including a fair number of connected texts.

Michelena brought to his investigations a magisterial and unprecedented command of all this material, and by applying to it his understanding of synchronic and diachronic phonology he was able to work out in stunning detail just what has happened to the phonology of Basque during the last two millennia.

In his book, Michelena proceeds by reasoning backwards from the evidence until he finally arrives at a reconstructed phonological system for the Basque of around 2,000 years ago. Here I find it more convenient to begin simply by stating that reconstructed system and then describing the various developments which have led to the modern language.

3.2 THE PRE-BASQUE PHONOLOGICAL SYSTEM

In this book I shall use the term *Pre-Basque* for the stage of the language to which Michelena's reconstruction applies. This is the time when Basque began to borrow words from Latin, and must therefore represent a period roughly 2,000 years before our own era.

For Pre-Basque Michelena reconstructs a rather sparse phonological

system of just sixteen consonants and five vowels. The consonants are divided into two groups of eight, called *fortis* and *lenis* by Michelena, with each fortis consonant having a lenis counterpart. These are shown in Table 3.1. Note that the evidence for **p** in Pre-Basque is decidedly scanty. The symbols for the consonants are chosen for orthographic convenience, and as a mnemonic guide to the modern reflexes of these segments; note that Pre-Basque emphatically did *not* have contrastive voicing. In Michelena's interpretation, the fortis consonants were distinguished from their lenis counterparts by greater tension and energy of articulation. He does not elaborate much on this view, but it seems clear that the two series were distinguished at the phonetic level by at least three characteristics:

1 *Duration*. The fortis consonants were longer than the lenis ones. The nasals and liquids were probably distinguished chiefly by duration (though note that **R** was a trill, **r** a tap).
2 *Degree of occlusion*. The fortis obstruents were completely occluded; the lenis obstruents typically were not, except in certain environments, notably after nasals.
3 *Voicing*. The fortis plosives were voiceless; the lenis plosives were typically voiced, though perhaps not exceptionlessly so.

Table 3.1 The phonemes of Pre-Basque

Fortis:	(p)	t	k	tz	ts	N	L	R
Lenis:	b	d	g	z	s	n	l	r
Vowels:	i	e	a	o	u			

Moreover, this fortis/lenis contrast was entirely confined to intervocalic position, being neutralized everywhere else, except that the two series of obstruents also contrasted word-medially after a nasal or a liquid. Word-initially, only lenis consonants occurred, and not all of those: a lexical item in Pre-Basque could begin only with a vowel or with one of **b g z s n l**. Note that no lexical item could begin with **d** or **r**, or with any fortis consonant. Word-finally, only fortis consonants occurred, and again not all of them: word-final **p** was absolutely impossible, and it is highly probable that no lexical item (and possibly no affix) could end in any plosive.

Note also that Pre-Basque had no *m*, no *f*, no *j* and no phonemic *h*; the sources of these modern consonants will be discussed below.

The absence of an **m** in Michelena's reconstruction has occasionally provoked sceptical comment from non-specialists, but see section 3.3 below for the evidence.

The fact that the fortis/lenis contrast was entirely confined to medial position led me to propose, in Trask (1985), that the fortis consonants were nothing more than geminates of the lenis ones. This interpretation, if

accepted, would reduce the Pre-Basque consonant system to the surprisingly small number of only eight consonants. As will be seen below, such a reinterpretation is supported by the fact that sequences of lenis consonants arising in word-formation and inflection always develop into fortis consonants, and further by the fact that the geminate sequences of Latin are invariably borrowed into Basque as fortis consonants. The reinterpretation is not, however, without problems; Hurch (1991) provides a critical view, and I am told that Michelena himself was planning to write a reply, but died before he could do so. In any case, nothing of what follows is affected by the reinterpretation, and I shall use Michelena's notation throughout the remainder of this book.

Consonant clusters were few and simple. Absolutely no clusters were permitted in word-initial position. Word-finally, a (neutralized) nasal or liquid could be followed by a fortis sibilant (an affricate), producing clusters like *-ntz*, *-rtz* and *-ltz*, but it is highly doubtful that the other final clusters of the modern language (like *-st* and *-rt*) were possible at the time, and even some of the instances of these three clusters appear clearly to result from syncope, such as *beltz* 'black' < **beletz*. Word-medially, the two-consonant clusters described in Chapter 2 (like *-st-*, *-rd-*, *-ng-*, *-nt-*, *-rl-*) were mostly present, and medial liquid–sibilant clusters like *-rz-* were also possible. Medial three-consonant clusters like *-rst-* and *-ndr-* were probably rather more frequent than they are in the modern language. There is ample evidence that a number of medial three-consonant clusters formerly present have been reduced. For example, R has *arsto* for common *asto* 'donkey'; common *(h)osto* 'leaf', a derivative of *orri* 'leaf', is attested in the seventeenth-century writer Tartas as *orsto*; northern *ozpin* 'lightning bolt', a compound of *ortzi* 'sky' and **bini* 'tongue', is attested in the sixteenth-century writer Sauguis as *orzpin*; western *aztu* 'forget' is *anztu* in old Bizkaian.

We simply don't know whether the modern series of palatal consonants (*tt dd tx x ñ ll*) was present in Pre-Basque. Most probably these segments were already present, but they never occurred in lexical items, but only in 'expressive' forms derived from ordinary lexical items or affixes by the replacement of another consonant with a palatal.

The vowel system of Pre-Basque is identical to that found in all modern varieties except Z and R, though, as we shall see below, there is clear evidence that the vowel system has not remained unchanged for 2,000 years but that it has rather first expanded and then contracted again.

In Michelena's interpretation, the word-accent in Pre-Basque fell most usually on the second syllable of a lexical item of two or more syllables. This accent was frequently accompanied by the presence of a phonetic aspiration [h] which had no phonological value, though in a number of cases this aspiration occurred on, or was later transferred to, the initial syllable. It was not possible for two aspirations to occur in a single word, or for any aspiration

to occur later than the onset of the second syllable. (Michelena's view of the word-accent has recently been challenged by Hualde (1992, 1995), discussed below, but Hualde does not question Michelena's account of the aspiration.)

In the remainder of this chapter, we shall examine the development of this ancestral system into the modern phonological system. For the moment, though, I shall point out that this secure reconstruction has direct consequences for the possible forms of lexical items in Pre-Basque, and hence immediately renders impossible many of the suggestions which have been made by non-specialists about alleged ancient loans from Basque into other languages, or about possible cognates in putative genetic relations.

Absolutely no native Basque lexical item of any antiquity can begin with any of *p*, *t*, *d* or *r*; all modern Basque words beginning with *p*, *t* or *d* are loan words or phonaesthetic formations (initial *r* is still impossible today). (For the special case of a few words like *talde* 'group' see section 3.19 below.) The evidence for word-initial *k* in Pre-Basque is scanty and doubtful. The monosyllable *ke* 'smoke' has everywhere a voiceless initial, as do a very few other words which look plausibly ancient, such as *koipe* 'oil' and *kirats* 'stench'. Indigenous formations with initial *k* are now common, especially in northern varieties, but virtually all such items postdate the Roman period and hence cannot be projected back into Pre-Basque. We may also note one or two regional variants such as *kar* 'flame', a variant of the more usual *gar*. Probably no word beginning with *m* existed in Pre-Basque with this initial, apart perhaps from one or two loans from Celtic, such as *mando* 'mule': most words with initial *m* today are loan words or recent formations, often 'expressive' in nature, while the rest had initial **b** in Pre-Basque (see below). Native words can begin with any vowel (or with *h* in the aspirating varieties), or with any of the consonants *b g z* (very common), *s l* (less common) or *n* (uncommon). No native word can begin with any consonant cluster at all.

Consider the consequences of these observations. Such words as *tutur* 'crest', *trikatu* 'rest', *kosko* 'acorn cap' and *muga* 'boundary', found also in neighbouring Romance languages, have often been regarded as loans from Basque. But the first two of these words would have been absolutely impossible in Pre-Basque, while the other two could only have existed in different forms (such as **gosko* and **buga*) which later underwent identical changes in both Basque and Romance. It has been an exceedingly common practice in some quarters to extract words from a modern Basque dictionary and to project them unhesitatingly back to the pre-Roman period, in the same form, often on the indefensible ground that 'nothing is known about the history of Basque'. One of the goals of this book is to put a stop to such practices.

3.3 PLOSIVES

At some stage after the Pre-Basque period, the contrasting fortis and lenis plosives, with their predictable voicing difference, came to be reinterpreted as contrastively voiceless and voiced plosives, respectively; the difference in duration was accordingly lost, though the difference in degree of occlusion has remained down to the present day: in modern Basque, *b d g* are usually incompletely occluded, except for some French Basque speakers, for whom the introduction of fully occluded realizations of these segments seems to have been a very recent development, since descriptions from a generation or two ago report incomplete occlusion even in French Basque, and indeed the incomplete occlusion of *b* is clearly described by Oihenart in the seventeenth century. It is likely, of course, that the voicing contrast found in all the neighbouring Romance languages was an important influence in the introduction of a voicing contrast into Basque. The most important consequence of this reinterpretation was that the voicing contrast could be introduced into word-initial position. Hence previously impossible contrasts like *kai* 'wharf, quay' (a Romance loan) and *gai* 'material' (a native word) became part of the language, though the functional load of this contrast is still exceedingly low in initial position today. Many words show regional variation in voicing (*kar* ~ *gar* 'flame'; *pake* ~ *bake* 'peace'), and recent borrowings from Romance sometimes fail to respect the voicing of the lending language (for example, Michelena (1957) reports that, in the Basque of his native Rentería, Spanish *corbata* 'necktie' is borrowed as *gorbata*, while Spanish *gabarra* 'barge' is borrowed as *kabarra*). Nevertheless, it is not true, as some commentators have asserted, that the voicing contrast in Basque is weak or variable: the contrast is fully established, and seemingly has been at least since the time of our earliest texts.

In native words, of course, the reinterpretation left only voiced plosives in initial position. Hence we find plenty of ancient native words throughout the country with forms like *bizi* 'alive', *buru* 'head', *gogo* 'soul' and *gatz* 'salt', but practically no non-loan words with initial voiceless plosives except those of more recent origin and mostly severely localized distribution (initial *d-* is not found in native words, of course, except in finite verb-forms, in which the prefix *d-* is common; see Chapter 4). Early loans from Latin and Romance show the same pattern, reflecting the fact that Pre-Basque had only its single series of lenis plosives to render both voiced and voiceless plosives in the lending language: hence *bake* 'peace' (< PACE), *bike* 'pitch' (< PICE), *Bortu(ak)* 'the Pyrenees' (< PORTU), *berna* 'calf, leg' (< PERNA), *bekatu* 'sin' (< PECCATU), *barkatu* 'forgive' (< PARCERE), *bazkatu* 'feed' (< PASCERE), *dorre* 'tower' (< TURRE or Spanish *torre*), *denbora* 'time' (< TEMPORA), *gela* 'chamber, room' (< CELLA), *gauza* 'thing' (< CAUSA), *gerezi* 'cherry' (< CERESEA), *garden* 'clear' (of liquids) (< CARDINU 'bluish'), *ganbara* 'room' (< CAMERA), *gertu* 'certain; prepared' (< CERTU), *gorputz* 'body' (< CORPUS); *balea*

'whale' (< BALLAENA), *bedeinkatu* 'bless' (< BENEDICERE), *done* 'saint' (< DOMINE), *diru* 'money' (Z *diharü*) (< DENARIU), R Aezk *dekuma* 'tithe' (< *DECUMA), *damu* 'regret' (< DAMNU), B *domeka* 'Sunday' (< (DIES) DOMINICA), *garau* 'grain' (< GRANU), *gura* 'desire' (< GULA), *gisu* 'plaster' (< GYPSU).

In a number of cases, however, we find an initial voiceless plosive either as a regional variant or as the most usual form. Thus, for *bake* 'peace' (< PACE) some areas have *pake*; beside common *gorputz* 'body', Z and R have *khorpitz* and *korpiz*, respectively; next to *bike* 'pitch' (< PICE), *pike* is about equally common; *titare* ~ *titara* 'thimble' is perhaps more widespread than *ditare* (< DIGITALE); *katea* ~ *katiña* 'chain' (< CATENA) is much more usual than *gate(a)*; *katu* 'cat' (< CATTU) is likewise more widespread than *gatu*; and *kaiku* 'wooden bowl' (< CAUCU 'drinking vessel') is the only attested form of this word. There are two reasons for this. First, Basque shows a sporadic but notable tendency to devoice the initial plosive if the following syllable contains a voiceless plosive (this observation would apply to all five of the examples just cited). Note that the native words *bihi* 'grain', *behi* 'cow' and *gurdi* 'cart' are nowhere attested with voiceless plosives, while their compounds *bikain* 'excellent' (< **bihi-gain*), *bekorotz* 'cow dung' (< **behi-korotz*) and *gurpil* 'cartwheel' (< **gurdi-bil*) are attested in places as *pikain*, *pekorotz* and *kurpil*, showing exactly such voicing assimilation. Second, the continuing influence of the neighbouring Romance languages, all of which generally retain the initial voiceless plosives of Latin, may have induced bilingual speakers to re-form the Basque words accordingly.

(Note: This is the position taken by the majority of vasconists, including Michelena and me, on the origin of initial voiceless plosives in some loans from Latin and early Romance. Other views are possible, however, and are preferred by some specialists: some speakers may have borrowed Latin words with initial voiceless plosives from the beginning, or there may have been a later period when Basque-speakers consistently borrowed Latin words with initial voiceless plosives regardless of the voicing in Latin. But the testimony of the native vocabulary, with its nearly categorical voicing of initial plosives, makes it impossible to entertain the view that Pre-Basque actually had a voicing contrast for initial plosives.)

Only a handful of apparently indigenous lexical items exhibit initial voiceless plosives: the items *ke* 'smoke', *kirats* (and variants) 'stench' (possibly a compound of *ke*) and *koipe* 'oil, grease' have everywhere a voiceless initial; the verb *kendu* 'remove' has a voiceless initial everywhere but in R, which has *gentu*; for 'bramble', *kapar* is more widespread than *gapar*, and for 'ignite' *piztu* is more widespread than *biztu* (this is a derivative of *bizi* 'alive'). A rare instance with *t* is *tu* 'saliva, spit', but the imitative origin of this item hardly needs pointing out. One more item is the extraordinary but virtually universal *kalte* 'injury, harm', a seeming derivative of *galdu* 'lose', doubly anomalous with its *-lt-* cluster; only R has *galte*, but the word means 'loss' in R. All of these words normally have aspirated plosives in the aspirating dialects. The reason for this handful of anomalous items is not known, though some of them may represent nothing more than voicing assimilation between plosives. (For the special case of words like *talde* 'group' and *toki* 'place', see section 3.19.2.)

Since the reinterpretation of fortis and lenis plosives as voiceless and voiced, before our earliest substantial texts, Romance words have most often been borrowed with retention of the original voicing, though, as the examples above illustrate, this is not invariably the case.

Consider first voiceless plosives: Occitan *coma* ~ *como* ~ *coumo* 'mane, horsehair' is borrowed into Basque as *kuma*; the Romance preposition *contra* 'against' is borrowed as a postposition *kontra*; a Romance form allied to Old Castilian *cocote* and Occitan *cogòt* 'nape' is borrowed as *kokot(e)*; Castilian *casco* 'skull' is borrowed as *kasko*; a Romance development of Latin PALU 'pole' is borrowed as *paru*; Castilian *pino* 'pine' or a related word is borrowed as *pinu* ~ *piñu*; a Romance word lying between Latin PUTEU and Spanish *pozo* 'well' is borrowed as *putzu*; some Romance form connected with Castilian *tosco* 'rough, crude, coarse' (of Latin origin) is borrowed as *toska* 'kaolin' (Z *toxka* 'clod'); Castilian *tabla* or Gascon *taule* 'plank, board' is borrowed as *taula*. (A few of these seem to have been borrowed very early, such as *porru* 'leek', whose vocalism suggests a source very close to Latin PORRU: cf. Castilian *puerro*, with Romance diphthongization.)

With initial voiced plosives: *bainu* 'bath' < Sp *baño*, *dantza* 'dance' < Sp *danza*, Fr *dance*, *gerra* ~ *gerla* 'war' < Sp *guerra*, Fr *guerre*.

The Romance voicing is not respected in cases like *pintza* 'membrane' (< Aragonese *binza*) and the two words cited above.

Word-medially, of course, the fortis and lenis plosives merely became voiceless and voiced, respectively. Hence words like *lepo* 'neck', *ate* 'door', *zoko* 'corner', *gabe* 'without', *sudur* 'nose' and *hagin* 'molar' simply continue the ancient plosives.

In loans from Latin, word-medial voiceless plosives, both simplex and geminate, were rendered by Pre-Basque fortis plosives and hence appear in modern Basque as voiceless plosives: *ipizpiku* 'bishop' < EPISCOPU, B *erripa* 'slope' < RIPA, *errota* 'wheel; mill' < ROTA, *aditu* 'hear, understand' < AUDITU, *nekatu* 'exhausted' < NECATU 'killed', *bake* 'peace' < PACE, *laket* 'be pleasing' < PLACET, *barkatu* 'forgive' < PARCERE, *apario* 'meal' < *APPARIU, *katu* 'cat' < CATTU, *bekatu* 'sin' < PECCATU, *zaku* 'sack, bag' < SACCU, *okela* 'morsel, meat' < BUCCELLA 'mouthful'; *zuku* 'juice, soup' < SUCCU.

In the same position, Latin simplex voiced plosives were borrowed as lenis plosives and have become voiced plosives: old G *zabau* 'tablecloth' < SABANU, *abere* 'large animal' < HABERE 'have' (or from a Romance development of this), *judu* 'Jew' < IUDAEU, *bago* 'beech' < FAGU, *lege* 'law' < LEGE, *errege* 'king' < REGE, *magi(ñ)a* 'sheath' < VAGINA. The rare Latin voiced geminates presented problems, but, being both long and completely occluded, they were most typically taken over as Pre-Basque fortis plosives:

apaez ~ *apaiz* ~ *apez* 'priest' (< ABBAS); B *zapatu* 'Saturday' (< SABBATU); the common element *Ap(h)at(a)-* (< ABBATE 'abbot') in toponyms like *Apatamonasterio* (Biz). The same occurs in the odd loan from Arabic: *atorra* 'shirt' < Arabic *ad-durrāʔa* 'type of woollen shirt' (cf. Old Spanish *adorra* 'shirt' < Arabic *ad-durrāʔa* 'type of woollen shirt'). (And R *repat(t)an* 'shepherd boy' reflects an Arabic source with two voiced geminates, but this may be merely a loan from Aragonese.) The general rule seems to be that each of the four types of medial plosive in Latin was rendered by the Basque plosive which matched it in at least two of the three properties of duration, degree of occlusion and voicing. Note also the interesting treatment of the clusters in *apal* 'humble' (< AD VALLE) and *gutizia* 'desire' (< Old Sp *cobdicia*); on this, see section 3.19.

In Pre-Basque, the contrast between the two series of plosives was maintained after nasals and liquids, but, in all dialects except Z and R, all plosives eventually came to be voiced after *n* or *l* (though not after *r*). Thus we have common *denbora* 'time' but eastern *t(h)enp(o)ra* < TEMPORA 'times' and common *aldare* 'altar' but eastern *alt(h)are* < ALTARE. Native words show the same developments: common *ongi* 'well' (adv.) is R *onki*, Z *hunki*, and common *alde* 'side' is eastern *alte*. Very occasionally B fails to show this voicing, as in B *denpora* 'time'; the reason for this is not known.

The history of Pre-Basque **b** requires special attention. In a number of cases **b** has developed into *f*, and this represents a major source of *f* in modern Basque. This process is sporadic, unpredictable and somewhat surprising, since Pre-Basque **p** does not normally yield *f*, but it no doubt reflects the facts that **b**, like the other lenis plosives, was usually incompletely occluded, and that voicing was genuinely non-distinctive in Pre-Basque. This development occurs most often intervocalically, but rarely also initially: *kabia* 'nest' (< CAVEA) is *kafia* in some regions; the verb *ibeni* ~ *ipini* (and other variants) is *ifini* in places; common *barre* 'smile, laugh' has a variant *farre* (rarely also *parre*). The word *zubi* 'bridge' is sometimes written as *Zufi* or *Zuffi* in medieval toponyms, and one medieval document mentions a certain *Nunno-falzahuri*, 'the town of Nuño the Black', in which *falza* represents *baltza* 'the black'. Particularly striking is the name of the province of Navarre: all the evidence suggests that *Nafarroa* 'Navarre' and *nafar* 'Navarrese' were the usual medieval forms in all areas, and the town of Nafarrate in Araba is recorded as *Naffarrate* as early as 1025.

(The very rare instances of *p* > *f* are confined to names: *Fadura*, a variant of *Padura* (and *Madura*) (< PADULE 'water-meadow') and *Fradu* and *Fradue*, house names (Biz) (< Sp *prado* 'meadow').)

The other peculiarity of **b** is that, in word-initial position, it often develops into *m*, and this is the principal source of initial *m* in native Basque words. This process too is highly sporadic, but it is clearly favoured by a following nasal in the same word: *mihi* ~ *min* 'tongue' (< **mini* < **bini*); *mahats* 'grapes' (< **banats*); *min* 'pain' (< **bin*). Very occasionally, the same

things happens in medial position: *zamau* 'tablecloth' (< SABANU), *zumel* 'holm oak' (< **zubel* < **zur-bel*), and possibly also in *ametz* 'gall-oak' and *amets* 'dream', for which the attested variants *amentx* and *aments* suggest original **abentz* and **abents*, respectively.

These developments suggest that Pre-Basque **b** had rather a wide range of phonetic realizations, with [m] perhaps being particularly frequent initially and [f] medially. (The suggestion of Martinet 1950, 1955 that Pre-Basque might have had a distinct phoneme /m^b/ (and also /n^d/) has found no support among vasconists.)

A third feature of **b** is that it is almost always lost in word-initial position before **o**, and very rarely before **u**. Thus, we have *ollo* ~ *oilo* 'hen' (< **bollo*, from a Romance development of PULLU 'chicken'), *ondo* 'bottom, side' (< **bondo*, from a Romance development of FUNDU 'bottom'), *okela* 'morsel, meat' (< **bokela*, from a Romance development of BUCCELLA 'mouthful'), *ostiko* 'heel; kick' (< **bostiko* < POSTICU 'posterior'), *otu* 'request, supplication' (< **botu* < VOTU), *horma* 'wall; ice' (< *borma*, attested in the seventeenth-century writer Oihenart in both senses, < FORMA). It is rarely possible to detect this development in native words, though the universal *on* 'good' may be an instance, if the Aquitanian *Bon(n)-* is the same item. Rare exceptions to this development are *bortz* ~ *bost* 'five' and *bortitz* 'strong, violent' < FORTIS. Before **u**, we have *urki* 'birch' alongside the less common *burki*, probably from Indo-European, though the rare B variant *turki* and old G *epurki* muddy the waters considerably, and the less common *buztarri* 'yoke' alongside the more widespread *uztarri*.

There is no trace of evidence that medieval Basque ever distinguished [b] from [v], in contrast to Old Spanish, which did.

In a number of cases, an original medial cluster **nb** has yielded modern *m*. Thus, *seme* 'son' surely derives from **senbe* (the form *Sembe* is attested in Aquitanian; see Chapter 6). It is probable that a number of other cases of medial *m* have a similar origin, such as *hamar* 'ten', *amu* 'hook' and possibly *ama* 'mother', if this is not merely a nursery word. There are even seemingly recent instances of this: northern *zonbat* 'how many?', a compound of *zoin* 'which?' and *bat* 'one', has a widespread variant *zomat*. The case of *ume* 'child' is particularly interesting, since both *Ombe* and *Vmme* are attested in Aquitanian, with the second looking very much like an intermediate form between an original **unbe* and modern *ume*. In one or two cases Basque *nb* continues Latin M, as in *ganbara* ~ *k(h)anbara* ~ *k(h)anbera* 'room' (< CAMERA), though this development might simply reflect a Romance development **cambra* (cf. French *chambre*).

The absence of an *m* from the Pre-Basque phonological system has sometimes attracted comment from non-specialists. But consider the evidence:

1 In modern Basque, the very frequent *m* is overwhelmingly confined to obvious loan words and to phonaesthetic formations of no great antiquity.

There are scarcely two dozen words with *m* which appear genuinely to date back to the pre-Roman period.

2 We have a large number of loan words in which modern *m* has indisputably developed from Latin or Romance *b* (or another labial borrowed into Basque as **b**): *magi(ñ)a* 'sheath' (< VAGINA), *mutil* 'boy' (< PUTILLU), *makila* 'stick' (< BACILLA), *muxika* 'peach' (< (MALA) PERSICA), *mendekatu* 'avenge oneself' (< VINDICARE), *zamau* 'tablecloth' (< SABANU), eastern *Mendekoste* 'Pentecost' (< PENTECOSTE), *mimen* ~ *mihimen* 'osier willow' (< VIMEN), Z *mühüllü*, R *mullu*, B G *millu* 'fennel' (< FENICULU), B G *mika* 'magpie' (< PICA), Z *mezpera* 'eve' (< VESPER), western *bañu*, central *mainu* 'bath' (< Sp *baño*), and many others. Some of these are attested only with *m*; others have *m* in some areas but *b* or *p* elsewhere.

3 A number of native words which normally have *b* appear with *m* in some varieties: common *biga(i)* 'two-year-old heifer' but HN LN *miga*, common *baneki* 'if I knew' (with the universal *ba-* 'if') but old B *manequi*, LN *mantxut* 'what did you say?' but common *badantzut* 'I hear', and so on. There are also cases in which *m* is common but *b* is attested locally, such as common *me(h)ar* 'narrow' (< **benar*) but B *berar*, with nasal dissimilation.

4 Ancient items with *m* behave in word-formation as though they had *b* instead. For example, *arpin* 'plantain' may be securely derived from **ardi-bini* 'sheep-tongue'; the development of this form to *arpin* would be absolutely regular in Basque (compare *bepuru* 'eyebrow', from *begi* 'eye' plus *buru* 'head', or *okin* 'baker', from *ogi* 'bread' plus *-gin* 'who makes', or *arpigae* 'young ewe', from *ardi* 'sheep' plus *bigae* 'heifer' (< **bigana*), or *abatei* 'knell', from *abade* 'priest' plus *dei* 'call'). Similarly, *ope* 'slender bread roll' is from *ogi* 'bread' plus *mehe* 'slender' (< **bene*); northern *o(r)zpin* 'lightning bolt' derives from **ortzi-bini* 'sky-tongue'; and *ozpin* 'vinegar' is from an unidentified first element plus the common second element *-min* 'spicy, hot', related to *min* 'pain', from **bin*, all three words showing the regular devoicing of a plosive after a voiceless sibilant. This confirms that *mihi* ~ *min* 'tongue' derives from **bini*, *mehe* 'slender' from **bene*, and *min* 'pain' from **bin*.

5 In the dialects retaining the aspiration, *h* can follow any liquid or *n*: *alhaba* 'daughter', *ilhun* 'dark', *erhi* 'finger', *urrhe* 'gold', *senhar* 'husband', *anhitz* 'many', *inhurri* 'ant' and so on. But there is not a single instance of *h* after *m*, just as there is no instance of *h* after *b* (or after any other voiced plosive).

6 While most of the other Basque consonants make an appearance somewhere in the rich inflectional morphology of the language, and *n* in particular is very frequent indeed, *m* is categorically absent from inflectional affixes.

7 Likewise, *m* is categorically absent from grammatical words like pronouns, conjunctions, subordinators, determiners, question words and postpositions.

8 Among the dozens of word-forming suffixes in Basque, *m* is absolutely lacking, save only in the abstract-noun-forming suffix *-men* ~ *-mendu*, a transparent loan from Latin -MENTU.

In sum, then, the evidence is overwhelming that *m* was absent from Pre-Basque and that modern *m* has developed from earlier *b* (or *nb*) where it has not been borrowed. (See also section 5.5 for a treatment of the use of *m* in expressive formations.)

In word-final position, it seems unlikely that Pre-Basque permitted any plosives at all, at least in lexical items, though a few inflectional suffixes may perhaps have ended in fortis plosives. The modern language has several very common inflectional suffixes ending in *-t* or *-k*. The evidence from alternations shows clearly that the first-person singular agreement-marker *-t* and the second-person singular agreement-marker *-k* derive from **-da* and **-ga*, respectively, while the absolutive plural-marker *-k* quite possibly derives from **-ge* (on all this, see Chapter 4). The ergative case-suffix *-k* and the essive suffix *-tzat*, which can never be followed by any other material, exhibit no alternations, and there is no way to tell if they have a similar origin. A few lexical items have acquired final plosives by various phonological developments. Western *bart* 'last night' is *barda* in the east, and clearly results from loss of the final vowel; the universal *bat* 'one' has an apparent combining form *bede-*, pointing to original **bade* or **bada*. The universal *bort* 'bastard' may have a similar origin. Western *bost* 'five' is derived from *bortz*, retained in the east, by the process described in section 3.12. A modest number of loan words from Occitan or Old Spanish show final plosives, such as *kokot* 'nape' (Occitan *cogòt*), *agot* 'leper' (Old Spanish *agot*), *kok* 'coke' (Spanish *coque*). Z *topet* 'flask' doubtless has a similar source. Today final plosives are most frequent in interjections and words of imitative origin: *karrak* 'crack!', *dzast* 'bang!', *dzart* 'smack!', *(k)ok* 'vomit, indigestion', *tak* 'tick, tap', and probably also B G *txit* 'very' and also '(a single) word'. Final *-p* is absolutely unattested except in one or two interjections like B *eup*, used in mockery.

In word-medial position, the ancient **p/b** contrast has generally been continued as *p/b*. There are, however, a few items in which both voiceless and voiced plosives are attested medially. The widespread *ebaki* 'cut' (B *ebagi*) appears in some eastern and western varieties as *ep(h)ai*, and such derivatives as *epaiki* 'shears for cutting iron' and *epaile* 'cutter' are found throughout the country. (See Chapter 4 for the variation *-ki/-gi* in participles.) The verb *igan* ~ *igo(n)* 'ascend' is attested as *ikai* in the seventeenth-century writer Oihenart. The word *ipar* 'north', Michelena suggests, is identical to *ibar* 'valley' and originates in the compound *ipar-haize* 'valley wind, north wind'. Common *eduki* 'hold' (B *edugi*) has an old B participle *ituten*. The verb *egotzi* 'throw' appears to be the source of the derivatives *ekoizpen*, *ekoizte* 'product, fruit'. Moreover, the first-plural marker *-gu* sometimes

appears as *-ku*, as in northern *zauku* ~ *zaiku* and B *jaku* ~ *yaku* 'it is to us', in contrast to central forms like *zaigu*, and the postposition *behe* 'under' often appears as *-pe* when suffixed, as in *lurpean* 'under ground'. The reason for such variation is not known.

Words with medial *d* in some varieties often have *r* elsewhere: *ideki* ~ *ireki* 'close', *edan* ~ *eran* 'drink', *eduki* ~ *eroki* 'hold' and others. Most of these seem to have had *d* originally, though a few may have had *r*.

Very occasionally, a Latin or Romance coronal plosive in initial position appears as *l*: *leka* 'pod' (< THECA) (but Z *theka*); eastern *lizifrina* 'discipline', *lanjer* 'danger', R *lantzatu* 'dance' (common *dantzatu*). And the modern town of *Larraga* (Nav) appears to be the town called *Tárraga* by Ptolemy.

In very rare and sporadic cases, a word-initial plosive, most often a velar, has been lost or added in post-Roman times, sometimes very recently. Examples: *kabia* 'nest' (< CAVEA) is *(h)abia* in places; *kamuts* 'blunted' (< Occ *camus* 'blunt') is *amuts* in places; *kokots* 'chin' is *okotz* in places; *kupa* 'barrel' is more commonly *upa*; B *tarro* 'gully' is *arro* in places; *putz* 'puff of breath, fart' is often *utz*; *poker* 'belch' is more commonly *oker*; Lat CUNA 'cradle' yields Sal *ua*. The Romance loan *kamaña* 'shepherd's bed' has a putative variant *amaña* listed in Azkue, but there is no independent confirmation of this form. Sp *carlinga* 'cabin, cockpit' appears to be the source of localized G *arlinga* 'place for the fireman/stoker on a train'. On the other hand, Sp *acero* 'steel' appears in some varieties as *galtzairu* (with an extra *l* as well). Another one is possibly *epel* 'lukewarm', if this derives from Latin TEPIDU, but this etymology is far from certain.

For the aspirated plosives of the northern dialects, see section 3.11.

In sum, Pre-Basque plosives were distributed as shown in Table 3.2. Note that **d-** was apparently not found outside of finite verb-forms and loan words.

Table 3.2 The Pre-Basque plosives

	-b-	-d-		-g-
b-		(d-)	g-	
	-p-	-t-		-k-

3.4 SIBILANTS

Generally speaking, fortis and lenis sibilants contrasted only intervocalically in Pre-Basque, except that the contrast was also maintained word-medially after a neutralized rhotic, a neutralized lateral or a neutralized nasal. Initially, only lenis sibilants occurred; finally, only fortis ones. It seems likely that the phonetic realizations of the sibilants in Pre-Basque were not very

different from their modern descendants: affricates for the fortis ones, fricatives for the lenis ones. The Pre-Basque distribution is still strongly represented in the modern language. Initially, only fricatives are usual: *su* 'fire', *zu* 'you', *sortu* 'be born', *zahar* 'old', *sai* 'vulture', *zaio* (an auxiliary verb-form). Some northern varieties, however, have the unique item *tzar* 'bad'; this is thought to be a specialized and anomalous development of *zahar* 'old', resulting from the reinterpretation of a postposed form as a free form (see section 3.19). Northern varieties also have a few 'expressive' items in *tz-*, such as L *tzurruntzuntzun* '(travelling) in an old cart', Z *tzüsto* 'rotten wood', LN *tzut* 'methodical, orderly', Z *tzipi-tzapa* '(walking) in small steps' and a few others in this vein. Moreover, both the western dialects and certain eastern varieties have almost completely replaced initial *x* by *tx*, so that *xirula* 'Basque flute' and *ximista* 'lightning' appear as *txirula* and *tximista* in these varieties, and similarly for other words. Word-finally, affricates are still the norm: *hotz* 'cold', *hots* 'shout', *beltz* 'black', eastern *bortz* 'five', *mahats* 'grape(s)'. Here, however, a number of exceptions have been introduced. Verb-stems ending in a sibilant may show a fricative in the radical: hence *ikusi* 'see' has radical *ikus*. The very common instrumental case-suffix is invariably *-z*, never **-tz*. A few imitative items end in fricatives, such as *piz* ~ *pix* 'piss' in some varieties. R, uniquely, has final *z* in a few ancient lexical items, such as *korpiz* 'body' (common *gorputz*). And recent loan words generally preserve final fricatives: hence *arroz* 'rice' (a loan from Spanish) now forms a minimal pair with native *arrotz* 'foreigner'.

In loans from Latin and early Romance, Latin s is almost invariably represented by Basque *z*: *ezpata* 'sword' < SPATHA, eastern *zamari* 'horse' < SAGMARIU 'pack-horse', *zela* 'saddle' < SELLA, *ziape* 'mustard' < SENAPE, *bazkatu* 'feed' < PASCERE, *zigilu* ~ *zigulu* 'seal' < SIGILLU, *zoru* 'soil' < SOLU, *zabau* ~ *zamau* 'tablecloth' < SABANU, *azeri* 'fox' < ASENARIU (a personal name), *meza* '(Catholic) mass' < MISSA, *gauza* 'thing' < CAUSA, *ozte* 'host, troop' < HOSTE, *iztupa* 'hemp, oakum' < STUPPA, *gerezi* 'cherry' < CERESEA, and very many others. After a nasal or a liquid, the outcome of Latin s is *tz*: *antzara* 'goose' < ANSERE. (The eastern dialects, however, which retain the ancient contrast between fricatives and affricates after a nasal, have *anzera*.) We might have expected the Latin geminate -SS- to yield Basque *-tz-*, but there are no certain cases of this; one possible example is G *pitzatu* 'crack' (v.), which may derive from an unattested **FISSARE*, but the etymology is not certain. In final position, we normally find *-tz*, as in *bortitz* 'strong, violent' < FORTIS and *gorputz* 'body' < CORPUS, but note the exceptional *apaez* (etc.) 'priest' < ABBAS and *maiz* 'often' < MAGIS 'more'. Only in a handful of seemingly ancient loans do we find Basque *s* for Latin s: *soka* 'rope' < SOCA, western *siku* 'dry' (eastern *ziku*) < SICCU, *gisu* ~ *kisu* 'plaster' < GYPSU, and a very few others. In an important article, Michelena (1965) concludes from this evidence that Latin s must therefore have been laminal, like English and French *s*, and not apical, like the *s* of the modern Iberian Peninsula. This is

the single respect in which the evidence from Basque sheds otherwise unavailable light on the pronunciation of Latin.

Our earliest written records of Basque, however, show Spanish *s* being borrowed as Basque *s*, while Basque *z* is reserved for the sibilant written *z* in Spanish; the Spanish consonant has been an interdental fricative since about the sixteenth century, at least in the north, but was earlier something different. It is difficult to tell just when Romance *s* acquired its apical pronunciation, but it was probably earlier rather than later, at least in the north, and the apical was certainly present by the time of the Arab occupation of Spain. Note for example eastern Basque *deus* ~ *jeus* 'anything', from a Romance development of GENUS 'kind'; this word was clearly borrowed after the palatalization of velars before front vowels had occurred in Romance, but before the Basque loss of intervocalic *n*. Note also the interesting case of *soro* ~ *solo* 'field, meadow', borrowed from a Romance development of SOLU (cf. Castilian *suelo* 'ground'); this is the same Latin word that was earlier borrowed directly as *zoru* 'soil'.

Recall from Chapter 2 that Basque has sibilant harmony: a word may contain only laminal sibilants or only apical ones: *zezen* 'bull', *izotz* 'frost', *itsaso* 'sea', *sasi* 'bramble'. This harmony has continued to apply to compounds and loan words down to fairly recent times: the compound of *zin* 'truth, oath' with *-etsi* 'consider' is attested as *zinetsi*, but today the form *sinetsi* ~ *sinitsi* is universal. The Spanish word *francés* 'French' was borrowed as *fran(t)zes*, but since the eighteenth century the usual form has been *fran(t)ses*.

In Bizkaian, we find textual evidence for confusion between *z* and *s* beginning to appear in the early seventeenth century; today this contrast has been totally lost (in favour of the apical) throughout Bizkaian and in much of Gipuzkoan. The spread of the merger across Gipuzkoan appears to involve leaps from town to town across intervening countryside; in Azpeitia, for example, older speakers born in the town lack the contrast, while those born in the surrounding farmhouses retain it (Joxe Mari Ibarguren, p.c.). The *tz*/*ts* contrast is lost (in favour of the laminal) in the same region.

There is a detectable tendency for *z* to develop into *s* before a stop consonant, especially a coronal. Many of the Araba toponyms recorded in 1025 with *z* have *s* today: *Hazteguieta* (*Astegieta*), *Bahaheztu* (*Maestu*), *Eztarrona* (*Estarrona*) and so on. This is probably the reason for the regional variation observed in a few words like *ezne* ~ *esne* 'milk'. The word *adiskide* 'friend' may be another example, if Lafon's celebrated etymology is correct: he proposes **adinez-kide* (*adin* 'age', *-(e)z* Instr, *-kide* 'fellow').

Very rarely and sporadically, an initial sibilant is lost by dissimilation, as in *Anso* 'Sancho' (< *Sanso*) and the toponym *Estabe* in Araba, Spanish *Cestafe*, recorded in 871 as *Zeztave*. The common imperative verb-forms *zatoz* and *zatozte*, both 'come!', with initial *z-* marking second person, appear in G as *atoz* and *atozte*.

Such dissimilation occasionally occurs internally. B has a word *gortaits* 'manure', apparently from *gorta* 'courtyard' and *sits* 'dung', parallel to *perusits* 'guano' (*Peru* 'Peru'). Common *isats* 'broom' (both senses) has an eastern variant *jats* (< **iats*).

Very occasionally an epenthetic sibilant is inserted before a plosive, as in B *plauta* ~ *plausta* 'popgun' (< FLAUTA) and B *mutur* ~ *mustur* 'snout'.

Z, uniquely, has acquired contrasting voiced sibilants partnering *z s x*. Its /ʒ/ is the regular Z development of |j|. Otherwise, *z* and *s* acquired voiced allophones in voiced environments, and then the variety borrowed some Romance words with initial voiced sibilants, thus introducing a voicing contrast whose functional load is close to nil.

In sum, Pre-Basque sibilants were distributed as shown in Table 3.3.

Table 3.3 The Pre-Basque sibilants

	-z-			-s-	
z-		-tz	s-		-ts
	-tz-			-ts-	

3.5 NASALS

Pre-Basque had the fortis and lenis nasals **N** and **n**; these contrasted only intervocalically. Initially, only **n** appeared, and finally only **N** (see below). At some time before the appearance of our first substantial texts, intervocalic **n** was categorically lost from all varieties (apart from two special cases, discussed below). The evidence from eleventh-century toponyms suggests that this loss must have occurred before AD 1000. The fortis **N**, no longer having a lenis counterpart, was then reduced to *n*. These processes are well illustrated by the Latin word ANNONA 'provisions', which was borrowed as **aNona*, but which appears in modern Basque as *anoa*.

In native words, intervocalic *n* generally continues Pre-Basque **N**: *anaia* (and variants) 'brother (of a man)' < **aNa-*, *ainara* 'swallow' (bird) < **aiNa-la*, *ene* 'my' < **eNe*, *arrano* 'eagle' < **aRaNo*. This fortis (or geminate) nasal is directly represented in medieval documents in a few instances, notably in the name *Annaya* (*anaia* 'brother').

A number of words, both native and borrowed, now exhibit stem alternations resulting from the fact that an original **n** was lost intervocalically but retained elsewhere. Thus, for example, **ardano* 'wine' appears as *ardao* ~ *ardo* (and other variants) today, but its combining form is *ardan-*, as in *ardandu* 'ferment' (*-tu* verb-forming suffix) and *ardantza* 'vineyard' (*-tza* noun-forming suffix), while Latin CATENA 'chain' appears as *katea* today but as *katen-* in word-formation, as in *katenbegi* 'link of a chain' (*begi* 'eye'). This shows clearly that the loss of final vowels in the first elements of

compounds and derivatives (see section 3.19) predates the loss of intervocalic **n**.

In some cases in which intervocalic *n* was preceded by *i*, or by an *e* which was raised to *i*, the nasal underwent the usual palatalization to *ñ* (see section 3.8), and hence remained when *n* was lost: thus, for example, *magia* 'pod' (< VAGINA 'sheath') has a regional variant *magiña*, and *katea* 'chain' (< CATENA) has a variant *katiña*.

In a very few cases in which intervocalic *n* was preceded by *u*, the *n* developed into *m* by labial assimilation: thus, for example, the widespread *kuma* 'cradle' (< CUNA). The same occurs in the native word for 'elm', which is *zumar* in the west but *zu(h)ar* ~ *zugar* in the east, from **zunar*. The word *artizkuna* 'place for milking sheep', which contains the common suffix *-(g)une*, appears as *artizkuma* in G. The town in Bizkaia called *Luno* in Spanish is popularly *Lumo* in Basque.

The loss of intervocalic **n** left behind nasalization of the adjacent vowels. That nasalization in most (not all) cases remains today in the eastern dialects Z and (more regularly) R, which therefore still have a full set of distinctive nasal vowels. (R shows the peculiarity that it normally retains the nasal vowels only in nouns and verbal nouns, but not in adjectives or other verbforms.) In B, we have the explicit testimony of two sixteenth-century writers (Garibay and Madariaga) that nasalized vowels were still present in that dialect, even though they were not overtly represented in the earliest Bizkaian texts. At some later time, probably not much later, nasalization was simply lost in B. In all the remaining dialects, nasalization was apparently lost too early to be recorded at all. Thus, for example, 'prudent' is *zu(h)ur* in most dialects but *zũhũr* (with front rounded vowels) in Z (< **zunur*), and 'slender, thin' is *me(h)e* in most dialects but *mẽhẽ* in Z (< **mene* < **bene*). A particularly good example is **zene* 'small, insignificant'. This is *zẽhẽ* in Z, *zehe* in L LN, and *ze* elsewhere; its palatalized form **xene* yields Z *xẽhẽ*, R *xẽ*, L LN *xehe*, elsewhere *(t)xe*. Sporadically, but very frequently, however, something different happened: nasalization of a vowel or diphthong was reinterpreted as representing a following *n*. Hence, for example, **zani* 'watchful, hopeful, expectant' yielded *zãĩ*, which remains today in the east, while the other varieties variously show *zai* by denasalization or *zain* by reinterpretation. This sort of variation is very common: *arrain* ~ *arrai* 'fish' (< **arrani*), *sehi* ~ *sein* ~ *sẽĩ* 'boy, servant' (< **seni*), *garau* ~ *garaun* 'grain' (< GRANU), *ihitz* ~ *ihintz* ~ *intz* 'dew' (< **initz* or possibly **inintz*).

Even R occasionally loses the nasalization. Thus Z *ãhãbe* 'bilberry' (< **anabe*) appears as *abi* in R; common (and Z) *mahats* 'grapes' (< **banats*) is *mats* in R. Some R words reported with nasal vowels by Bonaparte and Azkue had lost the nasalization by the 1950s. In a couple of words, R shows metathesis of the nasalization: *ãre* 'sand' < **arẽ* < ARENA; *gãzta* 'cheese' < **gaztã* < **gaztane*.

The twelfth-century pilgrim Picaud appears to record a nasal vowel when he writes *ardum* for 'wine': this must represent *ardũ*, a form attested today in the east.

A handful of words show an unexpected nasal vowel in R: *õla* 'hut' (Z *olha*), *õre* 'dough' (Z *orhe*), *ũr* 'hazelnut' (Z *hür*), *ũrzo* 'pigeon' (Z *urzo*). Some of these may result from contamination by near-homophones in which nasalization is normal; others are simply mysterious.

Occasionally the resolution of nasal vowels was affected by other regular or irregular processes. The word **ardano* 'wine', for example, developed to **ardãõ*. This merely becomes *ardao* in modern B, but all other dialects have reduced the unusual sequence *ao* to *o*, giving *ardo* in the central dialects. In Z, the same reduction occurred, followed by the usual raising of *o* to *u* in final syllables (see section 3.9), but the nasalization was not lost, and the Z form is *ardũ* (with final stress). In L and LN, the nasalization was exceptionally transferred to the preceding consonant, yielding *arno*. The Latin word BENEDICERE 'bless' was borrowed as **benedikatu*, which apparently underwent an early metathesis to **bedenikatu*. This in turn developed to **bedeĩkatu*, which by reinterpretation became modern *bedeinkatu* (this reinterpretation clearly happened too late for the newly introduced *n* to cause voicing of the following plosive; see section 3.3).

Very often the vowels flanking the lost nasal were protected from hiatus by the presence of the aspiration, or possibly sometimes by the insertion of an aspiration. Thus **bini* 'tongue' became first **mini* by the nasal assimilation discussed above and then **mĩĩ*. This gives *mĩhĩ* today in Z and *mihi* in the other northern dialects. Loss of the aspiration elsewhere has variously led to *mii* or (by vowel coalescence) *mi* or (most frequently, by reinterpretation) *min*. Not infrequently the dialects which have lost the aspiration have broken up the hiatus by inserting *g*, the consonant favoured for this purpose: hence the word for 'fleeing, flight' is *ies* ~ *ihes* elsewhere but *iges* in the west (< **ines*, with an unusual final fricative). Very occasionally, in the vicinity of the rounded vowel *u*, the consonant inserted is *b* rather than *g*, and even less frequently we find *d* or *r* inserted instead.

Latin N and NN were borrowed as **n** and **N**, respectively, with the same consequences as in native words: *ahate* ~ *ate* 'duck' (Z *ãhãte*, B *agate*) < ANATE, *azeri* ~ *azari* 'fox' (medieval *Azeari*, a surname) < ASENARIU, *balea* 'whale' < BALLAENA, *bilau(n)* 'peasant' < VILLANU, old G, Sout *zabau* 'tablecloth' < SABANU, *ohore* ~ *oore* ~ *ore* 'honour' (Z *õhõre*) < HONORE, *li(h)o* 'flax' < LINU, *anoa* 'provisions' < ANNONA, *mo(e)ta* 'kind, sort' < MONETA 'coin'.

Very occasionally an early dissimilation leads to a different result, as in *arima* 'soul' < ANIMA.

Basque *ba(h)e* 'sieve' cannot derive directly from Latin VANNU; it must reflect a Romance development **vane* (note French and Occitan *van*).

Initial **n** and final **N** both yield *n* in modern Basque: *neska* 'girl', *nabar* 'multicoloured; grey', *ni* 'I', *neke* 'fatigue' (< NECE 'death'), *nota* 'stain'

(< NOTA 'mark'), *min* 'pain' (< **bin*), *lan* 'work', *astun* 'heavy', *gizon* 'man', *on* 'good'. The neutralized nasal which preceded a consonant also comes down as *n*: *mendi* 'mountain', *handi* 'big', *antzare* (and variants) 'goose' (< ANSERE), *landa* 'heath' (< Rom), *ongi* 'well' (adv.) (< **onki*).

Medieval toponyms often show ⟨*-nn-*⟩ for final **N** when the article follows: *Hurigurenna* (Araba 1025; the second element is modern *guren* 'edge'), *Urrenguenna* (Bizkaia 1070; same second element). The same graphy occurs elsewhere, for example in the suffix *-no*: toponyms *Egganno* (Bizkaia 1082; modern *Echano*) and *Helcanno* (Gipuzkoa 1025; modern *Elkano*); personal names *Enneco* (modern *Eneko*), *Amunna* (*amona* 'grandmother'), *Annaya* (*anaia* 'brother'). This suggests that original **N** was still pronounced long in the eleventh century, and did not merge with **n** until later. Aquitanian stems ending in a nasal show the same phenomenon: *Belexenn-is*, *Bihoscinn-is*, *Sembetenn-is*, *Seniponn-is*, (*Herculi*) *Ilunn-o* (*Andose*), all with Latin genitive *-is* or dative *-o*. Vowel-final stems do not show this: *Andere*, dative *Andereni*, and so on.

The Latin cluster MN was resolved in two different ways. On the one hand, we have *damu* 'regret' < DAMNU, showing simplification of the cluster. On the other, we have *done* 'saint' < *DOMNE < DOMINE and B *autono* 'September' < AUTUMNU; these last two may represent loans from a stage in which Latin MN had already become Romance **nn*.

On occasion *n* has been lost before a cluster or an affricate: *aintzin* 'front' has a common variant *aitzin*, and old B G *anztu* 'forget' is *aztu* today (cf. northern *ahantzi*). In some such cases it may be difficult to determine if the *n* is conservative or epenthetic: *ikatz* ~ *inkatz* 'charcoal', *itze* ~ *untze* (and also *iltze*, *ultze*) 'nail'.

Very occasionally, a final *n* is lost. Such words as *orain* 'now', *egun* 'today' and *ondoren* 'consequence' occur locally as *orai*, *egu*, *ondore*. The *-n* which marks past tense is systematically lost in certain eastern varieties, and old B sometimes loses the *-n* which marks the hortative use of the subjunctive. The loss of *-n* before the relational suffix *-ko* is discussed in Chapter 4.

In sum, Pre-Basque nasals were distributed as shown in Table 3.4.

Table 3.4 The Pre-Basque nasals

	-n-	
n-		-N
	-N-	

3.6 LATERALS

Pre-Basque had fortis and lenis laterals **L** and **l**, which contrasted only intervocalically; initially, only **l** occurred, and finally, only **L**. At some stage

intervocalic **l** was converted to the tap *r*, merging with the existing **r**. This change was categorical in most dialects, but sometimes fails to occur in Z, which has, for example, *solo* 'field', *olio* 'oil' and *zelü* 'sky' for common *soro* (< SOLU), *orio* (< OLEU), *zeru* (< CAELU). The disappearance of intervocalic **l** left **L** without a lenis counterpart, and **L** was then reduced to *l*. Consequently, intervocalic *l* in native words must in most cases continue Pre-Basque **L**: *alu* 'vulva, vagina' < **aLu*, *alor* 'cultivated field' < **aLor*, *ola* 'forge, foundry; hut, cabin' < **oLa*, *ilun* 'dark' < **iLun*, and so on.

The fortis nature of a final lateral is indicated by the common spelling of *zabal* 'wide' as *zaball* or *zavall* in medieval toponyms: *Monnio Zaballa* (1062), *Zaballa* (monastery, 1087), *super S. Johannem de Zavalla* (945), *Harizavalleta* and *Harrizavallaga* (1025), *Lacizaballa* (1067). The same occurs with other words, such as with *apal* 'humble': *Domna Apalla* (1079). Particularly striking is the stream name *Larçabaig* in Béarn, which can only continue the common Basque toponym **Larzabal* (*larre* 'pasture'), with the usual Gascon treatment of final geminate *-ll*.

Latin intervocalic -L- and -LL- were borrowed as Pre-Basque **l** and **L**, respectively, and develop in the same way: *begiratu* 'watch, look at' < VIGILIARE (with a folk etymology from *begi* 'eye'), *borondate* 'will' < VOLUNTATE, *(h)aizkora* 'axe' < ASCIOLA, *gura* 'desire' < GULA, *gereta* 'rustic gate, enclosure' < Pre-Basque **geleta* < *CLETA; *balea* 'whale' < BALLAENA, *makila* 'stick' < BACILLA, plural of BACILLUM, *bilau(n)* 'peasant' < VILLANU, *gaztelu* 'castle' < CASTELLU, *angelu* 'soil' < ANGELLU, *zela* 'saddle' < SELLA.

Initial **l** and final **L** have both come down into modern Basque as *l*: *lan* 'work', *lau(r)* 'four', *lasai* 'calm', *laket* 'be pleasing' (< PLACET), *lapitz* 'slate' (< LAPIS 'stone'), *lore* 'flower' (< FLORE), *lege* 'law' (< LEGE), *argal* 'slender', *ohol* 'plank' (< **onol*), *epel* 'lukewarm', *mutil* 'boy' (< PUTILLU), *apal* 'humble' (< AD VALLE). The neutralized lateral which occurred before a consonant also comes down as *l*: *alde* 'side' (< *alte*, preserved in the east), *elge* 'cultivated field', *albo* 'side', *giltza* 'key'.

The development of intervocalic **l** to *r* again produced some striking alternations: *euskara* 'Basque language', combining form *euskal-*, as in *euskaldun* 'Basque-speaker' and *Euskal Herria* 'Basque Country'; *gari* 'wheat', combining form *gal-*, as in *galgorri* 'variety of wheat' and *galbae* 'sieve for wheat'. With loans from Latin, such as *haizkora* 'axe' (< ASCIOLA), combining form *haizkol-*, it is easy to be sure of an original lateral. With native words, this is far more difficult, because there has been rule inversion; for example, Latin SAGMARIU 'pack-horse' is the source of eastern *zamari* 'horse', whose combining form is *zamal-*, as in *zamaldun* 'horseman'. (Such rule inversion must be distinguished from cases of dissimilation like that found in *txolarre* 'sparrow', from *txori* 'bird' plus *arre* 'grey'.) Hence, even when intervocalic *r* alternates with *l*, we cannot be sure that *l* was the original segment without independent testimony. Only occasionally is this available, as with *hiri* 'town', which everyone is happy to identify with the

element *Ili-* found in so many town names in the Roman period.

With toponyms, we are sometimes more fortunate: *Gebara* (Ara) is Ptolemy's *Gébala*; *Padura* (Ara) is *Padule* in 1025; *Zuberoa* (the province) is apparently the medieval *Subola*; *Erronkari* (the valley) is *Roncali* in 1085; *Eskaroz* (Nav) is documented in the medieval period as *Escaloce* and *Escaloz(e)*.

In sum, Pre-Basque laterals were distributed as shown in Table 3.5.

Table 3.5 The Pre-Basque laterals

	-l-	
l-		-L
	-L-	

3.7 RHOTICS

In all likelihood, Pre-Basque **R** and **r** were a trill and a tap, respectively, and they probably contrasted only between vowels. Elsewhere the contrast was neutralized (though old Z exhibited a contrast between the two rhotics before another consonant). In modern varieties, the result of the neutralization is usually a trill, at least in careful speech, even in Romance loans like *brontze* 'bronze' and *krabelin* 'carnation', but there is evidence, as we shall see, that this may not always have been so.

The ancient contrast survives today, in the form of trilled *rr* and tapped *r*, in all Spanish Basque varieties. Minimal pairs are abundant: *ere* 'also' and *erre* 'burn', *gori* 'fiery' and *gorri* 'red' (these words are not related), *zori* 'luck' and *zorri* 'louse', *gora* 'up' and *gorra* 'deaf' (definite form), *hura* 'that one' and *hurra* 'hazelnut' (definite form) (⟨h⟩ = zero in these varieties).

In word-final position, a rhotic is almost always a trill in modern Basque, but there is a handful of words in which a final rhotic surfaces as a tap when a vowel-initial suffix is added: *zer* 'what?', *nor* 'who?', *ur* 'water', *zur* 'wood', *or* 'dog', the stem *ber-* 'self, same', the northern forms *haur* 'this', *hirur* 'three' and *laur* 'four' (which have lost their final rhotic south of the Pyrenees), and a handful of loan words, such as *plazer* 'pleasure', *tirader* 'drawer' and *erretor* 'parish priest'. Thus, the more typical *hur* 'hazelnut' has the definite form *hurra*, but *ur* 'water' has *ura*.

In Z, the tapped *r* has been almost entirely lost in ordinary speech, though it is sporadically retained by at least some speakers. Hence spoken Z has *büü* 'head' (common *buru*), *dia* 'they are' (common *dira*), *hie* 'your' (familiar) (common *hire*) and so on. The sole exception is the definite form of 'water', which is *ura* elsewhere but *hurra* in Z. The *r* is none the less usually written in Z. Note that the aspiration in Z forms like *erho* 'kill' does not prevent this loss: the modern form is *eho*.

In L and LN, something interesting has been happening. First, the former trill has become everywhere a voiced uvular fricative, somewhat resembling French /r/ but noticeably scrapier. Hence words like *erre* 'burn' and *hurra* 'the hazelnut' are invariably pronounced with a uvular, as are words like *zahar* 'old' and *zakur* 'dog', with their neutralized final rhotics. This uvular realization is spreading to the tapped *r*, and hence words like *hura* 'that one' and *hari* 'thread' are now also frequently heard with uvulars, though not invariably so. Though little research has been done on this, we appear to have a clear case here of lexical diffusion, with the uvular spreading to ever more words and with ever greater frequency. It is notable, for example, that *hura* 'that one' is far more frequently pronounced with a uvular than *ura* 'the water', and the purely anecdotal evidence at my disposal strongly suggests that the frequency of the uvular realization is increasing steadily (Yvan Labéguerie, p.c.). If this change goes to completion, then the contrast between the two rhotics will be lost, but this has not happened yet, in spite of some statements to the contrary in the literature (though it may have happened for some speakers).

The reason it is not certain that the result of the neutralization has always been a trill is the evidence from word-formation. For example, *bihar* 'tomorrow' is of course pronounced with a trill today, and this trill is retained in formations like *biharretik biharrera* 'from tomorrow to tomorrow' (i.e., 'forever', a phrase found in the seventeenth-century writer Axular). However, this word forms a seemingly ancient derivative *biharamun* ~ *biharamon* 'the next day' (probably *bihar* + *egun* 'day'), in which the rhotic shows up as a tap, suggesting that the neutralized rhotic was anciently a tap, or perhaps even that the rhotic contrast was anciently maintained in word-final position (though such a conclusion would be entirely at odds with all other fortis–lenis contrasts).

The presence of rhotics in consecutive syllables is tolerated if both are taps: *erori* 'fall', *hirurak* 'all three', B *arerio* 'enemy'. Trill–trill and trill–tap sequences do not appear to occur within single morphemes. But the case of tap–trill sequences is interesting: there is clear evidence that the tap in this case has generally been dissimilated, most often to /l/. The word for 'ear' is *belarri* in the west, *beharri* or *begarri* elsewhere, and Michelena reconstructs **berarri*. Likewise, 'snow' is commonly *elur* but appears as *edur* or *erur* in places, chiefly in B, and Michelena reconstructs **erur*. And 'breast, chest' is usually *bular*, but R has *burar* or *budar*, suggesting **burar*. Another case is 'grass', commonly *belar*, but *berar* or *bedar* in B, pointing to **berar*, and still another is *ilar* 'pea', B *idar*, suggesting **irar*. Particularly interesting is 'silver'. This is commonly *zilar* or *zillar*, but *zildar* is attested, suggesting **ziLar*. But both *zidar* and *zirar* are found in the east and in B, pointing instead to **zirar*.

Very occasionally a rhotic is lost by dissimilation, as in *adore* 'courage' < ARDORE.

R, uniquely, has acquired word-initial *r* in loans from Romance, such as *rezibi* 'receive', *reina* 'queen', *repat(t)an* 'shepherd boy'. But most such words show a prothetic vowel in R, as elsewhere.

In sum, Pre-Basque rhotics were distributed as shown in Table 3.6.

Table 3.6 The Pre-Basque rhotics

-r-		
	-R	
-R-		

3.8 PALATALIZATION

The six segments *tt dd x tx ll ñ* occupy a place apart in the Basque phonological system, even though their distinctive status is no longer quite so sharply marked as it was some centuries ago. In all likelihood, these 'palatal' segments (as I shall call them) were already present in Pre-Basque, but we cannot be sure of that, since they never appeared in ordinary lexical items, except possibly as allophones of other coronal segments in predictable circumstances. They were certainly present in the language by the medieval period, and the Basque personal names recorded as early as the fifteenth century demonstrate both their presence and their distinctive role as clearly as one could hope for (Valle Lersundi 1933–1934). The byname *Gutia*, probably 'the Small', is occasionally written as *Guchia* as early as the thirteenth century; this clearly represents either *tx* or *tt*.

Broadly speaking, the palatal segments arise in two very different circumstances:

1 'Automatic' or 'phonological' palatalization occurs when an ordinary coronal consonant is preceded by the high front vowel *i*.
2 'Expressive' palatalization occurs when an ordinary segment is deliberately replaced by a palatal consonant, or, vastly less frequently, when a palatal consonant is added to the beginning of a vowel-initial word.

These two processes have nothing at all in common except the introduction of palatal segments. I begin with automatic palatalization.

It seems clear that, at some ancient stage of the language, some (but not all) ordinary coronal consonants underwent simple palatalization when preceded by the vowel *i* (possibly with additional conditions; see below). The process was automatic, and the resulting palatal segments were effectively mere allophones of the ordinary coronals. This process produced the following results: *t* → *tt*, *z* → *x*, *s* → *x*, *tz* → *tx*, *ts* → *tx*, *l* → *ll*, *n* → *ñ*.

Note that the other coronal consonants, *d*, *r* and *rr*, apparently did *not* undergo automatic palatalization, thus posing an interesting little phonological problem. The results of this process are still plainly evident today in B and G in the west and in Z in the east; the central dialects for the most part do not exhibit such palatalization, for reasons which I shall discuss below. Table 3.7 shows some examples; the first column gives the usual forms in the central dialects, which largely correspond to the written forms in the standard orthography. As some of these examples show, the *i* which induces the palatalization is often absorbed into the following palatal consonant if it is preceded by a vowel. In modern B and G, palatalization usually occurs only if the consonant in question is also followed by another vowel; in Z, it often occurs even in the absence of such a following vowel. Generally speaking, a coronal does not undergo palatalization if it is followed by another consonant:

Table 3.7 Automatic palatalization

	Central	B, G, Z
	ditu (Aux verb-form)	B G *dittu*
	gizon 'man'	B (regional) *gixon*
	isil 'silent'	B G *ixil*
	itsaso 'sea'	B G *itxaso*
	(h)ila 'dead'	B G *illa*
	baino 'than'	B G *baiño* ~ *baño*
	oin 'foot'	B G *oin*, Z *huiñ*
	oina 'the foot'	B G *oiña* ~ *oña*, Z *huiña*
But:	*iduri* 'seem'	B G *iduri*, Z *üdürü*
	dira 'they are'	B G *dira*, Z *di(r)a*
	irrintzi 'mountain cry'	B G Z *irrintzi*

izter 'thigh'	B G *izter* ~ *iztar*
gaizki 'badly'	G *gaizki*
ispilu 'mirror'	B G *ispillu*

Both B and G, however, exhibit a striking extension of this type of palatalization:

e(g)in da 'it has been done'	B G *eiñ dda*
indaba 'string bean'	B G *iñddaba*
(h)il da 'he has died'	B G *illa*

That is, if a preceding *n* or *l* undergoes automatic palatalization, then *d* also becomes palatalized, even though *d* is never palatalized directly by a preceding *i*; moreover, the sequence *ll dd* merges to *ll*.

Automatic palatalization is still productive today in B and G, and apparently also in some eastern varieties, but there are two or three clearly ancient words which sometimes or always fail to undergo it, notably *aita* 'father' and

maite 'beloved'. It seems very odd having lexical exceptions to a productive process, especially when those exceptions are not recent loans.

The central dialects today lack such automatic palatalization (as does R), but there are grounds for believing that these varieties formerly had it as well. Such Romance loans as *ollo* 'hen' (< Sp *pollo* or something related), *muño* 'hill' (from a Romance development of **bunno*) and *bañu* 'bath' (< Sp *baño*) appear in the central dialects as *oilo*, *muino* and *bainu* ~ *mainu* (which in fact represent the modern standard orthography). Since it is difficult to avoid the conclusion that such words were borrowed with Romance palatal consonants, most specialists believe that the central dialects must have undergone depalatalization, with the earlier palatal segment being unpacked into a sequence of *i* plus a coronal consonant, and the same analysis is applied to indigenous words like central *baino*, peripheral *baño*, 'but'.

Such unpacking is occasionally attested directly. I have heard the recent Spanish loan *pañuelo* 'handkerchief' pronounced in B as *painuelo*. Some vasconists think that the universal *aita* 'father', which often fails to undergo automatic palatalization, may result from earlier **atta*, with a palatal plosive.

Expressive palatalization involves the replacement of another consonant by a palatal or the addition of a palatal to the front of a vowel-initial word. The consonant so replaced is most often a coronal, but not invariably so. Examples: *zezen* 'bull', expressive form *xexen*; *zakur* 'dog' and *(t)xakur*; *zoko* 'corner' and *(t)xoko*; *gozo* 'sweet, delicious' and *goxo*; *hezur* 'bone' and *hexur*; *labur* 'short' and *llabur* ~ *txabur*; *bero* 'hot' and *bello*; *Peru* 'Peter' and *Pello* (and *Txeru*); *nabar* 'many-coloured' and *ñabar*; *nabo* 'turnip' and *ñabo*; *tapa* 'step' and *ttapa*; *zuri* 'white' and *(t)xuri*; *guti* 'not much' and *gutti* ~ *gutxi*; *tipi* ~ *tiki* 'small' and *ttipi* ~ *txiki*; *tente* 'erect' and *ttentte*; *tu* 'saliva' and *ttu* ~ *txu*; *popa* 'poop (of a boat)' and *(t)xopa*; *maingu* 'lame' and Z *txainkü*; *inurri* 'ant' and *txinurri*; *onil* 'funnel' and *txonil*; *ingude* 'anvil' and *txingure* ~ *xinguri*; *Domiku* 'Dominic' and *Txomin*; *Martin* 'Martin' and *Txartin* ~ *Matxin*; *Madalen* 'Madeleine' and *Maddalen* ~ *Matxalen*; *Santiago* 'James' and *Xanti*; and many, many others. (See Chapter 5 for a discussion of this process in personal names.)

It seems clear that the original function of this process was to create a diminutive or affectionate form, and this function still survives in many pairs today. In a number of instances, however, the palatalized version has become the unmarked form, and the original form either has been lost or survives as either an augmentative or a pejorative variant. For example, *zakur* remains as the unmarked word for 'dog' in places, with *(t)xakur* meaning 'little dog' (but never 'puppy'); elsewhere, though, *(t)xakur* is the unmarked form, and *zakur* means only 'big dog' or is absent altogether. The word for 'calf' is everywhere *(t)xahal*, and we can't even tell what the original sibilant was. An unusually striking case of differentiation involves the word *zori*, which formerly meant 'omen'. The original form survives today as *zori* 'luck' (as in the all-purpose greeting *Zorionak!* 'Best wishes!', 'Merry Christmas!' and so on), while its diminutive *(t)xori* is now the universal word for 'bird' (from the former practice of seeing omens in the flight of birds). In words with intrinsically diminutive meanings, the palatalized form has often displaced the original form more or less entirely, so that *ttipi* ~ *txiki* 'small', *gutti* ~ *gutxi* 'not much' and *xehe* ~ *txe* 'small' have all but driven *tipi* ~ *tiki*, *guti* and *zehe* ~ *ze* out of the language, except in formal styles. The same has happened with *etxe* 'house', whose earlier form *etze* or *etse* (both are attested) survives in only a few localities. It is reported that, in the Basque spoken in the United States, the force of palatalization has been lost altogether, and so, for example, 'corner' is *zoko* or *xoko* indifferently.

Today such expressive palatalization seems to be no longer productive, or at best only weakly productive. The productive manner of forming diminutives is by the addition of a diminutive suffix like *-txo* ~ *-txu*, *-tto*, *-xka* or *-ño*: *ama* 'mother', *amatxo* 'mummy, mommy'; *Jon* 'John', *Jontxu* 'Johnny'; *liburu* 'book', *liburuxka* 'booklet'; *iturri* 'spring', *iturriño* 'small spring'. Such use of diminutive suffixes is itself old in Basque, and the modern ones clearly derive from the palatalization of earlier diminutive suffixes like *-ko*, *-to*, *-so* and *-no*, all attested in the medieval period and some of them apparently attested in Aquitanian.

In the western and eastern dialects, the segment *tx* in particular is so frequent in non-expressive formations that its expressive function is now minimal.

Naturally, loan words are sometimes borrowed with palatal consonants present in the donor language: southern *Txina* 'China' (< Sp *China*), northern *Otrixia* 'Austria' (< Fr *Autriche*), southern *muñeka* 'doll, dummy' (spelled *muineka* in the standard orthography) (< Sp *muñeca*) and so on. Such loans reinforce the new status of the palatal consonants among the ordinary phonemes of the language.

3.9 VOWELS

Pre-Basque had just the five vowels **i e a o u**. So far as we can judge, it probably also had the modern set of diphthongs, **ai ei oi ui au eu**. These diphthongs were not distinguished from sequences of the corresponding vowels, but the point is that they counted as single syllables for such processes as aspiration assignment (see section 3.11), whereas other vowel sequences counted as two syllables.

Any vowel or diphthong could occur in any position in a word, as is still the case today. However, a monomorphemic word normally contained a maximum of one diphthong. At least some modern instances of diphthongs derive from the loss of intervocalic consonants. For example, *hodei* 'cloud' has final stress in Z and must therefore derive from a trisyllable **odeCe* or **odeCi*, and the same is true of *izei* ~ *izai* 'fir' and a few other words. (Word-stress in Z falls on the ancient penult.)

The Pre-Basque vowels have been astonishingly stable; indeed, for the most part they have remained unchanged for some 2,000 years, as witnessed by loans from Latin like *lege* 'law' (< LEGE), *putzu* 'well' (< PUTEU), *merkatu* 'market' (< MERCATU), *garau* 'grain' (< GRANU), *erika* 'heather' (< ERICA), B *zapatu* 'Saturday' (< SABBATU) and *ohore* 'honour' (< HONORE). Even the Aquitanian forms agree in vocalism with the modern forms, as in *NESCATO*, *ANDERE* and *CISON* (modern *neskato* 'little girl', *andere* 'lady' and *gizon* 'man'). Only a few developments have disturbed the Pre-Basque vowels, almost all of them either sporadic or confined to particular varieties.

In a first syllable, *e* is usually raised to *i* if the second syllable contains a high vowel and a third syllable exists. Above all, this raising affects non-finite forms of verbs, which regularly contain the prefix **e-*. Thus: *ibili* 'go about, be active' (< *ebili* (attested)), *ikusi* 'see' (< *ekusi* (attested)), *ipini* (and variants) 'put' (< *epeini* (attested)), *ikuzi* 'wash' (< *ekuzi* (attested)). (Cf. *ekarri* 'bring', *egosi* 'cook', *edeki* 'open', *egin* 'do, make', *ekin* 'continue', *edan* 'drink', etc.) The unraised forms are more frequent in the earliest texts than they are now, and still exist today in some varieties, above all in R, suggesting that this raising was a fairly recent process. It must have been recent if, as proposed in Chapter 4, *-n* class verbs like *egin*, *ekin* and *edan* are derived from earlier **egini*, **ekini* and **edani*.

An extremely rare process is 'vowel-doubling'. Latin VIMEN 'osier willow' and Romance *saco* 'sack' appear in the aspirating dialects as *mihimen* and *xahako* 'leather bag for wine'. There are no other certain examples of this: the oft-cited *ahaire* 'air, melody' can be explained as a compound of *aho* 'mouth' and the loan *aire* 'air'. The common cases like *zahar* 'old' appear to continue genuine ancient disyllables; the form *SA.HAR* is actually attested on the Lerga stele (see Chapter 6).

In Z, original *u* is fronted to *ü* [ʏ] in most circumstances: *düt* 'I have it', *lagün* 'companion', *lüze* 'long', *zü* 'you', *güzi* 'all' and so on, all of which retain *u* in other varieties. This fronting is prevented by a following tapped *r* or a following fricative *s*, but not by a following trill *rr* or by a following *ts*, *z* or *tz*. Hence Z has *gü* 'we' but *gure* 'our' (now usually *gue*, by the recent loss of intervocalic *r*), *hür* 'hazelnut' but *hur* 'water', *güzi* 'all' and *hüts* 'empty' but *uste* 'opinion', *ikhusi* 'see' and *busti* 'moist'. The fronting is also blocked by a following cluster of *r* plus coronal plosive: hence *urde* 'pig', *urthe* 'year', but *bürkhi* 'birch', *khürlo* 'crane', *ürzo* 'pigeon', *ürpo* 'pile of dung'.

This regular pattern has been considerably disturbed by mutual assimilations among *i*, *ü* and *u*. So, for expected **burü*, we find *bü(r)ü* 'head'; for expected **burdiña* we find *bürdüña* 'iron'; for expected **itsü* we find *ütsü* 'blind'; and there are many other such cases.

Before another vowel, this *ü* is unrounded to *i*. Thus, *ütsü* 'blind' plus the article yields *ütsia*. In R, *u* is fronted to *i* in the same circumstances: *buru* 'head', definite form *buria*.

In Z and R, original *au* becomes *ai* except before one of *r rr s ts*. Hence common *gau* 'night' and *gauza* 'thing' are *gai*, *gaiza* in these varieties.

Assimilations involving high vowels, especially of *i* to *u*, are found sporadically, especially in B in the west and Z and (most notably) R in the east. Thus, common *ikuzi* 'wash' appears as *ukuzi* in all these varieties, while R has *urun* and *uturri* for common *irun* 'spin' and *iturri* 'spring'.

The vowel *u* is sporadically fronted to *i* in all varieties. Thus, common *umore* 'humour' appears as *imore* in some varieties of HN and LN; *kuma* 'mane' is *kima* in varieties of G; *errun* 'lay (eggs)' and *gorputz* 'body' are *errin* and *korpiz* in R; *muku* 'mucus' (< MUCCU) has a widespread variant *muki*.

In Z, original *o* is often raised to *u* in circumstances which are far from clear, but especially before *n*: Z *gizun* 'man', *hun* 'good', *huñ* 'foot', in the face of common *gizon*, *on*, *oin*. This raising does not feed the fronting of *u* and must be more recent than fronting.

In a number of words B has *u* for common *i*: B *ule* 'hair', *uri* 'town', *uger* 'swimming', *ultze* 'nail', *urten* 'go out', *uzen* 'name' and so on, in the face of common *ile*, *(h)iri*, *igeri*, *iltze*, *irten*, *izen*. The reason for this is not known, but in most cases the *u* appears to be an innovation. It is not clear whether B G *izu* 'trembling', in the face of central and eastern *izi*, represents an instance of this or whether the eastern form derives from the fronting of *u* mentioned above.

B also exhibits some peculiarities involving the vowels *a* and *e*. The final *e* of the other dialects often appears as *a* in B: *lora* 'flower' (common *lore*), *ota* 'gorse' (common *ote*), *andra* 'woman' (common *and(e)re* 'lady'). And many varieties of B have a kind of rudimentary vowel harmony, in which *a* is raised to *e* after a high vowel in the preceding syllable of a phonological word: *zaldi*[j]*e* 'the horse' (common *zaldia*, with the article *-a*), *egune* 'the day' (common *eguna*), *erri bet* 'a village' (common *(h)erri bat*, with the indefinite article *bat*), *etorri de* 'he/she has come' (common *etorri da*, with auxiliary *da*). Finally, a few words which elsewhere have internal *e* followed by a liquid have *a* in B: B *barri* 'new', *baltz* 'black', *garri* 'waist', in the face of common *berri*, *beltz*, *gerri*.

The Pyrenean dialects, especially Sal and R, exhibit extensive syncope of vowels, and syncope is also attested in some eastern varieties of HN, especially in the writings of Joaquín Lizarraga, a native of Elkano. Examples: Sal *aingru* 'angel' (common *aingeru*); R *bedratzu* Sal Aezk Lizarraga *bedratzi* 'nine' (common *bederatzi*); Sal *denbra* R *tenpra* 'time' (common *denbora*); R *tupla* 'onion' (common *tipula*); Lizarraga *atra* 'go out' (common *atera*); Sal R Lizarraga *abre* 'animal' (common *abere*). Forms of auxiliary verbs are affected, so that, for example, the present-tense plural forms of *izan* 'be' appear locally as *gra*, *zra*, *dra*, the second showing a unique initial cluster.

Vowels in hiatus require particular attention. When the vowels in hiatus occur in the first and second syllables of a word, various processes have

applied to resolve the hiatus, and these processes are not everywhere the same. The aspirating dialects place an *h* between the vowels: *zahar* 'old', *zuhur* 'prudent', *mahats* 'grapes' (< **banats*), *ahal* 'ability', *ohore* 'honour' (< HONORE), *ihintz* 'dew' (< **ini(n)tz*), *aho* 'mouth', *behar* 'necessity', *bihar* 'tomorrow', *ziho* 'tallow' (< **zino*), *oihan* 'forest', *eihar* 'dry', *uh(a)in* 'wave' and so on. In the dialects which have lost the aspiration, things are more complex. Sometimes, particularly when the vowels are identical or both non-high, the hiatus is simply retained, or else the two vowels coalesce into one: hence *zahar* 'old' is *zaar* or *zar*; *ahal* 'ability' is *aal* or *al*; *ahate* 'duck' (< *ANATE) is *aate* or *ate*; *ihintz* 'dew' is *intz*; *behar* 'necessity' is *bear*; *bihar* 'tomorrow' is *biar*; *aho* 'mouth' is *ao*; and so on. Frequently, however, a consonant is inserted to break up the hiatus. This consonant is most often *g*, particularly in the west, but instead we sometimes find *b* (especially next to *u* or less commonly *o*), *d* or *r*. Thus, *ohe* 'bed' is often *oge*; *bihar* 'tomorrow' is often *bigar*; *uharte* 'land between rivers' is *ugarte*; *zuhur* 'prudent' is *zugur*; *uhalde* 'riverbank, river' is *ugalde* or *ubalde*; *aho* 'mouth' is *ago* or *abo*; *zahar* 'old' is *zagar*; Spanish *bahía* 'bay' is borrowed as B *baida*; *ahate* 'duck' appears as *agate* or *arate* in some varieties of B; common *me(h)ar* 'narrow' appears as *medar* in some varieties of HN and LN; *sahats* 'willow' (probably < **sanats*) is *sagats* or *sarats* in places; and so on.

Vowels which can form recognized diphthongs usually do so: *lau* 'flat' (< PLANU), *maiz* 'often' (< MAGIS 'more'), *maizter* ~ *maister* 'tenant' (and 'master shepherd') (< MAGISTER 'master'), *deus* 'anything' (< GENUS 'kind'). The sequences *ae* and *oe* often become *ai* and *oi*, as in *haitz* 'crag' (< **anetz*), *arratoi* 'rat' (< RATONE), but note the unusual case of *moeta* ~ *mota* 'kind, sort' (< MONEDA 'coin').

The word for 'dinner' is exceptionally interesting. The most conservative form is *auhari*, preserved in LN, and, with the usual fronting of *au*, as *aihá(r)i* in Z. R has *aigari*. Sal and some western varieties have *abari*, apparently with strengthening of intervocalic *u* to *b*, and a few varieties of G have strengthened this further to *apari*. Most other varieties, however, have *afari*; here the unusual *f* is usually thought to result from a coalescence of *u* and *h* at an ancient stage, though it might simply derive from *b* in the familiar way.

Another unusual case is Latin MANICA 'sleeve'; instead of the expected **mai(n)ka*, we find *mahanka* ~ *mahanga* in the north and *mauka* in the south.

If the hiatus occurs later in a word, *h* cannot occur, of course, and the hiatus must be resolved in another way or retained. So, for example, **azenari* 'fox' (< ASENARIU) is *azeri* in most of the country today, but B has (or formerly had) all of *azegari*, *azagari*, *azeari* and *azari*. The word for 'elbow', a compound of *uko* 'forearm' and *ondo* 'bottom', appears in old B as the regular *ukaondo*, but the form *ukondo* is nearly universal today.

An important diagnostic is the treatment of the sequence **ae*, which arises in a number of circumstances in Basque morphology. Broadly speaking, this

sequence is resolved to *a* in the west but to *e* in the east, as in the ergative plural affix **-aek*, which gives western *-ak* and eastern *-ek*.

When the article *-a* is suffixed to a vowel-final word, the outcome varies with region. After *i*, most dialects do nothing. In the Pyrenean dialects, the *i* is reduced to a glide, so that *zaldia* 'the horse' is [saldja]. After the *i*, B and G insert a glide [j], which often undergoes some kind of strengthening. Thus *zaldia* 'the horse' is variously [saldija], [saldiʒa] or even [saldiʃa] in these varieties, or sometimes [saldije], etc., with the raising of *a* after a high vowel. After *u*, B and G often insert a labial glide which falls together with *b* (an approximant intervocalically, recall), and so *burua* 'the head' may be *buruba* in these varieties. L and LN insert [j] in this position, and hence the local pronunciation is *buru*[j]*a*. The Pyrenean varieties have variously *bur*[jw]*a* or *bur*[j]*a*.

If the vowel is *e* or *o*, this is raised in B and in some varieties of G. Hence *lorea* 'the flower' and *astoa* 'the donkey' are realized as *loria* and *astua* (three syllables) in these varieties. L and LN do something surprising: they convert the mid vowel to a glide, and so they have *lor*[j]*a* and *ast*[w]*a*. Observe that only mid vowels become glides, and not high vowels. This pronunciation was recorded by Gavel (1920) and is still current. The Pyrenean varieties do the same, but recall that they also reduce high vowels to glides.

If the vowel is *a*, the result in most varieties is coalescence, and so *neska* 'girl' has the definite form *neska*. In Z, in which the stress-accent marks the ancient penult, these are distinguished as *néska* and *neská*, respectively. But B is very different: some varieties have the definite form *neskea*, with raising to *e*, while others have *neskia*, with raising all the way to *i*. This last form can readily be interpreted as a case of rule reordering, with the raising of *a* to *e*, formerly later than the raising of *e* to *i* in the same context, now reordered to feed the earlier raising rule. On this, see Jacobsen (1971), de Rijk (1970).

Interestingly, this raising does not affect plurals, and the plural of *neska* is *neskak* in B as elsewhere. Moreover, the lowering of final *e* to *a* in B, discussed above, has led to reanalysis, and the plurals of *lora* 'flower', *ota* 'gorse' and *andra* 'woman' in B are *lorak*, *otak* and *andrak*, in place of the expected **loriak*, **otiak* and **andriak*, and in contrast to common *loreak*, *oteak* and *and(e)reak*.

Since the stress-accent of Z and R regularly falls on the ancient penult, the oxytone words of these varieties must continue forms in which vowels anciently in hiatus have coalesced. Above I mentioned cases like Z *elíza* church' and *elizá* 'the church', but there are other cases in which an ancient intervocalic consonant has seemingly been lost: *odéi* 'cloud', *izéi* 'fir', *etsái* 'enemy', *eztéi* 'wedding' and so on. These must derive from forms like **odeCe*; the identity of the lost consonant is unknown, though it might have been *h*, especially if these words are ancient compounds or derivatives, since the language anciently allowed the aspiration in the third syllable in such

formations. Such cases are parallel to others like R *artzái* 'shepherd' (< **artzani*) and recent Z *aihái* 'dinner' (< earlier *aihari*). The word *ibai* 'river' may be another of these, but the word has not survived in the east.

All ancient participles (and other non-finite verb-forms) contain a prefix **e-*. This prefix has developed variously. Before a following *a* or *o*, it becomes *j*, and this is the sole source of *j* in the native lexicon: *jarri* 'put' (< **earri*), *jan* 'eat' (< **ean*), *jautsi* 'go down' (< **eautsi*), *jo* 'hit' (< **eo*), *joan* 'go' (< **eoan*), *josi* 'sew' (< **eosi*). Before *i* or *u*, **e-* is lost: *izan* 'be' (< **eizan*; inflected forms like *naiz* 'I am' show that the *i* is part of the root), *ukan* 'have' (< **eukan* < **edukan*), *utzi* 'leave' (< *eutzi*, an attested form). There is no certain case of **e-* before *e*, but a plausible candidate is the verb meaning 'go out', for which the forms *jalgi* and *elki* are in complementary distribution, suggesting **eelki*, with mutual dissimilation in one case and coalescence in the other.

If **e-* is followed by a consonant, then, if the second syllable contains a high vowel and there is a third syllable, **e-* is raised to *i-*, as described above: *ikusi* 'see' (< *ekusi*), *ibili* 'go about' (< *ebili*), *iduri* 'seem' (< **eduri*), *iritzi* 'opine' (< *eritzi*), *ikuzi* 'wash' (< **ekuzi*); note that some of the ancestral forms are attested, and indeed some are still in regional use today. (This same development is widely attested in nouns: *ipizpiku* 'bishop' (< EPISCIPU), *tipula* (and variants) 'onion' (< CEPULLA), regional *iguzki* 'sun' for common *eguzki*, and so on.)

Otherwise, **e-* generally remains today: *egin* 'do, make', *eman* 'give', *ekarri* 'bring', *etzan* ~ *etzun* 'lie down, recline', *errun* 'lay (eggs)', *entzun* 'hear', *egosi* 'boil, cook', *ebaki* 'cut' and so on.

These are the most widespread and regular developments, but in practice there is a good deal of regional variation in the forms of these verbs. There are a few problematic forms, like *irauli* 'turn over' and *ikasi* 'learn, study', which may have acquired their *i-* by analogy, and *irun* 'spin (yarn)', which may derive from **eirun*.

The diphthongs have mostly remained stable, but there is considerable fluctuation between *ei* and *ai*: *izai* ~ *izei* 'fir', *gai* ~ *gei* 'material', *bait-* ~ *beit-* (verbal prefix), *eztai* ~ *eztei* 'wedding'. There are sporadic cases of diphthongs being levelled or reduced, especially before clusters: regional *alki* for common *aulki* 'chair', regional *arroltza* for *arraultza* 'egg', regional *ardiki* and *urthiki* for *aurdiki* 'throw'. In B, *ai* is regularly reduced to *a* in a final syllable before *n*: B *ezpan* 'lip', *gan* 'top', in the face of common *ezpain*, *gain*, etc.

Very occasionally a vowel appears to have undergone spontaneous diphthongization. The word *laster* 'quick' is attested in this form from the thirteenth century, but today it is *laister* ~ *laixter* in the central dialects, and a variant *lauster* is recorded in the north. Common *(h)andi* 'big' is recorded in this form since the medieval period, but modern B has *aundi*, a form attested no earlier than the late eighteenth century. (See also section 3.15 for further instances of such unetymological diphthongs in loan words.) Disliking a vague appeal to 'expressive' or 'augmentative' diphthongization, Michelena (1977a: 488–489) suggests a possible origin. He notes that, at some earlier stage of the language, diphthongs were often reduced before clusters, so that, for example, the noun *itaun* 'confession' (and other senses) forms a verb *itaundu* 'confess', which is widely attested as *itandu* in early texts. This process effectively neutralized pairs like *au/a* and *ai/a* before a cluster. This reduction was then lost, and the etymological diphthong was restored before clusters, so that the verb just cited, for example, is again *itaundu* in the modern language. Cases like *aundi* may therefore represent an overgeneralization of the restored diphthongs.

See Chapter 4 for the unusually radical reductions of diphthongs which have occurred in finite verb-forms.

3.10 GLIDES

Pre-Basque had no phonemic glides, but, at an early stage, it acquired the palatal glide [j] in certain circumstances. This [j] arose from word-initial **e** followed by **a** or **o** (or perhaps rarely by **e**). This circumstance arose chiefly, and perhaps solely, in non-finite verb-forms, in which the prefix **e-* occurs in all ancient verbs, and hence the resulting segment (orthographic ⟨j⟩) occurs only in non-finite verb-forms among native words in the modern language. (There are three exceptional nouns (*jaun* 'lord', *jabe* 'owner', *jai* 'festival'), but these are perhaps derived from ancient participles.) Examples include *jo* 'hit', *joan* 'go', *jan* 'eat' and *jakin* 'know'; the only case in which *j* is followed by any segment other than *a* or *o* is the eastern *jin* 'come', an irregular contraction of *jaugin*. The eastern verb meaning 'go out', which appears variously as *elki* or as *jalgi*, may represent two different resolutions of an original **eelki*; no other case is known of a possible initial **ee-*.

Loan words like *joko* 'game' (< IOCU) and *jende* 'people' (< Sp *gente*) have provided additional instances of this glide.

This palatal glide has had a colourful development in Basque, and that development provides a remarkable instance of a diaphone. In two widely separated areas, the coast of Lapurdi and the south of Bizkaia, the diaphone |j| is still today a palatal glide [j]. In all the rest of the country, however, it has undergone strengthening to some kind of consonant. In most of Lapurdi and Low Navarre, it has become a voiced palatal plosive [ɟ]. In the larger part of Bizkaia, it has developed into a voiced palato-alveolar affricate [dʒ], somewhat resembling the English consonant of *judge*, but slightly more palatalized. In all of Zuberoa, |j| has become instead a voiced palatal-alveolar fricative [ʒ], somewhat resembling the French consonant of *juge* 'judge', but markedly more palatalized: it is perhaps best described as an alveolo-palatal. In the southern part of the Basque-speaking region of Navarre, and in a small part of Gipuzkoa, it is likely that the same

thing happened, but we have no direct record of [ʒ]: instead, what we find is the *voiceless* fricative [ʃ]. In this region, and only in this region, the diaphone |j| ceases to be a separate phoneme, for here it has merged with the existing fricative /ʃ/, notated ⟨x⟩, and, while all other Basques write *jan* for the verb meaning 'eat', regardless of their pronunciation, the Navarrese have traditionally written *xan*.

That leaves a large area right in the centre of the country, including virtually all of Gipuzkoa and the eastern part of Bizkaia, and here something truly remarkable has happened. In this region, the diaphone has also developed into a voiceless fricative, but one which is velar or uvular: [x] or [χ]. Hence, in this region, *jan* is [xan] or [χan], a pronunciation which is startlingly at odds with the voiced palatals heard in most of the rest of the country, and a surprising destiny for a segment which started out (recall) as [e].

The developments in Navarre and Gipuzkoa are, of course, strongly reminiscent of developments in late medieval Castilian, in which /ʒ/ devoiced to [ʃ], thereby merging with the existing /ʃ/, and then the resulting /ʃ/ moved to the back of the mouth to become the famous Spanish *jota*, pronounced [x] or [χ], in such words as *rojo* 'red' and *jarra* 'jar' (not to mention *Jérez*, the name which is the source of English *sherry*; the English word preserves the older Castilian pronunciation). It is therefore usually thought that we are looking at an example of a borrowed sound change, since such a shift in pronunciation is decidedly rare in languages generally and seemingly somewhat unnatural. If this is so, however, it is surprising to find the backing only in the centre of the Basque-speaking area, in an area surrounded by varieties of Basque which have not undergone the same shift, rather than in the west and south of Bizkaia and Araba, an area which has been in direct contact with Castile itself for centuries.

Be that as it may, we now come to the feature of the Gipuzkoan shift which is truly astounding. Recall that, in Navarre, |j| has simply merged with *x*. For reasons which I shall explain in a moment, it seems clear that the same merger occurred in Gipuzkoa. However, when |j| was later shifted to the back of the mouth, *original x remained palato-alveolar and did not shift!* In other words, we appear to have a case of *reversal of merger*: instances of original *x* did not shift, while instances of *x* derived from original |j| did shift. Thus, for example, *jo* 'hit' and *jan* 'eat' now have back fricatives, while original *ximist* 'lightning' and *xistu* 'saliva' are *tximist* and *txistu* in modern G.

Reversal of merger is supposed to be impossible, and the reader might be inclined to suspect that no merger ever took place at all in Gipuzkoa: that the two voiceless fricatives somehow always remained phonetically distinct. But there is evidence against this. The loan words *xaboi* 'soap' and *baxera* 'dishes' have become *jaboi* and *bajera* in modern G; *axola* 'attention, care' has become *ajola*, all with back fricatives, even though it is certain that they

formerly had instances of *x* not derived from |j|. Interestingly, *gixaxo* 'poor fellow' has become *gixajo*, in which only one of the sibilants has been backed, though this retains an expressive variant *gixaxo*, and there are a few other such cases in which original *x* has exceptionally been backed.

Michelena therefore proposes that the merger genuinely took place, but that instances of original *x* mostly retained the expressive function typical of palatal segments and thus resisted the backing process, while instances of *x* derived from |j| and those found in loan words lacked this expressive function and were hence subject to backing. In a few cases, as with *axola* and the second *x* of *gixaxo*, the expressive force of the segment was no longer strongly felt, and hence these too underwent backing. If this explanation is correct, it represents an extraordinary instance of a phonological change obstructed by functional factors.

The word *joan* 'go' has in places undergone striking developments. In part of the area where |j| is [ʒ], this verb has become *gan*, while in part of the area in which |j| is [x], it has become *fan*, both forms resulting from the coalescence of the initial segment with the following *o*, probably pronounced as a glide [w].

In the preceding section we examined the strengthening in B and G of glides arising at morpheme boundaries. Something similar happens morpheme-internally in R, which has *lexo* 'window' and *batixatu* 'baptize' for common *leiho* and *bateiatu*, and similarly for other such words.

3.11 ASPIRATION

Pre-Basque had a phonetic aspiration which was both frequent and prominent. The occurrence of that aspiration within words was subject to severe constraints which are largely, but not entirely, understood. In the Aquitanian materials at our disposal, the aspiration is regularly written (as *H*) in approximately the same positions in which it occurs today in those dialects of Basque which retain the aspiration.

The history of the aspiration since the Pre-Basque period has essentially been one of loss. It appears that such loss began first in the central dialects: our earliest records from Gipuzkoa and Navarre show no trace of it. In Bizkaia and Araba, the aspiration is still written throughout most of the medieval period in Basque personal names and place names recorded in Spanish texts, and it is exceedingly abundant in the place names listed in the *Reja de San Millán*, compiled in Araba in 1025. It is generally believed that Castilian itself still retained its /h/ at this time, and hence it seems safe to conclude that orthographic ⟨h⟩ in these Basque names genuinely represents an aspiration. By the time the Bizkaian dialect came to be written down in the late sixteenth century, however, the aspiration had disappeared from this variety, and the same is true of our single document recording the southern dialect of Araba.

North of the Pyrenees, in contrast, the aspiration has survived down to the present day, except that it has been very recently lost from ordinary speech along the coast of Lapurdi. The frequency of the aspiration is not everywhere the same: it is today exceedingly frequent in Z but noticeably less frequent in LN and L.

The occurrence of the aspiration is subject to the following constraints:

1 The aspiration can occur no later than the onset of the second syllable.
2 There can be no more than one aspiration in a word, even in a compound both of whose members bear an aspiration in isolation.
3 The aspiration can occur word-initially only on a syllable which otherwise begins with a vowel or a voiceless plosive followed by a vowel; in the second syllable, it can occur only if the syllable otherwise begins with a vowel, a voiceless plosive (not preceded by a sibilant), a liquid or a nasal (but never /m/).

Note that the aspiration never falls on a syllable otherwise beginning with a voiced plosive, a fricative, an affricate or /m/, nor does it fall on a voiceless plosive preceded by a sibilant. For purposes of syllable-counting, a diphthong in the initial syllable counts as one syllable, as shown by such northern forms as *aiher* 'propensity', *gauherdi* 'midnight', *oihan* 'forest' and eastern *auhari* 'dinner'.

Constraints 1 and 2 above do not apply either to the Aquitanian materials or to the place names from Araba recorded in the *Reja de San Millán*; all these materials involve exclusively proper names, and the difference appears to be that both elements of a compound name can bear the aspiration independently, something which is impossible today. This suggests that the two elements of a compound retained their individual identity in the earlier language to a greater degree than is the case today. Otherwise, the medieval materials generally show the aspiration in the same positions in which it is found in the northern dialects today. Our earliest connected texts show essentially the modern state of affairs: all the constraints are obeyed.

Modern Z has a tiny number of exceptions to constraint 1, all of them compounds or derivatives. Z *artolha* 'shepherd's hut' and *sarjalkhi* 'entry and exit' reflect the aspiration of the independent words *olha* 'hut' and *jalkhi* 'exit', and *a(r)akhói* 'carnivorous' contains the suffix *-k(h)oi* 'fond of', which, exceptionally, is stressed in Z. But the localized Z form *baranthailla* 'February' is absolutely extraordinary and has no known explanation. Other such cases sometimes found in print are usually purely etymological spellings or out-and-out errors.

With just a tiny handful of possible exceptions, the Basque aspiration is not etymological – that is, *h* does not continue an earlier segment, and the aspirated plosives are not distinct in origin from the unaspirated voiceless plosives. Instead, the aspiration originated as a suprasegmental feature, possibly one correlated with the position of the word-accent in Pre-Basque. It is

this suprasegmental origin which is chiefly responsible for the restricted distribution of the aspiration.

A simple demonstration of the non-etymological origin of the aspiration is provided by the treatment of loans from Latin, which regularly acquire aspirations reflecting nothing in Latin but required by the phonology of Basque. Table 3.8 lists some examples from the dialects retaining the aspiration:

Table 3.8 Aspiration in loan words

ARENA > *harea* 'sand'	ARMA > *harma* 'weapon'
HONORE > *ohore* 'honour'	ANATE > *ahate* 'duck'
APTARE > *hautatu* 'choose'	ASCIOLA > *haizkora* 'axe'
LEONE > *lehoin* 'lion'	SOLU > *sorho* 'field'
ANNONA > Z *anhua* 'provisions'	PIPER > Z *phiper* 'pepper'
BACILLA > *makhila* 'stick'	CERTU > *gerthu* 'certain'

As a rule, the position of the aspiration in loan words is rather consistent, but there are exceptions. Latin COLPU 'bay, gulf' is borrowed variously as *golko* ~ *golkho* ~ *kolko* ~ *kholko*, but variation to this extent is unusual. (The Basque word means also 'space between one's chest and one's clothes'.)

Note in particular the regularity with which an aspiration represents the former position of a lost intervocalic *n*.

In both native words and loan words, there are clear rules for the placement of the aspiration. Broadly speaking, these rules are as follows; they are presented in apparent order of priority, from highest to lowest priority:

1 With only a tiny handful of exceptions in older texts, a finite verb-form never bears an aspiration, save only for the presence of the prefix *h-* which marks the intimate second-person singular, as in *haiz* 'you are'. This probably reflects the lack of stress on these items.
2 The definite article *-a* is never preceded by the aspiration, nor indeed is any vowel-initial suffix.

   Examples: *khe* 'smoke', def. *khea*; *hil* 'dead', def. *hila*; *lo* 'sleep', def. *loa*; *ur* 'water', def. *ura*; *hur* 'hazelnut', def. *hurra*.

3 An /h/ must appear at the onset of the second syllable to separate two vowels in hiatus if these vowels cannot form one of the six recognized diphthongs *ai ei oi ui au eu*.

   Examples: *behar* 'necessity', *aho* 'mouth', *ihintz* 'dew', *ohol* 'plank', *liho* 'flax', *xehe* 'small', *iharduki* 'argue', *ihes* 'flight', *eho* 'grind', *ohoin* 'thief', *uholde* 'torrent', *ahal* 'ability', *behor* 'mare', *ohe* 'bed', *ohar* 'notice', *zuhur* 'prudent', *zuhaitz* 'tree', *lehoin* 'lion', *ahate* 'duck', *ohore* 'honour', *puhullu*

'fennel', *kaholla* 'provisions', *kaheka* 'owl'. (A virtually unique exception is *joan* 'go'.)

4 If the first two syllables begin with voiceless plosives which could in principle be aspirated, the first one usually gets the aspiration. (Virtually all of these are loan words, of course.)

Examples: *khako* 'hook', Z *khorpuz* 'body' (< CORPUS), *pharkatu* 'forgive' (< PARCERE), Z *phike* 'pitch' (< PICE), *phiper* 'pepper' (< PIPER), Z *thiti* 'breast', *khate* 'chain' (< CATEA).

But: *kanpo* 'outside', *kantu* 'song' (< CANTU), *pitika* 'kid', *tapa* 'step', *kaka* 'shit', *kalte* 'injury'.

Note that the pattern **kakha*, in which the second of two voiceless plosives is aspirated, is virtually unattested.

5 Otherwise, a voiceless plosive at the onset of the second syllable is normally aspirated, so long as it is not preceded by a sibilant:

Examples: *bethe* 'full', *bethi* 'always', *urthe* 'year', *lekhu* 'place' (< LOCU), *zathi* 'piece', *dithi* 'breast', *aphal* 'humble' (< AD VALLEM), *gathu* 'cat' (< CATTU), *artho* 'millet, maize', *aurthen* 'this year', *ethorri* 'come', *okher* 'twisted', *muthil* 'boy' (< PUTILLU), *bothere* 'power' (< POTERE), *berthute* 'virtue' (< VIRTUTE), Z *inkhatz* 'charcoal'.

But: *bake* 'peace' (< PACE), *zutik* ~ *xutik* 'standing up', *miko* 'crumb' (< Occ), *lapa* 'limpet, burdock' (< LAPPA).

The exceptional participles like *hartu* 'take' and Z *heltü* 'arrive' (common *heldu*) probably represent late formations.
Some bisyllabic participles in *-tu* and *-ki* fail to take the aspiration, doubtless by analogy with the very much larger group of participles with three or more syllables, in which these endings cannot possibly be aspirated.

6 If none of the above applies, an initial voiceless plosive is usually aspirated. (Virtually all of these are loan words, and this rule has a large number of exceptions.)

Examples: *khako* 'hook', *khe* 'smoke', *khide* 'companion', *phesta* 'fiesta', *phutzu* 'well', *phixka* 'small amount', *phiko* 'fig', *theina* 'ringworm', *thema* 'obstinacy', *thiti* 'breast', *thona* 'stain', *thu* 'spit'.

But: *kaiola* 'cage', *kontra* 'against', *korapilo* 'knot', *portu* 'harbour', *porru* 'leek', *taula* 'plank', *tamal* 'misfortune'.

7 If none of the above applies, a liquid or nasal (other than /m/) at the onset of the second syllable is aspirated if the word is not vowel-initial. (This rule has so many exceptions it is barely worth stating.)

Examples: *senhar* 'husband', *zilhar* 'silver', *berhatu* 'augment', *belhar* 'grass', *belhaun* 'knee', *gurhi* 'fat, butter', *lanho* 'mist', *zorhi* 'ripe', *zelhai* 'plain' (geog.), *sorho* 'field', *urrhe* 'gold', Z *uñhũ* 'onion' (< Fr *oignon*)

Such forms, which are very numerous, suggest that an intervocalic liquid anciently formed part of the preceding syllable, and not of the following syllable.

But: *sare* 'net', *sari* 'prize', *larre* 'pasture', *galant* 'elegant', *bele* 'crow', *gona* 'skirt', *bainan* 'but'.

8 In all remaining circumstances, the aspiration /h/ is potentially contrastive, and may occur or not. Minimal pairs are possible. The /h/ is fully contrastive word-initially or between two vowels which can form a recognized diphthong; after a liquid or a nasal it may or may not occur but has little or no contrastive value.

Examples: *ari* 'busy', *hari* 'thread', *sehi* 'boy, child', *sei* 'six', *haur* 'child', *ahur* 'palm of the hand', *gai* 'material', *nahi* 'desire', *hur* 'hazelnut', *ur* 'water', *erhi* 'finger', *eri* 'illness', *hala* 'thus', *ala* 'or', *alhaba* 'daughter', *urrhe* 'gold', *hurren* 'nearest', *gorri* 'red', *inhurri* 'ant'.

In loan words, a prothetic vowel counts as a syllable for purposes of assigning aspiration, and hence words like *errota* 'wheel' (< ROTA), *arropa* 'clothing' (< Sp *ropa*) and *ezpata* 'sword' (< SPATHA) never contain aspirated plosives.

When a compound is formed from two words, each of which is aspirated in isolation, the results are variable. For example, the word for 'cemetery' is a compound of *hil* 'dead' and *herri* 'town', but **hilherri* is impossible, and the compound is variously *hilerri* or *ilherri*. The very rare spellings like *hilebethe* 'month', found in a few early authors, are almost certainly etymological only; in the example, the second element is *bethe* 'full'.

A compound does not necessarily respect the position of the aspiration in its elements. The word for 'siblings', a compound of *haur* 'child' and *-ide* 'fellow' is *haurride* in L but *aurhide* in Z, by rule 7. The word for 'moon', a compound of **hilV-* 'moon' and *argi* 'light', may appear either as *hilargi* or (more usually) as *ilhargi*, again by rule 7. (Cf. *hilabete* 'month', with *bethe* 'full'.)

The two words *aita* 'father' and *maite* 'beloved' never have aspirated plosives. Since these two words also fail to undergo automatic palatalization, they must have an unusual history. A favourite suggestion is that they continue original geminates: **atta*, **matte*. (This proposal is due to Nils Holmer, cited in Michelena 1951.)

The participle *jakin* 'know' never has an aspirated plosive, but of course this derives from **eakin*, in which the plosive was in the third syllable.

Interjections and phonaesthetic items are exempt from the ordinary rules

governing aspiration, except for the ban on multiple aspirations: Z *hupa!* 'let's go!', *hapataka* 'galloping noise'.

A crucial issue is the phonological status of the aspiration. In Z, it is indisputably contrastive: there are a number of minimal pairs with and without the aspiration, or with the aspiration in different places. Table 3.9 lists some examples from that dialect; the distinctive stress-accent of Z is written except in cases in which it is not clearly recorded, and some of these examples are taken from seventeenth-century texts but appear to be no longer current.

Table 3.9 Aspiration and minimal pairs in Zuberoan

har 'worm', *har* 'take' (radical) vs. *ar* 'male'
há(ṙ)i 'to that one' (Dat) vs. *á(ṙ)i* (progressive auxiliary)
hézi 'raise, domesticate' vs. *ézi* 'because, for'
auhen 'lamentation' vs. *auen* 'who has you' (relative)
hála 'thus' vs. *álha* 'eat' vs. *ála* 'or'
ha(ṙ)an 'valley' vs. *a(ṙ)hán* 'plum'
belhar 'grass' vs. *belar* 'forehead'
é(ṙ)hi 'finger' vs. *é(ṙ)i* 'sick'
ókher 'twisted' vs. *óker* 'belch'
húrte 'rainy spell' vs. *úrthe* 'year'

Observe that the contrastive value of the aspiration extends to voiceless plosives in this variety. Nevertheless, the functional load of the aspiration is not great, particularly on voiceless plosives: much more typical are cases like *khána* and *kána*, which are merely variants of the word meaning 'stick, cane' (a loan from Romance).

In LN and L, the contrastive value of the aspiration is much more marginal, with only a small number of minimal pairs, almost all of them listed in rule 8 above. In these varieties, the aspiration or lack of it on plosives is never contrastive, and using the 'wrong' value for the aspiration merely produces an atypical pronunciation. Even *h* is not infrequently facultative in LN and L: for example, *horri* and *orri* merely represent variant pronunciations of the word for 'leaf', and *anhitz* and *hanitz* are alternative pronunciations of the word for 'lots of'.

The absolute frequency of the aspiration (both *h* and the plosives) is much greater in Z than in LN or L, and, even though this variety has more minimal pairs than the others, it follows my eight rules far more consistently than do LN and L. Vasconists have often assumed that Z is conservative, that it retains the ancient state of affairs more reliably than LN or L, but this is open to question. There seems every reason to suppose that Z has in some cases extended the aspiration to words which did not historically have it. In particular, Z seems to have extended the aspiration to virtually every single monosyllable which can bear it. Hence Z forms like *hur* 'water', *hor* 'dog', *hun* 'good' and *huñ* 'foot', in the face of common *ur*, *or*, *on* and *oin*, very likely represent innovations.

Among the very few cases in which *h* may continue an earlier velar plosive are *harri* 'stone', which may be related to the substrate element **karr-* 'rock, crag' attested in Romance and Celtic, and the three demonstratives *hau(r)*, *hori*, *hura*, which appear as *kaur*, *kori*, *kura* or as *gau(r)*, *gori*, *gura* in the Pyrenean dialects. No one is quite sure what to make of these cases.

For further discussion of the aspiration, see Michelena (1950c, 1951, 1977a: Ch. 11).

3.12 CONSONANT CLUSTERS

Native words of any antiquity never contain initial clusters. Medial clusters in indigenous words are chiefly of the types (1) sibilant + plosive or (2) liquid or *n* + plosive, affricate or sibilant. The rare final clusters are mostly of the type liquid or *n* + affricate; at least some of these clearly derive from syncope, as with *beltz* 'black' < **beletz*. A handful of other types occur in words of some seeming antiquity: *rl* in *arlo* 'field' and *erle* 'bee' (both probably bimorphemic in origin), *rn* in *ernai* 'alert', *sn* in *esne* 'milk' (a contraction of earlier *esene*), and the extraordinary *sm* in *asmo* 'idea', an ancient loan from Romance. All these clusters generally survive today, with only three developments of note.

First, most western and central varieties have lost the affricate/fricative contrast after a liquid or *n*. Many eastern varieties still permit the seemingly ancient distinction between *rz* (as in *arzulo* 'cave') and *rtz* (as in *artzain* 'shepherd'); other varieties permit only *rtz*, and similarly for other such sequences.

A second, and particularly striking, development in Basque is the change of *rz* into *s* and the associated change of *rtz* into *st*. These two changes are most frequent in the west and least frequent in the east, but nowhere are they either categorial or unattested. Thus, eastern *urzo* 'pigeon, dove' = western *uso*; eastern *bortz* 'five' = western *bost*; eastern *-tarzun* '-ness' (a combination of *-tar* and *-zun*) = western *-tasun*; eastern *ortzegun* 'Thursday' = western *ostegun*; eastern *(h)ertze* 'intestine' = western *este*; eastern *bertze* 'other' = western *beste*. (But some of these 'western' forms are also found in varieties of Z and R.) Note also western *mesede* 'grace, favour' < Sp *merced*. The toponym *Satrústegui* (Nav) is attested in the fourteenth century as *Santurcegui* and clearly derives from *sancti Georgi* 'Saint George'; cf. modern *Santurtzi* (Sp *Santurce*) in Bizkaia. The toponym *Elosu* (Ara) derives from *elorri* 'hawthorn' + *-zu* 'full of'. The development is categorical if an apical *s* is nearby. Thus Sal has *-(t)arzun* for most abstract nouns but *osasun* 'health'. The widespread modern *sasi* 'bramble', attested as *'çarci* vel *sassi'* in the seventeenth-century writer Oihenart, probably derives from **zarzi* or **sarzi*.

The third development is the voicing of plosives after *l* or *n* (but not after *r*). This voicing applies categorically in all varieties except Z and R in the

east, which fail to undergo it. Thus, as explained in section 3.3, we have common *denbora* 'time' but eastern *t(h)enp(o)ra* < TEMPORA 'times' and common *aldare* 'altar' but eastern *alt(h)are* < ALTARE. Native words show the same developments: common *ongi* 'well' (adv.) is R *onki*, Z *hunki*, and common *alde* 'side' is eastern *alte*. One or two word-forming suffixes usually fail to undergo this voicing, notably *-kizun* 'future action', *-koi* 'fond of' and *-kor* 'tending to': *eginkizun* 'task' (*egin* 'do'), *ibilkoi* 'restless' (*ibili* 'go about'), *emankor* 'fertile' (*eman* 'give'). A couple of words in B exceptionally fail to undergo the voicing: B *denpora* 'time', *ilinti* 'firebrand'. Otherwise, the occurrence of a voiceless plosive after *l* or *n* is generally a reliable indication that the word containing it is of late origin in Basque.

The unique variation found in the verb 'say', western *esan* but eastern *erran*, perhaps derives from an unusual cluster, possibly **esran*.

A handful of medial three-consonant clusters exists: *andre* 'lady' (< *andere*), *baldres* 'slovenly' (a loan from Old Spanish), old B G *anztu* 'forget' (now reduced to *aztu*), R *arska* 'crib, manger' (common *aska*), R *arsto* 'donkey' (common *asto*), R *ainzto* ~ *aizto* 'knife'. Most of these derive either from word-formation or from syncope, and they tend strongly to be reduced, with R being notably conservative in retaining a number of them unreduced.

Place names of non-Basque origin and containing clusters not usual in Basque sometimes have local pronunciations in which the clusters have been metathesized to produce more familiar clusters. Thus, the Bizkaian cities of *Gernika* and *Zornotza* have localized versions *Gerrinke* and *Zorrontza*, and *Afrika* is often *Apirka* for older speakers.

For the treatment of impermissible clusters in loan words, see section 3.15.

3.13 WORD-ACCENT

The history of the word-accent is complex and controversial, and its study has been obstructed by the fact that, until very recently, we had little in the way of reliable information on the subject for the majority of varieties. Even some of the best descriptions of regional varieties fail to note the presence of the word-accent, and more than a few have gone so far as to deny the existence of the word-accent, a position which is remote from the truth. A further complication is that, in connected speech, the word-accents are largely overridden by the phrase-accent, so that words lose their individuality; this is true of all but the eastern dialects, in which the stress-accent of a word generally remains intact.

Earlier vasconists were inclined to think that the eastern accent-type represented the most direct continuation of the ancient state of affairs, largely because of the different development of initial and medial consonants in Basque, but few would accept such a view these days. Today, it is agreed by

probably all scholars that the stress-accents of the eastern varieties are late innovations, but there is no consensus on the western pitch-accent. In an important paper which has unfortunately never been published, Jacobsen (1975) argues that this pitch-accent is also an innovation, deriving chiefly from the coalescence of vowels in hiatus and then extended analogically in certain circumstances. This attractive idea accounts rather well for a striking fact about the western word-accent: while only a minority of non-plural forms are accented, all plurals are accented. Since many specialists believe that the plural endings generally result from the reduction of longer sequences, such as ergative plural *-ek* ~ *-ak* from something like **-agek* (see Chapter 4), and since it is well established that pitch-accents can arise from vowel coalescence, Jacobsen's analysis is a very economical one which deserves more attention than it has so far received, no doubt because of its unpublished status. It has, however, been further developed by Grundt (1977, 1980).

Martinet (1950, 1955), as part of his account of the frequent word-initial vowels of Basque, has proposed that some ancient stage of Basque, possibly the Pre-Basque of Michelena's reconstruction, must have had a stress-accent which fell on the initial syllable of a word. This proposal has attracted little support. It does, however, have the advantage of explaining why, in loan words in which the first two syllables have voiceless plosives, it is normally the first one which receives the aspiration in those dialects retaining it: we find the type *khaka*, but not the type **kakha*.

Michelena (1957–1958, 1972a, 1977a) does not reject Martinet's idea outright, but he suggests that, if valid at all, it must have applied to a stage of the language much earlier than the Pre-Basque of his own phonological reconstruction. Michelena, in contrast, argues that Pre-Basque had a word-accent, of unstated nature but most likely a stress-accent, which regularly fell on the second syllable of a polysyllabic word, except in compounds. In this analysis, the presence of the aspiration was closely associated with the location of this word-accent; recall that the aspiration most commonly falls on the second syllable of a word when that syllable is capable of bearing the aspiration. This view has the obvious advantage of explaining why the aspiration does not normally fall later than the second syllable and why there is normally only one aspiration per word. On the other hand, it has some difficulty with the word-initial aspiration found in a substantial number of cases, and it provides no basis for understanding the western pitch-accent. An interesting consequence is that the Type IV accentual system described above must be an almost undisturbed continuation of the Pre-Basque accent, while all other accentual systems must be innovations. Michelena goes on to propose that the penultimate stress accent of some eastern varieties results from the reanalysis of three-syllable forms, in which the accent, lying on the second syllable, could be equally regarded as occurring on the second syllable from the beginning or on the second syllable from the end.

In spite of some difficulties, Michelena's account is still perhaps the majority view among vasconists.

In two recent examinations, Hualde (1992, 1995) rejects all these interpretations in favour of a very different view. Dismissing the various appeals to aspiration, initial vowels and vowel coalescence as either irrelevant or of only minor significance, he notes that the western varieties exhibiting the pitch-accent system called Type I frequently retain unchanged the word-accent of Latin and early Romance; for example, the B variety of Getxo has *dénpora* 'time' (< TEMPORA), *ántzar* 'goose' (< ANSER), *sékula* 'never' (< SAECULA), but *autóno* 'September' (< AUTUMNU), *doméka* 'Sunday' (< DOMINICA) and *makílla* 'stick' (< BACCILLA). All these words belong to the accented class today, which is understandable for the first group but mysterious for the second: with an accent on the second syllable, they would have been regular in Michelena's system, and hence they should have developed into unaccented words today. (In a Type I accentual system, monomorphemic words are normally unaccented unless they are loan words retaining the marked accentual position of the source language.)

Examining a range of data, including some that have only recently been made available by published descriptions, Hualde finally concludes that Pre-Basque must have had a regular word-final accent. On this basis, he is able to interpret the western systems as resulting from the retraction of the accent towards the beginning of the word.

That is the state of play today. Each proposal has certain advantages and certain drawbacks, and there is no consensus. The word-accent remains the one aspect of Pre-Basque phonology on which we are still in the dark.

One final observation: Basque versification utterly ignores the word-accent in all varieties, even in Z, with its strong stress-accent. (A treatment of versification is beyond the scope of this book; see the various writings of Jean Haritschelhar, most notably Haritschelhar (1969).)

3.14 SUMMARY OF PHONOLOGICAL CHANGES

Here I present a brief summary of the principal phonological changes leading from Pre-Basque to the modern Basque dialects. With only a few exceptions, it is not possible to determine the order in which the various changes occurred, and I therefore group them by the same categories used above. The Pre-Basque system which underwent the changes was as shown in Table 3.10. The aspiration was phonetically present but not phonemic. Initial **d-** occurred only in finite verb-forms; **p** was rare at best; final **-t** and **-k** may perhaps have occurred in suffixes (only). The vowels were **i e a o u**, and the recognized diphthongs were probably, as today, **ai ei oi ui au eu**.

Plosives. The contrasting fortis and lenis plosives were reinterpreted as distinctively voiceless and voiced, respectively. The new voicing contrast was

Table 3.10 The Pre-Basque phonological system

	-b-, -d-, -g-	
	-z-, -s-	
b-, (d-), g-	-n-, -l-, -r-	
z-, s-		-tz, -ts
n-, l-	-p-, -t-, -k-	-N, -L, -R
	-tz-, -ts-	
	-N-, -L-, -R-	
i, e, a, o, u		

extended to word-initial position, and initial *t* and *d* became possible. Except in the eastern dialects, all plosives were voiced after *n* or *l*. The consonant *b* sometimes became *m* in initial position and sometimes developed to *f*, especially in medial position; initial *b* was very frequently lost before *o* and occasionally before *u*. At some stage, final *t* and *k* became possible.

Sibilants. The contrasting fortis and lenis sibilants became affricates and fricatives, respectively, which is quite possibly what they had always been anyway. In western and central dialects the affricate/fricative contrast was lost after a nasal or a liquid. Final sibilants became possible, so that the affricate/fricative contrast was extended to final position. In B, the *s/z* and *ts/tz* contrasts were lost in very recent times.

Nasals. Pre-Basque **n** was lost intervocalically, and then **N** was reduced to *n*, merging with the surviving instances of **n**. This loss of **n** led to the introduction of nasal vowels, which survive today only in the eastern dialects. Elsewhere nasalization was either lost or reinterpreted as a following *n*. The hiatus resulting from the loss of **n** was often resolved by the introduction of *h* or another consonant. Initial **b** frequently became *m*, especially before another nasal, and medial **nb** was frequently reduced to *m*; these developments, together with loans from Latin, led to the introduction of *m* into Basque.

Liquids. Lenis **l**, when intervocalic, developed to *r*, merging with existing **r**; then fortis **L** was reduced to *l*, merging with surviving instances of **l**. Fortis **R** and lenis **r** have generally survived as a trill *rr* and a tap *r* down to the present day in most varieties. In Z, however, intervocalic *r* has recently been lost, leaving Z with only one rhotic; in L and LN, *rr* has developed into a uvular fricative, and this uvular is now in the process of being extended to *r* by lexical diffusion.

Palatalization. Most coronals were formerly palatalized automatically after

i and before a following vowel; this palatalization remains today in the peripheral dialects but has been lost in the central ones. Otherwise, the palatal segments did not anciently occur in lexical items, but were found only in expressive variants of lexical items, a function which is still prominent but no longer productive. With the loss of some of the earlier unmarked forms, and with the effect of loan words, the palatal segments must now be regarded as ordinary consonant phonemes.

Vowels. The ancient five-vowel system is retained in most modern varieties; the contrastive nasalization introduced by the loss of intervocalic *n* has been lost everywhere except in Z and R, which retain distinctive nasal vowels. In Z, most instances of original *u* have developed into a front rounded vowel *ü*, with some consequent mutual assimilation of *i*, *ü* and *u* within words, and *o* has often been raised to *u*. In certain circumstances, mid vowels have been raised or lowered, and diphthongs have occasionally been levelled or reduced, especially before clusters.

Glides. Word-initial *e* became *j* [j] initially before *a* or *o*; this [j] has in most varieties been strengthened to some kind of consonant. Prevocalic *i* and *u* (and sometimes also *e* and *o*) have in some varieties been converted into glides [j] and [*w*], which in places have been strengthened to consonants.

Aspiration. The ancient aspiration was never either segmental or contrastive. It has been retained only in the northern varieties, and has probably been extended in Z; elsewhere, the aspiration has been lost, though it is attested in personal names and place names in medieval Bizkaia and Araba.

Word-accent. There is as yet no consensus on this. Martinet's word-initial accent has found no support; Michelena's second-syllable accent is perhaps the favourite interpretation; Hualde's final accent is still awaiting judgement.

Consonant clusters. Little has happened to the Pre-Basque clusters apart from the voicing of plosives after *n* and *l* in all but the eastern varieties. Medial three-consonant clusters have usually been reduced to two-consonant clusters. Loan words have introduced new clusters not formerly possible, especially in initial position.

Metatheses. Instances of metathesis, while always sporadic, are by no means rare in the history of Basque. Latin BENEDICERE 'bless' should have been borrowed as **benedikatu*, but this must have metathesized to **bedenikatu* to yield the attested *bedeinkatu*. Basque *mukuru* ~ *mukurru* 'height' is apparently borrowed from Latin CUMULU, with metathesis of the earlier **kumuru*. The verb meaning 'seem, appear' occurs both as *iduri* and *irudi* in various parts of the country, and no one is sure which is the more

conservative form. The western word for 'butterfly' is given only as *mitxeleta* in Azkue's dictionary, but the usual form today is *tximeleta*. Many other such cases could be adduced. An extraordinary case is the word for 'swallow' (the bird). This is *ainara* ~ *enara* in much of the country, but *elae* in some varieties of B. The first of these implies an original **aiNala* ~ **eNala*, but the second equally implies **eLana*. One form or the other, then, must result from a metathesis of the nasal and the lateral accompanied by an exchange of fortis and lenis character between these two segments.

Compound verbs. Compound verbs are common in Basque at all periods, and they have frequently undergone complex and irregular developments leading to substantial regional variation in form. A verb meaning 'wait for' appears variously as *itxeden*, *itxedan*, *itxedon*, *itxadon*, *itxaron*, *itxaran*, *itxodon*, *itxogon*, *itxon* and *itxoin* (at least), and this is thought to be an ancient compound of *hitz* 'word' and **edun* 'have'. The widespread verb *eskaini* ~ *eskeini* 'offer' looks very much like a compound of *esku* 'hand' and *ipini* ~ *ibeni* 'put'. Another verb for 'wait for' is *iguriki* ~ *iguruki* ~ *eguriki* ~ *eguruki*, which is probably a compound of *egun* 'day' and *eduki* 'hold'.

3.15 TREATMENT OF LOAN WORDS FROM LATIN AND EARLY ROMANCE

Latin nouns are usually borrowed in their accusative forms: *ahate* 'duck' < ANATE, *ohore* 'honour' < HONORE, *bake* 'peace' < PACE, *ingude* 'anvil' < INCUDE, *errege* 'king' < REGE, *lore* 'flower' < FLORE. But there are a few instances of borrowed nominatives: *gorputz* 'body' < CORPUS, *lapitz* 'slate' < LAPIS 'stone', *maizter* 'master shepherd' < MAGISTER, *apaez* (and variants) 'priest' < ABBAS 'abbot'. Adjectives are borrowed in the accusative singular masculine/neuter: *ziku* 'dry' < SICCU, *xahu* 'clean' < SANU 'healthy'. A rare nominative is *bortitz* 'strong, violent' < FORTIS.

Verbs are normally borrowed as the past participle, the result being a Basque perfective participle, commonly used as a citation form: *okupatu* 'occupy' < OCCUPATU, *perekatu* 'caress' < FRICATU, *aditu* 'hear' < AUDITU, *goitatu* 'look after, pay attention to' < COGITATU. Occasionally the Basque participle is built upon the Latin or Romance infinitive, particularly when the original participle is irregular: *eskiribatu* ~ *iskiribatu* ~ *izkiriatu* 'write', from SCRIBERE, not from SCRIPTU. A unique instance of a borrowed finite verb-form is the eastern participle *laket* 'be pleasing' < PLACET (re-formed as *laketu* in places). The Latin infinitive HABERE 'have' is the source of *abere* 'animal, cattle', but this merely reflects a widespread Romance development.

Latin /a e i o u/, both long and short, are borrowed as Basque *a e i o u*, as shown in Table 3.11.

As can be seen, Basque does not diphthongize Latin short /e/, /o/. In loans from Romance after the Romance diphthongization of these vowels, the

Table 3.11 The Latin vowels in Basque

FĀGU > *bago* 'beech'	LACU > *laku* 'lake'
LĒGE > *lege* 'law'	CERTU > *gertu* 'certain'
LĪNU > *liho* 'flax'	PICE > *bike* 'pitch'
FLŌRE > *lore* 'flower'	ROTA > *errota* 'wheel'
PLŪMA > *luma* 'feather'	LUCRU > *lukuru* 'avarice'

diphthong is usually reduced: *erregu* 'request' (< *ROGU; cf. Sp *ruego*), *leku* 'place' (< LOCU). Medieval documents, notably the *Reja de San Millán* (1025), reveal that diphthongized Romance forms and non-diphthongized Basque forms of toponyms often existed side by side along the boundary between the languages. However, the Basque vowel sequences *uo* and *oe*, arising in word-formation, were taken into Romance as the diphthong /ue/: for example, the medieval toponyms *Zalduhondo* (Ara) and *Zuloeta* (Nav) are modern Spanish *Zalduendo* and *Zulueta*.

We occasionally find an unexpected *ai* in place of *a* in a first syllable: *maingu* 'lame' (< MANCU), *haizkora* 'axe' (< ASCIOLA), *maindire* 'sheet' (< MANTILE), *aingeru* 'angel' (< ANGELU), *saindu* 'saint, holy' (< SANCTU).

The Latin diphthong /au/ is borrowed as *au*: *gauza* 'thing' < CAUSA, B *autono* 'September' < AUTUMNU 'autumn'. This *au* regularly develops to *ai* in eastern dialects (Z *gaiza* 'thing' (< CAUSA)), and the same development occurs sporadically in all dialects before a following *u*: common *kaiku* 'wooden bowl' < CAUCU 'drinking vessel'. The case of *aditu* 'hear' < AUDITU is exceptional. There is no trace of Latin /ae/ in Basque: *balea* 'whale' < BALLAENA, old Z *zekuru* 'century' < SAECULU.

Latin word-initial plosives are usually borrowed as Basque *b d g*: *balea* 'whale' < BALLAENA, *diru* (Z *diharü*) 'money' < DENARIU, *garau(n)* 'grain' < GRANU, *bake* 'peace' < PACE, *dorre* 'tower' < TURRE or Sp *torre*, *gela* 'room' < CELLA. Initial /w/ and Romance /v/ are borrowed as *b*: *begiratu* 'look at, watch' < VIGILIARE, *berde* 'green' < VERDE. Initial /f/ is borrowed as *b* or lost: *baba* 'bean' < FABA, *biko* ~ *iku* 'fig' < FICU. At some later stage, probably under continuing Romance influence, we often find Latin initial voiceless plosives rendered as voiceless plosives in Basque, and some earlier loans were seemingly re-formed: *gatu* ~ *katu* 'cat' < CATTU, *gatea* ~ *katea* 'chain' < CATENA, *gerezi* ~ *kerexa* 'cherry' < CERESEA, *biper* ~ *piper* 'pepper' < PIPER. Occasionally, by apparent hypercorrection, we find Basque voiceless plosives which are not etymological: *pintza* 'membrane' < Aragonese *binza*, regional *perde* 'green' for common *berde* < Rom *verde*.

Word-medial voiceless plosives, simplex or geminate, and geminate voiced plosives are borrowed as Basque *p t k*: *ziape* 'mustard' < SENAPE, Sout *lupu* 'wolf' < LUPU, *zeta* 'silk' < SETA, *ahate* 'duck' < ANATE, *biko* 'fig' < FICU, *lapa* 'burdock' < LAPPA, *gatu* 'cat' < CATTU, *zuku* 'juice, soup' < SUCCU, *ziku* 'dry' < SICCU, *okela* 'morsel, meat' < BUCCELLA 'mouthful', B *zapatu* 'Saturday' < SABBATU, *apaez* 'priest' < ABBAS 'abbot'. Word-medial simplex voiced plosives are borrowed as *b d g*: *baba* 'bean' < FABA, *abere* 'cattle' < HABERE 'have', *padura* 'water meadow' < PADULE, *zigulu* 'seal' < SIGILLU, *bago* 'beech' < FAGU.

The Latin sibilant /s/ is almost always borrowed as *z*: *zeta* 'silk' (< SETA), *zapore* 'taste' (< SAPORE), *zigilu* 'seal' (< SIGILLU), *ziape* 'mustard' (< SENAPE), *azeri* 'fox' (< ASENARIU), *meza* 'mass' (< MISSA), *izkribatu* 'write' (< SCRIBERE), *gerezi* 'cherry' (< CERESEA). In final position we usually find *tz*: *bortitz* 'strong' (< FORTIS), *gorputz* 'body' (< CORPUS), but note *apaez* 'priest' (< ABBAS). A very few words show *s* instead, such as *soka* 'rope' (< SOCA) and *gisu* 'plaster' (< GYPSU).

Initial clusters are always broken up in some way. Most usually, an epenthetic vowel is inserted: *boronde* 'forehead' < FRONTE, *garau(n)* 'grain' < GRANU, *gereta* 'rustic gate' < *CLETA. If the initial consonant is a labial, however, it is often lost: *lama* 'flame' < FLAMMA, *lore* 'flower' < FLORE, *luma* 'feather' < PLUMA, *lau* 'flat' < PLANU, *laket* 'be pleasing' < PLACET. (And note also B *laru* 'yellow' < CLARU 'clear'.) If the initial consonant is a sibilant, a prothetic vowel is added: *ezpata* 'sword' < SPATHA, *iztupa* 'hemp, oakum' < STUPPA. An initial /r/ always acquires a prothetic vowel: *arrosa* 'rose' < ROSA, *errege* 'king' < REGE, *Erroma* 'Rome' < ROMA.

Medial clusters existing in Pre-Basque are retained, apart from the usual voicing of plosives after *n* or *l*: *denbora* 'time' < TEMPORA 'times', *aldare* 'altar' < ALTARE, *ingude* 'anvil' < INCUDE, *maizter* 'master shepherd' < MAGISTER.

Impermissible clusters, particularly those beginning with a plosive, are broken up or reduced: *pundu* 'point' < PUNCTU, *liburu* 'book' < LIBRU, *lukuru* 'avarice, usury' < LUCRU, *zinu* 'sign' < SIGNU, B *autono* 'September' < AUTUMNU 'autumn', *eskiribatu* ~ *iskiribatu* ~ *izkiriatu* 'write' < SCRIBERE, *abendu* 'December' < ADVENTU, *ifernu* 'hell' < INFERNU, *el(e)iza* 'church' < ECCLESIA, *gisu* 'plaster' < GYPSU.

As can be seen from the examples above, Latin velars remain velars in Basque before front vowels. In later loans, the palatalization of velars before front vowels is sometimes visible in the forms of loans in Basque. Lat CEPULLA 'onion' appears variously as *kipula* ~ *tipula* ~ *dipula*, the last two variants perhaps reflecting Romance palatals. Lat GENUS 'kind' appears to be the source of French Basque *deus* ~ *jeus* 'anything'. Lat GENTE 'people' appears as *gente* in R but as *jente* ~ *jende* in the rest of the country. Lat ARCHIATER 'doctor' gives archaic *atxeter* ~ *altxeter*.

A few loans show the loss of an initial syllable: *sendo* 'robust, strong' (< EXEMPTU), *saiatu* 'try' (< Fr *ensayer*), *txertatu* 'insert' (< Sp *injertar*), *txukatu* 'wipe off, dry' (< Sp *enjugar*). Another possible case is *debekatu* 'prohibit', if this derives from Lat IMPEDICARE 'obstruct', but Michelena prefers to invoke Old Spanish *deviedo* 'prohibited'.

A few individual loan words exhibit unusually complex phonological

developments. B *bi(x)ao* 'siesta' derives from MERIDIANU 'of midday'. Common *poz* 'happiness' is thought to derive from GAUDIU 'joy'. Common *bedeinkatu* 'bless', from BENEDICERE, must continue a metathesized form **bedenikatu*.

Folk-etymology has affected the forms of a few loans. Lat VIGILIARE gives Basque *begiratu* 'watch, look at', whose form has been influenced by *begi* 'eye'. Sp *zanahoria* 'carrot', of Arabic origin, yields Basque *zainhori*, literally 'yellow-root', with the final *-a* interpreted as the Basque article. The Romance loan *estable* 'shed, shelter' (cognate with English *stable*), with its alien cluster, has been re-formed as *estalpe*, from *estali* 'cover' and *-pe* 'under'. Another possible case is *buztarin(a)* ~ *uzterina* ~ *uztare* ~ *uztain* ~ Z *üztari(a)* 'crupper; hindquarters', which surely continues Lat POSTILENA 'crupper'; the Basque form suggests an unattested Lat **POSTELINA*, but we may simply be seeing interference from *buztan* 'tail'.

As far as we can tell, the position of the Latin word-accent was generally retained in Basque.

3.16 THE PHONOTACTICS OF PRE-BASQUE

There has been no systematic study of the phonotactics of native Basque words of any antiquity, nor do we even have at present an agreed list of words which were certainly or probably in the language 2,000 years ago. There are, however, a few points which stand out upon even a casual inspection. Most prominent among these is the remarkable frequency of word-initial vowels.

It is obvious that vowel-initial lexical items are extraordinarily frequent in Basque. While I have seen no reliable figures, I would hazard a guess that close to 50 per cent of native nouns and adjectives begin with vowels (this includes those items which begin with *h* in the aspirating dialects). Just to provide some very rough confirmation, Table 3.12 shows the numbers of pages of entries beginning with particular letters in Azkue's 1905 dictionary, which excludes the most recent loans from Romance; the dictionary includes affixes as well as lexical items, a policy which noticeably inflates certain of the shorter sections.

The five vowels represent altogether 404 pages out of a total of 1,029 pages, or about 40 per cent (the dictionary actually contains 1,042 pages, but the 'missing' pages are devoted to C, H, J, Q, V, W and Y, for which there are no entries). The E and I sections are significantly inflated by the inclusion of a large number of verbs whose participles bear the prefix *e-* or *i-* (see Chapter 4), and the I section is further inflated by Azkue's practice of not distinguishing J from I. Still, the overall pattern is clear. Note that A, with 120 pages, is far and away the most frequent initial segment, followed by E, with 96. By comparison, the most frequent initial consonant, Z, occupies only 87 pages. Moreover, recall that *none* of the 160 pages devoted

Table 3.12 Initial segments in Azkue's dictionary

A:	120	T:	43
E:	96	N:	37
Z:	87	P:	36
I:	79	TX:	36
B:	73	D:	16
O:	62	R:	7 (affixes only)
G:	62	F:	5
M:	58	TZ:	5
K:	58	TT:	4
L:	48	Ñ:	3
U:	47	LL:	1
S:	46	TS:	0

to P, T, K, D and R can represent ancient lexical items, suggesting that the proportion of vowel-initial words in Pre-Basque was probably much closer to 50 than to 40 per cent.

The seemingly extraordinary frequency of initial vowels is the only issue in the phonotactics of Pre-Basque which has attracted any significant comment, but this topic has in fact attracted so much attention that I shall devote a separate section to it below (section 3.18). There is no point in trying to illustrate the vast frequency of initial vowels here; look at any page of this book containing words other than loan words.

Apart from the very frequent word-initial vowels, recall that lexical items in Pre-Basque could begin with one of only six consonants: **b g z s l n**. These are not of equal frequency. First, **b g z** are all exceedingly frequent in native words, and it would be easy to list hundreds of native words beginning with each. A modest sample: *buru* 'head', *begi* 'eye', *bihotz* 'heart', *beso* 'arm', *belar* 'grass', *behi* 'cow', *baso* 'woods', *burdina* 'iron', *berun* 'lead' (metal), *beltz* 'black', *berri* 'new', *bat* 'one', *bi(ga)* 'two', *bortz* 'five', *bai* 'yes', *baina(n)* 'but', *bezala* 'like, as'; *gerri* 'waist', *gibel* 'liver', *gizon* 'man', *giltza* 'key', *gar* 'flame', *gain* 'top', *gurdi* 'cart', *gudu* 'combat', *goroldio* 'moss', *gari* 'wheat', *goi* and *garai* 'high place', *gaitz* 'bad', *garbi* 'clean', *gizen* 'fat', *gorri* 'red', *gor* 'deaf', *gero* 'after, later', *gu* 'we'; *zubi* 'bridge', *zur* 'wood', *zakur* 'dog', *zaldi* 'horse', *zulo* 'hole', *zezen* 'bull', *zati* 'piece', *zabal* 'wide', *zikin* 'dirty', *zahar* 'old', *zuzen* 'straight', *zuri* 'white', *zazpi* 'seven', *zortzi* 'eight', *zehar* 'across', *zer* 'what?', *zu* 'you'.

In contrast, **s** and **l** are substantially less common, though still not rare, and one could easily find at least some dozens of native words with these initial segments. A sample: *sudur* 'nose', *sabel* 'belly', *soin* 'shoulder; body', *su* 'fire', *sagar* 'apple', *seme* 'son', *senar* 'husband', *suge* 'snake', *sarde* 'pitchfork', *sakon* 'deep', *sei* 'six'; *lepo* 'neck', *larru* 'skin', *larre* 'pasture', *lan* 'work', *lagun* 'friend', *lo* 'sleep', *leiho* 'window', *lur* 'earth', *luze* 'long', *labur* 'short', *latz* 'rough', *leun* 'smooth', *lehen* 'first', *lau(r)* 'four'. Finally,

initial **n** is not common at all; I can here give something close to an exhaustive list of native words beginning with this segment: *neska* 'girl', *neba* 'brother (of a woman)', *negar* 'tears', *negu* 'winter', *neurri* 'measure', *nahi* 'desire', *naba* 'high plain', *nagi* 'idleness', *nabar* 'multicoloured, dun', *nahasi* 'mixed', *nagusi* 'chief, principal', *nor* 'who?', *ni* 'I'. There are very few others.

The very few native words with initial *m-* all derive from earlier **b**. Thus, *mihi* ~ *min* 'tongue' is from **bini*; *mendi* 'mountain' must be from **bendi* (if this is not a loan word); *mahats* 'grapes' is undoubtedly from **banats*; *mahai(n)* 'table' is probably from **banane* or something similar. The great frequency of initial *m-* in modern Basque results chiefly from borrowing, but also from the tendency to use *m-* in coining onomatopoeic and phonaesthetic items.

In word-final position, native lexical items can again have any vowel, and again vowels are exceedingly frequent here. It appears that the only word-final consonants in Pre-Basque were the fortis sibilants **tz** and **ts**, plus **N** (the neutralization of the two nasals), **R** (the neutralization of the two rhotics) and **L** (the neutralization of the two laterals).

Examples of the affricates (the laminal one is far commoner): *hitz* 'word', *gatz* 'salt', *ihi(n)tz* (< **initz*) 'dew', *hotz* 'cold', *latz* 'rough', *hots* 'shout, cry', *mahats* 'grapes', *sits* ~ *sats* 'moth; ordure'. The affricate could apparently be preceded by a nasal or a liquid, though such forms are rare and almost always involve *-tz*: *zuntz* 'strand of wool' (this word may not be ancient), *bortz* 'five', *zurtz* 'orphan', *beltz* 'black' (probably from earlier **beletz*), *hartz* 'bear' (possibly a loan from Indo-European).

Of the resonants, **N** and **R** are fairly frequent in this position, while **L** is positively uncommon: *lan* 'work', *ozpin* 'vinegar', *gurin* 'animal fat, butter', *gizon* 'man', *min* 'pain', *haran* 'valley', *ezpain* 'lip', *oin* 'foot', *ozen* 'sonorous', *gizen* 'fat', *astun* 'heavy'; *zakur* 'dog', *hezur* 'bone', *abar* 'branch', *elur* 'snow', *ibar* 'water meadow, valley', *izar* 'star', *gogor* 'hard', *eder* 'beautiful', *hiru(r)* 'three', *lau(r)* 'four' (the *r* is original in both of these), *hamar* 'ten', *agur* 'good-bye'; *ahal* 'ability', *hil* 'dead', *zabal* 'wide'. It is possible that some of those in final *-n* derive from intervocalic **n**, just as *arrain* 'fish' derives from **arrani*.

There were absolutely no word-initial clusters in Pre-Basque, and probably no final clusters other than the resonant-plus-affricate clusters just described. The word-medial clusters were discussed in section 3.12.

It is interesting to consider the possible constraints upon consonants in successive syllables. (In what follows, I omit verbs from consideration.) To begin with, the sibilant harmony described in section 3.4 for the modern language was almost certainly present in Pre-Basque, since native words other than compounds never mix laminal and apical sibilants in a single word, and we find only forms like *zezen* 'bull', *izotz* 'frost', *zurtz* 'orphan', *itsaso* 'sea', *sats* ~ *sits* 'moth; ordure' and *itsusi* 'ugly'.

A striking fact is that, in monomorphemic native words of the pattern CVCV-, it is extremely rare for both consonants to be plosives (as noted by Lakarra 1995); in the few that exist, both plosives are usually voiced: *begi* 'eye', *bide* 'road', *biga* 'two', *bigun* 'soft', *bider* 'time, occasion', *bat* 'one' (< **bade* or **bada*), *gabe* ~ *baga* 'without', *gudu* 'combat' (suspected by some of being a loan word), *gogo* 'mind, soul' (possibly a reduplication), *gibel* 'liver' (a compound) and *gogor* 'hard' (almost certainly a reduplication). Exceptional are *beti* 'always' and *guti* 'not much'; it is highly probable that these are bimorphemic and contain the ancient adjective-forming suffix *-ti*. The highly anomalous *tipi* ~ *tiki* 'small', with its initial *t*, is undoubtedly an expressive formation; José Ignacio Hualde has suggested (p.c.) that this might represent a metathesis of an original **piti*, possibly also involving the suffix *-ti* and conceivably partly of Romance origin. It appears that Pre-Basque strongly disfavoured this pattern.

Words of this form, then, including those with initial *m*, are almost invariably loan words or expressive formations of no great antiquity. Consequently, it is generally safe to regard words like *beko* 'forehead; beak', eastern *moto* 'headscarf; hair bun' (and other senses), *kako* ~ *gako* 'hook' and *mutur* 'snout; extremity' as loan words, even though no Romance source has yet been securely identified for some of them, except in those cases in which an 'expressive' origin seems more likely, as perhaps with *mutur*.

Indeed, even CVCV- words in which only the second consonant is a plosive are not exactly abundant, though they are certainly more numerous than the last group: *lapur* 'thief', *labur* 'short', *zakur* 'dog', *negu* 'winter', *lepo* 'neck', *suge* 'snake', *zoko* 'corner', *labe* 'oven'. The same goes for VCV-words, though these are probably more frequent still: *hagin* 'molar', *idi* 'ox', *hego* 'south wind', *agur* 'good-bye', *eder* 'beautiful', *ikara* 'trembling', *ugatz* '(female) breast', *hegal* 'wing', *egun* 'day', *ugari* 'numerous', *aker* 'he-goat', *odol* 'blood', *oker* 'twisted', *ipurdi* 'buttocks', *hegi* 'ridge', *abar* 'branch', *ibar* 'water-meadow, valley' and quite a few others.

Overwhelmingly, though, a plosive at the beginning of the second syllable is preceded by a resonant or a sibilant. This pattern is so exceedingly frequent as to constitute virtually the stereotype for a Basque word. Here is just a tiny sample of what is really a very large number of words. First, with a resonant: *ardi* 'sheep', *argi* 'light, bright', *erdi* 'half, middle', *zaldi* 'horse', *barda* 'last night', *alte* ~ *alde* 'side', *urde* 'hog', *sarde* 'pitchfork', *galte* ~ *galde* 'question', *garbi* 'clean', *gurdi* 'cart', *oldar* 'moment', *urte* 'year', *arte* 'interval, until', *albo* 'side', *murgil* 'dive', *arto* 'millet, maize', *handi* 'big', *enda* 'breed, race', *indar* 'power, force', *mendi* 'mountain', *ondar* 'sand, beach', *andere* 'lady', *sendo* 'strong, vigorous'. With a sibilant: *aste* 'week', *azti* 'fortune-teller', *la(i)ster* 'quick, soon', *asto* 'donkey', *ezti* 'honey', *gazte* 'young', *listu* 'saliva', *aski* 'enough', *aska* 'cradle', *esku* 'hand', *ezpain* 'lip', *ozpin* 'vinegar', *azpi* 'below', *neska* 'girl', *izter* 'thigh', *asko* 'lots of; some',

azkar 'strong, vigorous', *azken* 'last', *oskol* 'shell', *oste* 'back', *oztopo* 'obstacle', *esker* 'thanks', *ezker* 'left (hand)', *uztarri* 'yoke' and so on – there are really very many of these.

Note, however, the following further observation. In a word of the form PVRPV-, where P is a plosive and R is a resonant, both plosives are voiced. We find plenty of words like *gurdi* 'cart' and *barda* 'last night', but seemingly none with a voiceless plosive, even after *r*, a position in which plosives have not undergone voicing. Hence there are no certain native words of the forms **bVrtV-*, **gVrkV-* and so on. The rare exceptions are almost certainly loan words, like eastern *barta* ~ *parta* 'mud, muddy place, swamp', which is also found throughout western Romance, and *burki* 'birch', a rare variant of *urki*, and often thought to be a loan from Germanic.

3.17 MORPHEME STRUCTURE

Monomorphemic Basque nouns and adjectives of any antiquity are overwhelmingly disyllabic; monosyllables and trisyllables, while by no means rare, are much less common. (The special case of verbs is discussed separately below.) Moreover, these ancient nouns and adjectives conform strongly to certain severe constraints on their structure. It is worth looking at these patterns in some detail; if for no other reason, recognition of these patterns can help in deciding whether a given lexical item is genuinely ancient in the language or not. It is convenient to begin with the most frequent disyllabic pattern. Ancient disyllables generally conform to the following canonical form:

$(C_1)V(C_2)C_3V(C_4)$

As explained above, C_1, when present, is one of *b g z s l n*, or occasionally *m*, when this derives from original **b*. C_4, when present, is one of *n l r tz ts*. If C_2 is absent, C_3 can be any consonant except *j*; if C_2 is present, then C_3 must be either a plosive (most often a coronal one) or an affricate; if C_3 is a plosive, C_2 must be one of *l r n z s*; if C_3 is an affricate, C_2 must be one of *l r n*. Either vowel may be a recognized diphthong, but not both. Here are some typical examples of each possible variant of this canonical form:

VC_3V
ate 'door', *euli* 'fly', *izu* 'trembling, fear', *atzo* 'yesterday', *atso* 'old woman', *idi* 'ox', *izai* 'fir', *ogi* 'bread', *ele* 'word, conversation', *alu* 'vulva', *ira* 'fern', *orri* 'leaf', *uda* 'summer', *urre* 'gold', *aita* 'father'

With *h* as C_3: *aho* 'mouth', *ohe* 'bed', *ihi* 'reed' (< **ini*)

VC_3VC_4
izar 'star', *izen* 'name', *adar* 'horn, branch', *ilar* 'pea', *itzal* 'shade, shadow', *abar* 'branch', *igel* 'frog', *elur* 'snow', *ibar* 'valley', *irin* 'flour', *ikatz* 'charcoal', *orein* 'deer', *usain* 'odour', *aker* 'male goat', *arotz* 'carpenter'

With *h* as C_3: *ahal* 'ability' (< **anal*), *ahur* 'palm', *ohoin* 'thief' (< **ono-*), *oihan* 'forest', *ohol* 'wooden plank' (< **onol*), *eihar* 'dry', *ehun* 'hundred'

VC_2C_3V
aste 'week', *asto* 'donkey', *alde* 'side', *enda* 'race, breed', *arto* 'millet', *azpi* 'below', *osto* 'side', *erbi* 'hare', *argi* 'light', *azti* 'fortune-teller', *albo* 'side', *esku* 'hand', *ezki* 'linden', *erdi* 'half', *untzi* 'vessel, container, *ardi* 'sheep', *arte* 'holm-oak', *ortzi* 'sky', *urde* 'pig'

$VC_2C_3VC_4$
izter 'thigh', *astun* 'heavy', *argal* 'thin', *ozpin* 'vinegar', *indar* 'force', *ezker* 'left (hand)', *esker* 'thanks', *ezpain* 'lip'

C_1VC_3V
gerri 'waist', *zuri* 'white', *gari* 'wheat', *zulo* 'hole', *neba* 'brother (of a woman)', *bero* 'hot', *negu* 'winter', *garai* 'height', *berri* 'new', *lasai* 'calm'

With *h* as C_1 or C_3: *hori* 'yellow', *harri* 'stone', *herri* 'country', *hego* 'south', *mihi* 'tongue' (< **bini*), *sehi* 'child, boy' (< **seni*), *lohi* 'mud', *nahi* 'desire', *behi* 'cow', *ziho* 'tallow' (< **zino*)

Note that, in this pattern, it is extremely rare for both consonants to be plosives; in the few such cases, both plosives are voiced, as in *bide* 'road' and *gogo* 'mind'. In fact, it is uncommon even for C_3 to be plosive here, and almost unheard-of for it to be a voiceless plosive, *lepo* 'neck' and *zati* 'piece' being rare exceptions.

$C_1VC_3VC_4$
berun 'lead', *nabar* 'mottled, dark', *zilar* 'silver', *sabel* 'stomach', *gibel* 'liver', *gurin* 'animal fat', *gizen* 'fat', *lizun* 'lascivious', *leizar* 'ash tree', *zezen* 'bull', *zakur* 'dog'

With *h* as C_1 or C_3: *higuin* 'disgust', *haran* 'valley', *hegal* 'wing', *mahats* 'grapes' (< **banats*), *zahar* 'old', *lehor* 'dry', *lehen* 'first', *bihotz* 'heart'

$C_1VC_2C_3V$
garbi 'clean', *sarde* 'pitchfork', *gurdi* 'cart', *zaldi* 'horse', *salda* 'broth', *gazte* 'young', *mendi* 'mountain' (< **bendi*?), *giltza* 'key', *sendo* 'robust', *galde* 'question', *bertze* 'other'

With *h* as C_1: *handi* 'big', *hertze* 'intestine', *hontza* 'barn owl'

Recall that, in this pattern, if C_3 is a plosive, it must be voiced, except after a sibilant, when it must be voiceless, or after *r*, when it may be either.

$C_1VC_2C_3VC_4$
sorgin 'witch', *laster* 'quick', *ziztor* 'icicle', *bazter* 'edge', *bizkar* 'back'

With *h* as C_1: *hondar* 'sand', *harbel* 'slate'

A few of these, like *astun*, *ozpin* and *harbel*, are certainly or probably compounds, but very ancient ones.

This, then, is what native Basque words look like. The much smaller group of monosyllables follows the usual rules for initial and final consonants, except that initial *b* is curiously almost absent:

su 'fire', *hotz* 'cold', *hots* 'cry, noise', *hitz* 'word', *lo* 'sleep', *lur* 'earth', *goi* 'height', *gau* 'night', *lan* 'work', *zin* 'oath', *gor* 'deaf', *sits* 'moth', *ar* 'male', *ur* 'water', *or* 'dog', *hur* 'hazelnut', *hil* 'dead', *latz* 'rough', *so* 'glance', *gar* 'flame', *soil* 'lone', *soin* 'body', *zur* 'wood', *haitz* 'oak', *gatz* 'salt'

A few monosyllables have final clusters: *hartz* 'bear', *beltz* 'black' (< **beletz*), *bortz* ~ *bost* 'five', *zurtz* 'orphan', *antz* 'resemblance', *bort* 'bastard', western *bart* 'last night' (< *barda*, preserved in the east). As can be seen, some of these clusters are secondary and very likely all of them are.

Trisyllables mostly look like disyllables with an extra vowel added at the end:

ardo 'wine' (< **ardano*), *itsaso* 'sea', *hodei* 'cloud' (< **odeCe*), *belarri* 'ear', *mahai(n)* 'table' (< **banane*), *udare* 'pear', *ugari* 'numerous', *uztarri* 'yoke', but *hezur* 'bone' (< **enazur*), *erreka* 'ravine', *ipurdi* 'buttocks' (possibly bimorphemic)

Four-syllable stems are all but unknown: *goroldio* 'moss' (possibly bimorphemic) and B *arerio* 'enemy' are virtually the only examples.

Verbal roots are very different in structure. For one thing, ancient verbal roots are very frequently monosyllabic; for another, they can begin with virtually any consonant, including a rhotic or a coronal plosive. Here is a sample; the ancient prefix **e-* and the participial suffix *-i* are segmented out:

e-karr-i 'bring', *j-arr-i* 'put', *e-ser-i* 'sit down', *e-ror-i* 'fall', *i-pin-i* 'put', *i-dur-i* ~ *i-rud-i* 'seem', *e-dan* 'drink', *i-zek-i* 'burn', *i-bil-i* 'go about', *e-man* 'give', *j-oan* 'go', *e-torr-i* 'come', *e-gin* 'do, make', *e-ntzun* 'hear' (< **e-nezun*), *e-kin* 'get busy', *i-raun* 'last', *j-aus-i* 'fall', *e-ba-i* 'cut', *(e)-utz-i* 'leave', *e-rrun* 'lay eggs', *i-tzul-i* 'turn'

This preponderance of monosyllabic verbal roots raises severe questions, as does the fact that any consonant at all can occur in root-initial position. It is possible, of course, that verbal roots in Pre-Basque were simply subject to different morpheme-structure rules from nouns and adjectives; after all, verbal roots are never found in isolation, but only combined with other morphemes in finite and non-finite forms. But Michelena, in several places, suggests that verbal roots preserve a more ancient phonotactics in which word-initial coronal plosives, at least, were possible in Basque. And Lakarra (1995) has chosen to attach still more weight to the verbal facts: he suggests that, at some exceedingly remote stage of the language, *all* lexical morphemes were monosyllabic, and that the dominant polysyllabic form of nouns and adjectives results from extensive compounding aided by a certain amount of reduplication. He has enjoyed a certain amount of success in identifying some ancient morphs with perhaps recognizable meanings, but it is too early to evaluate his research programme. If there ever was such a stage, it must have been long, long before the Pre-Basque of some 2,000 years ago reconstructed by Michelena.

For the rather distinctive phonology of 'expressive' formations, see section 5.5.

3.18 THE PROBLEM OF THE INITIAL VOWELS

This very great frequency of word-initial vowels has attracted a certain amount of comment and speculation. Broadly speaking, three views may be distinguished.

The null hypothesis. The frequent initial vowels result neither from an ancient addition of vowels nor from an ancient loss of initial consonants. Instead, they represent nothing more than a phonotactic preference among ancient speakers, and there is nothing to be explained. Though rarely or never defended explicitly, this is probably the *de facto* view of most vasconists. Interestingly, there are two striking pieces of evidence in favour of this view. First, as we saw in section 3.11, the presence of the aspiration in words like *senhar* 'husband' and *alhaba* 'daughter' suggests that intervocalic consonants in Pre-Basque, or some of them, were apparently assigned to the preceding syllable, leaving the following syllable vowel-initial; this suggests that Pre-Basque had a strong phonotactic preference for vowel-initial syllables. Second, recall from section 3.8 that coronal consonants were palatalized by a preceding high front vowel, but not by a following one, again suggesting that such consonants were more tightly bound to the preceding vowel and hence again that an intervocalic consonant was assigned to the preceding syllable. None of this is decisive, of course, but it is certainly suggestive.

The vowel-addition hypothesis. The initial vowels result from the ancient addition of vowels to lexical items which earlier began with consonants. The several defenders of this view have not generally favoured a merely phonological process of prothesis; instead, they have seen the vowels as fossilizations of ancient prefixes. Schuchardt (1893), for example, proposes that initial *e-* and *i-* in nouns often represent a fossilized ancient article. Bengtson (1991c, 1993) interprets almost all initial vowels instead as fossilized noun-class prefixes dating from a time when Basque had noun classes (see section 6.9). At present, there appears to be absolutely no direct evidence in favour of any such interpretation, but there is one point that should be noted. In ancient

verbs, in which there is no doubt at all that the prefix **e-* is uniformly present in non-finite forms (see Chapter Four), the majority of roots begin with consonants, and moreover a number of such verbs exhibit roots beginning with consonants which absolutely cannot occur initially in nouns or adjectives: *etorri* 'come', *eten* 'break, interrupt', *eduki* 'hold, have', *edan* 'drink', *iduri* ~ *irudi* 'seem' (it is not clear which of these is the earlier form), *irun* 'spin', *erein* 'sow', *ipini* 'put', *ekarri* 'bring', *ekin* 'get busy', and others. However, one should not make too much of this: the verbal prefix **e-* is easily reconstructible for Pre-Basque, while not a single vocalic prefix can be sensibly reconstructed for nouns or adjectives. (Schuchardt's proposal is probably accepted by no one today, and Bengtson has been unable to identify any semantic correlations for his putative prefixes.)

The consonant-loss hypothesis. The initial vowels are the result of a wholesale loss of certain initial consonants at some ancient stage of the language. Now the fact that Michelena's reconstruction of Pre-Basque sets up the surprisingly small figure of only six word-initial consonants might be taken as providing some considerable plausibility for the idea that the language had earlier lost some initial consonants, and scholars have not been slow to explore this possibility. Two versions of this hypothesis have been presented.

One version is the view of Theo Vennemann (1994a), who suggests that an ancestral form of Basque possessed a series of 'weak' consonants which, following the Indo-European tradition, he terms *laryngeals*. These consonants, he suggests, might have been [h], [ʔ] and various non-sibilant fricatives such as [x], [χ] and [f]. In Vennemann's account, all of these weak consonants were simply lost in initial position, leaving the language with a substantial number of newly vowel-initial words; he further proposes that, as in Indo-European, certain of these consonants, such as [χ], might have altered a following vowel to *a* before disappearing, thus accounting for the enormous frequency of initial *a* in Basque.

Vennemann's proposal, it seems to me, suffers from massive difficulties. For one thing, there is not the slightest shred of direct evidence for the earlier existence of these putative lost consonants. For another, these consonants, if they ever existed, would surely also have occurred in other than initial position, but there is no trace of them in medial or final position; hence Vennemann's account would apparently require that all these consonants were lost without trace in *all* positions, a view which involves a great deal of special pleading.

The other version has been defended most vigorously by the French linguist André Martinet in several publications, most notably in his 1955 book. (A somewhat similar view has recently been independently put forward by John Bengtson (p.c.), but Bengtson has so far neither developed nor published his ideas, and so I shall not consider them here.) Martinet's central thesis is simple. He proposes that the fortis (voiceless) plosives, confined to

non-initial positions in Michelena's reconstruction of Pre-Basque, were, at some earlier stage, very frequent in word-initial position. In this position, he argues, they were always aspirated, and these aspirated initial plosives underwent weakening, first to fricatives, then to *h*. This *h* has at least sometimes remained in the aspirating dialects, but otherwise it has simply been lost.

There is no doubting the intrinsic phonetic plausibility of Martinet's idea: comparable developments are well attested in a number of other languages. Against Martinet's proposal, of course, is the fact that it would destroy the perfect symmetry of Michelena's reconstruction, since the fortis/lenis contrast would be extended to initial position for plosives and only for plosives. But an argument from symmetry can hardly be decisive, especially since it is clear from the evidence of Latin loans, in which initial plosives are almost invariably retained, that the lenition process posited by Martinet, if it ever occurred, must have occurred at a time before the stage reconstructed by Michelena. The question, of course, is whether there is any evidence for the loss of ancient initial plosives.

Interestingly, the answer appears to be 'yes – but not much'. There are four pieces of evidence, all of them very sparse and fragmentary, which might be interpreted as providing some kind of support for Martinet's proposal.

The first and most striking evidence comes from the three demonstratives and their derivatives. In the aspirating dialects, these invariably show initial *h*: *hau* ~ *hon-* 'this', *hori* ~ *horr-* 'that', *hura* ~ *har-* 'that (over there)', *hemen* 'here', *hor* 'there', *han* '(over) there' and so on. The remaining dialects overwhelmingly have vowel-initial items: *au* ~ *on-*, *ori* ~ *orr-*, *ura* ~ *ar-*, *emen*, *or*, *an* and so on. But some of the Pyrenean dialects show initial plosives. R has *kau* 'this', *kori* 'that', *kura* 'that (over there)', *keben* 'here' and so on throughout all the forms. And Aezk similarly has *gau*, *gori*, *gura*, *geben* and so on. It is very difficult to explain these plosives except by assuming that they are original: it looks very much as though these three stems anciently began with *k*, which has everywhere been reduced to *h* or lost, except in R and Aezk, in which *k* has either been exceptionally retained or voiced to *g*. In several publications Michelena has expressed the view that the initial plosive is indeed original in these stems, and has therefore been weakened or lost in all other dialects – though not everyone agrees with him in this interpretation, and note that the article is *-a* in the Pyrenean dialects as elsewhere (this derives from the distal demonstrative).

Of course, the obvious question at this point is whether the Pyrenean dialects exhibit other examples of initial plosives which are elsewhere absent. But this time the answer is 'no': all the other words which begin with vowels in the other dialects also begin with vowels in the Pyrenees. This is not encouraging.

Second, there are one or two items which are attested in Romance with an initial plosive but which lack that plosive in Basque. Notable is the case of

the universal (*h*)*arri* 'rock'. A stem **karr-* 'rock, crag' is widely attested in western Romance; this has no Latin source, and is generally believed to be pre-Latin, possibly from Celtic, in which the stem is also attested. (See Corominas and Pascual (1980) under *alcarrila*.) There seems every reason to suppose that the Basque word is of the same origin, and hence that it too represents an instance of the lenition of an initial voiceless plosive. Moreover, the word for 'maple tree', which is *astigar* in most of the country and *aztigar* in G, appears as *gaztigar* in some eastern varieties. This leads Michelena (1959: 525; 1973a: 60; 1977a: 253) to suggest that this word is a compound whose second element is the obscure but securely attested *ihar* 'maple' (*Acer hispanicum*) and whose first element is **kast-*, attested in tree names in Latin and Romance and apparently of substrate origin.

Third, Basque exhibits a few morphemes which have variant forms with and without initial voiceless plosives. Most strikingly, there are a couple which function both as independent words and as derivational suffixes, and which show an initial voiceless plosive when they are suffixes, but not when they occur as free forms. Thus, for example, we have *ume* 'child', but *-kume* 'offspring' in such compounds as *katakume* 'kitten' (*katu* 'cat'), *arkume* 'lamb' (*ardi* 'sheep'), and probably also *emakume* 'woman', but 'girl' in Sal (*eme* 'female'). Similarly, we have *alde* 'side', but *-kalde* in a few derivatives like *sukalde* 'kitchen' (*su* 'fire'), and *ohi* 'habit, custom' but *-koi* 'fond of', as in *ardankoi* 'fond of wine' (*ard*(*a*)*o* 'wine'). In addition, there are quite a few derivational suffixes which have two forms, the one with the initial plosive being usually found after a consonant, the other after a vowel: *-tar* ~ *-ar* (ethnonymic), *-tasun* ~ *-asun* (abstract-noun suffix), *-tegi* ~ *-egi* 'place', *-kara* ~ *-ara* 'manner', *-kide* ~ *-ide* 'fellow' and a few others.

Finally, we have some tantalizing evidence from Aquitanian. As we shall see in section 6.9, the ancient Aquitanian language is the more-or-less direct ancestor of Basque. It is noteworthy that the Aquitanian names recorded in the northern part of Aquitania frequently show an initial *t-*, while names recorded in the south (near the Basque Country) do not. Moreover, there are one or two instances in which otherwise identical morphs are attested with *t-* in the north but with *h-* in the south: northern *Talsco-*, southern *Halsco-*. This could perhaps be taken as evidence that initial *t-*, while remaining in the extinct northern varieties of Aquitanian, had been lenited to *h-* in the southern varieties ancestral to Basque.

I am not aware of any further evidence for the ancient loss of initial plosives. Of course, it is entirely possible that Martinet is right, and that, with the few exceptions just noted, the loss of initial voiceless plosives was so early and so complete that no trace of them now remains. This is exactly the sort of problem on which we might reasonably expect some light to be shed if we could ever locate any distant relatives of Basque. Unfortunately, as we shall see in Chapter 6, the innumerable attempts to identify just such

relatives, quite apart from their many other shortcomings, fail on precisely this point: they *never* shed such light, and they never offer any fruitful hypotheses about the prehistory of Basque before the Roman period. As Michelena has often stressed, this fact alone is enough to render valueless all the conjectured relationships. In all probability, we will never know anything about the origin of the vowel-initial words of Basque.

(Interestingly, I am aware of no explicit attempts to test Martinet's hypothesis by the dedicated seekers after genetic connections whose work is discussed in Chapter 6. Indeed, very many of these scholars quite explicitly segment the inconvenient initial vowels of Basque into oblivion in their search for cognates, thereby effectively adopting the vowel-addition hypothesis mentioned above. Since this policy has manifestly led to no interesting results, perhaps one or two of these investigators might like to redirect their efforts with Martinet's thesis in mind.)

3.19 DISTINCTIVE PHONOLOGY OF WORD-FORMATION

Words obtained by compounding and derivation are, of course, subject to the same phonological constraints and processes as other words. In addition, however, such formations show the effects of certain additional processes not exhibited elsewhere, except occasionally in verbal morphology. Some of these additional processes involve the most complex phonological developments in the language, and, taken together with the ordinary phonological changes that have affected Basque, they produce some interesting alternations between free forms and combining forms.

Generally speaking, both compounds and derivatives formed by suffixation behave in exactly the same way, and I shall not distinguish the two types of word-formation here (recall that prefixes are not used in word-formation, apart from a handful of late calques and loans from Romance). There are, however, significant differences in the treatment of a lexical stem which occurs as the first element in a compound or a derivative and of a lexical stem which occurs as the final element in a compound, and there are further certain peculiarities exhibited by some word-forming suffixes. I shall discuss each of these in turn. Throughout this discussion, recall that Basque *h* is not a true consonant and does not behave like one; in Pre-Basque it was purely a suprasegmental, and its presence generally had (and has) no effect upon word-formation. The abbreviations *NFS* and *VFS* stand for 'noun-forming suffix' and 'verb-forming suffix', respectively.

There is one general point to be made. Haplology is the norm in Basque; it usually occurs whenever it can possibly do so. Thus, for example, *sagar* 'apple' plus *ardo* ~ *arno* 'wine' ought to yield **sagarrardo* ~ **sagarrarno* by the usual rules, but the word is *sagardo* ~ *sagarno* 'cider'. The compound of *eme* 'female' and *magiña* 'sheath' ought to yield **emamagiña*, but the word is *emagiña* 'vagina'. Syllable loss may occur even when the consecutive

syllables are not quite identical: *ume* 'offspring' plus *berri* 'new' yields *umerri* 'lamb'; *beko* 'forehead' plus *gorri* 'red' yields LN Z *bekorri* 'red mark on an animal's forehead'.

3.19.1 First elements

It is first elements which exhibit the most complex behaviour. These complexities derive from the action of a sizeable number of phonological changes, both language-wide and specific to word-formation; only in a few instances is it possible to determine the order in which these developments must have occurred, and here I shall adopt an order of presentation which, I hope, will be maximally illuminating. I begin with final vowels.

A final vowel in a first element behaves differently depending on which syllable it is in. If it occurs in the first syllable (that is, if the first element is a monosyllable), the vowel is unaffected. Since monosyllables are not common in Basque, there are only a few such cases. Examples:

ke 'smoke' + *zulo* 'hole' → *kezulo* 'chimney'
lo 'sleep' + *gela* 'room' → *logela* 'bedroom'
hi 'you' (intimate) + *-keta* Activity NFS → *hiketa* 'use of the intimate pronoun'
su 'fire' + *-gai* 'suitable for' → *sugai* 'fuel'
jo 'hit' + *aldi* 'occasion' → *joaldi* '(a) blow'
su 'fire' + *harri* 'stone' → *suarri* 'flint'

A final vowel in a third or later syllable is lost:

itsaso 'sea' + *gizon* 'man' → *itsasgizon* 'sailor' (*g* = [k])
iturri 'spring' + *buru* 'head' → *iturburu* 'fountainhead'
txapela 'beret' + *-dun* 'having' → *txapeldun* 'champion'
eliza 'church' + *-koi* 'fond of' → *elizkoi* 'pious'
pilota 'jai alai' + *-ari* Professional NFS → *pilotari* 'jai alai player'
itsaso 'sea' + *untzi* 'vessel' → *itsasuntzi* 'ship'

Recent formations are often exceptions:

pilota 'jai alai' + *leku* 'place' → *pilotaleku* 'jai alai court'
euskara 'Basque language' + *-tu* VFS → *euskaratu* 'translate into Basque'

An earlier formation would have yielded **euskaldu* for the second (see below for the complications).

In a second syllable, a final vowel is usually lost before a following vowel:

baso 'woods' + *urde* 'hog' → *basurde* 'wild boar'
erre 'burn' + *hauts* 'dust' → *errauts* 'ash'
baso 'woods' + *herri* 'habitation' → *baserri* 'farmhouse'
aita 'father' + *on* 'good' → *aiton* 'grandfather'

gerra 'war' + *aldi* 'time' → *gerraldi* 'wartime'
otso 'wolf' + *eme* 'female' → *otseme* 'she-wolf'
buru 'head' + *hezur* 'bone' → *burhezur* 'skull'
oilo 'hen' + *ar* 'male' → *oilar* 'cock, rooster'
aita 'father' + *ama* 'mother' + *-k* Pl → *aitamak* '(one's) parents'

But compare

seme 'son' + *alaba* 'daughter' + *-k* Pl → *seme-alabak* '(one's) children'

Recent formations often fail to observe this rule:

gero 'later' + *aldi* 'time' → *geroaldi* 'future tense' (grammar)

Otherwise, in a second syllable, a final vowel *i* is lost:

harri 'stone' + *-gin* 'maker' → *hargin* 'stonecutter'
herri 'country' + *beste* 'other' → *herbeste* 'foreign country'
harri 'stone' + **bel* 'black' → *harbel* 'slate'

It will become clear below that such loss of *i* was anciently a pervasive process. None the less, exceptions are not wanting:

harri 'stone' + *-tu* VFS → *harritu* 'surprise' (originally 'petrify'?)

Here there is an obvious functional explanation: the expected **hartu* would be homophonous with the common verb *hartu* 'take'. In fact, however, *harri* 'stone' forms a large number of derivatives in which the *i* is not lost, such as *harrikada* 'blow with a stone' and *harriola* 'stony place'; many of these are severely localized and may be late formations. But a number of other common words, such as *berri* 'new', *gorri* 'red' and *zuri* 'white' absolutely never lose their *i*. It is hard to know what to make of this. Moreover, *i* is not normally lost after a labial: *zubi* 'bridge' + *-gin*, *-gile* 'who makes' yields *zubigin*, *zubigile* 'bridge-builder'.

In a second syllable, final *-a*, *-o* and *-e* are all neutralized as *-a*:

lore 'flower' + *-tu* VFS → *loratu* 'bloom, blossom'
zulo 'hole' + *-tu* VFS → *zulatu* 'dig'
gona 'skirt' + *-dun* 'having' → *gonadun* '(one) who is wearing a skirt'
baso 'woods' + *jaun* 'lord' → *basajaun* 'Old Man of the Woods'
golde 'plough' + *-tu* VFS → *goldatu* 'plough' (v.)
otso 'wolf' + *porru* 'leek' → *otsaporru* 'wild leek'
sendo 'healthy' + *-garri* Instr NFS → *sendagarri* 'remedy'
etxe 'house' + *gain* 'top' → *etxagain* 'roof'
maite 'beloved' + *-tu* VFS → *maitatu* 'love' (v.)
beso 'arm' + *-pe* 'under' → *besape* 'armpit'
kanta 'song' + *-tu* VFS → *kantatu* 'sing'
aza- 'cabbage' + *lore* 'flower' → *azalore* 'cauliflower'

(The noun *azak* 'cabbage' is always plural in Basque.)

The treatment of *u* in this position is complex. Sometimes it is lost (like *i*), sometimes it is neutralized to *a* and sometimes it remains unaffected. Broadly speaking, loss of *u* is most frequent in the east (especially in Z) but becomes less common as we move towards the west, while lowering to *a* is most frequent in the west, but all three outcomes are widely attested. Here are a few examples:

gatu 'cat' + *belar* 'grass' → *gatubelar* 'catnip' (L)
katu 'cat' + *-kume* 'offspring' → *katakume* 'kitten' (B G)
katu 'cat' + *narru* 'skin' → *katanarru* 'catskin' (B)
buru 'head' + *-kide* '-fellow' → *burkide* 'comrade, colleague' (Z R LN, but attested in old B)
buru 'head' + *-ko* NFS → *buruko* 'cap' (Z LN HN), 'pillow' (G), but *burko* 'pillow' (B)
esku 'hand' + *zabal* 'wide' → *eskuzabal* 'generous' (common)

Several further instances of apparent lowering merely reflect the Romance origins of the words involved: *gertu* 'certain, ready' and *gertatu* 'prepare', *kantu* 'song' and *kantatu* 'sing', and so on.

Very few lexical items end in plosives, but the cases of vowel loss just discussed often bring plosives into final position in a first element. The treatment of such final plosives is complex. If such a plosive is followed by a vowel, then the plosive is usually converted to *t*:

begi 'eye' + *ile* 'hair' → *betile* 'eyelash'
begi 'eye' + *azal* 'skin' → *betazal* 'eyelid'
gurdi 'cart' + *abere* 'animal' → *gurtabere* 'draught animal'
ardi 'sheep' + *ile* 'hair' → *artile* 'wool'
argi 'light, bright' + *izar* 'star' → *artizar* 'morning star'
sagu 'mouse' + *itsu* 'blind' → *satitsu* 'shrew'
sagu 'mouse' + *and(e)re* ~ *and(e)ra* 'lady' → *satandre* ~ *satandera* 'weasel'
erdi 'half' + *hordi* 'drunk' → *ertordi* 'tipsy' (Z)

If a sibilant follows, the combination of *t* plus sibilant yields an affricate:

begi 'eye' + *sein* 'boy' → *betsein* 'pupil' (of the eye)
ogi 'bread' + *zare* 'basket' → B *otzara* 'basket'
bat 'one' + *-zu* Indef Pl → *batzu* 'several, some'
ardi 'sheep' + *-zai(n)* 'guardian' → *artzai(n)* 'shepherd'
ogi 'bread' + *sein* 'boy' → *otsein* 'servant'
errege 'king' + *zubi* 'bridge' → *Erretzubi* (habitation name)

Before a sonorant, the plosive is lost:

bat 'one' + *-na* Distributive → *bana* 'one apiece'
bait- Verbal prefix + *naiz* 'I am' → *bainaiz* 'since I am, I who am'

If the plosive is followed by another plosive, the cluster is resolved as follows: the cluster is reduced to a single plosive which has the place of articulation of the second original plosive and which is invariably voiceless:

begi 'eye' + *buru* 'head' → *bepuru* 'eyebrow'
begi 'eye' + *gain* 'top' → *bekain* 'eyebrow'
idi 'ox' + *-tegi* 'place' → *itegi* 'stable for oxen'
errege 'king' + *bide* 'road' → *errepide* 'highway' (this one is attested as *erret bide* in the Fuero General of Navarre)
gurdi 'cart' + *bide* 'road' → *gurpide* 'cartpath'
ogi 'bread' + *-gin* 'maker' → *okin* 'baker'
ogi 'bread' + *-bil* 'round' → *opil* 'bread roll'
ogi 'bread' + *mehe* 'slender' (< **bene*) → *ope* 'slender cake'
bat 'one' + *-tu* VFS → *batu* 'unite, unify'
lot- 'tie, fasten' + *-garri* Instr NFS → *lokarri* 'link'
polit 'pretty' + *-ki* adv. → *poliki* 'slowly, gently'
ardi 'sheep' + *-kume* 'offspring' → *arkume* 'lamb'
ardi 'sheep' + *mihi* ~ *min* 'tongue' (< **bini*) → *arpin* 'plantain'
Lapurdi (a province) + *-tar* 'who is from' → *lapurtar* 'person from Lapurdi'

Oddly, this process seems rarely or never to occur when the first element contains a voiceless plosive with a following vowel: for example, *erreka* 'stream' + *bide* 'road' does not yield **errepide*, but rather *errekabide* 'riverbed' (compare *errege* + *bide* above). And, as explained above, it also fails to occur if vowel loss would expose a labial: compare *zubigin* 'bridge-builder' with *okin* 'baker'.

These developments are attested in medieval toponyms – hence *Eleizpe*, *Elexpe*, *Elizpea*, *Elizpuru*, etc., all from *el(e)iza* 'church' plus *-pe* 'below' or *buru* 'head', and *Erret Ihera* and *Erret Zubi*, from *errege* 'king' plus *ihera* 'mill' and *zubi* 'bridge'. They are not, however, attested in Aquitanian, in which we find names like *Senitennis* and *Cissonbonis*, with the combining forms *seni-* 'boy' and *gizon-* 'man', in place of the expected *sen-* and *giza-*.

Michelena (1951) proposes the following explanation for these reductions. The first element of a compound bore the primary stress (leading to reduction of the second element: see the next section). This would, in isolation, have fallen on the second syllable (section 3.13), but, in a compound, the stress would have been retracted to the first syllable by a process not unlike the familiar 'thirteen-men rule' of English. And the resulting initial stress in the first element would have led to the observed reductions.

In certain cases, the first element in a formation exhibits an unexpected *t*, or a devoicing of a following plosive, even though there is no plosive to act as a source. Examples:

su 'fire' + *argi* 'light' → *sutargi* 'firelight'

behi 'cow' + *-zai(n)* 'guardian' → *betzai(n)* 'cowherd'
zohi 'sod' + *erdi* 'half, middle' → *zoterdi* 'fallow land'
eri 'finger' + *gain* 'top' → *erkain* 'fingertip'
harri 'rock' + *gaitz* 'big' → *harkaitz* 'boulder'
bihi 'vintage, fruit, grain' + *gain* 'top' → *bikain* 'cream; excellent'
oin 'foot' + *huts* 'bare' → *ortuts* 'barefoot'

(The unexpected *r* in the word for 'barefoot' is discussed below.) It is not clear what to make of this, but Michelena (1977a: 224) is inclined to see this merely as an overgeneralization of *t*-formation in first elements, and not as a direct continuation of some earlier phonological development.

Historically, all such cases of final vowel loss clearly preceded the loss of intervocalic *-n-*. As a result, we find some striking alternations:

ardao ~ *ardo* 'wine' (< **ardano*) + *-tza* NFS → *ardantza* 'vineyard'
katea 'chain' (< **katena*) + *begi* 'eye' → *katenbegi* 'link of a chain'
harea 'sand' (< **harena*) + *erloju* 'clock' → *haren erloju* 'hourglass'
mihi ~ *min* 'tongue' (< **mini*) + *-tzo* NFS → *mintzo* 'conversation'
artzai(n) 'shepherd' (< **artzani*) + *-tza* NFS → *artzantza* 'shepherding'
balea 'whale' (< **balena*) + *bizar* 'beard' → *balenbizar* 'whalebone'
bigai ~ *biga* 'heifer' (< **bigane*) + *-txa* Dimin → *bigantxa* 'young heifer'

You will recognize here the loans from Latin CATENA 'chain', ARENA 'sand' and BALLAENA 'whale'.

Very commonly, a final *n* in the first element remains even when it is intervocalic:

sehi 'boy' (< **seni*) + *ar* 'male' → *senar* 'husband'
on 'good' + *-etsi* 'consider' → *onetsi* 'approve of'
zin 'truth' + *-etsi* 'consider' → *sinetsi* 'believe' (archaic *zinetsi*)

Such words commonly bear the aspiration in northern dialects (*senhar*, *onhetsi*, *sinhetsi*). It appears that the morpheme boundary was enough to prevent the regular loss of intervocalic *n*. But sometimes the *n* is lost:

sehi 'child' (< **seni*) + *aska* 'cradle' → *seaska* 'cradle'
on 'good' + *-asun* '-ness' → old B *o(g)asun* 'goodness'
zezen 'bull' + *-aga* → *Zezeaga* (house name in Gip)

The regular change of intervocalic *l* into *r* has led to comparable alternations:

euskara 'Basque language' (< **euskala*) + *-dun* 'having' → *euskaldun* 'Basque-speaker'
gari 'wheat' (< **gali*) + *buru* 'head' → *galburu* 'head of wheat'
haizkora 'axe' (< **haizkola*) + *-ari* Professional NFS → *haizkolari* 'woodcutter'
joare 'bell' (< **joale*) + *-dun* 'having' → *joaldun* 'bellwether'

eri 'illness' (< **eli*) + *gaitz* 'bad' → *elgaitz* 'fever'
bazkari 'lunch' (< **bazkali*) + *-tu* VFS → *bazkaldu* 'eat lunch'

The word for 'axe' is of course a loan from Latin ASCIOLA.

It is clear that this second alternation has sometimes undergone inversion:

zamari 'horse' (< Lat SAGMARIU 'pack-horse') + *-dun* 'having' → *zamaldun* 'horseman'
merkatari 'merchant' (< Lat MERCATU 'market' + Lat -ARIU Professional NFS) + *-goa* Collective NFS → *merkatalgoa* 'commerce'
amore 'love' (< Lat AMORE) + *-tsu* 'full of' → *amoltsu* 'affectionate'

Naturally, it is possible that a few of the cases cited earlier might also represent instances of such inversion, rather than an original intervocalic lateral, but we have no way of knowing in the case of native words.

First elements which end in *-n* to begin with often show unexpected behaviour. On the one hand, the nasal is sometimes simply lost; on the other, it occasionally develops into *-r*. The first of these may trigger some of the other developments we have already seen. Examples:

egun 'day' + *erdi* 'middle' → *eguerdi* 'noon'
egun 'day' + *-zki* NFS → *eguzki* 'sun' (western)
egun 'day' + *-ki* NFS → *eki* 'sun' (eastern)
egun 'day' + *aldi* 'time' → *eguraldi* 'weather'
gizon 'man' + *bide* 'way' → *gizabide* 'conduct; gentlemanly behaviour'
gizon 'man' + *-koi* 'fond of' → *gizakoi* 'man-chasing woman'
gizon 'man' + *aldi* 'time' → *gizaldi* '(a) generation'
jaun 'lord' + *-egi* 'place' → *jauregi* 'palace'
jaun 'lord' + *-etsi* 'consider' → *jauretsi* 'adore' (archaic)
jaun 'lord' + *-goa* Collective NFS → *jaurgoa* 'seigneury' (archaic)
belaun 'knee' + *-ik* adv. + *-ko* Adjectival → *belauriko* 'on one's knees'
oin 'foot' + *huts* 'bare' → *ortuts* 'barefoot'
oihan 'forest' + *zabal* 'wide' → *Oyarzabal* (surname)

The frequency of these curious developments varies according to item. For *gizon*, the combining form *giza-* is categorical and appears in dozens of derivatives; for *egun*, we most often find *egu-* before a consonant and *egur-* before a vowel, but there are exceptions, though we almost always find one or the other in ancient formations. For *jaun*, we nearly always find *jaur-* in its few derivatives; *jaun-* appears in the diminutive *jauntxo* 'local political bigwig' (a very common word in the medieval period), but diminutives often fail to undergo otherwise regular processes. For the other words cited, the combining form in *-r* is uncommon and appears only occasionally; more usually, the *-n* is retained in word-formation. (The unexpected *t* in the word for 'barefoot' was discussed above.)

First elements which end in *-r* to begin with often lose the rhotic in

word-formation; this is most obviously the case with words which end underlyingly in the tapped rhotic, but the trilled rhotic is also lost in certain circumstances. First, examples with the tapped rhotic (*ur* 'water', Definite *ura*, and *zur* 'wood', Definite *zura*):

ur 'water' + *bide* 'road' → *ubide* 'irrigation canal'
ur 'water' + *alde* 'side' → *uhalde* ~ *ugalde* 'riverbank'
zur 'wood' + *haitz* 'oak' → *zuhaitz* ~ *zu(g)a(i)tz* 'tree'
zur 'wood' + *bihotz* 'heart' → *zubihotz* 'heart of oak'

There are very few Basque words ending in a rhotic which surfaces as a tap in intervocalic position, but these two very common ones almost always lose the rhotic in word-formation in formations of any antiquity. Formations which retain the *r*, like *urgain* 'surface of the water' (*gain* 'top') may in most cases be safely regarded as of recent origin (compare *uhain* 'wave', probably of identical formation).

Many frequent words ending in a trilled rhotic behave in the same way (*lur* 'earth', Definite *lurra*), and so on:

lur 'earth' + *bizi* 'living' → *lubizi* 'landslide'
lur 'earth' + *hartz* 'bear' → *luhartz* 'scorpion'
lur 'earth' + *gorri* 'red' → *lugorri* 'cultivated land'
adar 'branch' + *begi* 'eye' → *adabegi* 'knot' (in a tree)
hamar 'ten' + *bi* 'two' → *hamabi* 'twelve'
behor 'mare' + *-ka* Dimin → *behoka* 'filly'
izter 'thigh' + *-egi* NFS → *izte(g)i* 'groin'

This puzzling process is clearly ancient: it is generally believed that *ibai* 'river' is derived from *ibar* 'water meadow, valley' by the addition of a suffix, and that the province name *Bizkai(a)* is derived in the same way from *bizkar* 'elevation' (both *ibar* and *bizkar* have trilled rhotics). It is, however, dead: more recent formations do not show the loss of a rhotic, and we find instead forms like B *urtxakur* 'otter' (*ur* 'water' + *txakur* 'dog') and *lurrikara* 'earthquake' (*lur* 'earth' + *ikara* 'trembling').

More dramatic reductions may occur. The verb *egin* 'do, make' has a combining form *egi-* which is frequently reduced, as in the combination of its verbal noun *egite* 'deed' with the suffix *-une*; this should have yielded **egita-une*, but the attested forms in B are *itaune* ~ *itauna* ~ *itun* 'question; advice; confession'. B *egiune* 'agreement' appears to represent *egi-* plus *-une* directly.

3.19.2 Second elements

We have already seen that the initial segment of a second element is sometimes modified by the influence of the preceding first element. Beyond this, however, there are certain types of reduction which may affect a second element, particularly in the oldest compounds. Such reductions chiefly affect disyllables which have a medial aspiration in the aspirating dialects, such as *zahar* 'old': when such an item stands in second position, it must lose its aspiration, because *h*, recall, cannot occur later in a word than the onset of the second syllable. Thus:

zahar 'old' → *-zar*
behe 'below' → *-be*, *-pe*
mehe 'thin, slender' → *-me*
bahe 'sieve' → *-be*
gehi 'quantity, material' → *-gei*, *-gai*
ohi 'habit' → *-(k)oi*

Examples:

ardi 'sheep' + *zahar* 'old' → *artzar* 'old sheep' (Z LN)
or 'dog' + *zahar* 'old' → *ozar* 'big dog' (L)
zur 'wood' + *mehe* 'slender' → *zume* 'osier' (common)
zeta 'silk' + *bahe* 'sieve' → *zetabe* 'fine sieve' (Z)

The frequent noun-forming suffix *-tza* appears many times in the eleventh-century *Reja da San Millán* as *-zaha*, and even today *-tza* appears in places as *-tzai-* when followed by the article *-a*. We may reconstruct **(t)zaha* + *-a* > **-tzaea* > *-tzaia*.

Various other reductions take place in this position:

luze 'long' → *-luz*
gabe ~ *bage* 'without' → *-ga*, *-ge*, *-ke*
duen 'who has' → *-dun*

Examples:

egi 'hill' + *luze* 'long' → *Egiluz* (toponym in Araba)
indar 'force' + *bage* 'without' → *indarga* 'feeble' (B)
bizar 'beard' + *duen* 'who has' → *bizardun* 'bearded (person)'

In a number of cases, it is clear, the form of a word occurring in second position has been generalized as an autonomous word. The anomalous eastern word *tzar* 'bad' (western *txar*) is derived from *zahar* 'old' used as a second element, as in *artzar* 'old sheep' above. The original *gehi* 'quantity, material' is now little used except in derivatives like *gehiago* 'more'; the form *gai* ~ *gei* has been generalized as the ordinary word for 'material, topic'. The now nearly universal *bi* 'two' doubtless results from a reduction of earlier *biga* in second position; we have clear textual evidence that this numeral was formerly postposed, as it still is in B today. The word *beltz* 'black' is a derivative of the familiar element **bel-* 'dark'; it is probably a generalization of a reduced form of **beletz*, which appears to be attested in Aquitanian. The word *ertz* 'edge' has an attested variant *eretz* which is doubtless original,

ertz resulting from compounds like *itsas-ertz* 'seaside'. The suffix *-koi* 'fond of' has yielded in Z an independent adjective *khoi* 'inclined (to do something)', and the suffix *-kume* 'offspring' has in B given rise to a word *kume* 'offspring (of an animal)' which now contrasts with the historically related *ume* 'child'. Since the eighteenth century, the very common suffix *-tasun* '-ness' has been used as a noun *tasun* 'quality'; in the modern language this even forms derivatives like *tasunezko* 'qualitative'. The anomalously formed words *talde* 'group', *toki* 'place' and one or two others in this vein, with their initial voiceless plosives, likewise result from the generalization of what were originally bound second elements. This is a topic which needs further investigation.

3.20 PHONOLOGICAL PECULIARITIES OF PROPER NAMES

The formation of surnames and of place names is discussed in some detail in Chapter 5. A few phonological points are, however, more appropriately discussed here.

Proper names formed by compounding and derivation are, in general, subject to the usual rules of word-formation discussed in the last section. In addition, however, they sometimes exhibit two further peculiarities which are not shared by other instances of word-formation.

The first peculiarity is confined to habitation names and to the French Basque Country, but within these limits it is categorical: any name which would otherwise end in a consonant automatically acquires a final *-e*. Thus, for example, the village whose name derives from *ur* 'water' and *epel* 'warm' is *Urepele*, with the final vowel apparently added merely to prevent the occurrence of a final consonant. Likewise, the famous resort known in French and in English as *Biarritz* (an obviously Basque name) is locally known as *Miarritze*, showing the final *-e* as well as a slightly unusual instance of the development of initial *b-* to *m-*. The same final vowel appears in *Azkaine* and *Berrogañe*, two villages whose names obviously include the final element *gain* 'top', and indeed in every single other such case north of the Pyrenees. It may well be that this final *-e* results from the renalysis of inflected forms in which the *e* is required for phonological reasons.

It appears that toponyms other than habitation names are not subject to this rule. In fact, the overwhelming majority of river names and mountain names in the north also end in vowels, but in almost every case the final vowel appears to be organic or (often in the case of mountain names) to be the article *-a*. None the less, there is a river called *Aran* and a small but famous mountain called *Larrun*, so I assume that vowel addition has not in general applied to such names. Nor is there any sign of the process south of the Pyrenees, where such habitation names as *Zarautz*, *Lemoiz*, *Ipazter*, *Laukiz*, *Andoain*, *Oiartzun*, *Eibar*, *Elgoibar* and *Ituren* are exceedingly common.

The second peculiarity affects both habitation names and (more especially, perhaps) surnames on both sides of the Pyrenees, but it is sporadic and far from categorical. This is the loss of an *initial* vowel. For example, the northern city and town called *Baiona* and *Baigorri* appear to be transparently from **ibai ona* 'the good river' and **ibai gorri* 'red river'. These etymologies are widely accepted (though see Chapter 5 for a different view), and they appear to show a loss of an initial vowel which is virtually unknown in ordinary instances of word-formation, except in a few cases of unusual length, such as B *bariku(a)* 'Friday', from **abari-baga-ko(-a)* 'dinner-without-*ko*(*-a*)' '(day) without dinner'. In the south, we have examples like *Zeanuri*, from **azenari-uri* 'fox-city', via the loss of intervocalic *-n-* and dissimilation of the first *-r-* to *-n-*: **Azenaruri* > **Azearuri* > **Azeanuri* > *Zeanuri*. (Alternatively, the **a-* might have been lost early, when the name was still six syllables long.) Several towns, such as *Barkoxe*, *Bardoze* and *Barrundia*, have names whose first element looks like *ibar* 'water meadow, valley', but the etymologies are not certain. There are, however, no more than a handful of certain or probable instances of such vowel loss in habitation names, and perhaps no case among other types of toponym.

Such vowel loss is noticeably more frequent in surnames. The exceedingly common element *etxe* 'house', whose combining form as a first element is variously *etxe-* or *etxa-*, not infrequently loses its initial vowel. Hence, the surname *Etxaberri* 'new house' has a variant *Txaberri*; *Etxaide* 'house-road' (*bide* 'road') is sometimes *Txaide*; and there are other similar cases. (The male given name *Xabier* (= Spanish *Javier*, English *Xavier*) is thought to be a development of *Etxaberri*, with Romance mediation to explain the vocalism.) A few such cases produce the startling result of creating a surname with initial *r-*, something which is otherwise absolutely prohibited in Basque. Thus, *Errekalde* 'beside the stream' has a common variant *Rekalde*, and *Errotaetxe* 'mill-house' has a very common variant *Rotaetxe*, and there are quite a few others.

Given the phonological impossibility of this last group, and given the generally highly sporadic nature of the process, it seems difficult to avoid the conclusion that, at least with the surnames, we are here not looking at a true Basque development at all, but only at a process which has sometimes applied in Romance to names of Basque origin to produce forms which have then been taken back into Basque. There is certainly some evidence for such vowel loss in Romance: for example, the Spanish surname *Chávez* is thought by many to derive from the common Basque surname *Etxabeste* 'other-house', but no such form as **Txabeste* or anything similar appears to be attested in Basque. On the other hand, there seems less reason to suspect Romance intervention in the toponyms.

3.21 PHONOLOGICAL TREATMENT OF RECENT LOANS

The phonology of Basque has been much disturbed by the influx of loan words over the last 2,000 years, and it is now far more similar to the phonologies of the surrounding Romance languages, especially Castilian, than was formerly the case. Consequently, recent loans typically adhere very closely in form to their Romance models.

Romance plosives are usually taken over unchanged, even after a nasal; final *t* and *k* are now tolerated, and plosive–liquid clusters are tolerated in both initial and medial position. Examples: *kobre* 'copper' (< Sp *cobre*), *ganibet* 'knife' (< Occ *ganibet*), *preso* 'prisoner' (< Sp *preso*), *truk* 'exchange' (< Fr *troc*), *pinta* 'pint' (< Sp *pinta*).

Castilian ⟨s⟩ (apical /s/) and ⟨z⟩ (/θ/) are borrowed as *s* and *z* respectively: *sozialista* 'socialist' (< Sp *socialista*). French laminal /s/ is conventionally borrowed as *z*, and hence French Basques have traditionally said, and written, *zozializta* for the same word, but this policy has been rejected by the Academy in favour of the southern usage. French /z/ is also borrowed as *z*: *zinga* 'zinc' (< *zinc*). Final sibilants are now accepted: *arroz* 'rice' (< Sp *arroz*), which now forms a minimal pair with native *arrotz* 'foreigner'. After *n* or *l*, a Romance sibilant is borrowed as an affricate: *garbantzu* 'chick-pea' (< Sp *garbanzo*), *boltsa* 'bag' (< Sp *bolsa*). Spanish ⟨ch⟩ and French ⟨tch⟩ are borrowed as *tx*: *Txina* 'China' (< Sp *China*). French ⟨ch⟩ is usually borrowed as *x*: *Xina* 'China' (< Fr *Chine*). Both French ⟨j⟩ and Castilian ⟨j⟩ are commonly borrowed as *j*: *jeneral* 'general' (< Sp *general*, Fr *général*). The fricative *f* is now accepted both initially and medially: *falta* 'error, fault' (< Sp *falta*), *Afrika* 'Africa' (< Sp *Africa*, Fr *Afrique*). But ⟨v⟩ is borrowed as *b*: *greba* 'strike' (< Fr *grève*).

A curiosity is that the final /d/ of Spanish, which is phonetically [θ], is borrowed as [θ] by some Spanish Basques, as in *UNE*[θ]*eko* 'of UNED' (Universidad Nacional de Educación a Distancia). Such a Spanish [θ] is never borrowed as *z*, while Spanish /θ/ is invariably borrowed as *z*, and never as [θ].

Romance nasals pose few problems. The palatal nasal is usually taken over as *ñ*: *koñak* 'cognac' (< Fr *cognac*, Sp *coñac*). Some speakers, however, unpack the palatal: *koinak*. Final *m* is not tolerated: *uraniuma* (< Fr *uranium*).

Liquids likewise pose few problems, but word-initial *r* is still not tolerated by most speakers, and a prothetic vowel must be added: *erradio* 'radium' (< Sp *radio*), *arropa* 'clothes' (< Sp *ropa*).

The five Spanish vowels are generally taken over without change, except that final /o/ is not infrequently borrowed as *u*: *banku* 'bank' (< Sp *banco*). The additional vowels and diphthongs of French pose problems. French ⟨u⟩ (/y/) is borrowed as *u*, while the other front rounded vowels are rendered variously, and the diphthongs are converted into Basque diphthongs according to their French spelling: *boitura* 'car' (< Fr *voiture*). The final schwa of Fench (which is pronounced in the south) is borrowed as *a*, as in the last example. A French nasal vowel is usually rendered with a following *n*: *trein* 'train' (< Fr *train*).

Particularly interesting is the treatment of Spanish final *-ón*, which is consistently borrowed as Basque *-oi*: *kamioi* 'truck, lorry' (< *camión*), *abioi* 'airplane' (< *avión*), *patroi* 'boss' (< *patrón*). The equivalent treatment does not apply to other vowels: *atun* 'tuna' (< *atún*), *delfin* 'dolphin' (< *delfín*), *afan* 'zeal' (< *afán*). And it happens even though final *-on* is perfectly possible in Basque: *gizon* 'man', *gabon* 'Christmas', *inon* 'anywhere'. The reason for this is a historical one. The Romance ending was originally *-one*, and this developed in Spanish merely by losing its final *-e*, yielding the modern *-ón*. In Basque, borrowed words with *-one* underwent the usual loss of intervocalic *n*, producing first *-oe* and then *-oi*. So, Romance words like **ratone* 'rat' and **razone* 'reason, right' have come into Castilian as *ratón* and *razón*, but into Basque as *arratoi* and *arrazoi*. Basque-speakers have therefore clearly concluded that Basque *-oi* is the normal equivalent of Spanish *-ón*, and so they continue to adjust new loans in *-ón* accordingly.

For further discussion, see Hualde (1993).