

AN ANALYSIS OF THE MONOSYLLABLES OF EARLY EUSKARA

Jon David Patrick and M. Elena Santiago Martínez

(University of Sydney)

Introduction

This paper analyses the monosyllables of Euskara both in terms of identifying the vocabulary and presenting a characterisation of its phonological profile. The operating paradigm for collecting and analysis of the words is described in brief. The methods of analysis are documented here for the sake of completeness and are intended to demonstrate two things. The methods are aimed at maximising the coverage of the known words in the language and are applied in a systematic way with a minimum of exception handling by extensive use of computer processing. The analysis is based around Mitxelena's model of the Pre-Basque and later phonological developments. The complete set of monosyllables has been collected and evaluated against all of the major commentators and separated into three putative time periods. The statistical distributions of the phonological shapes over the three time periods are presented. The extent to which the distributional statistics support or suggest anomalies and changes to the model are presented.

1. Preparation of Data

The data collection began with the dictionary of R. M. Azkue (1905). This dictionary is a useful starting point because as well as being a very large compendium of Basque words (39,664 headwords) only overtaken in the last 20 years, it is devoid of the neologisms and borrowings of the twentieth century. Most importantly, it is annotated for every word Azkue considered to be monomorphemic. However, the dictionary is not without the faults and limitations which could be expected from a lexicographer of his time and situation. For example, Azkue was a priest and as such he omitted some words he found offensive. He also identified some words whose authenticity is now doubted. Obviously, it does not hold the wealth of knowledge of the history of the language that has been accumulated throughout the twentieth century. Nevertheless, it is well respected as a major publication and a generally reliable scholarly piece of work, with only a very small amount of it being considered as "suspect".

The preparation of data has been a major task, since it is important to ensure a number of conflicting criteria are fulfilled. For the sake of completeness and

consistency, computers were used to do as much data processing as possible. This has contributed to the quality of the work in two ways. Firstly, it ensures comprehensiveness in that it is possible to analyse every word available in the lexicon, and, secondly it ensures consistent methods are used in processing each and every data item.

1.1. Identifying usable words

Our own work proceeded by entering manually into a computer file every word that Azkue indicated as monomorphemic. This amounted to 9913 words. Once the initial list of monomorphemes was compiled, it was necessary to separate the words that are considered relevant in contemporary terms. Significant work has proceeded on the historical study of the language, with its standardisation into Batua, in the identification of historical precedents of regional variants. The standardisation of the language has been managed by Euskaltzaindia, the Academy of the Basque Language, and it is a form of the language intended to represent a unification of the seven dialects.

All modern dictionaries are published in Batua but some also include non-Batua word-forms and some do not include all words identified as Batua. The first task to complete so as to be able to use these dictionaries for verification of the Azkue word list was to convert Azkue's peculiar orthography to a Batua form. Dr. Xabier Artola provided the rules that should be applied consistently for converting the orthography. These rules were then applied computationally to the whole list to ensure uniformity. However, this process did not solve all the problems with the orthography.

1.2. H-words

The distinction between the northern and southern treatment of the letter *-h* presented some difficulty for data identification and collection. In the southern dialects, *-h* is mute: a letter used in writing, not a phoneme used in speech. This is also the case in other neighbouring languages such as French and Spanish. In the northern dialects, however, *-h* forms part of the phonemic inventories of the dialects and it is represented in the IPA as [h], just as its English counterpart. In these northern dialects, the aspiration represented by this letter can be commonly found in many words, usually occupying intervocalic positions, but can also be found word-initially in otherwise onset-less syllables. In many instances, Euskaltzaindia has chosen to retain the letter in the orthography to represent the northern pronunciation, even if the southern speakers do not pronounce it. Hence they have published a list of *h*-words which contains words both with and without *-h* and thus indicate the orthographies of these words and other words for which the *-h* has been dropped in the Batua form. This list was studied to identify each of the words in Azkue so as to create a mapping of the word equivalence between Euskaltzaindia's *-h*-words and Azkue's monomorphemes.

The investigation produced some results interesting in their own right. Of the 307 words in Euskaltzaindia's *h*-list, 214 are not found in Azkue in the same form and out of these, 53 cannot be found in any other likely equivalent form. This

result prompted an investigation of the use of Euskaltzaindia's list and so we compared it to all words in the Sarasola dictionary. As a result, we discovered that 75 could not be found in the exact form in Sarasola and that 37 of those could not be found in another likely form. However, the time elapsed between the publication of both dictionaries can account for some of the missing words since we found around 14 *-h* words referenced in Sarasola as having fallen out of use. Thus, it is possible that the author might have chosen not to list the words which he considered to be more obsolete. By comparison, in Azkue we found 15 of the 37 words not found in Sarasola. Consequently, the 307 words were inserted into Azkue's list after we generated the Batua forms of the original Azkue forms.

There is a second issue pertaining to Azkue's entries. Whenever a word had two or more forms whereby one form had an *h* and another form did not, Azkue did not enter both as different headwords, but rather he entered one of them as an alternative orthography or sometimes just as a comment within an entry. Hence we found it necessary to compile a separate table of all alternative orthographies, which we denote as the Azkue-*h*-word list. These were later inserted into the original Azkue list. The same computational process applied to Euskaltzaindia's *h*-list was also applied to the Azkue *-h*-list and the two lists together produced four duplicate entries.

1.3. Word Duplication

The primary list of Azkue's monomorphemic words, after the basic orthographic issues were taken into account, produced some duplicates in a list of words. The source of duplicates is two-fold. First, some words, with different meanings in Azkue, yield identical orthographies with another word. For example, *haur* 'child, infant' (#1061) and *ahur* 'palm of the hand' (#1064) have different meanings but both reduce to *aur*, which is the entry under which they appear in Azkue. Second, the same word may be present with two or more different spellings which were classified by Azkue as different entries, but which are reduced to the same orthography after applying the Batua conversion rules. For example, *erhi* and *ehi* are variants of the same word 'finger' and reduce to the same Batua orthography, *eri*, and composite variant, *er* (#4017). These duplicates are retained throughout all the processing stages, however, they are removed for the calculation of the phonological statistics.

1.4. Constructing the word tabulations

Once the list of all monomorphemic words was compiled a tabulation of the word list for scrutiny and later computation was generated. Table 1.1 lists a few of the entries in that compilation. The tabulation consists of five columns. The first column shows the word in its Batua form. The second column shows the syllabic structure of the word. The third column shows the conversion of the word into a modern orthographical equivalent. The fourth column shows a variant computed to show the orthography with single letters for each phone, and the last column shows the word in its original form in Azkue. This shows the processing performed

Table 1.1

Sample of extracted data and primary computations of Azkue monomorphemic word list
(Ao = Azkue alternative orthography, Ez = Euskalzaindia orthography)

Batua	Consonant-vowel diphthong sequence	Modern orthographical equivalent	Single letter for each phone	Alternative form	Original Azkue form
A	V	a	a	—	A
aha	VCV	aha	aha	—	AHA
ahabia	VCVCVV	ahabia	ahabia	—	AHABIA
ahabi	VCVCV	ahabi	ahabi	ahabi (Ez)	AHABIA
ahago	VCVCV	ahago	ahago	—	AHAGO
aharatz	VCVCCVC	aharzaZ	aharatz	aharatz (Ao)	AHARTZARTZ

on the data set and enables the reader to reconstruct the sequence and rules of processing, from column five to one. This table is the basis of all analytical processing and the records are preserved in this form throughout all other procedures. Orthographic variants of the same words are removed manually at a later stage in the processing.

2. Data Analysis

2.1. Identifying monomorphemic words

The complete Azkue dictionary consists of 39,664 headwords of which 9913 are annotated as monomorphemic. There are an additional group of 1113 orthographic variants of monomorphemic headwords, yielding a total of 11026 word forms. The next stage of processing identified the words that are accepted by modern lexicographers as current, either by use or because of historical relevance. Words not identified in this group have fallen from favour because of antiquity alone or because of suspicions by latter day experts as to their authenticity. We divided the list into two groups which we call Common, for the words found in *at least* one modern lexicon and Uncommon for the words not found in *any* modern lexicon. The process used to create this subdivision was to compare the list of Azkue's monomorphemic words in their Batua orthography with each of the lexicons of four modern dictionaries, hereto labelled as Aulestia (1984), Kintana (1990), Morris (1998) and Sarasola (1995). A word that could be found in any one of these lexica was added to the Common list, otherwise it was placed in the Uncommon list. A word that is placed in the Common list has its orthographic variants removed from the Uncommon list. Later statistical processing has been performed on both lists in an attempt to identify structural differences should there be any.

Table 2.1 indicates the number of words from Azkue's monomorphemic Common list which can be found in each lexicon. The number of matches for each dictionary can be seen in the fourth column. These matches have been attached to

Table 2.1

Frequency of Common words represented in each lexicon and combination of lexica

Dictionary	Headwords	Unique words	%	Total words
Aulestia	49,600	275	4.40	3,407
Aulestia & Kintana		521	8.33	
Aulestia & Morris		17	0.27	
Aulestia & Sarasola		66	1.05	
Aulestia & Kintana & Morris		133	2.13	
Aulestia & Kintana & Sarasola		605	9.67	
Aulestia & Morris & Sarasola		68	1.09	
Aulestia & Kintana & Morris & Sarasola		1,790	28.61	
Kintana	43,553	1,561	24.95	5,407
Kintana & Morris		235	3.75	
Kintana & Sarasola		311	4.97	
Kintana & Morris & Sarasola		251	4.01	
Morris	23,373	58	0.93	2,647
Morris & Sarasola		95	1.52	
Sarasola	30,688	271	4.32	2,851
Total		6,257	100,00	

their entries in the database record. The resulting computation produced a list of 6257 Common words and 4772 Uncommon words.

As one of the objectives of this work is to identify the potentially older extant words of Basque, it is necessary to deal with the problem of the movement of modern phonology away from earlier forms. The work of Mitxelena (1985) and others give some indication of the orthographic elements that are likely to indicate early phonology, and particularly those elements that represent more recent developments and should therefore indicate late phonology, such as *f*, *j*, *ñ*, *tx*, *x* and *y*. Using this orthography the Common and Uncommon lists were divided further, yielding a subdivision of the data into old phonology and modern phonology.

Furthermore, there is another subdivision that can be made. The work of Mitxelena has led to a description of the structural forms of words in early Euskara, i.e. the period around the time of Roman contact, which Trask (1997) calls Pre-Basque. This structural description constitutes a set of rules which enables one to divide the data into word sets that do and do not conform to it. The application of these rules then divide the Old-Phonology sets into two sets, each of which we have labelled EARLY and LATE indicating structures in early-old-Basque and late-old-Basque respectively. We have performed the same statistical analysis on each of the word datasets, four old and two modern, with the aim of identifying structural consistencies and differences between them. The frequencies of the datasets for all of Azkue's monomorphemic words are shown in table 2.2.

Table 2.2

Subdivisions of Azkue's monomorphemic words into operational datasets

Datasets periods	Common	Uncommon
Early (old)	3337	2034
Late (old)	2210	1858
Modern	707	880
Total	6,254	4,772

At this point, the data can be processed automatically into one more categorisation, that is, the number of syllables in each word. Each cell in Table 2.2 can be thought of as made up of smaller cells of words grouped by number of syllables. While most words fall within the limit of three syllables, there are words with as many as eight syllables. Working progressively up through the syllable sizes is important for future research as compounding is a productive process of word formation in contemporary Basque, and seemingly this has always been the case. This is evidenced by the ratio of monomorphemic to compound and derivative words in Azkue's dictionary of about 1:4. Hence, to be able to produce a more accurate analysis of the list of Azkue's monomorphemic words, those which have fewer syllables need to be identified first.

We do not wish to imply that every word in the classes of Modern Phonology are necessarily recent additions to the lexicon, but only that *prima facie* they are not old words because of their contemporary orthography. With careful study, some of these words may well have earlier forms that can be included in the other lists. Nevertheless, we have processed both the Modern and the Uncommon lists as far as it is possible according to the same process as the Old-Phonology, Late and Early lists.

2.2. Syllable Analysis-Identifying Monosyllabic Words

Determination of the set of monosyllabic words in each of the datasets was done readily with a computer program. However, at this point a significant amount of human processing was introduced. In particular the orthographic form of a word belies many words of different meanings and forms. So the specific contents in the Azkue dictionary and other historical sources were scrutinised for relevant information about the phonological and orthographical shape of each word.

The two principal references for this work are the incomplete etymological dictionary of Agud and Tovar (1989) and Mitxelena's *Fonética histórica vasca (FHV)* (1985). Relevant comments from both were added as data, such as historical derivations and reconstructions as well as any extra information on alternative orthographies. This data forms the core data of the monosyllable analytical system and is stored in an XML format. It has 456 words with 1129 senses and runs to over 100 pages of printed text and is available from the authors.

The usefulness of computer processing has also been brought to bear on illuminating the range of monosyllabic words. From the phonological rules of early Basque all possible words supposedly available can be generated. A total of 912 putative monosyllabic words are derivable from these rules and presented in Appendix A1. They can act as a reference to the analysis of the etymology of known words and provoke questions as to why some phonological combinations were unused.

2.3. Monosyllabic words found in other Word Classes

There are other relevant word groups independent of the basic groups already derived from Azkue as a source of monosyllabic words. Four such groups are the synthetic verbs, the *i*-verbs, the *n*-verbs and the *tu/du*-verbs. The latter three are the verbs that end in *-i*, *-n* or *-tu/du*.

The list of synthetic verbs consist of 26 verbs which still retain a conjugational system, although it is clear that there was a larger number in the past and nowadays even some of the known forms have become obsolete. These verbs are easily identified (Table 2.3) and some of their stems can reasonably be used as another monosyllable dataset for this analysis. For a complete table see the Appendix A2.

Table 2.3
Synthetic Verbs

Synthetic verbs		
Atxiki	esan	iritzi
eduki	etorri	irudi
egin	etzan	izan
egon	ezagutu	jakin
ekarri	ibili	jario
eman	ihardun	jarraiki
entzun	ikusi	joan
erabili	irakin	ukan
eraman	iraun	

A purportedly older phenomenon is verbs with an *-i* ending. It is speculated with some conviction that this is a dative ending once productive for verbs, but no longer, which has become lexicalised for some verbs (Trask 1997). Removing the ending also reduces the number of syllables, rendering some of the verbs monomorphemic. These words are identified by a manual analysis of each of Azkue's senses within the entries. However, some of those senses are not verbs. Table 2.4 shows the list of verbs ending in *-i* and which were identified as monosyllabic within these constraints.

The historical analysis of Basque indicates that the verbalising suffix *-tu/du* was borrowed from Latin. Hence, the deletion of this suffix in a bi-syllabic word will

Table 2.4
Monosyllabic Verbs ending in *-i*

<i>-i</i> -VERBS				
babi	deitzi	eutzi	hazi	jauntsi
bazi	ebili	giri	hertsi	sari
bitzi	elki	hartzi	huzi	sarri
daitzi	euki	hasi	jarri	zezi
dauzi	eurki	hautsi	jaun (jaurri)	zoli

render the word-stem monosyllabic. However, as with the *-i* verbs, some such words are not verbs and manual analysis of each candidate was made to insert them into this list.

Table 2.5
Monosyllabic Verbs ending in *-tu/du*

<i>-tu/du</i> -VERBS		
aitu	heldu/heltu	gertu
hantu	histu	peitu
hartu	galdu	zutu
bortu	gandu	

There is one other verb class known to be old but which has uncertain interpretation. This class is the verbs with an *-n* ending, which is thought to represent some sort of nominaliser. There is only one monosyllable in this class, *jan*.

2.4. Explanations for the Distribution of Common & Uncommon Words

One Basque Lexicographer has offered this assessment for the membership of words in the Uncommon list (Morris, priv. comm, 2000). There are four principle reasons which could account for the rather large “uncommon” list.

1. Some are pure dialectal forms: *arroltza* = *arrautza*; *zuzna* = *susma*; *aztore* = *aztore* (even though the meaning is different).
2. Some of the words are mere localisms that may not even exist in the place that Azkue cites: *auteresti* (boast), *berru* (salamander)
3. Many of the words are from the extinct Erronkari dialect, hardly worthy of inclusion in a modern Basque dictionary destined for use throughout the Basque Country or obscure words from the small Zuberoan dialect.
4. Some of the words are so steeped in rural knowledge and hence so obscure, that their inclusion in modern dictionaries was thought unnecessary, such as words for a chestnut burr, etc. or words for an obscure, probably forgotten piece of farmer ingenuity such as words for a hinge or padlock.

2.5. The Compiled list of Monosyllables

The lengthy process of extracting words in their appropriate form produced a total of 517 monosyllabic words. This includes 506 from Azkue and another 11 derived from all the other sources already described. The distribution of words according to the six primary classes are shown in Table 2.7. This tabulation is not the final distribution of words, but only what is deducible from careful matching of headword entries from Azkue and the various modern dictionaries of Aulestia, Kintana, Morris and Sarasola. This table is modified by the analysis of the detailed etymology of each word. A list of the monosyllables that are omitted from the Azkue list but found in Sarasola's dictionary can be found in the Appendix (A3).

Table 2.7

Distribution of monosyllabic words over 6 classes prior to etymological analysis

Datasets periods	Common	Uncommon
Early (old)	211	26
Late (old)	97	52
Modern	81	50
Total	389	128

3. Phonological Profile of Monosyllabic Basque

This analysis covers the compiled list of monosyllables in their broad historical categories and discusses to a lesser extent the distributions across parts of speech.

3.1. The Compilation of Word Classes

The monosyllables of Early Euskara along with their translations into English and the internal cross-referencing of senses have been compiled in a table. However, this table is too large in size to appear in this paper. Suffice to say here that entries have been indexed into time periods as well as divided into parts of speech and that three different sources aforementioned have been consulted to determine whether the entry can or cannot be considered a monosyllable. These sources also present the evidence necessary to make a decision as to the form the word would or would not have had in early historical times. These categorical decisions in effect changed the distribution of monosyllabic words across categories to the revised version of Table 2.7, which is presented in Table 3.1. The word lists for each class are presented in Appendix A4.

Table 3.1

Distribution of monosyllabic words in 6 classes after etymological analysis

Datasets periods	Common	Uncommon
Early (old)	123	11
Late (old)	102	48
Modern	39	17
Total	264	76

3.2. Phoneme Distributions in Historical Classes of Monosyllabic Words

The distribution of the broad phonemic inventory over the 6 word classes has been compiled into a summary in A5 and an abbreviated summary in Table 3.2. Detailed inventories for each class are presented in the Appendix A6. The interpretation of these results must be treated carefully lest one falls into the trap of arguing the results show independent verification of the Mitxelena model of Pre-Basque. One should remember that the structure of Early, Late and Modern classes has been built to represent the model. Therefore, firstly we must analyse the results to determine if the distributional characteristics are consistent with the heuristic assessment of commentators (as no quantitative study of words in Basque has been published previously). And secondly, the results also be analysed contrastively to assess if the Uncommon partners reveal variations in the generating model, thereby indicating

Table 3.2

Percentage frequency of phoneme groups in historical groups of Basque monosyllabic words

Position in the Syllable	Phonemic Groups	Common Early	Common Late	Common modern	Uncommon early	Uncommon late	Uncommon modern
ONSET	Zero	14	11	10	9	10	6
	Labial	14	30	23	36	48	41
	Coronal	36	39	41	18	25	35
	Dorsal	37	20	26	36	17	18
MEDIAL	Vowel	72	78	72	54	71	82
	Diphthong	28	22	28	46	29	18
FINAL	Zero	33	13	26	18	19	12
	Labial	0	5	3	0	4	0
	Coronal	67	72	67	82	71	88
	Dorsal	0	11	5	0	6	0
Number of Words		123	102	39	11	48	17

its lack of coverage. Commentary on the results below is aimed at indicating elements of the model that are supported as well as variations that appear strong enough to warrant its reassessment.

3.2.1. *Common Early Words*

The onset position in these words is distributed as follows, coronals (36%), dorsals (36%), labials (14%) and zero (14%) dominated by *h*, *g*, *z*, *s* with frequencies of 20%, 12%, 15% and 11% respectively. In syllable final position, consonants are only coronals (67%) or zero (33%) dominated by *r*, *n*, *tz* and *ts* at 15%, 14%, 11% and 8% respectively. The medial position has a vowel to diphthong distribution of 72% to 28%.

3.2.2. *Common Late Words*

This class is dominated in the onset position by coronals (39%), and unlike the Common Early group, labials (30%) while dorsals are less frequent at 20%. The final word position has an increase in coronals (72%), dorsals (11%) and labials (5%), with a large reduction in zero (13%). In respect to the medial position, the results for this period are similar to the ones obtained in the previous period, vowels dominate at 78% and diphthongs follow at 22%.

3.2.3. *Common Modern Words*

These words have a distribution somewhat in between the Common Early and Common Late word classes. The onset position is dominated by coronals with 39%, whilst labials and dorsals appear about equal with 23% and 26% respectively. Coronals dominate the coda position with 67%, while zero is of a lesser frequency at 26%.

3.2.4. *Uncommon Early Words*

This class has a small number of words (11), hence the statistics are likely to have more uncertainty than the other classes. The most frequent consonant groups at onset are both labial and dorsal at 36% each. Only coronal and zero items are used in the final position at 82% and 18% respectively. The vowel-diphthong contrast is markedly different to all other groups with ratio of 54% to 46%.

3.2.5. *Uncommon Late Words*

At onset, the labials are the most frequent group with 46%, followed by coronals (25%) and dorsals (17%). The vowel-diphthong contrast is similar to other classes at 71% to 29% respectively. The coda position is dominated by coronals at 71% and followed by zero at 19%.

3.2.6. *Uncommon Modern Words*

Like the Uncommon Late word group the onset consonants are dominated by the labials (41%) and coronals (35%), the final consonants are mostly coronals (88%), and the vowels (88%) dominate over the diphthongs (12%).

3.3. Phoneme Distribution in Word Position

The phonemic profile of the monosyllabic words of Basque is presented in Appendix A5 cross-tabulated by historical class.

3.3.1. Phoneme Distribution in Word Initial Position

Considering the Common classes, it is evident that over time there has been a reduction in the usage of zero at onset, while there is a major shift in the labial consonants from initial *b* to *p*. In the modern era, the emergence of *f* is equally substantiated in the classes of Uncommon words. The development of the consonant pair clusters *pl*, *pr*, *bl*, *br*, *tr*, *dr*, and *dz* appear from onomatopoeic words and it is notable that *r* is present as the second member of each cluster.

The coronal consonants in word initial position also show a trend over the classes with *n*, *s*, *z*, and *l* being preferred in the early period and *t* and *z* in the later period. The presence of *tx* and *x* in the Modern classes is not notable in itself because it is a requirement of the model. However, its presence to the exclusion of all other coronals is notable. The dominant labial is *f*.

The dorsal consonants are dominated by *b*, *g* and *k* in the Early classes with consonant clusters more abundant in the Late classes and greater usage of *k*, with no further development in the Modern classes. The use of *k* is to be contrasted to Trask's comment "the evidence for word-initial *k* in Pre-Basque is scanty and doubtful" (1997: 128). In these results *k* is not only present in 4% of the Common Early set but 18% in the Uncommon Early set, only exceeded by the labial *b* (27%). Other results likewise offer a mild challenge to Trask's comment of the unlikely use of *m* in word initial position with frequencies of 5% and 9% in the Common and Uncommon Early classes respectively.

The distribution of *b* and *m* shows some disparateness with the medial vowels where although there are four examples of *be-* there are none of **me-*. This should be viewed in the context that *e* is used very sparingly in all word classes. Another distribution where *m* differs from *b* is that it is found with the diphthong *au* whereas *b* is not. These two facts give some support against the proposal that *m* is a later development from *b*. The evidence goes further to support the replacement of *b* by *f* as there is a total exclusion of *b* in the two classes of Common and Uncommon Modern where *f* makes its only appearance with 10% and 24% frequency respectively. Contrary to Trask's assertion (1977: 134) there is one word in grammatical categories with *m* initial, the pronoun *mu*.

3.3.2. Phoneme Distribution in Word Final Position

The most frequent termination of monosyllabic words over all classes is zero, consistent with the universal language dislike of codas. Out of the closed syllables, only coronals appear word finally in the following order: *r* (15%), *n* (14%), *tz* (11%) and *ts* (8%). There are even higher rates for the latter three at 27%, 18%, and 27% in the Uncommon class with frequency reductions to zero (18%) and *r* (0%). Further differences in the Uncommon Early class are the disappearance of *l*, *ltz*, and *rtz*.

The distributions of *l* and *r* do not appear to enlighten the problem of the merging of *l* into *r*. In the Common Early class, *r* is twice as frequent as *l*, but in the two later Common classes they are both infrequently represented. For the Uncommon classes

neither is present in the Early class and they are mutually exclusive in the later two classes. The one hint that they are mutually exclusive is the quite different distribution of their preceding vowel/diphthong where *l* has *i*, *ai* and *oi* and *r* has *o*, *u* and *au*, although both have *a* and *e* in the Common Early class.

The Common and Uncommon Late classes show an appearance of consonant clusters not present in the Early classes in this position. Noticeable increases in Common Late are *nk* (8%), *sk* (10%), *k* (11%), *z* (9%), *st* (9%) and *t* (8%). There is also the appearance of the labial *p* (5%).

The Common Modern class stands out for its high use of zero (26%) and *n* (15%). Uncommon Modern is distinctive for its relatively small use of zero (12%) and high use of *st* (12%), *tx* (29%), and *t* (18%).

The most frequent consonant cluster pairs are *sk*, *st*, *nk*, *np*, *nt*, *rt* and *st*. It is notable that *t* is used in combination with all other consonants and that all first members of the clusters are coronals which is to be contrasted to the collection of clusters at onset with an equal distribution of labials and coronals.

The status of the consonant clusters *rtz*, *ntz*, and *ltz* are uncertain in Basque etymology with debate as to whether they are single phonemes or consonant clusters. The statistical distribution over the sequence of the classes in this study shows some marked variation warranting interpretation. The cluster *rtz* is only found in Common Early and Common Late classes and *ltz* only in Common Early. On the other hand *ntz* is found well distributed across all 6 classes suggesting it has always been productive whereas the others fell from productive use after the Early period. Furthermore the variant cluster of *nts* appears in the Uncommon Late and Modern classes, and *rts* only appears in the Uncommon Modern class with no *lts* variant in any class. Given that all these forms are permissible under the Mixelena model of Pre-Basque (Trask 1997: 127) it may be worthwhile to include them in the inventory of Early Common words. These results could be interpreted as supporting the thesis that *ntz* is a true phoneme, and that *ltz* and *rtz* are the result of occasional syncope, while *nts* and *rts* remain enigmatic.

3.3.3. Vowel and Diphthong Distribution in word Medial Position

A summary picture of the usage of vowels and diphthongs in each historical class is presented in Table 3.4 and more details on the relationships between vowels and diphthongs and onset and final consonants is presented in tables A6-A8 in the Appendix. They show that for Common Early the most prevalent vowels are *a* (18%), *o* (16%), *u* (15%), and *i* (14%). The Common Late results show a preponderance of *a* (28%), *i* (20%), *u* (16%), *o* (10%) but a very low score for *e* (4%). The Common Modern results indicate the strong role of *a* (26%), and *i* (21%). Uncommon Early has only a few words and shows a trend for most usage of *a* (27%). Uncommon Late shows a slight shift in this pattern with *i* the most frequently used at 29%, *a* at 17%, and the diphthong *au* 17%. Uncommon Modern follows the same trend with *i* (24%), *u* (24%), *a* (18%), and *o* (18%).

The outstanding fact in the usage of vowels and diphthongs is the restricted exploitation of *e*, which is the least frequently used vowel in all classes while *a* and *i* are the most frequent. The diphthongs are thinly represented in most classes except in Common Early and Common Late but dominated by *au* and to a lesser extent *ai*.

Table 3.4
Frequencies of vowels and diphthongs in monosyllables for historical classes of Basque

Phonemes	Common Early	Uncommon early	Common Late	Uncommon late	Common modern	Uncommon modern
<i>a</i>	22	3	29	8	10	3
<i>e</i>	11	—	4	3	—	—
<i>l</i>	17	1	20	14	8	4
<i>o</i>	20	—	10	4	5	3
<i>u</i>	19	2	16	5	5	4
<i>ai</i>	11	2	2	2	5	—
<i>au</i>	9	2	12	8	3	1
<i>ei</i>	3	—	3	1	1	1
<i>eu</i>	1	—	1	1	1	—
<i>oi</i>	7	—	4	1	1	—
<i>ui</i>	3	1	1	1	—	1
Number of words	123	11	102	48	39	17

3.4. Review of Phonemic Distribution by Parts of Speech

The previous sections have concentrated on reviewing the overall distribution of phonemes, yet within categories of parts of speech there are distributions at variance with the overall patterns. We have reproduced the detailed tables representing the figures in Table 3.5 in Appendix A6 for each cell with a frequency greater than 10 as any smaller values are less likely to be indicative of systematic trends.

Table 3.5
The frequencies of part of speech classes in the monosyllabic words of Basque for 6 historical classes

Part of Speech	Common Early	Common Late	Common Modern	Uncommon Early	Uncommon Late	Uncommon Modern
Nouns	91	36	21	7	22	10
Pronouns	14	6	4	1	5	—
Adjectives	26	14	5	2	10	3
Adverbs	9	4	1	1	1	—
Determiners	8	2	—	—	—	—
Verbs	17	7	3	1	2	—
Interjections	20	17	11	—	7	4
Onomatopoeia	4	42	6	—	5	3

3.4.1. Nouns

The largest inventory of words is nouns and the phoneme distributions are produced in detail in Appendix A6.1-6 for the 6 periods. For Common Early nouns, at onset coronals and dorsals are most represented with 37% of each and labials and zero with 13% and 12 % respectively. Within those groups, the most frequent phonemes are *h* (22%), *z* (13%) and *s* (12%). The word final consonants are *n* (15%), *r* (14%), *tz* (13%), *ts* (10%). The medial position is represented by 67% vowels and 33% diphthongs with the most common vowel being *a*, followed by *i*, *o*, *u* and the diphthong *ai*.

The set of only seven Uncommon Early nouns is not large enough to make reliable assessments of the overall distributions. However, two obvious differences are the higher representation of *b* as initial onset at 29% and *ts* as word final with 43%.

Common Late nouns are dominated at onset by *p* (25%) and *t* (22%). The word final consonants are thinly distributed across 16 clusters with zero (17%), *t* (14%) and *k* (11%) the most dominant. The most frequent vowels are *a* and *u* and there is a dispersion of diphthongs over *ai*, *au*, *ei*, *oi*.

Uncommon Late nouns show a marked preference for *p* (27%) at onset and zero (18%) in the word final position. The preference for vowels is *i* and *a* and for the diphthong *au*.

Common Modern shows its strongest preference at onset for *tx* (28%) and *j* (33%). In the word final position the preferences are zero (24%), *n* (19%) and *tx* (14%). The most common vowel and diphthong are *a* and *ai*.

Uncommon Modern is a set of only 10 words, and as such trends are not very reliable. The most common onset phonemes are *f* (20%) and *tx* (20%). The most preferred word final consonant is *tx* (30%) and the most frequent medial vowels are *i*, *a* and *u*.

3.4.2. Pronouns

The only class of pronouns where trends may be reliable is the Common Early, with a set of 14 words (A6.7). At onset *h* (29%), *g* (21%), *n* (21%) and *z* (14%) are the most frequent. Only 3 items in word final position are present and they are zero (57%), *n* (21%) and *r* (21%). The vowels are evenly used along with the diphthongs *au*, *eu*, and *oi*.

3.4.3. Adjectives

There are only 3 sets of adjectives with sufficient class size to provide a reasonable picture of trends. They are Common Early (26), Common Late (14) and Uncommon Late (10) (Tables A6.8-10).

The Common Early class uses *h* (23%), *g* (12%), *s* (23%) and *z* (15%) at onset and as word final *ts* (23%), *tz* (19%), *r* (15%), *n* (12%), *l* (12%), and zero (15%). All vowels are used but dominated by *i*, and diphthongs *ai*, *au* and *oi* are used.

The Common Late class prefers at onset *t* (29%) *tz* (14%), *z* (14%), *d* (14%), and *p* (14%). The word final position most frequently uses *t* (21%) and *rt* (14%) with a single occurrence of 9 other consonant clusters. The most frequent vowels are *u* and *i* and the diphthong *oi*.

The Uncommon Late class prefers at onset *p* (20%) and *m* (20%) and at word final *ts* (20%) and *s* (20%). The vowels are all used and the diphthongs *ai*, *au* and *eu*.

3.4.4. Verbs

The only class of verbs of sufficient size fin which to observe trends is Common Early (A6.11). At onset *b* (29%) and *g* (18%) are the most prevalent, and for word final, *r* (35%) and *l* (18%). The vowels *a*, *e*, and *u* are used equally frequently and the only diphthong used is *au*.

3.4.4.1. Comparison to Synthetic Verbs

The phoneme distribution of the Synthetic verbs is shown in comparison with Common Early words and verbs distributions in Table 3.6 below. The onset consonants are zero (18%), labials (24%), coronals (24%) and dorsal/gutturals (35%), with the most frequent phonemes of *b* (12%), *m* (12%), *k* (24%), and *g* (24%). This shows significantly higher use of *k* and the labials compared to Common Early words and the verbs, with a concomitant reduction in the use of *b*. Clearly, this further contrasts with Trask’s (op. cit.) view that *k* is used sparingly in Pre-Basque.

Table 3.6

Comparison of the relative frequencies (%) of the most frequent phonemes for Synthetic Verbs and Common Early words and Verbs

	Preceding Consonants									Vowels	Succeeding Consonants					
	Zero	<i>b</i>	<i>M</i>	<i>n</i>	<i>s</i>	<i>z</i>	<i>k</i>	<i>H</i>	<i>g</i>	V	Zero	<i>ts</i>	<i>tz</i>	<i>n</i>	<i>r</i>	<i>l</i>
Synthetic	18	12	12	0	6	0	24	0	12	100	6	0	6	53	12	12
CE Words	11	10	5	4	11	15	3	21	13	71	31	9	11	15	16	7
CE Verbs	6	6	0	6	12	6	0	29	18	88	24	6	6	6	35	18

The most frequent consonants in word final position are distributed in *n* (53%), *l* (12%) and *r* (12%). Thus *n* is the dominant phoneme relative to Common Early words and verbs but there is also usage, even if low, of zero and *r*.

The most frequent vowels are *a*, *i* and *u* with no usage of *e* nor diphthongs whatsoever. Only vowels are used in synthetic verbs which is 12-29% higher than the Common Early classes.

The results from Table 3.7 compared to Table 3.8 indicate *-i* and *-n* are mutually exclusive for monosyllabic verbs because after a final *-i* is removed there are no final *-n* consonants. So *-i* is not attached after *-n* but is attached after *-d*, *-k*, *-l*, *-r*, *-s*, *-tz*. These distributions support Trask’s thesis (1997: 212) for the role of *-i* as creating a perfective particle and the unexpected appearance of *-i* in their combining forms in northern dialects.

Table 3.7

Frequency of final phones of 26 synthetic verbs

<i>i</i>	<i>n</i>	<i>o</i>	<i>u</i>
10	14	1	1

Table 3.8

Frequency of final phones of 26 Synthetic verbs after final *i* is removed

<i>d</i>	<i>k</i>	<i>l</i>	<i>n</i>	<i>o</i>	<i>r</i>	<i>s</i>	<i>tz</i>	<i>u</i>
1	3	2	14	1	2	1	1	1

3.4.5. Summary

The comparison of the various part of speech classes reveals a number of interesting variations.

In Table 3.9 it can be seen that zero is about twice as frequent at onset for nouns than for other word types. No pronouns begin with *b* while it is equally frequent for nouns, adjectives and verbs. The use of *n* at onset restricted almost solely to pronouns is distinctive along with the absence of *s*, which is preferred mostly for adjectives at twice the rate for nouns and verbs. Compared to all three classes of nouns, pronouns and adjectives, *z* is used at half the frequency for verbs. Generally, *b* is the most frequent phoneme used for all word types with *g* the second most common apart from adjectives.

Table 3.9
Comparison of the relative frequencies (%) of the most frequent phonemes
for Common Early word classes by Part of Speech type

	Preceding Consonants							Vowels	Succeeding Consonants					
	Zero	<i>b</i>	<i>n</i>	<i>s</i>	<i>z</i>	<i>h</i>	<i>g</i>	V	Zero	<i>ts</i>	<i>tz</i>	<i>n</i>	<i>r</i>	<i>l</i>
Nouns	12	8	1	12	13	22	12	67	28	10	13	15	14	6
Pronouns	7	0	21	0	14	29	21	77	57	0	0	21	21	0
Adjectives	8	8	0	23	15	23	12	77	15	23	19	12	15	12
Verbs	6	6	6	12	6	29	18	88	24	6	6	6	35	18

The medial phonemes show a greater preference for vowels over diphthongs for verbs (88%) but to a much less degree for nouns (67%).

The final word position shows the most distinctive bias in pronouns with 57% zero, while the other classes are less than half that amount. Pronouns also show a distinct bias for only coronals *n* and *r*, whereas adjectives show the highest preference for *ts* and *tz*. While nouns have the most balanced distribution across all phonemes, the verbs have a distinct preference for final *r*, *l* and zero.

3.5. Review of Minimal Lists

The role of Interjections and Onomatopoeia is uncertain in defining the phonemic repertoire of a language and so a further analysis is performed with them excluded from the word inventory. To further simplify the statistical profile of the inventory any word duplicates are omitted in computing Table 3.10. (Whole tables shown in Appendix A7).

The position of onset has more phonemic variation than that of the final position. While zero is used relatively constantly varying from 4-11% across historical classes, the labials, coronals and dorsals have greater variation. In the move from Common Early to Common Late, there is an increase in the use of labials and coronals with a concomitant decrease in dorsals with a shift in the opposite direction for all three in moving to Common Modern. The same trend is

found in moving from Uncommon Early to Uncommon Late. However in moving from Uncommon Late to Uncommon Modern, the dominant use of coronals increases, while the use of dorsals continues to decrease.

In the word final position the trends are more stable. Labials and dorsals are used only infrequently in all classes. The zero in final position drops considerably from Common Early to Common Late, but rises again in Common Modern; whereas it falls across the three Uncommon classes. This position is dominated by the coronals with relatively stable use across the three Common classes and increasing use across the Uncommon classes.

In the medial position the same trend for vowels contrasted to diphthongs is seen as with many other cases. There is a rise in the relative frequency of vowels from Common Early to Common Late, which falls back significantly in moving to Common Modern. However, there is a continuous increase in the use of vowels relative to diphthongs moving from Uncommon Early to Modern.

Table 3.10

Relative frequencies (%) of phonemes of monosyllables in historical Basque, excluding Interjections, Onomatopoeia and duplicates (Counts are presented in brackets)

Word Position	Phonemic Groups	Common Early	Common Late	Common modern	Uncommon early	Uncommon late	Uncommon modern
ONSET	Zero	11 (13)	4 (2)	3 (1)	9 (1)	8 (3)	8 (1)
	Labial	15 (17)	33 (17)	24 (7)	36 (4)	47 (17)	33 (4)
	Coronal	37 (43)	50 (26)	41 (12)	18 (2)	31 (11)	42 (5)
	Dorsal	37 (43)	13 (7)	31 (9)	36 (4)	14 (5)	10 (2)
MEDIAL	Vowel	71 (82)	79 (41)	66 (19)	55 (6)	75 (27)	83 (10)
	Diphthong	29 (34)	21 (11)	34 (10)	45 (5)	25 (9)	17 (2)
FINAL	Zero	31 (36)	13 (7)	24 (7)	18 (2)	14 (5)	8 (1)
	Labial	0 (0)	2 (1)	0 (0)	0 (0)	0 (0)	0 (0)
	Coronal	70 (80)	77 (40)	72 (21)	82 (9)	78 (28)	92 (11)
	Dorsal	0 (0)	8 (4)	3 (1)	0 (0)	8 (3)	0 (0)
Number of Words		116	52	29	11	36	12

4. Conclusions

The various analyses presented in this paper show that there is a certain amount of the monosyllabic lexicon of Basque that is conformant with Mitxelena's model

of Pre-Basque. However it points to a number of problems in the comprehensive analysis of the lexicon as well as inconsistencies between the distributional characteristics and the heuristics devised by commentators up to this point in time.

Our analysis of the monosyllabic words is not intended here as a definitive assessment or otherwise of the Mitxelena model but rather the beginnings of a statistical examination that is thorough in its methods and application so that no evidence is left unassessed. The use of computer processing has assisted at many levels in the process. Firstly, to verify and merge the various sources of the lexicon. Secondly, to present the many different evidences in a systematic and coherent manner, in a HTML table, that ensures the coverage of sources is comprehensive and search and retrieval of the evidences is maximally facilitated. Finally the computation functions of spreadsheets are used to generate tabulations of all configurations of the data so as to assist multiple viewpoints from which to scrutinise the data.

Our analysis points to the need for a small number of innovations in the model of the development of modern Basque from its ancient forms. Namely:

- the usage of the phonemes *b* and *m* may be more complicated than a simple pathway of *b* giving way to *m*.
- the distributional differences between the usage of phonemes between synthetic verbs and finite verbs requires explanation.
- the paucity of the vowel *e* in all word classes.
- the difference in the usage of diphthongs against vowels in all word classes except nouns.
- the assessment that the consonant cluster of *ntz* for word final position is more likely to be a single consonant whereas *rtz* and *ltz* are less likely to be so.
- Stronger evidence for the use of *k* at word onset in early Basque.
- Stronger evidence for the use of *h* at word onset in early Basque.

This analysis only touches in the simplest way the potential to mine and understand of the etymology of Basque from the monosyllabic word list using computer supported methods. The authors hope that with provision of their data in computer readable form and their methods in a reproducible description that others can take on the larger task from their simple beginnings.

5. Acknowledgments

The authors would like to thank everybody whose contribution, regardless of size, aided the completion of this project and indeed made this project possible. Our academic colleagues, members of the IXA talde: Xabier Artola, Iñaki Alegria, Eneko Agirre and Xabier Arregi who provided many electronic resources. Linguists Jose Ignacio Hualde, Beñat Oyharçabal, Jon Ortíz de Urbina, the late Larry Trask, Mikel Morris, Joseba Lakarra, Xabier Ormazabal and Myriam Uribe-Etxebarria for their continued encouragement. Much of the programming effort was injected by Eva Heinrich, Qiao Ngyuen and Robert Munro.

6. References

- Agud, M. & Tovar, A., 198), *Diccionario etimológico vasco*. Gipuzkoako Foru Aldundia: Donostia.
- Atkins, B. T. et al., 1996 (3rd ed.), *Collins French concise dictionary*. HarperCollins Publishers: Glasgow.
- Aulestia, G., 1984, *Basque-English dictionary*. University of Nevada Press: Reno.
- Azkue, R. M., 1984, *Diccionario vasco-español-francés*. Euskaltzaindia: Bilbo.
- Delbridge, A. et al (eds.), 1982, *The Macquarie dictionary*. Macquarie Library Pty Ltd: Dee Why.
- ARTFL, French-English Dictionary. Url: http://humanities.uchicago.edu/forms_unrest/FR-ENG.html.
- ARTFL Project: Webster's Revised Unabridged Dictionary, 1913 Edition. Url: http://humanities.uchicago.edu/forms_unrest/webster.form.html.
- García-Pelayo y Gross, R., 1986, *Larousse: nuevo diccionario manual ilustrado*. Printer Industria Gráfica S.A.: Barcelona.
- Glare, P. G. W. (ed.), 1973, *Oxford Latin Dictionary*. London: Oxford U.P.
- Larousse, 1998, *Diccionario español-francés/français-espagnol*. Larousse Editorial, S.A.: Barcelona.
- Mitxelena, K., 1985, *Fonética histórica vasca*. Exma. Diputacion Foral de Gipuzkoa: San Sebastian.
- Morris, M., 1998, Morris Student Plus: *Euskara-ingeles/English-Basque*. Klaudio Harluxet Fundazioa: Donostia
- Styles Carvajal, C. & Horwood, J. (eds.), 1996, *The concise Oxford Spanish dictionary*. Oxford U.P. Oxford,
- Sarasola, I., 1995, *Euskal histegia*. Kutxa Fundazioa: Donostia.
- Trask, R. L., 1997, *The history of Basque*. Routledge: London.
- U.Z.E.I., 1990, *Biologia, landare eta animalien izenak izendegia*. Elkar S.A: Donostia.

APPENDIX

A1. List of all possible monosyllables from the rules of Mitxelena's model of Pre-Basque (the Early period)

ail	ailtz	ain	aintz	air	airtz	aits	aitz	al	altz
an	antz	ar	artz	ats	Atz	aul	aultz	aun	auntz
aur	aurtz	auts	autz	ba	Bai	bail	bailtz	bain	baintz
bair	bairtz	baits	baitz	bal	Baltz	ban	bantz	bar	bartz
bats	batz	bau	baul	baultz	Baun	bauntz	baur	baurtz	bauts
bautz	be	bei	beil	beiltz	Bein	beintz	beir	beirtz	beits
beitz	bel	beltz	ben	bentz	Ber	bertz	bets	betz	beu
beul	beultz	beun	beuntz	beur	beurtz	beuts	beutz	bi	bil
biltz	bin	bintz	bir	birtz	Bits	bitz	bo	boi	boil
boiltz	boin	bointz	boir	boirtz	boits	boitz	bol	boltz	bon
bontz	bor	bortz	bots	botz	Bu	bui	buil	builtz	buin
buintz	buir	buirtz	buits	buitz	Bul	bultz	bun	buntz	bur
burtz	buts	butz	eil	eiltz	Ein	eintz	eir	eirtz	eits
eitz	el	eltz	en	entz	Er	ertz	ets	etz	eul
eultz	eun	euntz	eur	eurtz	Euts	eutz	ga	gai	gail
gailtz	gain	gaintz	gair	gairtz	gaits	gaitz	gal	galtz	gan
gantz	gar	gartz	gats	gatz	Gau	gaul	gaultz	gaun	gauntz
gaur	gaurtz	gauts	gautz	ge	Gei	geil	geiltz	gein	geintz
geir	geirtz	geits	geitz	gel	geltz	gen	gantz	ger	gertz
gets	getz	geu	geul	geultz	geun	geuntz	geur	geurtz	geuts
geutz	gi	gil	giltz	gin	gintz	gir	girtz	gits	gitz
Go	goi	goil	goiltz	goin	gointz	goir	goirtz	goits	goitz
gol	goltz	gon	gontz	gor	gortz	gots	gotz	gu	gui
guil	guiltz	guin	guintz	guir	guirtz	guits	guitz	gul	gultz
gun	guntz	gur	gurtz	guts	gutz	ha	hai	hail	hailtz
hain	haintz	hair	hairtz	hais	haitz	hal	haltz	han	hantz
har	hartz	hats	hatz	hau	haul	haultz	haun	hauntz	haur
haurtz	hauts	hautz	he	hei	heil	heiltz	hein	heintz	heir
heirtz	heits	heitz	hel	heltz	hen	hentz	her	hertz	hets
hetz	heu	heul	heultz	heun	heuntz	heur	heurtz	heuts	heutz

Hi	hil	hiltz	hin	hintz	hir	hirtz	hits	hitz	ho
hoi	hoil	hoiltz	hoin	hointz	hoir	hoirtz	hoits	hoitz	hol
holtz	hon	hontz	hor	hortz	hots	hotz	hu	hui	huil
huiltz	huin	huintz	huir	huirtz	huits	huitz	hul	hultz	hun
huntz	hur	hurtz	huts	hutz	il	iltz	in	intz	ir
irtz	its	itz	ka	kai	kail	kailtz	kain	kaintz	kair
kairtz	kaits	kaitz	kal	kaltz	kan	kantz	kar	kartz	kats
katz	kau	kaul	kaultz	kaun	kauntz	kaur	kaurtz	kauts	kautz
Ke	kei	keil	keiltz	kein	keintz	keir	keirtz	keits	keitz
kel	keltz	ken	kentz	ker	kertz	kets	ketz	keu	keul
keultz	keun	keuntz	keur	keurtz	keuts	keutz	ki	kil	kiltz
kin	kintz	kir	kirtz	kits	kitz	ko	koi	koil	koiltz
koin	kointz	koir	koirtz	koits	koitz	kol	koltz	kon	kontz
kor	kortz	kots	kotz	ku	kui	kuil	kuiltz	kuin	kuintz
kuir	kuirtz	kuits	kuitz	kul	kultz	kun	kuntz	kur	kurtz
kuts	kutz	la	lai	lail	lailtz	lain	laintz	lair	lairtz
laits	laitz	lal	laltz	lan	lantz	lar	lartz	lats	latz
lau	laul	laultz	laun	launtz	laur	laurtz	lauts	lautz	le
lei	leil	leiltz	lein	leintz	leir	leirtz	leits	leitz	lel
leltz	len	lentz	ler	lertz	lets	letz	leu	leul	leultz
leun	leuntz	leur	leurtz	leuts	leutz	li	lil	liltz	lin
lintz	lir	lirtz	lits	litz	lo	loi	loil	loiltz	loin
lointz	loir	loirtz	loits	loitz	lol	loltz	lon	lontz	lor
lortz	lots	lotz	lu	lui	luil	luiltz	luin	luintz	luir
luirtz	luits	luitz	lul	lultz	lun	luntz	lur	lurtz	luts
lutz	ma	mai	mail	mailtz	main	maintz	mair	mairtz	maits
maitz	mal	maltz	man	mantz	mar	martz	mats	matz	mau
maul	maultz	maun	mauntz	maur	maurtz	mauts	mautz	me	mei
meil	meiltz	mein	meintz	meir	meirtz	meits	meitz	mel	meltz
men	mentz	mer	mertz	mets	metz	meu	meul	meultz	meun
meuntz	meur	meurtz	meuts	meutz	mi	mil	miltz	min	mintz
mir	mirtz	mits	mitz	mo	moi	moil	moiltz	moin	mointz
moir	moirtz	moits	moitz	mol	moltz	mon	montz	mor	mortz

mots	motz	mu	mui	muil	muiltz	muin	muintz	muir	muirtz
muits	muitz	mul	multz	mun	muntz	mur	murtz	muts	mutz
na	nai	nail	nailtz	nain	naintz	nair	nairtz	naits	naitz
nal	naltz	nan	nantz	nar	nartz	nats	natz	nau	naul
naultz	naun	nauntz	naur	naurtz	nauts	nautz	ne	nei	neil
neiltz	nein	neintz	neir	neirtz	neits	neitz	nel	neltz	nen
nentz	ner	nertz	nets	netz	neu	neul	neultz	neun	neuntz
neur	neurtz	neuts	neutz	ni	nil	niltz	nin	nintz	nir
nirtz	nits	nitz	no	noi	noil	noiltz	noin	nointz	noir
noirtz	noits	noitz	nol	noltz	non	nontz	nor	nortz	nots
notz	nu	nui	nuil	nuiltz	nuin	nuintz	nuir	nuirtz	nuits
nuitz	nul	nultz	nun	nuntz	nur	nurtz	nuts	nutz	oil
oiltz	oin	ointz	oir	oirtz	oits	oitz	ol	oltz	on
ontz	or	ortz	ots	otz	sa	sai	sail	sain	sair
saits	sal	san	sar	sats	sau	saul	saun	saur	sauts
se	sei	seil	sein	seir	seits	sel	sen	ser	sets
seu	seul	seun	seur	seuts	si	sil	sin	sir	sits
so	soi	soil	soin	soir	soits	sol	son	sor	sots
su	sui	suil	suin	suir	suits	sul	sun	sur	suts
uil	uiltz	uin	uintz	uir	uirtz	uits	uitz	ul	ultz
un	untz	ur	urtz	uts	utz	za	zai	zail	zailtz
zain	zaintz	zair	zairtz	zaitz	zal	zaltz	zan	zantz	zar
zartz	zatz	zau	zaul	zaultz	zaun	zauntz	zaur	zaurtz	zautz
ze	zei	zeil	zeiltz	zein	zeintz	zeir	zeirtz	zeitz	zel
zeltz	zen	zentz	zer	zertz	zetz	zeu	zeul	zeultz	zeun
zeuntz	zeur	zeurtz	zeutz	zi	zil	ziltz	zin	zintz	zir
zirtz	zitz	zo	zoi	zoil	zoiltz	zoin	zointz	zoir	zoirtz
zoitz	zol	zoltz	zon	zontz	zor	zortz	zotz	zu	zui
zuil	zuiltz	zuin	zuintz	Zuir	zuirtz	zuitz	zul	zultz	zun
zuntz	zur	zurtz	zutz						

A2. The Synthetic Verbs

Infinitive form	Combining morpheme	Removing final <i>-I</i>	Removed leading morpheme	English meaning	Example
atxiki	-txe-	atxik-	-txik-	grab, hold, bite, cling	datxikio
eduki	-u-	eduk-	-duk-	have	du
egin	-gi-	egin	-gin	do/make	dagi
egon	-go	egon	-gon	be (temporary states)	dago
ekarri	-kar-	ekar-	-kar-	bring	dakar
eman	-ema-	eman	-man	give	bema
entzun	-ntzu-	entzun	-ntzun	listen	dantzun
erabili	-rabil-	erabil-	-bil-	use, move, be busy doing sth	darabil
eraman	-raman-	eraman	-man	carry	darama
esan	-io-	esan	-san	say	dio
etorri	-tor-	etor-	-tor-	come	dator
etzan	-tza-	etzan	-tzan	lay down	datza
ezagu(tu)	-zagu-	ezagu	-zagu-	know, meet (people...)	dazagu
ibili	-bil-	ibil-	-bil-	walk, be busy doing sth.	dabil
ihardun	-ihardu-	ihardun	-hardun	be busy doing sth	dihardu
ikusi	-kus-	ikus-	-kus-	see	dakusa
irakin	-iraki-	irakin	-kin	boil, ferment	diraki
iraun	-irau-	iraun	-un	last, endure, suffer	dirau
iritzi	-ritz-	iritz-	-itz-	have an opinion	deritzo
irudi	-irudi-	irud-	-ud-	seem, look like	dirudi
izan	-i-	izan	izan	be	da,dira
jakin	-ki-	jakin	-akin	know	daki
jario	-ri-	jario	-ario	flow	dario
jarraiki	-rrai-	jarraiki-	-arraik-	follow	darra
joan	-oa-	joan	-oan	go	doa
ukan	-u-	ukan	-kan	have	du

A3. Monosyllables from Sarasola not found in Azkue

Dictionaries: M- Morris, S- Sarasola, K- Kintana, A- Aulestia

Word	Included as	Dictionaries	Date	English/notes	Word	Included as	Dictionaries	Date	English/notes
ba		SKA/Azl	XIX		leun	leun	MSKA	1627	
bals		S	1847		lez		MSKA	1896	
bar		MSK	1950		lits		MSKA	1745	
bats	Bats	*S	1596		lord		SK	1977	
bit		S	1991		luis		S	1757	
blai		MSKA	1885		ments	ments	*MSK	1666	nothing, empty. Var of MENS
blok		MSK	ohar kaiera		mi		SKA	1824	Music note
bon		MSK	1945		mintz		MSKA	1802	
bortz	bortz	*MSKA/ Azl	1571	Var. BOST	net		MSK	1657	
boz	boz	*MSKA/Az	1562		neu	neu	MSKA	1745	1643
brixt		*S	1925	brist 1657 (Onom.)	ohm		MSK	1935	ohm
dan		SKA	1842		or	hor	MSKA	1562	dog
de		S	1896		plan		MSKA	1800	plan
deun		MSKA	1895		plat		S	1571	plate
do		SK	1977		plaun		SK	1643	plain(geog.)
don		S	XVI	on jaun (Contr.)	plus		SK	1975	
eurt		S	~1820		prest	prest	MSKA	1545	ready
fax		MS		See telekopia	re		MS	1824	note (mus.)
film		S	1954		rock		MS	1977	rock
fin		SKA	1545		ron		MS	1895	rum
fin		S	1571		san		SK	1596	saint
flan		S		See burdin	si		SK	1824	note (mus.)
flux		S	1859	flox 1664	sol		MSK	1824	note (mus.)
frai		*MSK	1643	friar. not in Azkue	sos	sos	MSKA	~1620	value of a coin 20 pounds
frak		SK	1907		sost	sost	MSK	1745	repent
gai	gai	*MSKA/ Azl	1643	material	stress		S	1992	stress
gas		MSKA	1933		te		SKA	1745	tea
ge		S	1761	G of alphabet	tenk		SK	1929	wait
geu	geu	*MSKA/ Azl	1596	intensive GU	test		SK	1973	test
geurtz	geurtz	*MSKA/ Azl	1657		tren		SKA	1880	train
gontz		S	1746		trust		SK	1935	trust (financial)
gris		*MSKA	1640		tsar		SK	1977	tzar
hain		*SKA/Azl	C15th	also	txo		SKA	1902	word of insult

Word	Included as	Dictionaries	Date	English/notes	Word	Included as	Dictionaries	Date	English/notes
hain	hain	* S	1571		txoil		SK	1782	a lot
han	han	*MSKA/ Azl	XV		txol		S	1916	
hax		S	XVII		txotx		SK	1745	toothpick
heu	heu	MSKA	1803		volt		MS	1977	volt
hor	hor	*MSKA/ Azl	1545	dog (Var of ora)	watt		S	1935	watt
huntz	huntz	*MSKA/ Azl	1643	UNTZ is a vari. Shrub Hedera halix	xar		*SK/Azl	1630	small
jazz		MSK	1960		xut	zut	MSK	1664	zut
klan		MSK	1977		zank		S	1924	
klar		*MSK	1617		zast		S	zist (1905)	
klask		SKA	1857		zen		MSKA	1858	
klik		MSKA	XIX		zeu	zeu	MSKA	1638	intensive of zu
klub		MSKA	1852		zink		SKA	1950	
la		SK	1824		zirt		SKA	1880	
lant		S	landu 1571		zuin		MSKA	1905	
latx		S	1968		zurtz	zurtz	*MSKA/ Azl	1571	orphan
latz			1562						

A4. Tables of Monosyllables for each historical word class

o - onomatopoeia, *i* - interjection

A4.1. Common Early monosyllabic words of Basque

a	ai	antz	au	auntz	aur	bai	bal	bartz	be
beltz	ber	bertz	bi	bits	botz	butz	e	gain	gaitz
gal	gan	gantz	gar	gatz	gau	gen	goi	gon	gor
gu	gun	gur	haitz	haltz	har	hartz	hats	hatz	hauts
hein	hel	heu	hi	hil	hin	hits	hitz	ho	hoin
holtz	hortz	hots	hotz	hu <i>i</i> .	huntz	hur	huts	<i>i i</i> .	itz
kar	kau	kaur	ke	kui	lai	lan	latz	laur	lei
lo	lor	lur	ma	mau	min	motz	mu	muin	nar
ni	no	non	nor	o	oi	ots	sai	sail	sar
sats	sei	sen	sits	so	soi	soil	soin	sor	su
u <i>i</i> .	ui	ur	ut <i>i</i> .	uts	uts	zai	zail	zail	zain
zain	zan	ze	ze	ze	zer	zer	zi	zil	zin
zin	zintz	zitz	zo <i>i</i> .	zoi	zotz	zu	zuntz	zur	zurtz

A4.2. Common Late monosyllabic words of Basque

aiz	ants	aup	ausk <i>o.</i>	az	bat	blink <i>o.</i>	bort	bortz	borz
brast <i>o.</i>	brau <i>o.</i>	dank	dart	daunb <i>o.</i>	dei	doi	drak <i>o.</i>	drank	drausk
drin <i>o.</i>	duin	dzanp <i>o.</i>	dzart	dzast	ep	et	eup	glask	goiz
grask <i>o.</i>	haz	izt	kask <i>o.</i>	kausk <i>o.</i>	kik	kink	klak <i>o.</i>	klausk <i>o.</i>	klink <i>o.</i>
kluk <i>o.</i>	kok	kosk <i>o.</i>	krak	krask <i>o.</i>	krik <i>o.</i>	krisk <i>o.</i>	mist	miz <i>o.</i>	mus
must <i>o.</i>	ñau <i>o.</i>	noiz	ñu	ok	op	pa	par	part <i>o.</i>	pik <i>o.</i>
pintz	pit	pits	piz	pla <i>o.</i>	plast <i>o.</i>	plaut <i>o.</i>	plunp <i>o.</i>	pot	pots
prei	printz	prizt <i>o.</i>	pu	punp <i>o.</i>	putz	sast <i>o.</i>	set	tai	tak <i>o.</i>
tank <i>o.</i>	taup	teink	ten	tink	tint	to	toil	traust	truk
tta	tu	tunt	tut	tzar	tzut	zart <i>o.</i>	zaust <i>o.</i>	zizt <i>o.</i>	zunp <i>o.</i>
zurt	zut								

A4.3. Common Modern monosyllabic words of Basque

antz	atx	fu	jaun	jeurt	jín	jo	motx	mox <i>o.</i>	pa
pits	txa	txai	txak <i>o.</i>	txar	txau	txil	txin <i>o.</i>	txint	txist
txit	txiz	txoin	txut	uf	ux	xai	xo	xotx	

A4.4. Uncommon Early monosyllabic words of Basque

aun	bats	bitz	bui	guntz	hain	kain	kutz	lats	mauts
na									

A4.5. Uncommon Late monosyllabic words of Basque

as	aurt	bart	beunt	blai	blau <i>o.</i>	blaust <i>o.</i>	brai	bri	brist <i>o.</i>
brus	but	draul	dsats <i>o.</i>	eit	has	hep	hink	hint	hup
irt	iz	kek	lazt	maus	mens	mintz	musk	ñi	ño
pau	pauts	pints	pis	pizt	plau <i>o.</i>	plust	pok	potz	satz
tast	ti	tols	troil	truin	tsats	ttik	zits		

A4.6. Uncommon Modern monosyllabic words of Basque

fast <i>o.</i>	feit	flus	furts	fut <i>i.</i>	jui	Katx	litx	mitx <i>o.</i>	otx
plox	sitx	txat	txintz	txost <i>i.</i>	xur	Xur			

A4.7. Word frequencies of Part of Speech classes by historical classes for monosyllabic words of Basque

	Common Early	Common Late	Common modern	Uncommon early	Uncommon late	Uncommon modern
NOUN	91	36	21	7	22	10
VERB						
(EX SYNTHETIC)	17	7	3	1	2	—
PRONOUN	14	6	4	1	5	—
ADJECTIVE	26	14	5	2	10	3
ADVERB	9	4	1	1	1	—
DETERMINER	8	2	—	—	—	—
CONJUNCTION	3	1	—	—	—	—
INTERJECTION	20	17	11	—	7	4
ONOMATOPOEIA	4	42	6	—	5	3

A5. Percentage frequencies of phonemes in monosyllable words of Basque by historical classes

Position	Phonemic groups		Common early	Uncommon early	Common late	Uncommon late	Common modern	Uncommon modern
ONSET	Null		14	9	11	10	10	6
	LABIAL	b	9	27	4	6	—	—
		p	—	—	13	15	5	—
		m	5	9	4	8	5	6
		pl	—	—	4	4	—	6
		pr	—	—	3	8	—	—
		br	—	—	2	—	—	—
		bl	—	—	1	6	—	—
		f	—	—	—	—	10	24
	fl	—	—	—	—	3	6	
	CORONAL	d	—	—	6	—	—	—
		tr	—	—	2	4	—	—
		ts	—	—	—	2	—	—
		tt	—	—	1	2	—	—
		dr	—	—	4	2	—	—
		dz	—	—	3	—	—	—
		n	4	9	1	—	—	—
		s	11	—	2	2	—	6
		tz	—	—	2	—	—	—
		t	—	—	13	6	—	—
		z	15	—	6	2	—	—
		l	6	9	—	2	—	6
		ds	—	—	—	2	—	—
		tx	—	—	—	—	33	18
	x	—	—	—	—	8	6	
	DORSAL	k	4	18	6	2	—	6
		kr	—	—	4	—	—	—
		gr	—	—	1	—	—	—
		n~	—	—	2	4	—	—
		gl	—	—	1	—	—	—
		kl	—	—	4	—	—	—
		h	20	9	1	10	—	—
		g	12	9	1	—	—	—
	j	—	—	—	—	26	12	

Position	Phonemic groups		Common early	Uncommon early	Common late	Uncommon late	Common modern	Uncommon modern
MEDIAL	Vowel		72	54	78	71	72	82
	Diphthong		28	46	22	29	28	18
FINAL	Null		33	18	13	19	26	12
	Labial	p	—	—	5	4	—	—
	CORONAL	st	—	—	9	8	3	12
		rtz	4	—	1	—	—	—
		nb	—	—	1	—	—	—
		nk	—	—	8	2	—	—
		np	—	—	4	—	—	—
		ns	—	—	—	2	—	—
		nt	—	—	2	4	3	—
		rt	—	—	6	6	3	—
		ts	8	27	2	8	5	—
		rz	—	—	1	—	—	—
		tz	11	18	1	4	—	—
		tx	—	—	—	—	8	29
		l	7	—	1	4	3	—
		ls	—	—	—	2	—	—
		ltz	2	—	—	—	—	—
		nts	—	—	1	2	3	—
		zt	—	—	3	4	—	—
		n	14	27	3	2	15	—
		s	—	—	1	10	3	6
		sk	—	—	10	2	—	—
		r	15	—	2	—	3	6
		rts	—	—	—	—	—	6
		t	1	—	8	4	8	18
	x	—	—	—	—	5	6	
	z	—	—	9	2	5	—	
	ntz	5	9	2	2	3	6	
	DORSAL	k	—	—	11	6	3	—
		n~	—	—	—	—	3	—
Number of Words			123	11	102	48	39	17

A6. Phoneme frequencies of monosyllabic words in Basque for Part of Speech Classes by Historical classes

A6.1. Phonemic Distributions of Common Early Monosyllabic Nouns of Basque

Ø	Preceding Consonant Clusters										Medial			Succeeding Consonant Clusters										D/G		
	Labials		Coronals					Dorsal/Guttural			V/D	Tot	Ø	Lab	Coronals											
	b	m	n	s	z	l	Tot	k	h	g					Tot	rtz	ts	tz	l	lʒ	n	r	t		ntz	Tot
1	2	1	3	1	2	0	2	5	0	5	4	9	18	1	0	2	2	3	1	1	2	4	0	2	17	0
0	2	0	2	0	1	1	0	2	1	0	0	1	5	3	0	0	0	0	0	0	1	0	0	0	2	0
1	1	1	2	0	1	5	0	6	0	3	0	3	12	1	0	0	3	1	0	3	0	0	0	1	11	0
2	1	0	1	0	1	1	2	4	0	5	1	6	13	4	0	1	2	3	0	1	0	2	0	0	9	0
2	1	1	2	0	1	3	1	5	0	2	2	4	13	2	0	1	1	1	0	0	1	5	0	2	11	0
1	0	0	0	0	2	3	1	6	0	1	2	3	10	4	0	0	2	2	0	2	0	0	0	0	6	0
2	0	1	1	0	0	0	0	0	1	1	1	3	6	2	0	0	1	0	0	0	2	0	0	1	4	0
0	0	0	0	0	1	0	1	2	0	2	0	2	4	2	0	0	0	0	0	0	2	0	0	0	2	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	3	1	0	4	0	1	1	2	7	4	0	0	0	1	0	2	0	0	0	0	3	0
1	0	1	1	0	0	0	0	0	1	0	0	1	3	2	0	0	0	0	0	0	1	0	0	0	1	0
6	7	3	10	1	6	10	5	22	1	15	7	23	61	11	0	5	8	10	2	2	7	11	0	5	50	0
5	0	2	2	0	6	4	2	12	2	5	4	11	30	14	0	1	2	3	0	7	2	0	1	16	0	
11	7	5	12	1	12	14	7	34	3	20	11	34	91	25	0	5	9	12	5	2	14	13	0	6	66	0
													Tot V Tot D Tot V/D													
7	8	3	11	1	7	11	6	24	1	17	8	25.3	67	12	0	6	9	11	2	2	8	12	0	6	55	0
6	0	2	2	0	7	4	2	13	2	6	4	12.1	33	15	0	0	1	2	3	0	8	2	0	1	18	0
12	8	6	13	1	13	15	8	37	3	22	12	37.4	100	28	0	6	10	13	6	2	15	14	0	7	73	0

A6.2. *Phonemic Distributions of Uncommon Early Monosyllabic Nouns of Basque*

Ø	Preceding Consonant Clusters						Medial			Succeeding Consonant Clusters						
	Labials		Coronals		Dorsal/Guttural		Total	V/D	Total	Lab	Coronals			Dors/Gut		
	b	m	N	l	k	h	g		Tot	Tot	ts	tz	n	ntz	Tot	
0	1	0	1	0	0	0	0	2	a	2	2	0	0	0	2	0
0	0	0	0	0	0	0	0	0	e	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	i	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	o	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	1	u	1	0	0	0	1	1	0
0	0	0	0	0	1	0	0	1	ai	1	0	0	1	0	1	0
1	0	1	0	0	0	0	0	2	au	2	1	0	1	0	2	0
0	0	0	0	0	0	0	0	0	ei	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	eu	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	oi	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	1	ui	1	0	0	0	0	0	0
0	1	0	1	0	0	0	1	3	Tot V	3	0	2	0	0	1	3
1	1	2	0	0	1	0	0	4	Tot D	4	1	1	0	2	0	3
1	2	1	0	1	1	0	1	7	Tot V/D	7	1	3	0	2	1	6
0	14	0	14	0	0	0	14	42.9	%V	42.9	0	29	0	0	14	43
14	14	14	29	0	0	0	14.3	57.1	%D	57.1	14	14	0	29	0	43
14	29	14	43	0	14	14	28.6	100	%Total	100	14	43	0	29	14	86

A6.4. *Phonemic Distributions of Uncommon Late Monosyllabic Nouns of Basque*

Ø	Preceding Consonant Clusters																	Medial					
	Labials						Coronals						Dorsal/Guttural			Total	V/D						
	b	P	m	pl	br	bl	Tot	tr	ts	tt	dr	s	t	z	l			ds	Tot	k	ñ	h	Tot
1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	4	0	0	0	0	5	a
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	e
1	0	2	1	0	0	0	3	0	0	1	0	1	0	0	0	0	2	0	1	1	2	8	i
0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	o
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	u
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	ai
0	0	2	1	0	0	0	3	0	0	1	0	0	0	0	0	0	1	0	0	0	0	4	au
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ei
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	eu
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	oi
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	ui
1	4	1	0	0	0	0	6	0	1	0	1	2	0	0	1	0	6	0	1	1	2	15	Total V
0	2	1	0	1	0	4	2	0	0	1	0	0	0	0	0	0	3	0	0	0	0	7	Total D
1	6	2	0	1	0	10	2	1	1	1	2	0	0	1	0	0	9	0	1	1	2	22	Total V/D
5	18	5	0	0	0	27	0	5	5	0	5	9	0	5	0	27	0	5	5	5	9.1	68.2	%V
0	9	5	0	5	0	18	9	0	0	5	0	0	0	0	0	14	0	0	0	0	0	31.8	%D
5	27	9	0	5	0	46	9	5	5	5	9	9	0	5	0	41	0	5	5	5	9.1	100	%Total

Medial	Succeeding Consonant Clusters																											
	Ø	Labial		Coronals																Dors/Gut								
		P	Tot	st	nk	ls	ns	nt	rt	ts	tz	l	nts	zt	n	s	sk	t	z	ntz	Tot	k	Tot					
5	0	0	1	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	5	0	0					
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
8	2	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	5	1	1						
2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1						
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
4	1	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	3	0	0						
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0						
1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0						
15	2	0	1	0	0	0	1	1	1	2	0	1	1	0	1	0	0	0	1	11	2	2						
7	2	0	0	0	0	0	0	0	1	0	2	0	0	1	1	0	0	0	0	5	0	0						
22	4	0	1	0	0	0	1	1	2	2	2	1	1	1	2	0	0	1	16	2	2							
68.2	9	0	5	0	0	0	5	5	5	9	0	5	5	0	5	0	0	5	5	50	9	9.1						
31.8	9	0	0	0	0	0	0	0	5	0	9	0	0	5	5	0	0	0	23	0	0	0						
100.0	18	0	5	0	0	0	5	5	9	9	9	5	5	5	9	0	0	5	5	73	9	9.1						

A6.5. *Phonemic Distributions of Common Modern Monosyllabic Nouns of Basque*

Ø	Preceding Consonant Clusters						Succeeding Consonant Clusters																			
	Labials			Coronals			D/Gut			Coronals													Dors/Gut			
	p	m	f	tx	X	Tot	nt	st	rt	rs	rz	tx	l	ns	n	s	r	t	x	z	ns	Tot	k	ñ	Tot	
																										fl
V/D		Total																								
1	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	4	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	2	0	2	1	1	3	0	0	0	0	1	0	0	0	0	0	0	3	0	0	0	
0	0	2	0	0	1	0	0	0	3	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	
0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	
0	0	0	1	0	1	1	2	2	5	3	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	
0	0	0	0	0	1	0	1	1	2	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	
0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	2	1	0	3	1	4	4	12	0	1	0	0	1	2	1	0	0	1	1	0	11	0	0	0	
0	0	0	1	1	2	3	1	4	9	0	0	0	0	0	2	0	0	1	0	1	4	0	1	1		
1	0	2	2	1	5	6	2	8	21	5	0	0	1	0	4	1	0	1	1	2	0	15	0	1	1	
5	0	10	5	0	14	5	19	19	57.1	5	0	0	5	0	5	0	0	5	0	5	0	52	0	0	0	
0	0	5	5	0	10	14	5	14	42.9	19	0	0	0	0	0	0	0	5	0	5	0	19	0	5	5	
5	0	10	10	5	24	29	10	38	100	24	0	0	5	0	14	0	5	19	5	0	5	71	0	5	5	

A6.7. *Phonemic Distributions of Common Early Pronouns*

				Preceding Consonant Clusters										Medial			Succeeding Consonant Clusters									
Ø	Labials		Tot	Coronals			Dorsal/Guttural				Total	V/D	Total	Ø	Labial	Coronals						Tot	Dors/Gut	Tot		
	b	m		n	s	z	l	Tot	k	h						g	Tot	rtz	ts	tz	l				lɛz	n
0	0	0	0	0	0	0	0	0	1	1	1	a	0	0	0	0	0	0	1	0	0	1	0	0	0	
0	0	0	0	0	1	1	0	0	1	1	2	e	2	0	0	0	0	0	1	0	0	1	0	0	2	0
0	0	0	0	1	0	1	0	1	0	1	2	i	2	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	2	0	0	2	0	1	3	3	o	3	1	0	0	1	1	0	0	0	1	0	0	2	0
0	0	1	1	0	0	1	0	1	1	3	3	u	3	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	ai	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	1	au	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	ei	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	eu	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	1	1	oi	1	0	0	0	0	0	1	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	ui	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	1	3	0	2	0	5	0	2	3	TotV	11	6	0	0	0	0	2	3	0	0	0	0	5	0
1	0	0	0	0	0	0	0	0	2	0	2	TotD	3	2	0	0	0	0	1	0	0	0	0	0	1	0
1	0	1	1	3	0	2	0	5	0	4	3	TotV/D	14	8	0	0	0	0	3	3	0	0	0	0	6	0
0	0	7	7	21	0	14	0	36	0	14	21	%V	78.6	43	0	0	0	0	14	21	0	0	0	0	36	0
7	0	0	0	0	0	0	0	0	0	14	0	%D	21.4	14	0	0	0	0	7	0	0	0	0	0	7	0
7	0	7	7	21	0	14	0	36	0	29	21	%Total	100	57	0	0	0	0	21	21	0	0	0	0	43	0

Medial		Succeeding Consonant Clusters																											
		Coronals																										Dors/Gutt	
		Labial		st	rtz	rrz	tz	l	nts	zt	n	s	sk	r	t	z	ntz	Tot	k	Tot									
V/D	Total	Ø	p	Tot	ts	rt	nt	np	nk	nb	rtz	tz	l	nts	zt	n	s	sk	r	t	z	ntz	Tot	k	Tot				
a	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0			
e	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
i	3	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	0			
o	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0			
u	5	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	3	0	0	5	0	0	0			
ai	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
au	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
ei	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
eu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
oi	3	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	2	0	0	0			
ui	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0			
TotV	10	0	0	0	0	2	0	1	0	0	1	0	0	0	0	0	0	0	1	3	0	1	10	0	0	0			
TotD	4	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	3	0	0	0			
TotV/D	14	1	0	0	0	2	0	1	0	0	1	1	0	0	0	1	0	0	1	3	1	1	13	0	0	0			
%V	71.4	0	0	0	0	14	0	7	0	0	7	0	0	0	0	0	0	0	7	21	0	7	71	0	0	0			
%D	28.6	7	0	0	0	0	0	0	0	0	0	0	7	0	0	7	0	0	0	0	7	0	21	0	0	0			
%Tot	100	7	0	0	0	14	0	7	0	0	7	7	7	0	0	7	0	0	7	21	7	7	93	0	0	0			

A6.10. Phoneme Distribution of Uncommon Late Adjectives of Basque

Ø	Preceding Consonant Clusters																Medial							
	Labials						Coronals										Dorsal/Guttural			Total				
	b	p	m	pl	br	bl	Tot	tr	ts	tt	dr	s	t	z	l	ds	Tot	k	ñ		h	Tot	V/D	Total
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	a	1
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	e	1
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	i	1
0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	2	o	2
0	0	0	1	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	u	2
0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	ai	1
0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	au	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ei	0
0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	eu	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	oi	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ui	0
0	0	1	2	0	1	0	4	0	0	0	0	0	1	1	0	0	2	0	0	1	1	Tot V	7	
0	1	1	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Tot D	3	
0	1	2	2	0	1	1	7	0	0	0	0	1	1	0	0	0	2	0	0	1	1	Tot V/D	10	
0	0	10	20	0	10	0	40	0	0	0	0	10	10	0	0	0	20	0	0	10	10	%V	70	
0	10	10	0	0	0	10	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	%D	30	
0	10	20	20	0	10	10	70	0	0	0	0	10	10	0	0	0	20	0	0	10	10	%Total	100	

Medial		Succeeding Consonant Clusters																					
		Coronals															Labial		Dors/Gut				
		Ø	p	st	sr	nk	ls	ns	nt	rt	ts	tz	l	nts	zt	n	s	sk	t	z	niz	Tot	Tot
V/D	Total	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
a	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
e	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
i	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
o	2	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0
u	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2	0
ai	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
au	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
ei	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
eu	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
oi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ui	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tot V	7	0	0	0	1	1	0	0	1	1	0	0	0	0	0	2	1	0	0	0	0	7	0
Tot D	3	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0
Tot V/D	10	1	0	0	1	1	1	0	2	1	0	0	0	0	0	2	1	0	0	0	0	9	0
%V	70	0	0	0	10	10	0	0	10	10	0	0	0	0	0	20	10	0	0	0	0	70	0
%D	30	10	0	0	0	10	0	10	0	10	0	0	0	0	0	0	0	0	0	0	0	20	0
%Total	100	10	0	0	10	10	10	0	20	10	0	0	0	0	0	20	10	0	0	0	0	90	0

A6.11. Phoneme Distribution of Common Early Verbs of Basque

Ø	Preceding Consonant Clusters						Medial			Succeeding Consonant Clusters														
	Labials			Coronals			Dorsal/Guttural			V/D	Total	Coronals										Dors/Gut		
	b	m	Tot	n	s	z	l	Tot	k			h	g	Tot	rrz	ts	tz	l	lcz	n	r		t	ntz
0	0	0	0	0	1	0	0	0	2	1	3	4	a	4	0	0	1	0	0	2	0	0	4	
1	1	0	1	0	0	0	0	0	1	1	2	4	e	4	0	0	1	0	1	1	0	0	3	
0	0	0	0	0	1	0	1	0	1	0	1	2	I	2	0	0	1	0	0	0	0	1	2	
0	0	0	0	1	1	0	2	4	0	0	0	4	o	4	0	0	0	0	2	0	0	2	0	
0	0	0	0	0	0	0	0	1	1	1	1	1	u	1	0	0	0	0	1	0	0	1	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	ai	0	0	0	0	0	0	0	0	0	0	
0	0	1	1	0	0	0	0	1	1	0	1	2	au	2	1	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	ei	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	eu	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	oi	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	ui	0	0	0	0	0	0	0	0	0	0	
1	1	0	1	1	2	1	2	6	0	4	3	7	Total V	15	3	0	1	3	0	1	6	0	12	0
0	0	1	1	0	0	0	0	0	0	1	0	1	Total D	2	1	0	0	0	0	0	0	0	1	0
1	1	1	2	1	2	1	2	6	0	5	3	8	Total V/D	17	4	0	1	3	0	1	6	0	13	0
6	6	0	6	6	12	6	12	35	0	24	18	41.2	%V	88.2	18	0	0	6	18	0	6	35	0	71
0	0	6	6	0	0	0	0	0	0	6	0	5.9	%D	11.8	6	0	0	0	0	0	0	0	0	6
6	6	6	12	6	12	6	12	35	0	29	18	47.1	%Total	100	24	0	6	18	0	6	35	0	77	0

A7.1. Phoneme Distribution of Common Early monosyllabic words of Basque excluding Onomatopoeia & Interjections

Ø	Preceding Consonant Clusters																	Medial				
	Labials				Coronals										Dorsal/Guttural			Total	V/D			
	b	p	m	f	Tot	d	n	tx	S	tz	t	z	x	l	Tot	j	k			h	g	Tot
1	2	0	1	0	3	0	1	0	2	0	0	1	0	2	6	0	0	5	5	10	20	a
1	4	0	0	4	0	0	0	1	0	0	2	0	0	0	3	0	1	1	1	3	11	e
1	2	0	1	0	3	0	1	0	1	0	5	0	0	0	7	0	0	5	0	5	16	i
2	1	0	1	0	2	0	3	0	2	0	0	1	0	2	8	0	0	5	2	7	19	o
2	1	0	1	0	2	0	0	1	0	0	4	0	1	6	0	0	0	3	3	6	16	u
1	1	0	0	1	1	0	0	0	2	0	3	0	1	6	0	0	1	2	3	3	11	ai
3	0	0	1	0	1	0	0	0	0	0	0	0	1	1	1	0	2	1	1	4	9	au
0	0	0	0	0	0	0	0	1	0	0	0	0	1	2	0	0	0	1	0	1	3	ei
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	eu
1	0	0	0	0	0	0	0	3	0	0	1	0	0	4	0	0	0	1	1	2	7	oi
1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	3	ui
7	10	0	4	0	14	0	5	0	7	0	13	0	5	30	0	1	19	11	31	82	Total V	
6	1	0	2	0	3	0	0	0	6	0	4	0	3	13	0	3	5	4	12	34	Total D	
13	11	0	6	0	17	0	5	0	13	0	17	0	8	43	0	4	24	15	43	116	Total V/D	
6	8.6	0	3.4	0	12.1	0	4.3	0	6	0	11.2	0	4.3	25.9	0	0.9	16.4	9.5	26.7	70.7	%V	
5.2	0.9	0	1.7	0	2.6	0	0	0	5.2	0	3.4	0	2.6	11.2	0	2.6	4.3	3.4	10.3	29.3	%D	
11.2	9.5	0	5.2	0	14.7	0	4.3	0	11.2	0	14.7	0	6.9	37.1	0	3.4	20.7	12.9	37.1	100	%Total	

Medial		Succeeding Consonant Clusters																							
		Labial										Coronals										Dors/Gut			
		∅	P	f	Tot	st	rtz	ts	rz	tx	tz	l	nts	zt	ltz	n	r	t	x	z	ntz	Tot	k	ñ	Tot
V/D	Total	1	0	0	0	0	0	0	0	2	0	0	0	1	3	4	0	0	0	0	2	19	0	0	0
A	20	4	0	0	0	0	0	0	0	1	0	0	0	1	2	2	0	0	0	0	7	0	0	0	
E	11	4	0	0	0	0	0	0	0	2	0	0	0	0	3	0	0	0	0	0	12	0	0	0	
I	16	5	0	0	0	0	0	0	0	3	0	0	0	0	3	0	0	0	0	1	14	0	0	0	
O	19	4	0	0	0	0	0	0	0	4	0	0	0	1	2	4	0	0	0	0	14	0	0	0	
U	16	4	0	0	0	0	0	0	0	0	0	0	0	0	1	5	0	0	0	2	12	0	0	0	
Ai	11	5	0	0	0	0	0	0	0	1	2	0	0	0	2	0	0	0	0	0	6	0	0	0	
Au	9	4	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1	5	0	0	0	
Ei	3	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	
Eu	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Oi	7	4	0	0	0	0	0	0	0	0	1	0	0	0	2	0	0	0	0	0	3	0	0	0	
Ui	3	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	
Total V	82	18	0	0	0	0	0	0	0	11	5	0	0	3	11	15	0	0	0	5	64	0	0	0	
Total D	34	18	0	0	0	0	0	0	0	2	3	0	0	0	6	3	0	0	0	1	16	0	0	0	
Total V/D	116	36	0	0	0	0	0	0	0	13	8	0	0	3	17	18	0	0	0	6	80	0	0	0	
%V	70.7	15.5	0	0	0	0	0	0	0	9.5	4.3	0	0	2.6	9.5	12.9	0	0	0	4.3	55.2	0	0	0	
%D	29.3	15.5	0	0	0	0	0	0	0	1.7	2.6	0	0	0	5.2	2.6	0	0	0	0.9	13.8	0	0	0	
%Total	100	31	0	0	0	0	0	0	0	11.2	6.9	0	0	2.6	14.7	15.5	0	0	0	5.2	69	0	0	0	

A7.2. Phoneme Distribution of Common Late monosyllabic words of Basque excluding Onomatopoeia & Interjections

Ø	Preceding Consonant Clusters												Dorsal/Guttural						Medial									
	Labials						Coronals						j k kr gr ñ g l kl ll h g						Total	V/D								
	b	p	m	f	pl	pr	br	fl	bl	d	tr	ts	tt	dr	dz	n	ɲ	s	sz	t	z	x	l	ds	Tot	Tot		
1	1	2	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	6	3	13	a	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0	2	c	
0	0	4	1	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	2	1	9	i	
1	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	7	o	
0	0	1	1	0	0	0	0	0	2	0	1	0	0	3	2	0	0	0	0	0	0	0	0	7	1	10	u	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	ai	
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	au	
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	3	ei	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	eu
0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0	4	oi	
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	ui	
2	4	9	2	0	0	1	0	0	16	2	1	0	0	1	2	0	0	1	2	6	2	0	0	17	6	41	Tot V	
0	0	0	0	0	1	0	0	0	1	3	1	0	0	0	1	0	0	0	4	0	0	0	0	9	1	11	Tot D	
2	4	9	2	0	2	0	0	0	17	5	2	0	1	2	1	0	1	2	10	2	0	0	0	26	7	52	Tot V/D	
3.8	7.7	#	3.8	0	2	0	0	0	30.8	4	2	0	0	2	4	0	0	1.9	4	12	3.8	0	0	32.7	11.5	78.8	%V	
0	0	0	0	0	2	0	0	0	1.9	6	2	0	0	0	1.9	0	0	0	8	0	0	0	0	17.3	1.9	21.2	%D	
3.8	7.7	#	3.8	0	4	0	0	0	32.7	#	4	0	0	2	4	1.9	0	1.9	4	19	3.8	0	0	50	13.5	100	%Total	

A7.5. Phoneme Distribution of Uncommon Late monosyllabic words of Basque excluding Onomatopoeia and Interjections

Ø	Preceding Consonant Clusters														Medial																																	
	Labials					Coronals										Dorsal/Guttural					Total	V/D																										
	B	p	m	f	pl	pr	br	fl	bl	Tot	d	tr	ts	tt	dr	dz	n	tx	s	tz			t	z	x	l	ds	Tot	j	k	kr	gr	n	g	kl	ll	h	g	Tot									
1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	7	a					
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	e				
2	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	12	i			
0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0				
0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0					
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0					
0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0					
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
3	1	5	3	0	1	0	1	0	0	0	1	0	3	1	0	1	0	0	0	0	0	0	0	0	0	0	8	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	5	27	Total V
0	1	2	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	Total D
3	2	7	4	0	1	0	2	0	1	0	0	1	0	3	1	0	1	0	0	0	0	0	0	0	0	0	11	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	36	Total V/D
8.3	2.8	#	8.3	0	3	0	3	0	0	30.6	0	0	3	3	0	0	0	0	0	0	22.2	0	2.8	0	0	0	2.8	0	2.8	0	8.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	75	%V
0	2.8	6	2.8	0	0	3	0	0	3	16.7	0	6	0	0	3	0	0	0	0	0	8.3	0	0	0	0	0	8.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	%D	
8.3	5.6	#	11	0	3	0	6	0	3	47.2	0	6	3	3	3	0	0	0	0	0	30.6	0	2.8	0	0	0	2.8	0	8	2.8	0	8.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	%Total

Medial		Succeeding Consonant Clusters																																				
		Labial														Coronals										Dors/Gut												
		∅	p	f	Tot	st	strz	nb	nk	ls	lp	ns	nt	rt	ts	rz	tx	tz	l	nts	zt	ltz	n	s	sk	rrs	r	t	x	z	ntz	Tot	k	ñ	Tot			
V/D	Total	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
A	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
E	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
i	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
O	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
U	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
ai	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
Au	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
ei	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Eu	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
oi	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ui	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total V	27	2	0	0	0	2	0	0	1	2	0	0	1	2	2	0	2	0	1	2	0	0	4	1	0	0	0	0	1	1	22	3	0	3	0	3		
Total D	9	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total V/D	36	5	0	0	0	2	0	0	1	2	2	2	1	2	2	0	2	2	1	2	0	1	5	1	0	0	0	1	28	3	0	3	0	3	0	3		
%V	75	5.6	0	0	0	6	0	0	3	6	5.6	0	3	6	5.6	0	5.6	0	3	6	0	0	3	0	0	0	0	3	2.8	61.1	8	0	8.3	0	8.3	0	8.3	
%D	25	8.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
%Total	100	13.9	0	0	0	6	0	0	3	6	8.3	0	3	6	8.3	0	5.6	5.6	3	6	0	2.8	3	0	0	0	0	3	77.8	8	0	8.3	0	8.3	0	8.3		

A6. Frequency for phonemes for synthetic verbs of Basque

Ø	Preceding Consonant Clusters						Medial			Succeeding Consonant Clusters															
	Labials		Coronals				Dorsal/Guttural		V/D	Total	Ø	Labial					Coronals					Dor			
	b	p	m	Tot	d	n	s	tz				t	z	Tot	p	Tot	ts	tz	l	n	s	r	t	z	Tot
0	0	0	2	2	0	0	0	2	6	a	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	e	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	0	0	2	0	0	0	1	5	i	5	0	0	0	1	2	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	2	o	2	0	0	0	0	1	0	1	0	0	0	0	2	0	0
2	0	0	0	1	0	0	0	1	4	u	4	1	0	0	0	0	1	1	0	0	0	2	1	1	0
0	0	0	0	0	0	0	0	0	0	ai	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	au	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	ei	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	eu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	oi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	ui	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2	0	2	4	1	0	1	1	17	Total V	17	1	0	0	1	2	9	1	2	0	0	15	1	1	1
0	0	0	0	0	0	0	0	0	0	Total D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2	0	2	4	1	0	1	1	17	Total V/D	17	1	0	0	1	2	9	1	2	0	0	15	1	1	1
18	12	0	12	24	5.9	0	5.9	5.9	100	%V	100	5.9	0	0	5.9	12	52.9	6	12	0	0	88	5.9	5.9	
0	0	0	0	0	0	0	0	0	0	%D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18	12	0	12	24	5.9	0	5.9	5.9	100	%Total	100	5.9	0	0	5.9	12	52.9	6	12	0	0	88	5.9	5.9	