

## Hypercorrections in a Western Basque text: What do they tell us about sibilant mergers?<sup>1</sup>

Dorota Krajewska\*

University of the Basque Country (UPV/EHU)

**ABSTRACT:** This paper analyses two 18th century texts and focuses on hypercorrections found in the spelling of sibilants, i.e. examples where an etymological apical fricative is represented with spelling elsewhere used for laminal fricatives. The texts used are Lubieta's dictionary (written in the central variety of Donostia/San Sebastián) and sermons composed by Bizente Sarria (written in the western variety of Etxebarri). As for Lubieta's text, in previous research hypercorrections have been interpreted to be caused by an ongoing merger in which the apico-alveolar fricative merges with the apico-laminal, and the remaining sound is the laminal. In Sarria's speech, however, another merger pattern was present, in which laminal and apical fricatives merge in favour of the latter. Because of that, hypercorrections found in that text are hypothesised to be purely graphical. Using quantitative methods, the paper compares the patterns found in spellings in the two texts in order to better understand merger processes involved (and especially hypercorrections), but also to reflect on the methodology which can be used to analyse such issues.

**KEYWORDS:** merger; sibilants; fricatives; Western Basque; Central Basque.

---

<sup>1</sup> The research reported in this paper was funded by the following projects: "Monumenta Linguae Vasconum VI" (PID2020-118445GB-I00), "Modern approaches to diachronic phonology applied to Basque (MADPAB)" (ANR-20-CE27-0007) and "Diachronic Linguistics, Typology and the History of Basque" (IT1534-22). Data and code used in this paper are available at <https://osf.io/5yt34/>.

\* **Corresponding author:** Dorota Krajewska. Hizkuntzalaritza eta Euskal Ikasketak Saila - Letren Fakultatea (UPV/EHU). Unibertsitateen ibilbidea, 5 (01006 Gasteiz). – [dorota.krajewska@ehu.eus](mailto:dorota.krajewska@ehu.eus) – <https://orcid.org/0000-0002-6909-985X>

**How to cite:** Krajewska, Dorota (2023). «Hypercorrections in a Western Basque text: What do they tell us about sibilant mergers?», *ASJU*, 57 (1-2), 565-581. (<https://doi.org/10.1387/asju.25923>).

Received: 2022-12-09; Accepted: 2023-03-24.

ISSN 0582-6152 - eISSN 2444-2992 / © UPV/EHU Press



This work is licensed under a

Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

## 1. Introduction

The study of Basque sibilants has seen a significant progress in the last decades. José Ignacio Hualde has contributed to this topic especially with his 2010 paper, which offers interesting insights on various aspects of sibilants, such as their acoustic properties, the textual evidence for mergers which occurred in Basque, the chronology of changes and the role of bilingualism in the changes. Since then, various studies have added to our understanding of sibilants and changes they have undergone (see, among others, Beristain 2021; Egurtzegi & Carignan 2020; Iribar Ibabe, Pagola Petrirena & Túrrez Aguirrezábal 2020; Muxika-Loitzate 2017; Zuloaga 2020).

The Basque sibilant system consists of lamino-alveolar, apico-alveolar and postalveolar pairs of fricatives and affricates (Michelena 1977; Hualde 2003). In some varieties it has undergone various changes, which resulted in reductions of the number of sibilant phonemes, i.e. in mergers (see Hualde 2010 for an overview). This paper focuses on two processes: the Western merger and the Central merger, schematically shown in Figure 1, and especially on the changes in the fricative series.

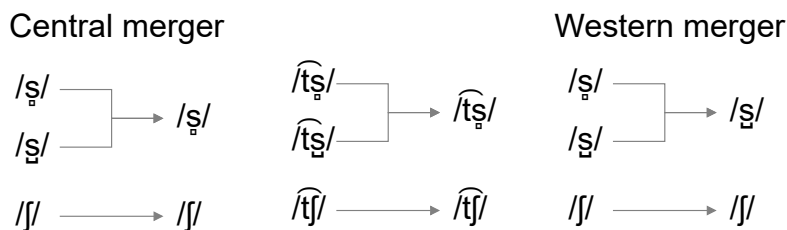


Figure 1

The Central and Western mergers

In the Western merger laminal and apical fricatives merge in favour of the latter. In affricates the remaining sound is the laminal. This change started in the Western varieties of the language, but extended at least to some extent to other varieties as well, e.g. to the coastal areas of the peninsular Basque (Zuloaga 2020). As already observed by Michelena (1977), the change started as a contextually conditioned neutralisation. Zuloaga (2020) shows that initially (starting in the 16th century), neutralisation in fricatives was limited to the word-final and preconsonantal positions (and especially before dental stops). Since the 18th century the merger extended to all phonological contexts.

In the Central merger, which started in some varieties of Central Basque, the direction of change in affricates is the same as in the Western merger, but the result is different in fricatives, as they merge into the laminal sound. In the modern language this pattern of merger is found in the coastal areas of Guipuscoa and in Urola valley (Camino 2000; Sagarzazu 2005; Beristain 2018; Beristain 2021).

Importantly, in historical texts the Central merger always co-occurs with the Western merger (Zuloaga 2020). Because of that, Zuloaga (2020) proposed that the Central merger could have originated as hypercorrection in reaction to the spread-

ing Western merger. Examples of hypercorrection are common in old texts (also in Western Basque), but the first attestations of what could be related to the Central merger are found in the early 17th century (Zuloaga 2020: 399).

Building on Zuloaga's (2020) ideas, Krajewska, Zuloaga and Egurtzegi (2022) analysed the so called Lubieta's dictionary to provide additional support for the hypercorrection hypothesis using quantitative methods. One of the arguments in favour of this explanation is related to the different influence of phonological context on the two mergers: the prevocalic context favours the appearance of non-etymological spelling of apical sibilants in Lubieta, which is uncommon in the preconsonantal position, i.e. in the most common environment of the Western merger:

The Central neutralisation shows its biggest incidence in the intervocalic context, a position where we would expect a phonetically-based neutralisation of place to have the least strength, given that place-related acoustic cues are likely best perceived intervocalically. The most straightforward way to account for an intervocalic apical to laminal change early in the merger process is through non-phonetic means, such as hypercorrection. (Krajewska *et al.* 2022: 25)

In general, the conclusions reached in Krajewska *et al.* (2022) are based on the idea that the distribution of etymological and non-etymological spellings can be more or less directly translated into the phonological system. Put differently, it is argued that the most salient patterns and generalisations found in spelling (e.g. that non-etymological spelling is particularly frequent in some phonological environment) are due to the ongoing sound change.

However, examples of hypercorrection (i.e. representing the etymological apical with spelling used for the laminal) are not only found in texts written in varieties which eventually merged fricatives into the laminal. Indeed, in many Western texts such hypercorrections are common (see examples in Zuloaga 2020), though they tend to be infrequent. In a situation of a merger in progress, when both laminal and apical fricatives are still present in the given variety, they might reflect speakers' uncertainty about their distribution. This would not be very different from what happened (according to the hypercorrection hypothesis) in Central Basque.

There are also a few Western Basque texts which show an important number of hypercorrections, even though scholars agree that the merger process must have been very advanced or completed in the variety and at the time they were written. This paper focuses precisely on such a situation. What happens in the spelling when an author attempts to represent the more conservative system no longer used in their variety but present in some other varieties? What factors influence the occurrence of hypercorrections when the language spoken by the author has (almost) completed the merger? Do we still find regularities similar to those described for Lubieta's manuscript? These questions are important, because the answers can not only add to our understanding of the Western merger, but they could also either support the conclusions presented in studies such as Krajewska *et al.* (2022) or pose a problem for the reasoning behind them.

The paper is organised as follows. Section 2 describes the data used in the study. Section 3 explains the objectives of the paper and formulates the predictions. Section 4 gives general descriptive statistics of the data. Section 5 presents the results of

statistical modelling. The paper closes with a discussion of results and conclusions in Section 6.

## 2. Data

Two texts were selected for this study. The goal was to have a more or less representative example of the (advanced) Western merger and another of the progressing Central merger. Moreover, both texts had to be long enough to allow a quantitative analysis.

The first text is *Diccionario en castellano y basquenze que sirve para la enseñanza de la vascongada* (Bilbao 2012). It is a bilingual manuscript from 1728 written by Joseph Domingo Lubieta for a merchant from Donostia so that he could learn Basque. As usual in such texts at that time, it contains a dictionary, elements of grammar, dialogues and other short texts. It is one of the oldest and most extensive sources which document the variety of Donostia. This text was analysed in Krajevska *et al.* (2022), and the same data set is used in this study.

The second source are manuscript sermons written by Bizente Sarria (1767-1835) (Etxebarria & Apraiz 2009). He was native of Etxebarri, located in the western part of the Western Basque dialectal area, where the mergers generally show more strength than in the eastern part (Zuloaga 2020: 413). The first three sermons were selected as a sample for this study.

The exact dating of the stages of mergers is notoriously problematic, and it is not always possible to know for sure whether or not it was completed in a given variety at a given time. As observed by Zuloaga (2020), various factors can obscure the picture, such as authors' particular choices of spelling or the fact that learned authors tended to reflect the etymological distinction very well regardless of their own variety. As for the texts used in this study, the analysis of Lubieta's manuscript in Krajevska *et al.* (2022) suggests that the Western merger was quite advanced in the variety of Donostia, and the Central merger was starting to gain strength. As for Sarria, Zuloaga (2010; 2020) affirms that his writings suggest that the merger was advanced in his variety and that it extended to all phonological contexts.

As regards data processing and annotation, all examples containing sibilants were extracted from the texts into a database. Data were then annotated for the following variables:

- SPELLING: *ConSpell* (conservative, etymological spelling) or *InnSpell* (innovative spelling).
- Etymological place of articulation (PLACE\_ETYM): *apical* vs *laminal*.
- Manner of articulation (MANNER): *fricative* vs *affricate*.
- CONTEXT: prevocalic word-medial prevocalic position (*\_V*), word-initial position (*#\_*), word-final position (*\_#*) and preceding a consonant (*\_C*).
- LOAN: *yes* / *old* (loanwords showing historical sound change) / *no* (recent loanwords).
- LEXEME: a standardised lexeme was assigned to each example (following entries from OEH). For this study (unlike in Krajevska *et al.* 2022), if the sibilant appears in a series of common affixes, the affix was used as lexeme, not the lex-

ical base. The suffixes in question are: *-z* (instrumental), *-gaz* (sociative), *-rantz* (directive), *-tze* (nominalization), *-tasun* (suffix deriving abstract nouns).

The data obtained in this way was then filtered. Firstly, examples for which etymological sibilant could not be determined were excluded (e.g. *zerbaist* ‘something’, which is likely to come from a laminal, but it is not certain). Secondly, affricates and recent loanwords were filtered out. Western and Central mergers differ in the fricative series, and thus affricates are not particularly relevant for the research questions posed in this paper. As for loanwords, it has been shown in Krajewska *et al.* (2022) that recent borrowings are much less likely to show innovative spelling than native lexicon and older loanwords. This is because the Spanish spelling was more normalised at that time, and writers were schooled in Spanish. Thus, in this paper I only focus on the part of the lexicon most likely to show the change. Nevertheless, some general observations on the whole dataset are also given in Section 3.

As regards the implementation, R was used for the analysis (R Core Team 2022), and the statistical models were fitted with the package *brms* (Bürkner *et al.* 2022).

### 3. Objectives and predictions

The goal of this paper is to analyse spelling in Sarria’s and Lubieta’s texts, focusing especially on etymologically apical fricatives. The main issue is whether there is evidence suggesting that in Sarria hypercorrections were purely graphical, unlike in Lubieta, where it has been argued that they correspond to the ongoing apical to laminal merger.

The first specific question concerns the influence of phonological context in both texts. It might be similar for both authors for etymologically laminal sibilants, with evidence of the merger stronger before consonants and word-finally than in the prevocalic position. For etymologically apical sibilants, however, the texts are expected to show differences. Previous research has shown that the prevocalic context favours apical to laminal neutralisation in Lubieta’s text, and that this can be interpreted as a representation of hypercorrections actually occurring in the variety of Donostia. If hypercorrections in Sarria are graphical and do not correspond to an ongoing sound change, it is unlikely that any clear contextual pattern will emerge or, if it does, it will have an explanation not related to phonology.

The second question is to what extent the spelling is conditioned by the lexeme itself and to what extent by other factors (such as the phonological context). We can predict that the choice of spelling will be particularly strongly determined by the lexeme in Sarria’s text, where the merger was completed or very advanced.

## 4. Data exploration

### 4.1. General observations

The two texts contain in total 16 830 tokens of sibilants, of which 13 956 are fricatives. Table 1 provides more details on the distribution of the sibilants in the corpus. As explained earlier, for the main analyses affricates and recent loanwords were excluded, which gave a total of 9 543 tokens, pertaining to 408 distinct lex-

emes. Table 2 shows the number of examples and distinct lexemes in the main corpus for each text and place of articulation, as well as the number of tokens after limiting the corpus as described.

**Table 1**

The number of examples for each sibilant in the whole corpus

MANNER	PLACE_ETYM	Text	Tokens
Fricative	Laminal	Lubieta	4017
		Sarria	4145
	Apical	Lubieta	1250
		Sarria	2459
Affricate	Laminal	Lubieta	890
		Sarria	629
	Apical	Lubieta	159
		Sarria	407

**Table 2**

The number of examples and distinct lexemes in the main corpus

PLACE_ETYM	Text	Tokens	Lexemes
Laminal	Lubieta	3714	129
	Sarria	3799	156
Apical	Lubieta	660	108
	Sarria	1370	126

**Table 3**

The number of distinct lexemes for each frequency range

Number of occurrences	Number of lexemes
1	150
2-10	171
11-50	55
51-100	12
101-300	12
301-1000	6
1001-	1

In the main corpus the mean number of tokens for each lexeme is 32 for laminals and 11 for apicals. However, there is huge variation (standard deviation is 134 for laminals and 30 for apicals). Thus, as expected in a corpus study, there are few extremely frequent lexemes (which, in the case of Basque, tend to have a laminal sibilant), and many infrequent ones. For example, as shown in Table 3, there are 150 lexemes with just one token and only 19 with more than 100. Table 4 lists 15 most frequent lexemes, among which are also a few suffixes. As can be seen, for many lexemes there are important differences in frequencies between the two texts, caused by dialectal differences or by the topic or textual genre.

Table 4  
15 most frequent lexemes in the main corpus

LEXEME	Total	Lubieta	Sarria
<i>izan</i>	1532	1163	369
<i>*edun</i>	807	436	371
<i>ez</i>	644	255	389
INS (-z)	495	110	385
<i>zu</i>	447	344	103
<i>guzti</i>	362	29	333
<i>zuek</i>	318	293	25
<i>zer</i>	267	70	197
<i>esan</i>	253	71	182
<i>*ezan</i>	226	226	0
SOC (-gaz)	211	0	211
<i>gauza</i>	204	42	162
<i>beste</i>	175	28	147
<i>hasi</i>	149	130	19
<i>asko</i>	136	31	105

## 4.2. Spelling

The writers use slightly different spelling to represent sibilants (Figure 2). Lubieta employs <s,c,z> for both fricatives and affricates, following the tradition of the Southern Basque Country before Larramendi (Mounole & Gómez López 2018: 493). As for Sarria, for the fricatives, he additionally uses <ss> and occasionally also <tz> and <ts>, though only the former appears in the fragment analysed (Zuloaga 2020: 378). He also distinguishes (though with various exceptions) fricatives and affricates.

As regards innovative vs conservative spelling, before focusing on the smaller dataset, some observations about the whole corpus are in order. Figures 3 and 4 show the proportion of innovative spelling for fricatives and affricates, respectively. It is particularly high for etymological apical affricates, and especially in Sarria (66% of all the examples, compared to 39% in Lubieta), suggesting that the merger was more advanced in affricates than in fricatives, where the proportion of innovatively spelled examples does not exceed 34%. However, given the differences between native words

(or old borrowings) and recent loanwords, the incidence of innovative spelling increases when the latter are excluded, as shown in Figure 5. On the whole, both texts show a significant proportion of sibilants spelled non-etymologically for both apical and laminal fricatives.

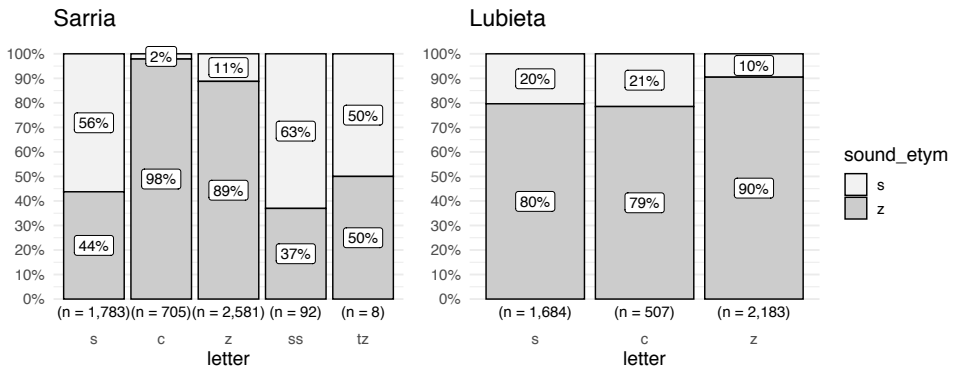


Figure 2

The spelling used in Lubieta’s and Sarria’s text (the main corpus)

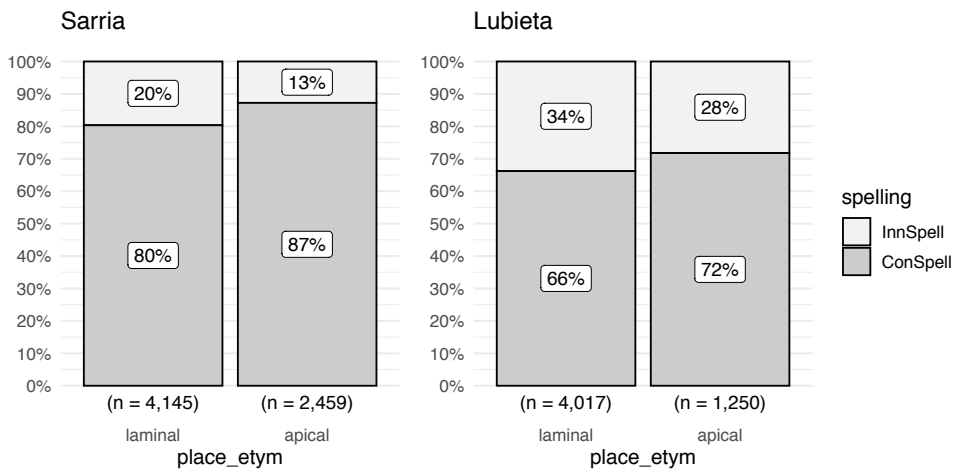


Figure 3

The proportions of innovative and conservative spelling for laminal and apical fricatives in the whole corpus



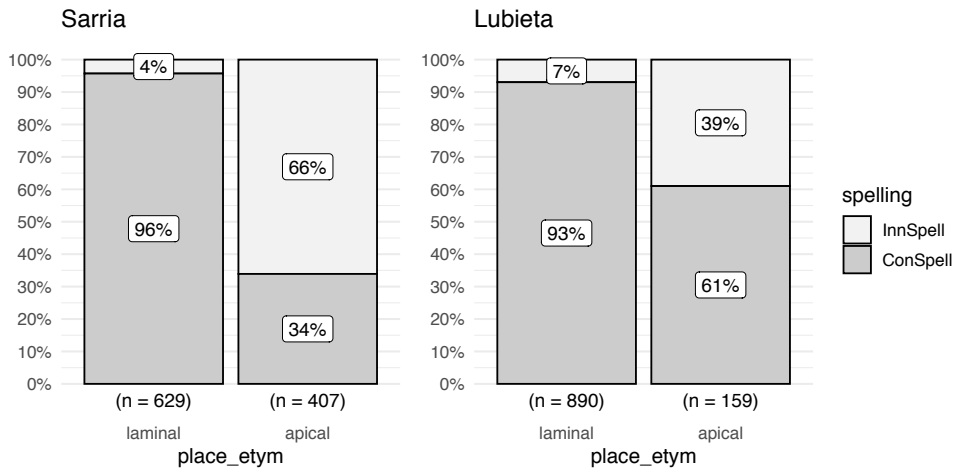


Figure 4

The proportions of innovative and conservative spelling for laminal and apical affricates in the whole corpus

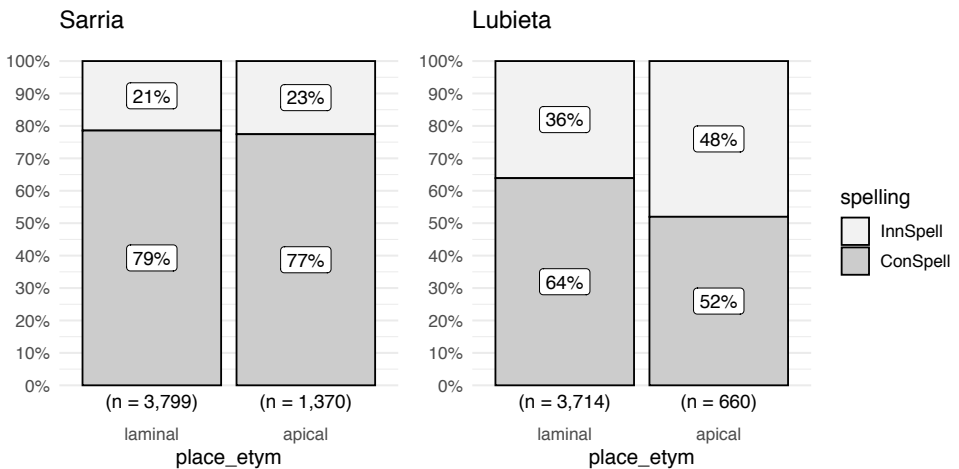


Figure 5

The proportions of innovative and conservative spelling for laminal and apical fricatives in the main corpus

Finally, even though it will be the object of statistical analysis in the next section, we can have a look at raw distributions of spellings across phonological contexts (Figures 6 and 7), and, especially for Sarria, give some examples (see Krajewska *et al.* 2022 for examples from Lubieta).

Innovative spelling in Sarria is found in all contexts, but with varying frequencies. For etymological laminals, it is most common before consonants and vowels (especially in lexemes such as *ezagutu* ‘know’, *gauza* ‘thing’, *adierazi* ‘express’). In other contexts innovative spelling is uncommon, but attested. Word-initially we have examples such as *siquindu* ‘get dirty’, *saspi* ‘seven’ or *sarratu* ‘close’. Word-finally we find it in various words and expressions containing the instrumental affix (e.g. *venturas* ‘maybe’, *obras* ‘by the work’, *legues* ‘like’, *gustis* ‘totally’) or in verbal forms (*dagos* ‘they are’ or *dozus* ‘you have them’).

For apicals, the proportion of InnSpell is highest word-finally, but notice the very small number of examples (13), which pertain to just three lexemes: *iguez* ‘escape’, *jolaz* ‘play’, and *dacuz* ‘he sees’. It is also quite high before consonants. The most common examples involve /t/ as the following consonant: *uste* ‘opinion’, *oste* ‘back’ *aste* ‘week’ and forms of the verbs *erakutsi* ‘show’, *irakatsi* ‘teach’ or *hasi* ‘start’. There are generally less examples with the sibilant before /k/, but there are instances of InnSpell there as well, in lexemes such as *esker* ‘thank’, *puskatu* ‘break’ or some forms of *edutsi*, the ditransitive auxiliary verb. Before word-internal vowels the nominalising suffix *-tasun* is the item most frequently spelled innovatively (102 out of 105 times). Other quite common examples include the following lexemes: *hasi* ‘start’, *erosi* ‘buy’ or *nagusi* ‘major’. There is also a handful of examples of innovative spelling word-initially, e.g. *zugue* ‘snake’, *zabelian* ‘in the womb’, *zortu* ‘create’ or *zuba* ‘fire’.

All these tendencies, however, must be taken with caution, as they show raw frequencies and do not take into account the effect of lexeme, and some frequent lexemes can easily skew the results. For example, the preconsonantal laminal in the lexeme *guzti* ‘all’ is spelled innovatively 313 times out of 333. The negation *ez*, however, appears in the preconsonantal position 223 times, but only 13 examples of the lexeme are spelled *es*.

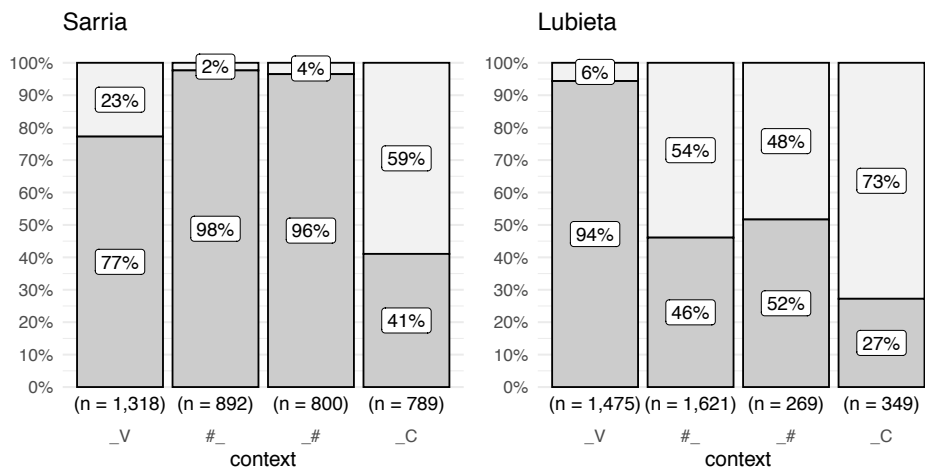


Figure 6

Innovative and conservative spelling across contexts for laminal sibilants

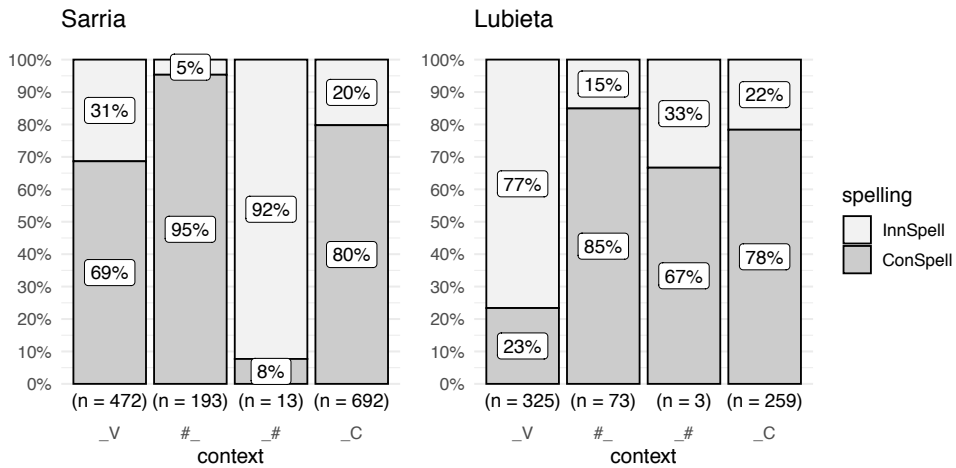


Figure 7

Innovative and conservative spelling across contexts for apical sibilants

## 5. Statistical modelling

### 5.1. Models

The goal of the statistical analysis is to model the influence a series of predictors on a binary output variable (innovative vs conservative spelling). This can be achieved with mixed-effects logistic regression, which has proven to be useful for analysing various topics in linguistics. Recently it has been used to study the sibilant mergers in Lubieta's text (Krajewska *et al.* 2022). The approach chosen here is similar to that paper, but Bayesian framework is employed instead. One of the reasons for this decision is the fact that complex hierarchical models are easier to define and fit in the Bayesian framework than in the frequentist one (Nicenboim & Vasishth 2016: 592-593; Schoot & Depaoli 2014: 79). Another aspect which makes this approach interesting for historical corpus studies is that it relies less on large samples (Schoot & Depaoli 2014: 79)

Three models were built to answer the questions that are the focus of this paper.

First of all, two models were fitted separately for Lubieta and Sarria with the following formula:  $\text{SPELLING} \sim \text{PLACE\_ETYM} * \text{CONTEXT} + (1 | \text{LEXEME})$ . That is, the model has both population-level effects (PLACE\_ETYM and CONTEXT and an interaction between them) and one group-level effect (LEXEME), which captures the fact that the examples are not completely independent (some lexemes appear more than once), and that not all lexemes show the same behaviour with regard to spelling. These models will be especially helpful to determine the importance of the effect of LEXEME as compared to other variables in the two texts.

Secondly, to analyse the importance of context the following model was fitted with data from both texts:  $\text{SPELLING} \sim \text{PLACE\_ETYM} * \text{CONTEXT} * \text{TEXT} + (\text{TEXT} | \text{LEXEME})$ . Thus, the model quantifies the effect of the variables PLACE\_ETYM, CONTEXT

and TEXT on spelling, as well as that of two-way interactions (PLACE\_ETYM \* CONTEXT, PLACE\_ETYM \* TEXT, CONTEXT \* TEXT) and a three-way interaction (PLACE\_ETYM \* CONTEXT \* TEXT). Additionally, the model includes the group-level effect of LEXEME, with varying slope for each text. That is, the model also captures the fact that there might be differences between texts in the way each lexeme is spelled.

Weakly informative priors were used in the models (see details in the supplementary materials). Such priors restrict the model in a way that unrealistically strong effects are less likely, but still allow for a wide range of values, if the data support it (McElreath 2016: 214).

Posterior predictive checks were performed for each model to ensure the models' adequacy. The diagnostics raised no concerns.

## 5.2. The effect of LEXEME

The first question concerns the importance of population-level effects (PLACE\_ETYM and CONTEXT) versus that of group-level effects (LEXEME).

One way to make such a comparison is with the help of the coefficient of determination  $R^2$ , which measures the proportion of total variance explained through all effects (conditional  $R^2$ ) or only population-level effects (marginal  $R^2$ ). The value of  $R^2$  was obtained separately from the model for Lubieta's text and from the model for Sarria's text. For Lubieta, marginal  $R^2$  is 0.15 (95% credible interval [0.11, 0.20]) and conditional  $R^2$  is 0.58 (95% CI [0.57, 0.59]). For Sarria, marginal  $R^2$  is lower (0.06; 95% CI [0.01, 0.14]), but conditional  $R^2$  is much higher (0.81; 95% CI [0.80, 0.82]). This suggests that the variables PLACE\_ETYM and CONTEXT (and their interaction) have more limited effect on spelling in Sarria's text than in Lubieta's. On the contrary, grouping at the level of lexeme is more relevant in Sarria than in Lubieta.

An additional useful measure is the interclass correlation coefficient (ICC), which can be used to find out how similar are the observations in each group (it is closer to zero if the measurements in the group do not resemble each other and it is closer to one if they do). For Lubieta the values of ICC is 0.51 and for Sarria it is 0.85. This means that there are less differences in the way a given lexeme is spelled in Sarria's sermons than in the other source.

More comparisons between the texts can be done using the model fitted with data from both texts. First, the standard deviations of the effect of LEXEME for each text are different: for Sarria it is 4.4 (95% CI [3.7, 5.2]), and for Lubieta it is 1.9 (95% CI [1.5, 2.4]). This means that there is greater variability between the different lexemes in Sarria than in Lubieta (i.e. some lexemes are much more likely than the mean to show innovative spelling, and others are much less likely).

This is visualised in Figure 8, which shows the probability of InnSpell for the most frequent lexemes in each text (more than 12 tokens in both texts taken together). Thus, these are the lexemes for which the model had the most data to calculate precise estimations (though some lexemes appear frequently enough only in one of the sources; in such cases the estimations for the other text are imprecise). It appears that the spelling is more consistent in Sarria's text: there are a few lexemes very likely to show InnSpell and many very unlikely to do so. In Lubieta's data there are

more lexemes which show middle values, and there are fewer items which show extremely high probability of InnSpell.

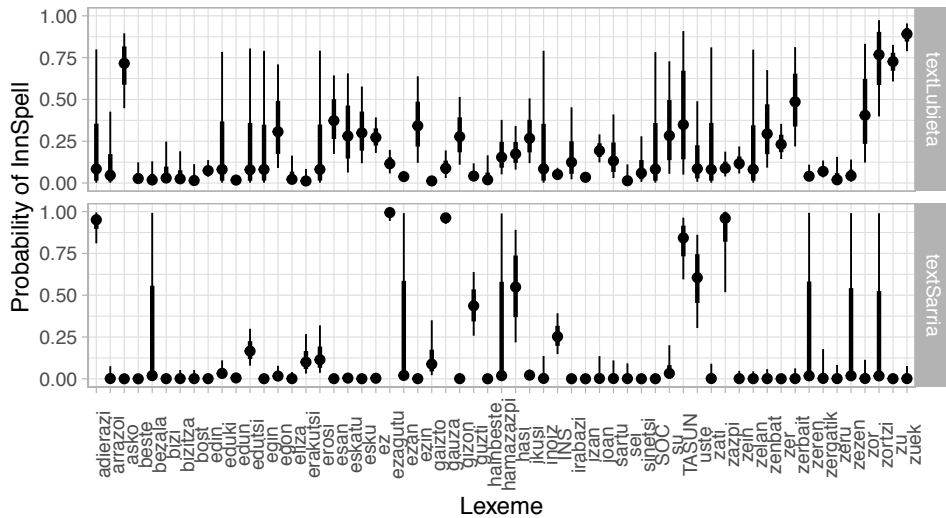


Figure 8

The probability of InnSpell for the most frequent lexemes in the texts

### 5.3. The effect of CONTEXT

The second question concerns the effect of phonological context on spelling. To answer it I will use the model built with the data from both texts.

Figure 9 shows the posterior distribution of the probability of InnSpell for etymologically apical and laminal sibilants, separately for the two texts and the four phonological contexts. The points represent the mean, bold lines are 66% credibility intervals and thin lines are 95% credibility intervals.

Starting with Lubieta, the results are (unsurprisingly) similar to those reported in Krajewska *et al.* (2022). For laminal sibilants the probability of InnSpell follows the hierarchy proposed in the previous literature:  $_C \ \& \ \_ \# \ > \ \# \_ \ \& \ \_V$ . The patterns for apical fricatives are different. Word-final apicals are usually spelled with <z> (but there are very few native words with an apical sibilant in word-final position). Innovative spelling is most probable in the prevocalic position, and least so in word-initial and preconsonantal context.

In Sarria's text, the overall probability of InnSpell is lower. For etymological laminals, it is highest in the preconsonantal position, but very low in the remaining positions (even though some examples are found in all contexts). For apical sibilants, the word-final context, similarly to what happens in Lubieta, show highest probability of InnSpell, but it must again be kept in mind that the amount of examples with word-final apical sibilant is very small. Leaving the word-final context apart, InnSpell appears most in preconsonantal and prevocalic positions. Before consonants we find

both etymological laminals and apicals spelled innovatively quite often. Before word-internal vowels, however, while laminals tend to be spelled conservatively, apicals show higher probabilities of being spelled innovatively. Word-initially, sibilants are mostly spelled etymologically.

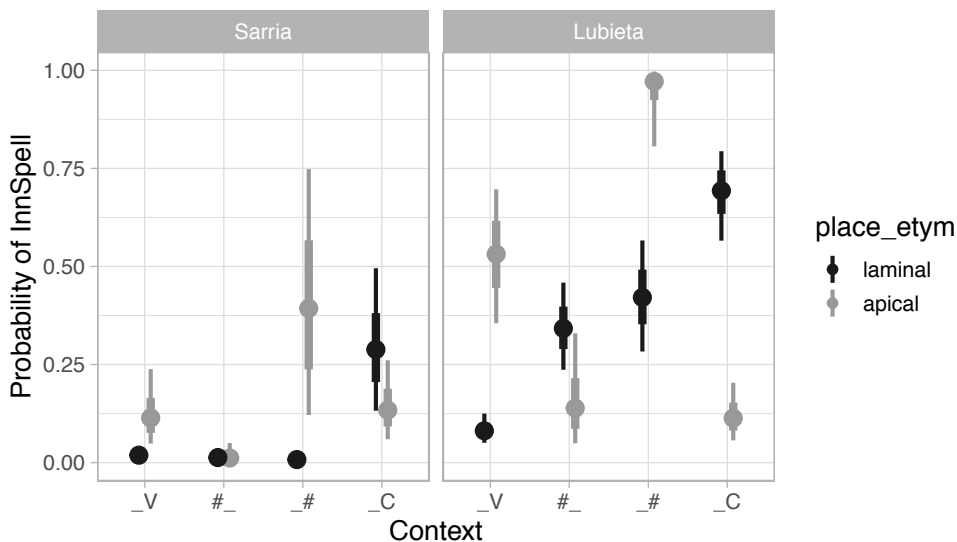


Figure 9

The probabilities of innovative spelling in the four phonological contexts

## 6. Conclusions

The analysis of spelling and its relation to the phonological system of any given time is notoriously problematic (Minkova 2015; Stenroos 2002). As regards Basque sibilant mergers, Zuloaga (2020) observes that in many Western texts the spelling is completely or to very large extent etymological, even though other available data suggest that the merger was very advanced in the author's variety. Sarria's text, the main object of this study, does show an important proportion of sibilants spelled innovatively, but the data is not necessarily straightforward to interpret.

As regards the issue of the influence of phonological context on the distribution of innovative and conservative spellings, for data from Lubieta's dictionary it has been argued in Krajewska *et al.* (2022) that the contexts which favour non-etymological spelling are different for apical and laminal fricatives, because of the way the Western merger and the Central merger developed and spread. As the change from laminal to apical fricative started to extend beyond the preconsonantal position, the merger was pushed in the opposite direction, and this is why we find hypercorrections in the prevocalic position.

Moving to Sarria, the prediction was the following: since the Western merger was completed or very advanced in his variety, we do not expect the context to play a very important role. Nevertheless, it is possible that patterns similar to those found

in Lubieta appear for etymologically laminal fricatives. As regards hypercorrections (i.e. apicals spelled with <z> or <c>), unlike in Lubieta, they certainly do not occur due to the laminal fricative being used in the spoken language by Sarria. Because of that, my prediction was that they will not be especially related to any given position. In general, instead of being driven by the context, the distribution of non-etymological spelling will be determined by other factors, such as lexeme.

Given the data and statistical models presented in the paper, I think that these predictions have been largely confirmed, but there are a few unexpected results as well.

Generally speaking, the role of phonological context is rather limited for Sarria's data as suggested by the very low proportion of variance explained by the variable of context. Instead, the way in which a word is spelled is more related to the lexeme: it appears that the writer decided to write a given lexeme in some way, regardless of the etymological value, and then stuck to it (though exceptions are possible).

Nevertheless, I have also shown that in Sarria's text there is an important difference between the preconsonantal and other contexts for laminal fricatives: only preconsonantal sibilants are frequently spelled innovatively. This is a common pattern in most Western texts (see Zuloaga 2020). Also apical fricatives are often spelled non-etymologically in the preconsonantal position, as are prevocalic sibilants (but not word-initial ones). Thus, for apical fricatives the phonological environment does play some albeit limited role.

The most puzzling result emerges when we compare the tendencies in apical and fricative sibilants: in the preconsonantal context both are frequently spelled non-etymologically, but in the word-initial position there are very few examples of that. This could suggest that the merger was indeed advanced, but not completed in Sarria's variety: the word initial position can be seen as perceptually the most salient position and likely to be affected last in a merger process. This is an interesting result which calls for further study on the chronology of Western merger.

In summary, using quantitative methods I have shown in this paper that Lubieta's and Sarria's text show differences in the distribution of hypercorrections, which are consistent with the different linguistic situation at the time they were written in the Central and Western Basque, respectively.

## References

- Beristain, Ander. 2018. Azpeitiko euskararen hiztegi dialektal bat – I. *FLV* 125. 7–54. <https://doi.org/10.35462/FLV125.1>.
- Beristain, Ander. 2021. Spectral properties of anterior sibilant fricatives in Northern Peninsular Spanish and sibilant-merging and non-merging varieties of Basque. *Journal of the International Phonetic Association* 52(3). 421–452. <https://doi.org/10.1017/S0025100320000274>.
- Bilbao, Gidor. 2012. Lubieta, izen berri bat euskal testuen historian. In Joseba A. Lakarra, Joaquín Gorrochategui & Blanca Urgell (eds.), *2nd Conference of the Luis Michelena Chair - Koldo Mitxelena Katedraren II. Biltzarra - II. Congreso de la Cátedra Luis Michelena*, 621–649. Vitoria-Gasteiz: UPV/EHU.

- Bürkner, Paul-Christian, Jonah Gabry, Sebastian Weber, Andrew Johnson, Martin Mo-drak, Hamada S. Badr, Frank Weber, Mattan S. Ben-Shachar & Hayden Rabel. 2022. brms: Bayesian Regression Models using “Stan.” <https://CRAN.R-project.org/package=brms> (2024/01/13).
- Camino, Iñaki. 2000. Abiaburu bat Donostiako euskera aztertzeko. In Koldo Zuazo (ed.), *Dialektologia gaiak*, 53-84. Vitoria-Gasteiz: Arabako Foru Aldundia.
- Egurtzegi, Ander & Christopher Carignan. 2020. An acoustic description of Mixean Basque. *The Journal of the Acoustical Society of America* 147(4). 2791-2802. <https://doi.org/10.1121/10.0000996>.
- Etxebarria, Nagore & Ainara Apraiz. 2009. *Bizente Sarria (1767-1835) sermoitegia: XVIII. eta XIX. mendeetako euskal oratoria sakratua* (Jarein). Bilbo: Labayru Ikastegia & Bilbo Bizkaia Kutxa.
- Hualde, José Ignacio. 2003. Segmental phonology. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.), *A grammar of Basque*, 15-64. Berlin: De Gruyter Mouton.
- Hualde, José Ignacio. 2010. Neutralización de sibilantes vascas y seseo en castellano. *Oihernart: cuadernos de lengua y literatura* 25. 89-116.
- Iribar Ibabe, Alexander, Rosa Miren Pagola Petrirena & Itziar Túrrez Aguirrezábal. 2020. Caracterización articulatoria de las fricativas sibilantes alveolares del euskera: Primera parte. *Loquens* 7(2). e073. <https://doi.org/10.3989/loquens.2020.073>.
- Krajewska, Dorota, Eneko Zuloaga & Ander Egurtzegi. 2022. Sibilant mergers in 18th-century Basque: a quantitative study. *Phonological Data and Analysis* 4(5). <https://doi.org/10.3765/pda.v4art5.67>.
- McElreath, Richard. 2016. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton: CRC Press.
- Minkova, Donka. 2015. Establishing phonemic contrast in written sources. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 72-85. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199232819.013.024>.
- Michelena, Luis. 1977. *Fonética histórica vasca*. 2nd edn. Donostia: Gipuzkoako Foru Aldundia. (1st edn., 1961; repr. in *OC*, vol. 6).
- Michelena, Luis. 2011. *Obras completas (Supplements of ASJU 54-68)*, 15 vol. San Sebastián & Vitoria-Gasteiz: “Julio Urkixo” Basque Philology Seminar, Gipuzkoa Provincial Council & UPV/EHU (ed. by Joseba A. Lakarra & Iñigo Ruiz Arzalluz).
- Mounole, Céline & Ricardo Gómez López. 2018. Euskara Zahar eta Klasikoa. In Joakin Gorrotxategi, Iván Igartua & Joseba A. Lakarra (eds.), *Euskararen historia*, 469-541. Vitoria-Gasteiz: Eusko Jaurlaritza.
- Muxika-Loitzate, Oihane. 2017. Sibilant merger in the variety of Basque spoken in Amorebieta-Etxano. *Languages* 2(4). 25. <https://doi.org/10.3390/languages2040025>.
- Nicenboim, Bruno & Shravan Vasishth. 2016. Statistical methods for linguistic research: Foundational Ideas—Part II. *Language and Linguistics Compass* 10(11). 591-613. <https://doi.org/10.1111/lnc3.12207>.
- R Core Team. 2022. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/> (2024/01/13).
- Sagarzazu, Txomin. 2005. *Hondarribiko eta Irungo euskara*. Irun: Alberdania.
- Schoot, Rens van de & Sarah Depaoli. 2014. Bayesian analyses: Where to start and what to report. *European Health Psychologist* 16(2). 75-84.



- Stenroos, Merja. 2002. Free variation and other myths: interpreting historical English spelling. *Studia Anglica Posnaniensia: International Review of English Studies* 38. 445-469.
- Zuloaga, Eneko. 2010. Nerbioi ibarreko euskara XVIII eta XIX. mendeetan: (I) Bizente Sarriaren bizkaierazko sermoiak. *ASJU* 44(2). 393-435. <https://doi.org/10.1387/asju.7337>.
- Zuloaga, Eneko. 2020. *Mendebaleko euskararen azterketa dialektologiko-diakronikorantz*. Vitoria-Gasteiz: UPV/EHU doctoral dissertation. <https://addi.ehu.es/handle/10810/49846>.

