

Final version published as:

Krajewska, D. (2022). The marking of spatial relations on animate nouns in Basque: A diachronic quantitative corpus study. *Journal of Historical Linguistics*, 12(2), 241–281.
<https://doi.org/10.1075/jhl.20061.kra>

The marking of spatial relations on animate nouns in Basque: a diachronic quantitative corpus study

Dorota Krajewska

University of the Basque Country UPV/EHU
dorota.krajewska@ehu.eus

Abstract

This corpus-based study examines the diachrony of differential place marking in Basque. In spatial cases, animate nouns in Basque exhibit heavier morphological forms than inanimate ones, but, under some circumstances, they can also be marked as inanimate. The data for the study comprises 66 sixteenth-to-twentieth-century texts (9,791 examples). A generalised linear mixed-effects model was fitted to analyse factors influencing the choice of marking. It is shown that animate nouns are sensitive to different aspects of the extended Animacy Hierarchy. The strongest effect is that of number (singular nouns prefer animate marking), followed by referentiality (pronouns are more prone to take animate forms than other nominals), and definiteness (definite nouns show animate marking more often than indefinite ones). The analysis also shows that animate marking became more widespread, and that there are dialectal differences. Moreover, more factors were relevant for the alternation in the earliest data (number, referentiality, definiteness, person and case) than in the most recent texts, where number is the most important.

Key words: Basque; animacy; morphology; spatial cases; diachrony

1 Introduction

The ways in which animacy interacts with different aspects of grammar have been extensively studied in linguistic typology (for example in Comrie 1989, Dahl & Fraurud 1996, Yamamoto 1999, Kittilä, Västi, & Ylikoski 2011, De Swart & de Hoop 2018, Santazilia 2019, 2020), but less often from a diachronic perspective (but see Cristofaro 2013 or Igartua & Santazilia 2018a).

This paper is a case study in the diachrony of what has been called by Haspelmath (2019) “differential place marking”. High animacy nouns, especially those referring to human beings, are not prototypical locations or targets/sources of movement, and, in some languages, this makes them incompatible with spatial cases, or they need additional morphology to take those cases (Aristar 1997, Creissels 2009, Creissels & Mounole 2011, Haspelmath 2019). Haspelmath (2019: 315) expressed this phenomenon in the form of the following universal regularity: “Deviations from usual associations of

role meanings and properties of referring expressions tend to be coded by longer grammatical forms”. Examples of languages in which this regularity operates, provided by Aristar (1997), include some Australian languages (e.g. Dyirbal), Dravidian languages (e.g. Kannada), and Basque.

In Basque, animate nouns show heavier morphological marking when used in spatial cases: the morpheme *-gan* (1) or the postposition *baita-* (2) is added (the choice depends especially on the dialect, as will be discussed later). Inanimate nouns (4) attach spatial endings directly. However, under certain circumstances, animate nouns can also appear without bridge morphemes (3).

- (1) *emakume-a-gan-dik*
 woman-DEF-ANIM-ABL
 ‘from the woman’
- (2) *emakume-a-ren baita-rik*
 woman-DEF-GEN ANIM-ABL
 ‘from the woman’
- (3) *emakume-eta-ra*
 woman-PL-ADL
 ‘to the women’
- (4) *etxe-tik*
 house-DEF.ABL
 ‘from the house’

This study is based on a corpus of sixteenth- to twentieth-century texts. I adopt a quantitative approach by using logistic regression and, more specifically, generalised linear mixed-effects model, a method not previously applied to Basque diachronic data. The goal is to analyse the factors which influence the choice of marking: the dedicated animate marking with *-gan/baita-* or the more general marking, and how this conditioning changes in time and space.

I focus especially on the question of which aspects of the (extended) Animacy Hierarchy (5) play a role. Following, Croft (1993: 130–132), I consider the subhierarchies of person, referentiality, animacy and definiteness. I also take into account the difference between singular and plural, included in the discussion of

animacy by some authors, for example, by Timberlake (1975) or Langacker (1991: 308).¹

- (5) Animacy: human < animate < inanimate
Person: first, second < third
Referentiality: pronoun < proper name < common noun
Definiteness: definite < indefinite
Number: singular < plural

The paper is structured as follows. Section 2 introduces the marking of spatial relations in Basque in inanimate and animate nouns, and summarises what we know about its diachrony. Section 3 poses the questions the study attempts to answer. Section 4 deals with methodology, corpus, and variables. Section 5 presents the results. Finally, Section 6 answers the research questions and comments on the applicability of methods used here to study the diachrony of a language such as Basque.

2 Basque spatial cases and animacy

In Basque, as observed by Igartua and Santazilia (2018b), the distinction between animate and inanimate nouns is especially important for the coding of direct objects (in some dialects) and spatial cases.

Basque has several spatial cases which encode location (inessive, *-(e)an*), movement from a source (ablative, *-tik/-rik*),² movement towards a goal (allative, *-ra(t)*), direction of movement (directional allative, *-rantz*) and endpoint of the movement (terminative, *-raino*) (Ibarretxe-Antuñano 2004, Creissels & Mounole 2011: 168).

There are various ways in which animate nouns can appear in spatial cases. In the general marking (GenM henceforth), the noun takes the endings inanimate nouns do: *-(e)an* for the inessive, *-ra(t)* for the allative, and *-tik/-rik* for the ablative (plural is marked with *-eta-* and indefinite with *-ta-*). There are two options for the animate marking (AnimM): the suffix *-gan* or the postposition *baita-* (*baitha-* in dialects with

¹ See Santazilia (2019: 49–50) for a list of subhierarchies affecting extended animacy considered in the literature.

² The archaic ablative *-rean* was also taken into account for this study (see Lakarra 1984 for a discussion of ablative endings in old texts). The diachronic and dialectal aspects of the allomorphy in local cases are discussed in Santazilia (2013: 247ff).

aspirated stops). Eastern dialects (Labourdin, High Navarrese, Low Navarrese, Souletin) use *-gan* and *baita-*, and only the former is found in western varieties (Biscayan, Guipuscoan and Alavese), both in the modern language and in the historical sources (Creissels & Mounole 2011: 172).

With *baita-*, we have *-n* in the inessive, *-ra(t)* in the allative, and *-tik/-rik* in the ablative. As for *-gan*, the morpheme of animacy is most probably *-ga*. It might be related to the ergative *-k*, as suggested by Lakarra (2005). It is also present in the western Basque comitative suffix *-ga-z* (*-z* is the instrumental case), e.g. *lagun-a-gaz* ‘with the friend’ [friend-DEF-COM]. In the inessive the suffix is *-gan*, to which *-a(t)* is added in the allative and *-dik* or *-ik* in the ablative. Both *-gan* and *baita-* can be attached to the nominal in the genitive (*-(r)en*) or in the absolutive, with the genitive more common in the modern language.

Table 1 shows the different options for the coding of animate nouns for the definite (singular and plural) and indefinite forms (which do not distinguish number).³ The table is based on the corpus analysed for this study and attempts to represent the different possibilities found in texts, but not all forms are used in a given dialectal variety at a given time, and some options are very uncommon. This variability will be briefly commented on in the remainder of this section and more thoroughly throughout the paper.

Table 1. Spatial cases of *emakume* ‘woman’.

Case	GenM	AnimM	
		<i>-gan</i>	<i>baita-</i>
Singular			
Ines.	emakume-an	emakume-a-gan emakume-a-ren-gan	emakume-a baita-n emakume-a-ren baita-n
All.	emakume-ra(t)	emakume-a-gan-a(t) emakume-a-ren-gan-a(t)	emakume-a baita-ra(t) emakume-a-ren baita-ra(t)
Abl.	emakume-tik/-rik	emakume-a-gan-dik/-ik emakume-a-ren-gan-dik/-ik	emakume-a baita-tik/-rik emakume-a-ren baita-tik/-rik
Plural			
Ines.	emakume-eta-n	emakume-ak-gan emakume-en-gan	emakume-ak baita-n emakume-en baita-n
All.	emakume-eta-ra(t)	emakume-ak-gan-a(t) emakume-en-gan-a(t)	emakume-ak baita-ra(t) emakume-en baita-ra(t)
Abl.	emakume-eta-tik/-rik	emakume-ak-gan-dik/-ik emakume-en-gan-dik/-ik	emakume-ak baita-tik/-rik emakume-en baita-tik/-rik
Indefinite			

³ Generally speaking, definite forms are used to refer to concrete nouns known by the speaker and the hearer, and indefinite forms are employed when the referent is not concrete or not determined (Santazilia 2013: 224).

Ines.	emakume-ta-n	emakume-gan emakume-ren-gan	emakume baita-n emakume-ren baita-n
All.	emakume-ta-ra(t)	emakume-gan-a(t) emakume-ren-gan-a(t)	emakume baita-ra(t) emakume-ren baita-ra(t)
Abl.	emakume-ta-tik/-rik	emakume-gan-dik/-ik emakume-ren-gan-dik/-ik	emakume baita-tik/-rik emakume-ren baita-tik/-rik

The class of nouns taking AnimM comprises primarily humans, but also beings conceptualised as humans, for example, gods and other supranatural beings. Nevertheless, the range of nouns employed with the markers in question is poorly described (Euskaltzaindia 1991: 35). Plants are treated as inanimate, but there is more variability with animals, especially in the most recent data, where animals occasionally appear to be conceptualised similarly to human beings: for instance, the noun *zaldi* ‘horse’ shows GenM in (6), but AnimM in (7).

- (6) *Durduzatu-a erori da, trunko bat bezala, zaldi-tik.*
 confuse-DEF fall AUX.3SG log one like horse-DEF.ABL
 ‘Confused, he fell from the horse like a log’ (HiriartUrruti)⁴
- (7) *Zaldi-a-gan-a zetozen euli-ak arpegi ondo-tik*
 horse-DEF-ANIM-ALL come.PST.REL fly-DEF.PL face around-DEF.ABL
zebilzkion Txomin-eri.
 walk.PST.3SG Txomin-DAT
 ‘The flies which were coming to the horse were flying around Txomin’s face.’
 (AnabitarteDonostia)

There are a few other exceptions, e.g. the reciprocal pronoun *elkar* ‘each other’ which is often found with AnimM even with inanimate nouns. De Rijk (2008: 62) mentions several other types of occasional exceptions:

The concepts “animate” and “inanimate” that we have used denote grammatical categories, not biological ones. The boundaries of these categories are culture-specific and liable to individual variation. In Basque, on the one hand, plants and the smaller animals belong to the inanimate category, as do the terms *arima* ‘soul’, *gorputz* ‘body’, and *izpiritu* ‘spirit’. On the other hand, words like *eguzki* ‘sun’, *ilargi* ‘moon’, and *euskara* ‘the Basque language’ may occasionally be personified and then take the animate forms: *eguzkiagandik* ‘from the sun’, *euskaragana* ‘to the Basque language’.

⁴ Texts and sources are listed in the appendix.

On the whole, as Creissels and Mounole (2011: 169–170) observe, most exceptions can be explained “in terms of de-personification of animate nouns and personification of inanimate nouns”.

As for the usage, similarly to what happens in other languages (Creissels & Mounole 2011), animate nouns with spatial cases are very often used to express metaphorical and not strictly speaking spatial functions:

In Basque, the spatial forms of animate nouns are not only characterized by a relatively heavy morphological marking: they also tend to be avoided in the expression of genuine spatial relationships, and are mainly found in contexts in which spatial cases fulfil non-spatial functions that have only an etymological link with their primary spatial function. (Creissels & Mounole 2011: 170)

For example, the inessive is often found with predicates such as *sinetsi* ‘believe’ (8), or when the location is in the conceptual rather than physical space (9-10).

(8) *Nor baitha-n sinhes-ten duk hi-k?*
 who ANIM-INES believe-IPFV AUX.2SG_[A].3SG_[O] you-ERG
 ‘In whom do you believe?’ (LeizarragaAbc)

(9) *Ni baita-n du-ke-zu adiskide bat*
 I ANIM-INES have-FUT-2SG_[A].3SG_[O] friend one
 ‘You will have a friend in me’ (EtxepareBPrimitiae)

(10) *cegaiti ene bioç-a Dorido-gan dago*
 because my heart-DEF Dorido-ANIM.INES be.3SG
 ‘because my heart belongs to Dorido’ (Lazarraga)

Cases encoding movement are more likely to be used with spatial meaning, especially the allative, which often express something like ‘to the place where N stands’ as in (11) (Creissels & Mounole 2011: 171). The ablative with *-gan* is more restricted, according to Creissels and Mounole (2011: 171), and in the modern language it is most commonly found with *urrundu* ‘to move away from’. This situation is already found in the oldest texts, with the allative expressing physical (11) and metaphorical movement (12), and the inessive restricted to metaphorical uses. The ablative is rarely attested in purely spatial function in texts, but the oldest example comes already from the sixteenth century (13).

(11) *Nihaur sekretuki nator zu-gan-a*

- I secretly come.1SG you-ANIM-ALL
 ‘I come to you secretly’ (EtxepareBPrimitiae)
- (12) *dadukat amore zu-gan-a*
 have.1SG_[A].3SG_[O] love you-ANIM-ALL
 ‘I have love towards you’ (EtxepareBPrimitiae)
- (13) *Iakes-gan-ik batzu ethor zitezen baino lehen*
 James-ANIM-ABL some come.RAD AUX.AOR.3PL than before
 ‘before some came from James’ (LeizarragaTesta)

In varieties which use both markers, there are a few differences between them. The first concerns the spatial relations which tend to be encoded with each: *baita-* is most common in the inessive, and *-gan* is almost only found in cases expressing movement (Euskaltzaindia 1991: 237). This is related to subtle differences in meaning: *baita-* refers more to the inside of a person, especially in a figurative sense, and *-gan* refers to a person’s vicinity. This difference can be seen clearly when *baita-* is employed in the ablative (Euskaltzaindia 1991: 237). In (14), with *-gan*, the movement is not from inside the person, but from the space around him or her, while in (15), with *baita-*, the figurative movement is conceptualised as coming from inside the person.

- (14) *Urrun zaitez ni-gan-ik*
 move.away AUX.IMP.2SG I-ANIM-ALL
 ‘Move away from me.’ (Euskaltzaindia 1991: 237)
- (15) *Ni baita-rik sortu da pentsamendu hau*
 I ANIM-ABL create AUX.3SG thought this
 ‘This thought comes from me.’ (Euskaltzaindia 1991: 237)

Since the earliest texts, *baita-* is also often used in several metaphoric expressions, for instance, with predicates of saying or thinking, e.g. *esan/erran bere baitan* ‘say to oneself’, literally ‘say in oneself’ (16) or in *ez egon bere baitan* ‘be out of one’s wits’, literally ‘not be in oneself’ (17). When used in the allative, *baitara* can often be translated with ‘towards’ (18).

- (16) *Eskribe-ta-rik batzuk errai-ten zuten bere baita-n*
 scribe-INDEF-ABL some talk-IPFV AUX. 3PL_[A].3SG_[O] their ANIM-INES
 ‘some scribes were saying to themselves...’ (LeizarragaTesta)

- (17) *zeren bait-ziakiten gizon koleratu-a (...) etzego-ela*
 because SUB-know.PST.3PL_[A].3SG_[O] man angry-DEF NEG.be.PST-COMP
bere baitha-n
 his ANIM-INES
 ‘because they knew the angry man was not in his wits’ (AxularGero)
- (18) *Baina zertifika-tzen du bere karitate-a gu baitha-ra*
 but demonstrate-IPFV AUX.3SG_[A].3SG_[O] his love-DEF we ANIM-ALL
Iainko-a-k
 God-DEF-ERG
 ‘But God demonstrates his love towards us’ (LeizarragaTesta)

Nevertheless, *baita-* can also express physical location, especially with reference to someone’s usual residence:

An important difference in the uses of *gan* and *baita* is however that, contrary to *gan* (see Section 4.3), *baita* is attested with the meaning ‘at N’s (a person) usual residence’, as in (18) from a 19th century manuscript quoted by Mitxelena (1987–2005).

- (18) *Anaia baitha-ra doha.*
 brother *baitha-ALL* go.PRS.3SG
 ‘He is going to his brother’s.’

This use of *baita* is consistent with the fact that *baita* is also found as the second formative of oiconyms, for example *Petrikobaita* ‘Peter’s’ (Biriadou), *Beñatbaita* ‘Bernard’s’ (Urrugne). (Creissels & Mounole 2011: 172)

The Basque *baita-* is in some respects similar to the French *chez* ‘at someone’s house’, which developed from the Latin noun *casa* ‘hut, house’ (Longobardi 2001, Harrison & Ashby 2003). It is an example of ‘house/home’ > locative grammaticalisation path (Kuteva *et al.* 2019: 233, 235–236). Harrison and Ashby (2003: 391–393) describe the semantic change in French in terms of metonymy and semantic generalization: from house (permanent or temporary lodging, (19)) to place of business (20), and eventually to people in a more figurative sense (21). As shown by Longobardi (2001), similar developments were initiated in other Romance languages, but did not proceed as far as in French. In particular, in the varieties of Occitan which were traditionally in contact with Basque, prepositions unrelated to *chez* were used.

- (19) *Je suis allé chez vous.*

- I AUX.1SG go.PTCP chez you.PL
 ‘I went to your place’
- (20) *Aller chez le coiffeur.*
 go chez the hairdresser
 ‘go to the hairdresser’s’
- (21) *Chez lui, c’est devenu une habitude.*
 chez him it.AUX.3SG become.PTCP a habit
 ‘It has become a habit for him’⁵

There are various parallels between the French *chez* and the Basque *baitan*: both can refer to someone’s residence or place of business. The difference is that the starting point in Basque was not a noun meaning ‘house’, but, as suggested by Creissels and Mounole (2011), rather a construction used to talk about people’s residence unrelated to any noun meaning ‘house’.⁶ Even though the topic requires a more detailed analysis, it seems to me that another difference is that the semantic extension have proceeded further in Basque than in French. This can be seen already in the sixteenth-century Basque translation of the New Testament (most probably translated from French): *baita-* corresponds to various French prepositions, the most common being *en* ‘in’ (as in (22), which is the French equivalent of (16)). Others include *vers* ‘towards’ or *entre* ‘between’. As for *chez*, it is found as the equivalent of *baita-* when someone’s lodging is being referred to.

- (22) *aucuns des Scribes disoyent en eux-mesmes*
 some of scribes were.saying in themselves
 ‘some scribes were saying to themselves’ (*Le Nouveau Testament* 1563)

As mentioned, animate nouns allow GenM in some circumstances. In the modern language, *-gan* and *baita-* are strongly preferred on singular nouns, but not necessarily

⁵ Examples (19)-(21) are from *Dictionnaire de l’Academie* (Academie française 2019: s.v. *chez*); glosses are mine.

⁶ Nevertheless, some scholars (e.g. Azkue 1923: 303–304) suggested that *baitan* could be a loanword from Romance (e.g. *baita* ‘tent’ or ‘hut’ in Piedmontese), but this explanation seems unlikely, especially because there are no traces of *baitan* used as a common noun (Trask 1997: 208, see Mitxelena 2011 [1970]: 264). Creissels and Mounole (2011) mention the possibility, already suggested by Mitxelena (2011 [1970]: 264), that *baita-* could have its origin in a relative clause with the verbal prefix *bait-*, but conclude that it is problematic, because, if this was the case, we would not expect genitive marking on the noun (e.g. *nire baitan* ‘in me’, with the genitive *nire*). However, this is not a problem since in the earliest texts the absolutive (e.g. *ni baitan* ‘in me’) actually predominates.

on plural or indefinite nouns (Euskaltzaindia 1991: 235). Nevertheless, there are important differences between dialects. In the Northern Basque Country GenM is possible with singular nouns for some speakers, as shown with examples from recent surveys: in (23) the speaker used *-gan*, but in the same context another speaker opted for GenM (24). Souletin Basque, as argued by Padilla-Moyano (2017: 763) is the most conservative as regards the coding of animate nouns, with GenM being more common than in other dialects.

(23) *Egorri ditut ene haurr-ak errienta-ren-gan-at*
 send AUX.1SG_[A].3PL_[o] my child-DEF.PL teacher-GEN-ANIM-ALL
 ‘I sent my children to the teacher’ (Oyharçabal, Epelde, & Salaberria 2009: A44)

(24) *Errienta-rat igorri ditut aurr-ak*
 teacher-DEF.ALL send AUX.1SG_[A].3PL_[o] child-DEF.PL
 ‘I sent children to the teacher’ (Oyharçabal *et al.* 2009: A44)

Meaning is also relevant for the choice of marking:

It is also important to keep in mind that the locative case allows an animate head noun to occur without the morph *-gan*. In these cases, the meaning is not ‘location (in, on, at)’ but ‘among’. So, for example, *mutil-en-gan* [boy-GEN.PL-gan.LOC] would mean ‘in/on the boys’, and *mutil-eta-n* [boy-PL-LOC] would be ‘among the boys’. (Ibarretxe-Antuñano 2004: 272)

A special case is the construction used to refer to a subset of a set, such as *gizonetarik batzuk* ‘some of the men’, in which the noun is marked with the ablative, and *-gan* and *baita-* are never used.

The diachronic changes in the use of animacy markers were studied by Creissels and Mounole (2011). Both markers are already found in the oldest Basque texts, but they appear to be innovations. Creissels and Mounole (2011) posit that the extension of *baita-* and *-gan* proceeded according to the Animacy Hierarchy: earlier in pronouns than in nouns and earlier in the definite singular than in definite plural or indefinite. They also found that in the earliest texts AnimM was more common in the allative and ablative than in the inessive. However, their study is based on a small corpus (two sixteenth-century and five eighteenth-century texts), which does not allow to quantify the results. Moreover, the size of the corpus is important for rarer forms. For instance, only three examples of *baita-* are found in the earliest data they analysed. Another

example are singular nouns with GenM, such as *gizonean* ‘in the man’, unattested according to Creissels and Mounole (2011: 176), but actually found in a few sources (25).

- (25) *dohain natural-ak gizon-ean hagitz korrupitu eta ezeztatu*
 gift natural-DEF.PL man-DEF.INES much corrupt and destroy
izan dira
 be AUX.3PL
 ‘The natural gifts have been corrupted and destroyed in the man.’
 (LeizarragaAbc)

Thus, previous research suggests that the diachronic changes in the marking of animate nouns happened along the Animacy Hierarchy, with important differences between dialects and between the spatial cases. The present study aims to provide a more detailed account of these changes.

3 Research questions

The goal of this paper is to analyse factors influencing the choice of marking on animate nouns, AnimM vs GenM. In particular, the study focuses on the following questions:

1. Is the marking of nouns referring to animate entities sensitive to the different aspects of the Animacy Hierarchy, as suggested by Creissels and Mounole (2011)? That is, are entities on the left side of hierarchies in (5), repeated in (26), more likely to be marked with AnimM, and those on the right more likely to appear with GenM? Which is the most important factor?
- (26) Animacy: human < animate < inanimate
 Person: first, second < third
 Referentiality: pronoun < proper name < common noun
 Definiteness: definite < indefinite
 Number: singular < plural
2. How does the spatial case influence the choice? In particular, is AnimM introduced earlier with cases expressing movement, i.e. allative and ablative, than with inessive? How could we explain this fact?

3. Does the proportion of nouns marked with AnimM as compared to those with GenM increase over time? Do the factors conditioning the choice of marking change?
4. Are there any dialectal differences in the marking of animate nouns? In particular, are eastern dialects more conservative than western dialects, and do they use GenM more extensively?

4 Methodology

4.1 Generalised linear mixed-effects models

Logistic regression is used to quantify the effects of various predictors on a binary dependent variable. Generalised linear mixed-effects models, apart from the usual predictors, referred to as “fixed effects”, additionally incorporate “random effects”, which allow models to better deal with unbalanced and interconnected data. In corpus linguistics, various data points usually come from the same source, thus violating the assumption of independence, necessary in the ordinary logistic regression (Gries 2015a: 99). Mixed-effects models, however, can take into account the variation related to e.g. author or specific item. Because of that, they are particularly adequate for corpus studies (Gries & Hilpert 2010, Gries 2015b,a), and have proven fruitful in diachronic corpus studies as well (see, for example, Gries & Hilpert 2010, Wolk *et al.* 2013, Barteld, Hartmann, & Szczepaniak 2016, Geleyn 2017, De Smet & Van de Velde 2020)

As for the building of the model, I use the top-down or backward model selection strategy (Zuur *et al.* 2009: 121–122, Gries 2013: 259–261). The first step is to fit a “beyond optimal” model, with all independent variables and as many interactions as possible or feasible. The next step is to evaluate whether the model should be simplified. The decision on whether or not discard a predictor, an interaction, or a random effect will be based on significance testing (an element should be discarded if the deletion does not make the model significantly worse). The effect of this procedure is the minimal adequate model.

I use *R* (R Core Team 2020) for the statistical analyses as well as the libraries *lme4* (Bates *et al.* 2015) and *ggeffects* (Lüdtke 2018). For plotting I use *sjPlot* (Lüdtke 2020) and *ggplot2* (Wickham 2016). The index of concordance C and the condition number for collinearity were calculated with *JGmermod* library (Grafmiller 2019), and

VIF values with *car* (Fox & Weisberg 2019). R^2 was obtained with the library MuMIn (Bartoń 2020).

4.2 Corpus

The analysis is based on tokens retrieved from a sample of sixteenth- to twentieth-century Basque texts from two major dialectal groups: dialects in which *baita-* and *-gan* are used (eastern varieties) and dialects in which only *-gan* is possible (western varieties). The sources are listed in the appendix. The sample used reflects the characteristics of the Basque historical corpus, in which religious texts (e.g. catechisms or sermons) predominate, especially before the nineteenth century. In general, more data is available for eastern dialects and therefore less examples were gathered for western dialects, especially for earlier periods.

There is no morphologically annotated corpus for the historical Basque, and so the data extraction was done largely manually. The first step consisted of searching for all words containing one of the relevant character sequences (among others, *-gan*, *baita-*, or the ending *-an*). The second step was to manually filter the results according to the following criteria:

- Only nouns referring to humans and animals were retained. Examples with inanimate nouns were discarded (168 examples in total).⁷
- As mentioned earlier, *baita-* can be used to form oikonyms. Examples found in the corpus (around 20) were not taken into account for further analysis.
- The subset construction, which uses GenM (*gizonetarik batzuk* ‘some of men’), was excluded (1,088 examples), especially because the construction has no counterpart with *-gan* or *baita-*.
- A few lexicalized adverbs with spatial cases were left out, such as *haurrean*, *haurretan*, *umetan* ‘in childhood’ (27), or *gatibutan* ‘enslaved’ (28). They do not refer to an animate entity but to a period of time or manner.

(27) *Gathibu-ta-n* *egoitia* *hala pena* *gaiz-en-a*
 slave-INDEF-INES be.NMLZ.DEF DEM suffering bad-SUPER-DEF
 ‘Being enslaved in this way is the worst suffering’ (EtxepareBPrimitiae)

⁷ This is because examples with GenM were also extracted for all lexemes found with AnimM. For instance, there are few tokens of the noun *gauza* ‘thing’ with AnimM. Retaining them for further analysis would require to add all examples of this noun marked with GenM, which could easily skew the results.

(28) *Iesus Iaun-a ere igan zen Hierusalem-a haurr-ean*
 Jesus lord-DEF also go AUX.PST.3SG Jerusalem-ALL child-DEF.INES
 ‘Lord Jesus also went to Jerusalem in childhood’ (EtxebZibuManual)

This procedure yielded 9,791 tokens, distributed over time periods and main dialectal areas as shown in Table 2. The number of tokens extracted from each text varies: the mean is 148 examples, the standard deviation is 230, the minimum is 10, and the maximum 1649 (the last figure comes from the sixteenth-century translation of the New Testament by Leizarraga). 78% of tokens show AnimM and 22% GenM.

Table 2. Data used in the study.

Period	Area	Tokens
1500-1600	East	2031
	West	85
1600-1750	East	2818
	West	108
1750-1876	East	1886
	West	603
1876-1940	East	1598
	West	662

4.3 Variables

The data was annotated with respect to the dependent variable MARKER with two levels: AnimM (-*gan* or *baita-*) and GenM.

The following extralinguistic predictors were chosen:

1. PERIOD. Four periods were distinguished (see Lakarra 1997, Gorrochategui, Igartua, & Lakarra 2018 on the periodisation of the history of Basque):
 - (1) Archaic Basque (1400-1600)
 - (2) Classical Basque (1600-1750)
 - (3) Early Modern Basque (1750-1876)
 - (4) Late Modern Basque (1876-1968; the latest texts analysed here are from the 1940s)
2. DIALECT. Two levels: *east* vs *west*. Two major dialectal areas were distinguished: eastern dialects (which use both *-gan* and *baita-*) and western (which only use *-gan*). More fine-grained dialectal distinctions would be problematic because of lack of data for some dialects and some periods.

Further predictors are related to linguistic properties: Animacy Hierarchy (1-4) and spatial case (5):

1. REFERENTIALITY. Three levels: *pronoun* vs *proper-div* (i.e. *proper noun* + *divinity*) vs *common noun*. Nouns referring to God and similar entities are extremely common in the Basque historical corpus, and it is reasonable to keep them apart from common nouns. They were grouped with proper nouns, which are not very common in the texts (and many proper names refer to Jesus).
2. NUMBER. Two levels: *singular* vs *plural*. Semantic value was considered. For example, *erregegana* ‘to the king’ is morphologically indefinite, but semantically singular. In turn, *milla presunatan* ‘in thousand people’ is also morphologically indefinite, but semantically plural.
3. PERSON. Two levels: 1/2 vs 3. The first and second person were grouped especially because of their much lower frequency as compared to the third person.
4. DEFINITENESS. Three levels: *definite* vs *definite_sem* (i.e. *semantically definite*) vs *indefinite*. *Definite* includes morphologically definite noun phrases, which follow the definite paradigm or have demonstratives:

- (29) *emazte-a-ren-gan-ik*
 women-DEF-GEN-ANIM-ABL
 ‘from the woman’ (BarbierXokoan)
- (30) *emakume-eta-ra*
 women-PL-ALL
 ‘to the women’ (ZabalaSermoiakI)
- (31) *gizon horr-en baita-rik*
 man that-GEN ANIM-ABL
 ‘from that man’ (HaranederEbanjelioa)
- (32) *zu-eta-ra*
 you-PL-ALL
 ‘to you’ (LeizarragaTesta)

The second value, *definite_sem*, includes nominals which are definite due to their semantics, even though their form is indefinite, e.g. first- and second-person pronominal forms (33) (except second person plural, which follows the

definite paradigm (32)), proper nouns or nouns which have unique referent and do not require definite article (34):

(33) *ni-ta-n*
I-INDEF-INES
'in me' (EtxepareBPrimitiae)

(34) *errege-gan-a*
king-ANIM-ALL
'to the king' (BeraiainMeza)

All other NPs are treated as indefinite (for instance (35), cf. (30)). Additionally, examples formally ambiguous between plural and indefinite are coded as indefinite (36), unless accompanied by a determiner, such as a possessive pronoun, in which case they are classified as definite. This is not optimal, but distinguishing between definite plural and indefinite forms is often problematic in historical texts.

(35) *emazte-ta-ra*
woman-INDEF-ALL
'to women' (MaterreDotrina)

(36)	<i>gizon-e-ta-n</i>	<i>gizon-eta-n</i>
	man-EPENTH-INDEF-INES	man-PL-INES
	'to men'	'to the men' (LeizarragaTesta)

5. CASE. Three levels: *inessive* vs *allative* (including also the directional and terminative) vs *ablative*.

Furthermore, certain interactions between variables were also added to the model:

1. Since the role of the linguistic variables might vary in time, interactions between all linguistic variables and the variable PERIOD are taken into account: REFERENTIALITY-PERIOD, NUMBER-PERIOD, PERSON-PERIOD, DEFINITENESS-PERIOD, and CASE-PERIOD.
2. Since the role of the linguistic variables might differ in the two dialectal areas, interactions between all linguistic variables and dialect were added: REFERENTIALITY-DIALECT, NUMBER-DIALECT, PERSON-DIALECT, DEFINITENESS-DIALECT, and CASE-DIALECT.

3. Finally, the PERIOD-DIALECT interaction is meant to capture the overall change in the behaviour of the dialectal areas in time.

The model also includes two random effects: varying intercepts for TEXT and LEXEME. The motivation for the first one is that various data points come from the same text, and there might be author-specific differences. The second effect adjusts the intercept for each lexeme appearing in the corpus.⁸ Random slopes (e.g. for TEXT and linguistic predictors) were not fitted not to overly complicate the model and also because not all texts have enough tokens to be able to calculate slopes for each predictor.

5 Analysis

5.1 Descriptive statistics

Out of 9,791 tokens analysed, 78% show AnimM and 22% GenM. In what follows, by building a regression model we will explore the results in more detail. Nevertheless, some introductory observations can be made on the basis of Error: no se encontró el origen de la referencia, which plots the proportion for GenM and AnimM for each independent variable in the study:

- PERSON: the first and second person are slightly more frequent with AnimM than the third person (84% vs 75%).
- NUMBER: GenM is not common in the singular (9%), but slightly more than half of the plural nouns appear with GenM.
- REFERENTIALITY: the proportion of GenM is similar for proper/divinity nouns and pronouns (around 12%), but much higher for common nouns (over 50%).
- DEFINITENESS: indefinite nouns show GenM in two thirds of examples. Definite nouns are mostly marked with AnimM. Semantically definite but formally indefinite nouns have AnimM more often than other definite nominals (88% vs 77%).
- CASE: allatives are slightly more common with AnimM than ablatives and inessives (83% vs 75-76%).
- DIALECT: texts from the eastern varieties have more GenM than those from the western dialects (24% vs 10%).

⁸ A standardised lexeme from the *General Basque Dictionary* (Euskaltzaindia 2019) was provided for each example.

- PERIOD: the highest proportion of GenM is found in the first time period, i.e. in the sixteenth century (35%). Later the proportion of GenM is roughly stable (16%-22%).

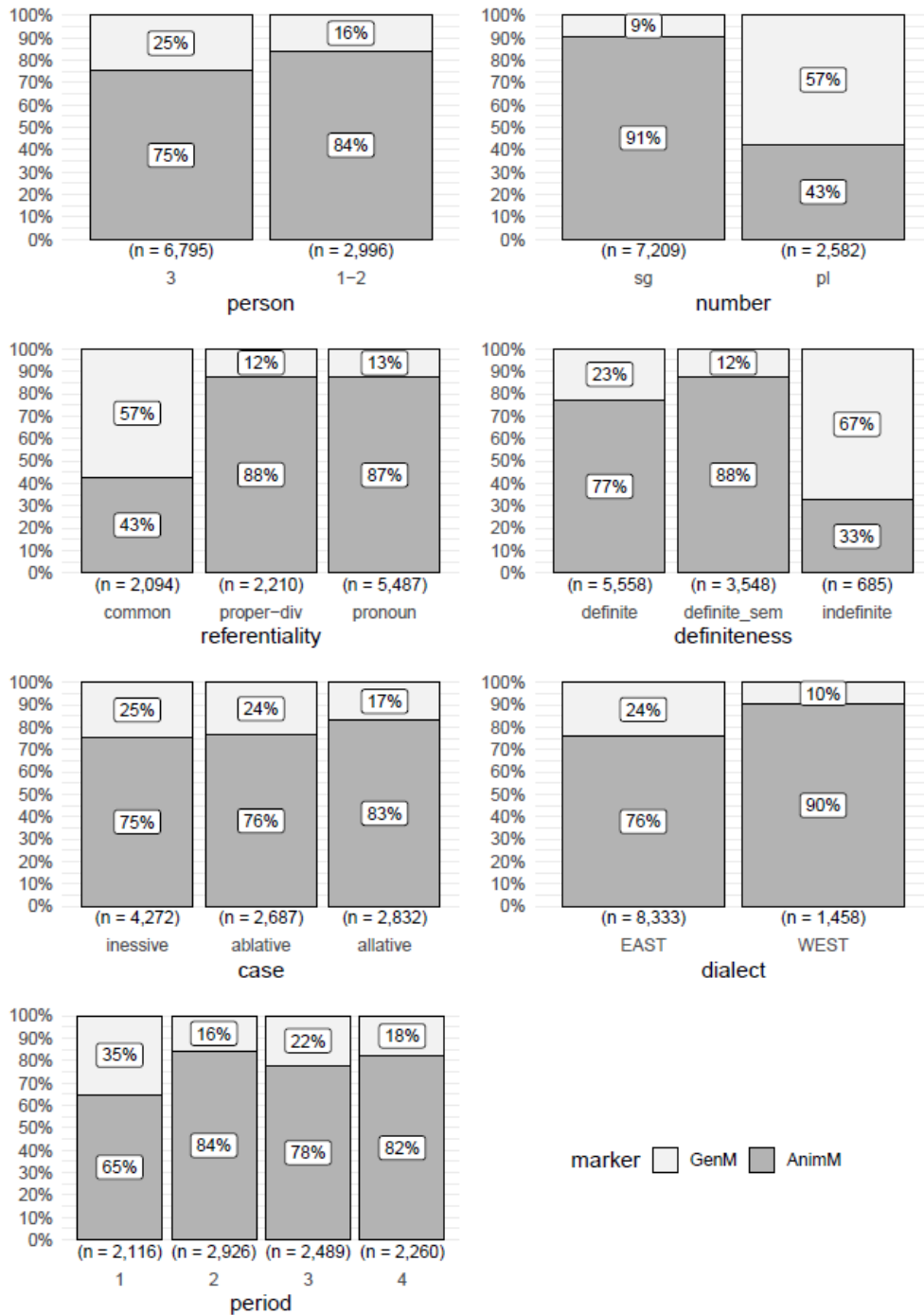


Figure 1. Proportion of GenM and AnimM for the variables analysed in the study.

5.2 Model fitting and model selection

The first model consisted of the two random effects, all independent variables and interactions as specified above.

As for the random effects, the inclusion of LEXEME improves the model. A likelihood ratio test shows that the model with the random effect is significantly better than the model without ($\chi^2(1) = 831.7, p < .0001$). The same applies to the random effect TEXT ($\chi^2(1) = 473.3, p < .0001$).

Moving to the fixed effects, a likelihood ratio test indicates that the exclusion of interactions PERSON-DIALECT and CASE-DIALECT results in a model which is not significantly worse ($\chi^2(3) = 1.01, p = 0.78$). Further deletions result in worsening of the model's fit. Thus, the final minimal model consists of the following:

- random effects of TEXT and LEXEME
- variables of CASE, DEFINITENESS, NUMBER, PERSON, REFERENTIALITY, PERIOD and DIALECT
- interactions: CASE-PERIOD, DEFINITENESS-PERIOD, NUMBER-PERIOD, PERSON-PERIOD, REFERENTIALITY-PERIOD, DEFINITENESS-DIALECT, NUMBER-DIALECT, REFERENTIALITY-DIALECT, PERIOD-DIALECT

The results of this model are given in Table 3 and Table 4 and will be discussed in the next section. Before that I will comment on the model's goodness of fit.

Table 3. Fixed effects in the final mixed-effects logistic regression model. Predicted odds are for GenM. The variables' reference levels are the following: singular, definite, inessive case, 3rd person, common noun, eastern dialect, 1st period.

Predictor	Estimate (log odds)	SE	z value	<i>p</i>
(Intercept)	-0.40	0.67	-0.59	0.557
case [ablative]	-2.29	0.22	-10.52	< 0.001
case [allative]	-2.80	0.22	-12.49	< 0.001
definiteness [definite sem]	-0.58	0.31	-1.87	0.062
definiteness [indefinite]	2.27	0.61	3.71	< 0.001
number [pl]	5.62	0.29	19.24	< 0.001
person [1-2]	2.63	1.23	2.13	0.033
referentiality [proper-div]	-1.81	0.42	-4.35	< 0.001
referentiality [pronoun]	-3.52	0.59	-5.98	< 0.001
period [2]	-1.07	0.79	-1.35	0.176
period [3]	-0.72	0.91	-0.79	0.427
period [4]	-2.73	0.91	-2.99	0.003
dialect [WEST]	-3.25	1.58	-2.06	0.039

case [ablative] * period [2]	0.33	0.37	0.89	0.371
case [allative] * period [2]	0.38	0.42	0.90	0.370
case [ablative] * period [3]	0.71	0.31	2.27	0.023
case [allative] * period [3]	1.05	0.33	3.19	0.001
case [ablative] * period [4]	2.85	0.33	8.57	<0.001
case [allative] * period [4]	2.71	0.35	7.74	<0.001
definiteness [definite sem] * period [2]	0.84	0.47	1.78	0.075
definiteness [indefinite] * period [2]	2.05	0.78	2.61	0.009
definiteness [definite sem] * period [3]	1.65	0.37	4.47	<0.001
definiteness [indefinite] * period [3]	-0.10	0.70	-0.15	0.882
definiteness [definite sem] * period [4]	3.15	0.54	5.82	<0.001
definiteness [indefinite] * period [4]	0.04	0.70	0.06	0.951
number [pl] * period [2]	0.33	0.38	0.87	0.382
number [pl] * period [3]	-1.17	0.33	-3.51	<0.001
number [pl] * period [4]	-0.89	0.34	-2.60	0.009
person [1-2] * period [2]	-1.57	0.57	-2.73	0.006
person [1-2] * period [3]	-2.18	0.46	-4.76	<0.001
person [1-2] * period [4]	-3.85	0.59	-6.50	<0.001
referentiality [proper-div] * period [2]	-1.36	0.52	-2.64	0.008
referentiality [pronoun] * period [2]	-1.21	0.56	-2.15	0.032
referentiality [proper-div] * period [3]	-1.19	0.47	-2.52	0.012
referentiality [pronoun] * period [3]	-0.36	0.46	-0.79	0.432
referentiality [proper-div] * period [4]	-1.24	0.72	-1.72	0.085
referentiality [pronoun] * period [4]	0.61	0.49	1.26	0.207
definiteness [definite sem] * dialect [West]	-4.03	1.17	-3.44	0.001
definiteness [indefinite] * dialect [West]	-0.55	0.70	-0.79	0.427
number [pl] * dialect [West]	2.54	0.62	4.10	<0.001
referentiality [proper-div] * dialect [West]	-1.81	0.83	-2.19	0.028
referentiality [pronoun] * dialect [West]	1.11	0.59	1.87	0.061
period [2] * dialect [West]	2.21	1.95	1.13	0.257
period [3] * dialect [West]	1.18	1.75	0.68	0.500
period [4] * dialect [West]	-2.18	1.78	-1.23	0.219

Table 4. Random effects in the final mixed-effects logistic regression model.

Random effect	<i>N</i> of groups	SD
LEXEME	517	2.20
TEXT	66	1.50

To begin with, the model's index of concordance C , which measures how well the model predicts the data, is 0.97 (if it is 0.5 the predictions are random, 1 is perfect correlation between the predictions and the data). Another measure of goodness of fit we can use is pseudo- R^2 (Nakagawa & Schielzeth 2013, Nakagawa, Johnson, & Schielzeth 2017). The marginal R^2 measures the variance explained by the fixed effects, and the conditional R^2 express the variance explained by the entire model, i.e. including

random effects. For the model fitted here, the marginal R^2 is 0.62 and the conditional R^2 is 0.88. Thus, on the whole the model's fit is acceptable.

However, the model has some collinearity issues, and one of the assumptions of regression models is precisely the lack of it: "When doing a multiple regression analysis, collinearity describes situations where one predictor can be predicted by other predictors. Collinearity frequently arises from highly correlated predictors, and it makes regression models harder to interpret" (Winter 2020: 112). Nevertheless, Levshina (2015: 272) observes that "logistic regression is quite robust with regard to some correlation between predictors". In my model the condition number with the intercept included is 38.7, which is higher than the level of 30, which indicates collinearity. The following predictors and interactions have Variance Inflation Factor values higher than 10, which is often considered a level suggesting problems with collinearity: CASE, DEFINITENESS, DIALECT, CASE-PERIOD, DEFINITENESS-PERIOD, PERSON-PERIOD, REFERENTIALITY-PERIOD and PERIOD-DIALECT. Taking into account the value corrected for the degrees of freedom, the highest value is that of DIALECT (3.2). It is unclear to me whether the collinearity is related to the intrinsic correlation between the linguistic variables, or because of a smaller amount of data for some combinations of variables related to the characteristics of the historical Basque corpus.

5.3 Effects of the predictors

Model's coefficients are listed in Table 3 and also represented visually in Figure 2, which orders the predictors by the value of coefficients predicted by the model. Items to the right of the line at 0 show an increased preference for GenM, and those to the left show a decreased preference for it. Those with the confidence interval crossing the line at 0 have no significant effect. We can easily see that, among the main effects, NUMBER and REFERENTIALITY have the strongest influence: plural nominals have the highest positive value (and thus the highest likelihood of GenM), and pronouns are placed at the other extreme with the lowest coefficient (and thus the lowest likelihood of GenM). The interpretation of coefficients is nevertheless not straightforward in a rather complex model as the one discussed here, and because of that I will use Figure 3 for the main variables and Figure 4 for the interactions. For each level of the predictors, they plot

probabilities of GenM (expressed in percentages) computed while keeping all other factors in the model constant.⁹ The lines represent 95% confidence intervals.

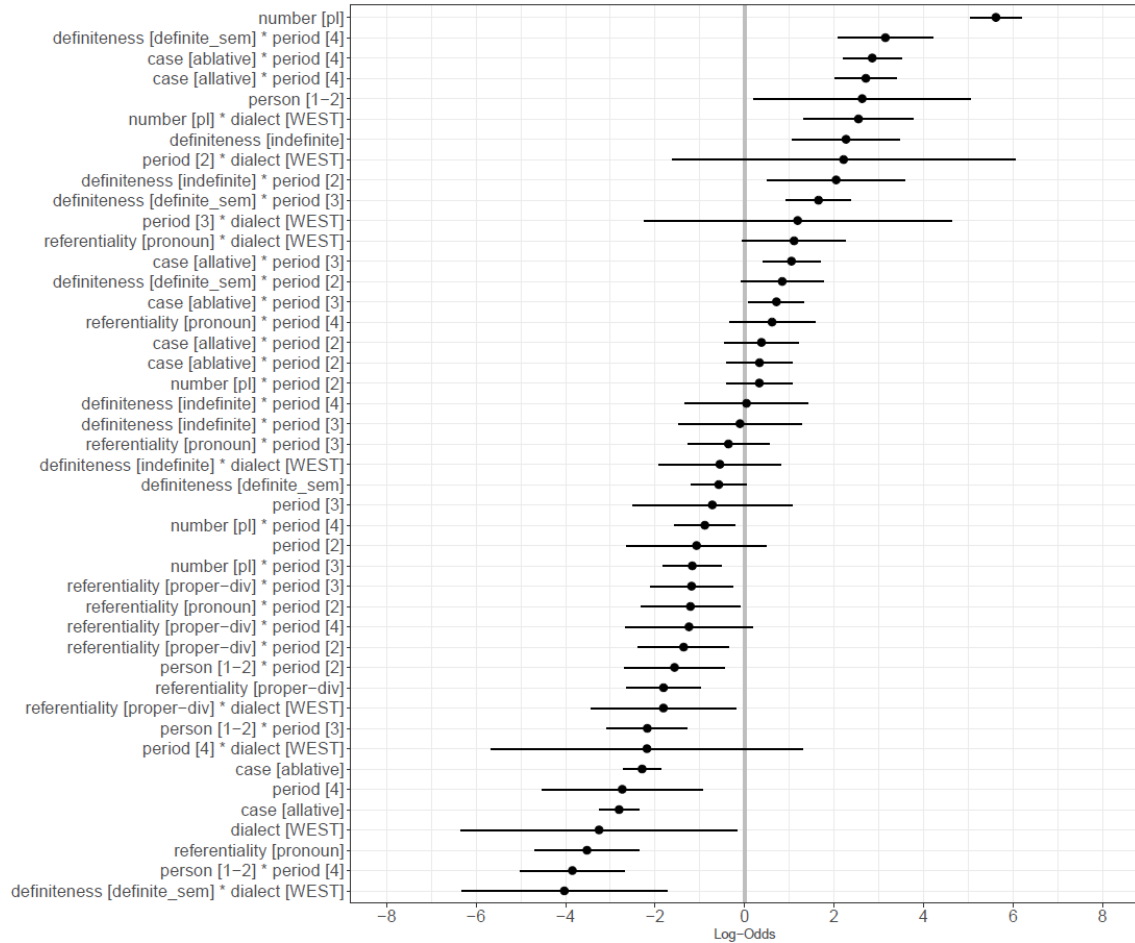


Figure 2. Log-odds of general spatial marking. Dots indicate the value predicted for each predictor, with horizontal lines showing 95% confidence intervals.

I will start with the main variables, leaving the interactions for the moment. The strongest effect is that of NUMBER, followed by the effects of REFERENTIALITY and DEFINITENESS. Plural nouns are much more often marked with GenM than singular: the probability is 0.62 for plural and 0.01 for singular. Pronouns show GenM less frequently than proper/divinity nouns, and much less so than common nouns: the probability of GenM is 0.29 for common nouns, 0.02 for proper/divinity nouns and 0.01 for pronouns. Indefinite nouns have stronger preference for GenM than definite nouns (0.25 vs 0.02). There is no significant difference between the two classes of definite nouns (those semantically definite but formally indefinite, and those semantically and

⁹ Calculated with the *ggeffects()* function in the *ggeffects* library (Lüdtke 2018).

formally definite). The effect of CASE is less pronounced: inessive is slightly more inclined to appear with GenM than ablative and allative. Finally, there is no significant difference between the first/second person and the third person.

Moving to the extralinguistic predictors, nominals are slightly more commonly marked with GenM in the earliest data (the first period) than in the later data, but the confidence intervals overlap, which does not allow to conclude that the differences are significant. Finally, eastern dialects show a slightly increased preference for GenM as compared to western dialects.

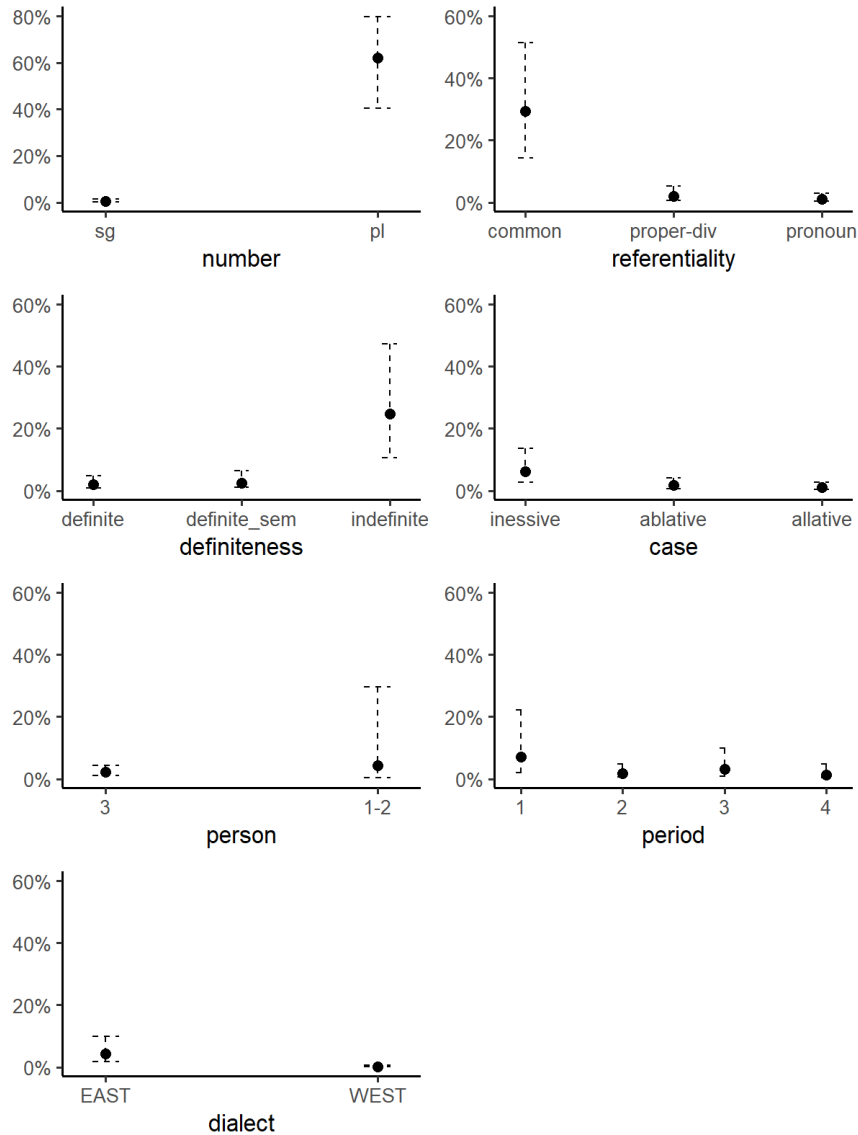


Figure 3. Effects of the predictors NUMBER, REFERENTIALITY, DEFINITENESS, CASE, PERSON, PERIOD, and DIALECT (the y axis shows probabilities of GenM expressed in percentages).

In order to understand properly the effect of each predictor, it is necessary to take into account the interactions between them (Figure 4). Let us start with interactions between the linguistic variables and the time period. In general, the probability of GenM decreases in time with all the variables. Plural nouns carry GenM more commonly in the sixteenth century than later: the probability of GenM is 0.86 in the first period, and 0.65, 0.52 and 0.38 for the subsequent ones. The probability of indefinite nouns showing GenM is around 0.5 in the first two periods, but falls to 0.15 in the third period

and 0.05 in the fourth. In REFERENTIALITY, however, the most important change happens between the third and the fourth period. Until then common nouns are much more likely to appear with GenM than proper/divinity nouns or pronouns, e.g. in the first period the probability of GenM is 0.45 for common nouns and 0.09 for proper-divinity nouns and 0.03 for pronouns. In the most recent data, however, the difference is much smaller. Proper and divinity nouns pattern similarly to pronouns (low preference for GenM), but in the earliest data they are slightly more prone to appear with GenM than pronouns. As for CASE, there is a significant difference between inessive and the other two values, allative and ablative, in the earliest data, with inessive being more prone to take GenM (0.25 vs 0.02-0.03). In the subsequent periods the difference is much smaller, and disappears in the most recent texts. In the variable PERSON, even though the confidence intervals are too big to draw a firm conclusion, the first/second person appears to be more inclined towards GenM than the third person in the earliest data.

Finally, I have to mention an exception to the downwards diachronic trend in the appearance of GenM visible in Figure 2: three predictors have quite high positive coefficients in the most recent data: semantically definite nouns and nouns in the ablative or allative. In Section 5.5 I will return to this issue, as well as the problem of the influence of the variable PERSON.

Moving to interactions with the variable DIALECT, there are slight but interesting differences in NUMBER, DEFINITENESS, and REFERENTIALITY. First, in both dialectal areas plural nouns have a strong tendency for GenM, but in the eastern texts it is even stronger. Second, in the eastern dialects indefinite nouns exhibit a greater preference of GenM than definite nouns, but in the western varieties definite and indefinite nouns behave in a similar fashion (both have low probability of GenM). As for REFERENTIALITY, in the western dialects there is no difference between the different types of nominals, but in the eastern varieties common nouns are much more prone to take GenM than other nominals. Finally, the interaction between DIALECT and PERIOD indicates that the eastern varieties favour GenM more than the western ones and that they diverge most in the sixteenth century.

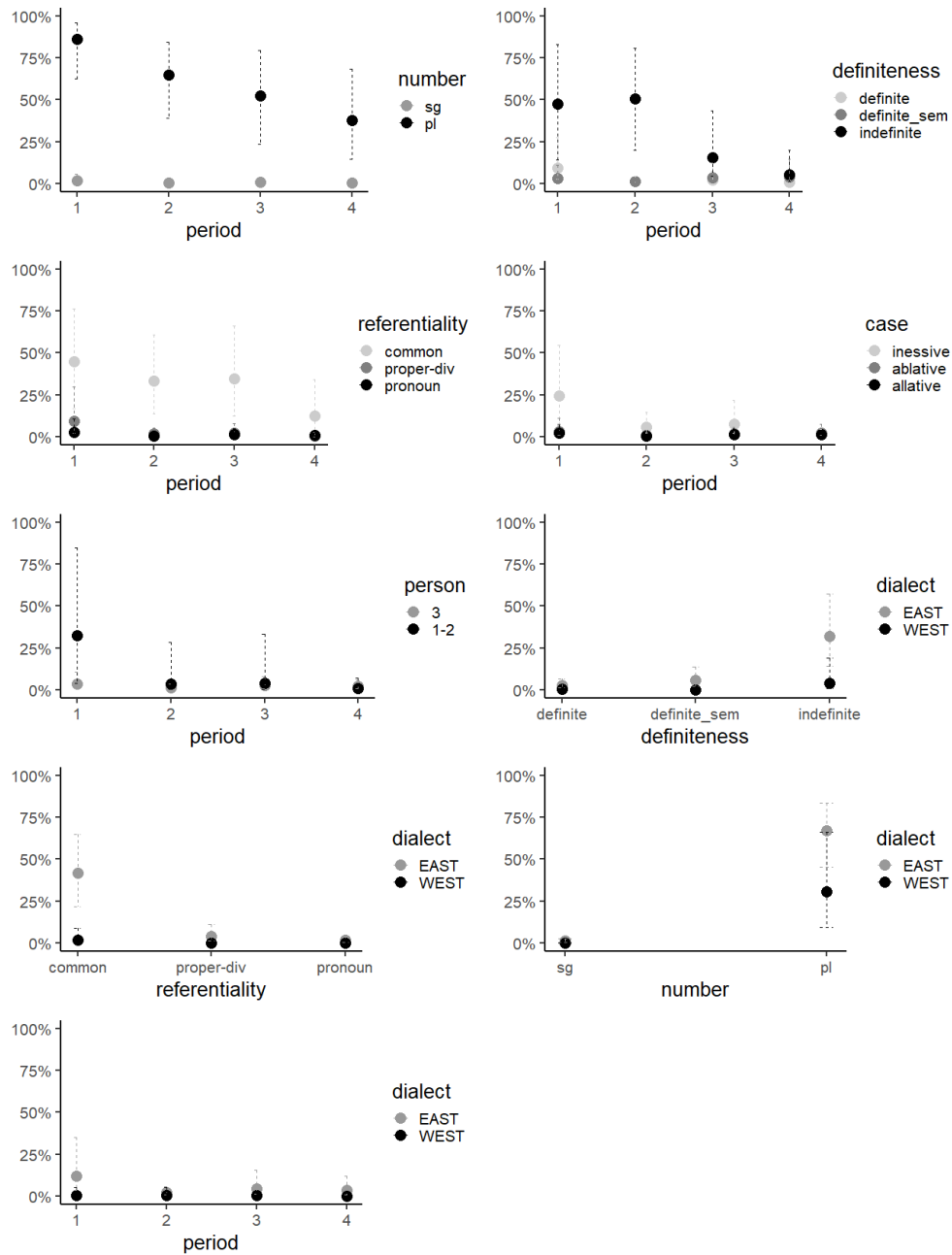


Figure 4. Interactions between predictors in the model (the y axis shows probabilities of GenM expressed in percentages).

5.4 Random effects

Finally, let us have a look at the two random effects in the model, TEXT and LEXEME. As indicated earlier, the random effects' contributions to the final model are highly significant. Table 4 above shows that the standard deviation is 1.5 for TEXT, and 2.2 for LEXEME, which means that there is greater variability between the different lexemes that there is between the authors.

Figure 5 plots the intercept adjustments for each text and the number of examples drawn from each text. Texts which appear above the line show higher tendency for GenM than those below the line. Additionally, the label's colour is different for eastern (black) and western dialects (grey). Some differences between the texts reflect dialectal variation. The texts marked with an ellipse have the biggest adjustments, which means that they favour GenM most. They were all written in the seventeenth and eighteenth centuries in the Souletin dialect. Most western texts appear rather close to the 0 line with only a few further down in the plot. Eastern texts, however, are quite scattered, and several show quite high negative adjustments (higher than for any western text), i.e. an increased preference for AnimM.

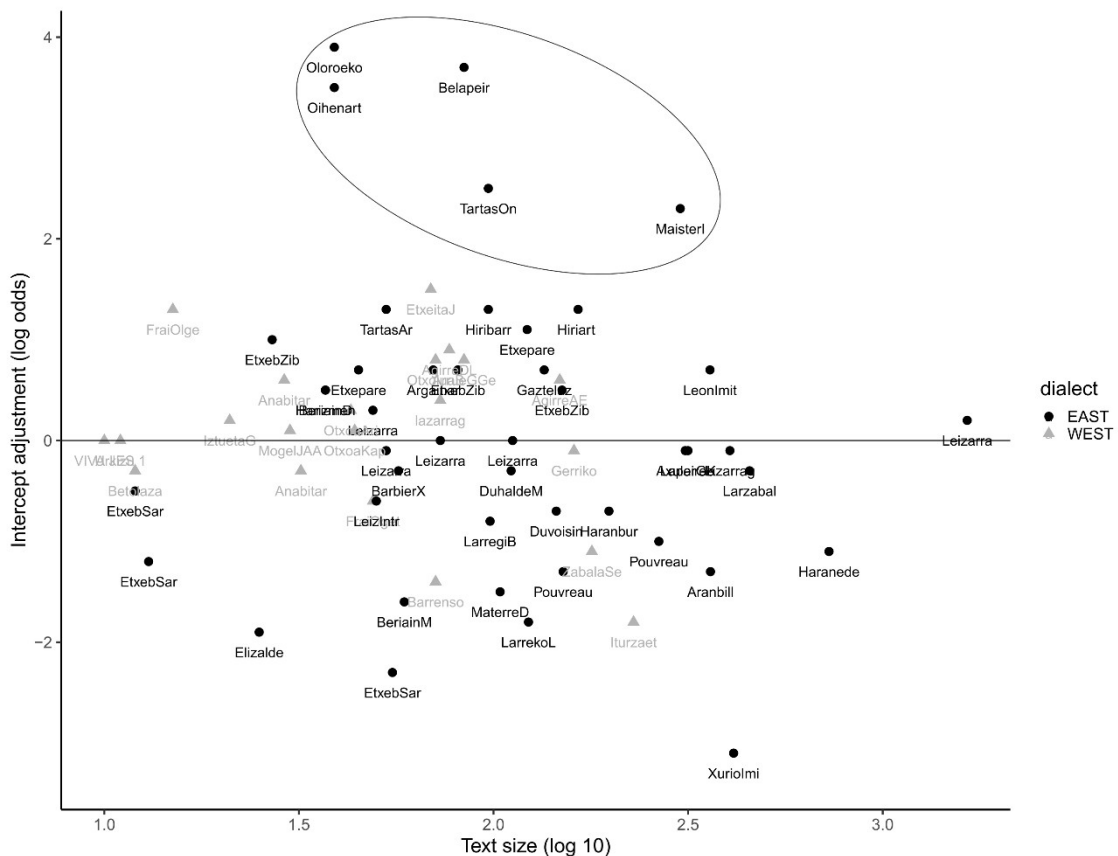


Figure 5. Intercept adjustments for the texts in the corpus.

As regards the effect of LEXEME, Figure 6 plots frequency and the intercept adjustments for lexemes with 20 or more tokens. Again, those with positive values exhibit a greater preference for GenM. Starting with those preferring GenM, the highest value has the noun *zaldi* ‘horse’. Other nouns with higher inclination for GenM are

related to religion, for example *santu* ‘saint’, *jaungoiko* ‘God’, the proper noun Jesus Christ, a few words usually referring to God (*jaun* ‘lord’ or *kreatzaile* ‘creator’), and supernatural beings such as *aingeru* ‘angel’ or *espiritu* ‘spirit’. As for nouns less related to religion, *haur* ‘child’ is more likely to carry GenM than *emazte* ‘woman’ or *gizon* ‘man’. In personal pronouns, *ni* ‘I’ and *zu* ‘you (sg)’ have small positive adjustments, *zuek* ‘you (pl)’, *elkar* ‘each other’ show small negative adjustments and *gu* ‘we’ and *bera* ‘s/he’ have quite high negative values.

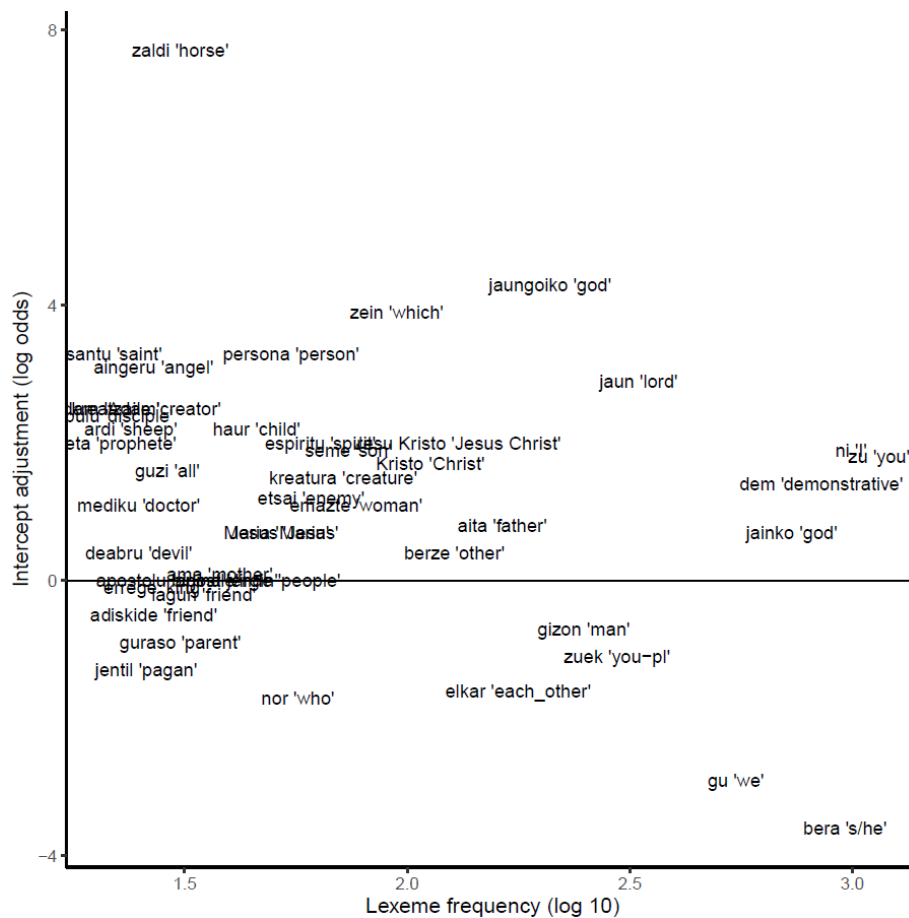


Figure 6. Intercept adjustments for lexemes with more than 20 tokens in the corpus.

5.5 Further analyses

Finally, I would like to return to certain issues mentioned in Section 5.3.

The first concerns the variables of PERSON and CASE and their influence on marking in the earliest data. In the sixteenth century, first- and second-person pronouns have higher preference for GenM, even though the effect does not appear to be statistically

significant in the model. In terms of raw data, 48% ($N = 582$) of first- and second-person pronouns show GenM in the first time period, compared to 9% ($N = 458$) of third-person pronouns. In the later data, the proportion of GenM is similar for both classes of pronouns. A high proportion of first/second-person pronouns marked with GenM is only found in the texts written in the Souletin dialect in the seventeenth and eighteenth century. These facts are related to the variable of CASE: in the sixteenth-century data two classes of nominals in the inessive case exhibit a stronger tendency towards GenM as compared to the ablative or allative: (a) first- and second-person pronouns, both in singular and plural, and (b) proper/divinity nouns (all examples refer to God: for example, *Iesus Kristean* ‘in Jesus Christ’, *gure Aita egiazkotan* ‘in our true Father’, *Iesus Jaunean* ‘in the Lord Jesus’).

Table 5 lists the proportions of GenM for pronouns in the earliest data separately for the inessive and other cases. Except for second-person plural, inessive has a much higher proportion of GenM than ablative or allative. Allative forms with GenM, such as *nitara* ‘to me’, are especially uncommon with pronouns: only nine examples were found for the first person and second person singular in the corpus analysed.

Table 5. GenM in first- and second-person pronouns in the sixteenth century.

Number	Person	Inessive		All/Abl	
		N	GenM (%)	N	GenM (%)
Singular	1	69	59,4	85	1,2
	2	41	63,4	76	1,3
Plural	1	82	84,1	54	14,8
	2	89	58,4	86	91,9

The second-person plural pronoun diverges from other pronouns in the oldest texts: in the allative and ablative GenM clearly predominates. The plural *zuek* ‘you (pl.)’ is quite recent: it developed when the pronoun formerly used in the second-person plural, *zu*, turned into singular pronoun (Martínez-Areta 2013: 302), and the plural of *zu* was created. The pronoun takes definite forms unlike other pronouns (cf. *zu-eta-n* ‘in you (pl)’ with the definite plural *-eta-* vs *zu-ta-n* ‘in you (sg)’ with the indefinite *-ta-*). It is thus morphologically quite like a plural of a common noun, and it appears that it behaves as such also as regards the marking of spatial relations. The outcome is that in the earliest data in the plural we have a contrast between the first and the second person: ‘to us’ or ‘from us’ are rather marked with AnimM, while ‘to you’ or ‘from you’ are

more commonly found with GenM. Example (37) illustrates this tendency: the two pronouns in the clause are marked differently, ‘from you’ with GenM and ‘to us’ with AnimM.

- (37) *ethorri zen-ean* *Timotheo zu-eta-rik* *gu-re-gan-a*
 come AUX.PST.3SG-DEF.INES Timothy you-PL-ABL we-GEN-ANIM-ALL
 ‘when Timothy came from you to us’ (LeizarragaTesta)

The second issue concerns an apparent increment in the probability of GenM in semantically definite noun phrases and ablative/allative case in the most recent data, reflected in positive coefficients predicted by the model. An examination of the data reveals that it is the reciprocal pronoun *elkar* ‘each other’ which causes this increase. Until the late nineteenth century, *elkar* is always marked with AnimM in my corpus (38), but in the most recent eastern texts around half of the examples with the pronoun show GenM (39). Moreover, all examples of *elkar* in the corpus except one are in the allative or ablative, which is why the model predicts a greater inclination towards GenM with allative and ablative forms in the fourth time period.

- (38) *behar dirate* *elkar-gan-a* *itzuli*
 must AUX.3PL each.other-ANIM-ALL return
 ‘they must return to each other’ (LeizarragaForma)
- (39) *elgarr-e-ta-rik* *urrun-tzen* *dira* *bi* *gizon-ak*
 each.other-EPENTH-INDEF-ABL move.away-IPFV AUX.3PL two man-DEF.PL
 ‘the two man move away from each other’ (BarbierXokoan)

6 Discussion and conclusions

In this paper I have analysed the marking of animate nouns in spatial cases in Basque. The first question was whether it is sensitive to various aspects of the Animacy Hierarchy: animacy itself, referentiality, number, person and definiteness.

The effect of animacy (human vs non-human animates) was not assessed directly, but, as shown in the discussion of the random effect of LEXEME, animals are more likely to be marked as inanimate than human nouns. In the whole corpus, out of 191 nouns referring to animals, only 21 exhibit AnimM. The most common of animal nouns, *zaldi* ‘horse’ is clearly conceptualised as inanimate. Some supranatural beings also show a

tendency to be marked with GenM more often than prototypical animate referents, especially in the earliest data.

I have found a strong effect of number and a slightly less prominent effect of referentiality and definiteness: (a) plural nouns favour GenM, contrary to singular nouns, (b) common nouns favour GenM, contrary to pronouns and proper/divinity nouns, and (c) indefinite nouns favour GenM, contrary to definite nouns. As for definiteness, there is no significant difference between nominals which are formally and semantically definite and those which are formally indefinite but semantically definite. This shows that the semantic feature of definiteness is important. The remaining factors, grammatical person and case, were only relevant for the marking of animate nouns in the early texts.

Thus, as far as referentiality, definiteness, and number are concerned, the results conform well with what the Animacy Hierarchy predicts. Among the different ways of conceptualising phenomena related to animacy, it appears that the most relevant one is “individuation” (Timberlake 1975, see also Hopper & Thompson 1980): “the degree to which the participant is characterized as a distinct entity or individual” (Timberlake 1975: 124). According to Timberlake (1975), proper, animate, concrete, singular, count, referential, and definite nouns can be characterised as highly individuated. Results presented here show that these nouns are the most likely to receive animate marking in spatial cases in Basque.

The marking of animate nouns varies across time and space. As regards diachronic changes, the proportion of nouns carrying AnimM was the lowest in the sixteenth century. I have also shown that the conditions under which general or animate marking is chosen change in time. The factors analysed can be ordered according to their diachronic persistence: number, referentiality, definiteness, case, and person (represented in Table 6). The effect of number is still pertinent in the most recent data. The effect of referentiality (common vs other nouns) is very strong until the late nineteenth century, but weakens later. Definiteness starts to lose relevance since the mid-eighteenth century. Person and case are only important for the earliest data. The expansion of animate marking, thus, has advanced between the sixteenth and the twentieth century, but the general marking is still possible on animate nouns in the modern language.

Table 6. The influence of difference factors on the marking of animate nouns.

Factor	Period			
	1400-1600	1600-1750	1750-1876	1876-1960
Number	+	+	+	+
Referentiality	+	+	+	(+)
Definiteness	+	+	(+)	(+)
Person	(+)	-	-	-
Case	(+)	-	-	-

As for the effect of person in the oldest data, the corpus study shows that in the sixteenth century more first- and second-person pronouns were marked with GenM than third-person pronouns. However, this only happens with inessive pronominal forms, and not those carrying allative or ablative, which did prefer AnimM (in general, AnimM extended earlier to allative/ablative forms). Thus, at least in the inessive, AnimM appears to have generalised first in the third person pronouns, and only later in the first and second person. This result goes against what Animacy Hierarchy predicts. A possible explanation could be that special animate marking is not so necessary on first/second-person pronouns, which are always animate. The distinction might be more relevant for the third-person pronouns, which can refer to animate and inanimate referents.

As regards dialectal variation, eastern Basque has stronger preference for GenM than western Basque. Additionally, some factors that are relevant for the marking in the eastern dialects do not operate in the western texts or are much weaker. On the one hand, the differences between (a) definite and indefinite nouns and (b) common nouns and pronouns exist in the eastern varieties but not in the western ones. On the other hand, the association between plural and GenM is stronger in the eastern varieties. It is possible that these differences reflect different chronologies: the western varieties generalised AnimM earlier than eastern ones, and only the last traces of the process can be seen in the western texts (the effect of number, which is particularly strong, remains, but it is weakened). This goes in line with previous research which argues that eastern dialects use GenM more extensively until later (as observed by Padilla-Moyano 2017: 763). Especially conservative is Souletin Basque, which is reflected in the adjustments

to the intercepts predicted by the model for Souletin texts. Nevertheless, some eastern texts show the opposite tendency, with a proportion of AnimM higher than western texts. At least some of those texts are seventeenth- to eighteenth-century Labourdin texts, which appear to be quite innovative. This issue would require more research.

Finally, I would like to bring forward a few more specific issues related to case. Similarly to Creissels and Mounole (2011), I have found that in the oldest texts inessive nominals were more prone to appear with GenM than ablative or allative nominals. This difference disappears in the subsequent periods. That the inessive is more common with GenM in the oldest texts might appear striking if we consider the compatibility of animate nouns to be used with spatial cases: they are not typical targets or sources of movements, but they are even less typical locations. I would like to propose that the difference between cases attested in the Basque data might nevertheless be related to the semantic (in)compatibility: the differential marking (i.e. morphologically heavier forms) is expected when the location, or target or source of movement is not prototypical. The allative and ablative can express physical movement towards or from around an animate being. The inessive, on the other hand, rarely encodes physical location with animate nouns in Basque: *nitan* ‘in me’ (with GenM) refers to conceptual, metaphorical space (as with the predicate such as ‘believe in someone’), and thus there is no real need for special marking. The exception is when someone’s residence is referred to with *baitan*. In fact, in Basque there are contexts where, even though spatial cases are employed, animate marking is impossible because the meaning is not related to spatial relations. Some examples include the subset construction (*gizonetarik batzuk* ‘some of the men’) or manner adverbs marked with the inessive, such as *haurretan* ‘in childhood’. AnimM eventually extended to the most typical uses of the inessive as well, probably because the spatial cases, in general, tend to behave in the same way in Basque (e.g. in the way plural or indefinite are formed).

Furthermore, the different behaviour of spatial cases might be related to the properties and diachrony of the two animacy markers, *-gan* and *baita-*: in the varieties which employ both, the former tends to be used with cases of movement, and the latter with location. The marker *baita-* appears to be an innovation of the eastern dialects, and its semantics is strongly related to location, as explained in Section 2: (1) usual residence, and (2) metaphorical, rather than physical location. Thus, a plausible

hypothesis, also suggested by Creissels and Mounole (2011), is that the oldest function of *baita-* was to talk about people's residence. As mentioned in Section 2, this change could have parallels in the grammaticalisation paths which the noun 'house/home' can undergo: 'house/home' > locative (Kuteva *et al.* 2019: 233, 235–236). I would like to propose that *baita-* entered the paradigm through the inessive, probably to refer to people's residence, and from there extended to other uses. The semantic change in Basque could have been a case of metonymy from house to a person living in it and further to the inside of a person (in a figurative sense), or a metaphorical mapping between the inside of a house and the inside of a person (i.e. "mind"). As a result, *baita-* has become a postposition referring especially to personal or intimate metaphorical space.

An issue for further research would be to analyse in more detail the differences between the two ways of marking animate nouns, *-gan* and *baita-*. A known factor is case: as already mentioned, *baita-* is preferred in the inessive, but both forms are possible with cases expressing movements, though with subtle differences in meaning (Lafitte 1991 [1944]: 168–170). The second question which needs more study is the alternation between genitive and absolutive as the base to which animacy markers are added: *ni-gan-a* 'to me' [I(ABS)-gan] or *ni-re-gan-a* 'to me' [I-GEN-ANIM-ALL]. The question is what factors influence the choice of absolutive or genitive. Referentiality appears to be relevant (but not, for example, number): in the data analysed here, the proportion of genitive marked common nouns appears to increase in time, especially in the eastern varieties (from just 1% in the first time period to 65% in the most recent sources). Pronouns show a higher proportion of genitives already in early sources.

The results presented here are comparable to those previously obtained in more qualitatively oriented approaches (Creissels & Mounole 2011), but the greater amount of data analysed allows for more fine-grained distinctions, and quantification of the effects that the different variables have. I believe that the methodology applied here, namely generalised linear mixed-effects models, which has not been previously used to analyse Basque data, has proven useful to analyse the diachrony of the language, despite the problems that the Basque historical corpus has, such as its relatively small size, limited choice of genres, or unequal geographic coverage.

Abbreviations

1	first person
2	second person
3	third person
A	ergative subject
ABL	ablative
ABS	absolutive
ALL	allative
ANIM	animacy marker
AOR	aorist
AUX	auxiliary verb
COMP	completive
DAT	dative
DEF	definite
EPENTH	epenthesis
ERG	ergative
FUT	future
GEN	genitive
IMP	imperative
INDEF	indefinite
INES	inessive
IPFV	imperfective
NEG	negation
O	transitive object
PART	partitive
PL	plural
PST	past
PTCP	participle
RAD	radical
REL	relative
SG	singular
SUB	subordinate

Appendix: Texts analysed in the study

Unless stated otherwise texts were taken from the *Euskal Klasikoen Corpusa* (Euskara Institutua 2013).

Abbreviation	Author and title	Area	Tokens
Period 1 (1400-1600)			
EtxepareBPrimitiae	Etxepare, <i>Linguae Vasconum Primitiae</i> (1545)	EAST	45
LeizarragaAbc	Leizarraga, <i>ABC edo Kristinoen instrukzioea</i> (1571)	EAST	53

LeizarragaForma	Leizarraga, <i>Othoitza eklesiastikoen forma</i> (1571)	EAST	73
LeizarragaKateximea	Leizarraga, <i>Kateximeia</i> (1571)	EAST	112
LeizarragaKonfesionea	Leizarraga, <i>Konfesionea</i> (1571)	EAST	49
LeizarragaTesta	Leizarraga, <i>Iesus Krist Gure Iaunaren Testamentu Berria</i> (1571)	EAST	1699
Lazarraga	<i>Lazarraga eskuizkribua</i> (texts A and B) (1567-1604) (Bilbao et al. 2020)	WEST	73
BetolazaDoktrina	Betolaza, <i>Doktrina Kristiana</i> (1596)	WEST	12
Period 2 (1600-1750)			
BeriainMeza	Beriain, <i>Tratazen da nola enzun bear den meza</i> (1621)	EAST	59
MaterreDotrina	Materre, <i>Dotrina kristiana</i> (1623)	EAST	104
BeriainDotrina	Beriain, <i>Dotrina kristioarena euskaras</i> (1626)	EAST	37
EtxebZibuManual	Etxeberrri Ziburukoa, <i>Manual debozionezkoa</i> (1627, 1669)	EAST	81
EtxebZibuNoelak	Etxeberrri Ziburukoa, <i>Noelak</i> (1630)	EAST	27
EtxebZibuElizara	Etxeberrri Ziburukoa, <i>Elizara erabiltzeko liburua</i> (1636)	EAST	150
HaranburuDebozino	Haranburu, <i>Debozino eskuarra</i> (1635)	EAST	198
AxularGero	Axular, <i>Gero</i> (1643)	EAST	311
VivaJesus	<i>Viva Jesus</i> (17th cent.) (Ulibarri 2010)	WEST	10
OtxoaKapanagaDotrin ea	Kapanaga, <i>Dotrinea</i> (1656)	WEST	44
OihenartNeurtitzak	Oihenart, <i>O.ten gaztaroa neurtitzetan</i> (1657)	EAST	39
HarizmendiOfizioa	Harizmendi, <i>Ama birjinaren ofizioa</i> (1658)	EAST	37
PouvreauFilotea	Pouvreau, <i>Filotea</i> (1664)	EAST	151
ArgainaratzDeboten	Argainaratz, <i>Deboten brebiarioa</i> (1665)	EAST	70
TartasOnsa	Tartas, <i>Onsa hilzeko bidia</i> (1666)	EAST	97
PouvreauImitazionea	Pouvreau, <i>Iesusen imitazionea</i> (1669)	EAST	266
TartasArima	Tartas, <i>Arima penitentearen okupazione debotak</i> (1672)	EAST	53
AranbillagaImitazio a	Aranbillaga, <i>Jesu Kristoren Imitazionea</i> (1684)	EAST	361
GazteluzarEgia	Gastelizar, <i>Egia katolikak</i> (1686)	EAST	135

BelapeireKatexima	Belapeire, <i>Katexima labürra</i> (1696)	EAST	84
Oloroeko kat.	<i>Oloroeko katixima</i> (1706) (Padilla-Moyano 2015)	EAST	39
OtxoaArinExpplikazioa	Otxoa Arin, <i>Doktrina kristianaren expplikazioa</i> (1713)	WEST	43
EtxebSaraGazteriari	Etxeberri Sarakoa, <i>Eskual-Herriko gazteriari</i> (c. 1718)	EAST	13
EtxebSaraHatsapenak	Etxeberri Sarakoa, <i>Eskuararen hatsapenak</i> (c. 1718)	EAST	55
EtxebSaraLauUrduri	Etxeberri Sarakoa, <i>Lau-Urduri gomendiozko karta, edo guthuna</i> (1718)	EAST	12
XurioImitazioea	Xurio, <i>Jesu-Kristoren imitazioea</i> (1720)	EAST	414
ElizaldeFApezendako	Elizalde, <i>Apezendako dotrina kistiana uskaraz</i> (1735)	EAST	25
Urkizu	Urkizu, <i>Liburu virgina santissimien errosario santuena</i> (Zuloaga 2020)	WEST	11
Period 3 (1750-1876)			
HaranederEbanjelioa	Haraneder, <i>Jesu Kristoren ebanjelio saildua</i> (1740)	EAST	728
MaisterImitazionia	Maister, <i>Jesü-Kristen imitazionia</i> (1757)	EAST	302
LarregiBerriko	Larregi, <i>Testamen berriko historia</i> (1777)	EAST	98
LizarragaZenbaitSand u	Lizarraga, <i>Zenbait sanduen biziak asteaz datozinak</i> (1793-1813)	EAST	405
MogelJAAbarka	Mogel, <i>Peru Abarka</i> (1802, 1881)	WEST	30
AgirreAErakusKonfesi o	Agirre, <i>Konfesioko eta komunioko sakramentuen gañean erakusaldiak</i> (1803, 1823)	WEST	148
Gerriko	Gerriko, <i>Kristau doktriña guztiaren espelikazioaren saiakera</i> (the first 26 chapters) (1805)	WEST	161
DuhaldeMeditazioneak	Duhalde, <i>Meditazioneak gei premiatsuenen ganean</i> (1809)	EAST	111
FraiOgetabost	Frai Bartolome, <i>Ogeta bost lezinoe ta sermoe bat</i> (1807)	WEST	49
FraiOlgeeta	Frai Bartolome, <i>Euskal Errijetako olgeeta ta dantzeen neurrizko gatz-ozpinduba</i> (1816)	WEST	15

ZabalaSermoiakI	Zabala, <i>Sermoiak-I</i> (1816-1833)	WEST	179
IztuetaGipuzkoa	Iztueta, <i>Gipuzkoako dantza gogoangarrien kondaira</i> (1824)	WEST	21
DuvoisinTelemake	Duvoisin, <i>Telemake</i> (c. 1833)	EAST	145
HiribarrenEgia	Hiribarren, <i>Eskaraz egia</i> (1858)	EAST	97
Period 4 (1876-1968)			
ArrueGGenobeba	Arrue, <i>Santa Genobebaren bizitza</i> (1885)	WEST	84
Hiriart Urruti	Hiriart Urruti, <i>Gontzetarik jalgiaraziak, Mintzaira, aurpegia, gizon</i> (1891-1914)	EAST	165
LapeireKredo	Lapeire, <i>Credo edo Sinhesten dut esplikatu</i> (1891)	EAST	316
AgirreDLorea	Agirre, <i>Auñemendiko lorea</i> (1898)	WEST	77
IturzaetaAzalduerak	Iturzaeta, <i>Ikasbide kristinaukorraren azalduera laburrak</i> (1899)	WEST	229
EtxeitaJaioterri	Etxeita, <i>Jaioterri maitia</i> (1910)	WEST	69
EtxepareJ	Etxepare, <i>Buruxkak</i> (1910)	EAST	122
LarrekoLekuko	Larreko, <i>Lekukotasuna</i> (1916-1936)	EAST	123
BarbierXokoan	Barbier, <i>Supazter xokoan</i> (1924)	EAST	57
LeonImitazionea	Leon, <i>Jesu-Kristoren imitazionea</i> (1929)	EAST	360
Larzabal	Larzabal, <i>Hitzaldi eta mintzaldi, Ipuin eta istorio, Roxali, Matalas, Ibañeta, Suedako neskatxa, Lana eri, Malentxo, alargun, Senperen gertatua, Paper mende, Hiru ziren, Bordaxuri, Etxahun, Mugari tiro, Antzerki laburrak</i> (1930-1970)	EAST	455
AnabitarteUsauri	Anabitarte, <i>Usauri</i> (1931)	WEST	32
AnabitarteDonostia	Anabitarte, <i>Donostia</i> (1932)	WEST	29
OtxoluaBertolda	Otxolua, <i>Bertolda eta Bertoldin</i> (1932)	WEST	71
BarrensorozUztaro	Barrensoroz, <i>Uztaro</i> (1937)	WEST	71

Funding information

The research for the paper was made possible by the grants from the Spanish Ministry of Science and Innovation (PGC2018-098995-B-I00 and PID2020-118445GB-I00). Support given by the research group on historical linguistics (IT1344-19) funded by the Basque Government is also acknowledged.

Acknowledgements

I would like to thank Iván Igartua, Paweł Krajewski, Manuel Padilla, Ekaitz Santazilia and the anonymous reviewers for their valuable comments and suggestions.

References

- Academie française. 2019. *Dictionnaire de l'Académie française (9e édition)*. <https://www.dictionnaire-academie.fr/>.
- Aristar, Anthony Rodrigues. 1997. Marking and hierarchy types and the grammaticalization of case-markers. *Studies in Language* 21:2.313–368. <https://doi.org/10.1075/sl.21.2.04ari>.
- Azkue, Resurrección María. 1923. *Morfología Vasca*. Bilbao: Editorial Vasca.
- Barteld, Fabian, Stefan Hartmann & Renata Szczepaniak. 2016. The usage and spread of sentence-internal capitalization in Early New High German: A multifactorial approach. *Folia Linguistica* 50:2.385–412. <https://doi.org/10.1515/flin-2016-0015>.
- Bartoń, Kamil. 2020. *MuMIn: Multi-Model Inference*. <https://CRAN.R-project.org/package=MuMIn>.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1.1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Bilbao, Gidor, Ricardo Gómez López, Joseba A. Lakarra, Julen Manterola, Céline Mounole & Blanca Urgell. 2020. *Lazarraga Eskuizkribuaren edizioa eta azterketa*. Bilbao: University of the Basque Country UPV/EHU.
- Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. Chicago: University Of Chicago Press.
- Creissels, Denis. 2009. Spatial cases. *The Oxford handbook of case* ed. by Andrej Malchukov & Andrew Spencer, 609–625. Oxford: Oxford University Press.
- Creissels, Denis & Céline Mounole. 2011. Animacy and spatial cases: Typological tendencies, and the case of Basque. *Case, animacy and semantic roles* ed. by Seppo Kittilä, Katja Västi & Jussi Ylikoski, 157–182. Amsterdam / Philadelphia: John Benjamins.
- Cristofaro, Sonia. 2013. The referential hierarchy: Reviewing the evidence in diachronic perspective. *Languages across boundaries: Studies in the memory of Anna Siewierska* ed. by Dik Bakker & Martin Haspelmath, 69–93. Berlin: Walter de Gruyter.

- Croft, William. 1993. *Typology and universals*. Cambridge: Cambridge University Press.
- Dahl, Östen & Kari Fraurud. 1996. Animacy in grammar and discourse. *Reference and referent accessibility* ed. by Thorstein Fretheim & Jeanette K. Gundel, 47–64. Amsterdam / Philadelphia: John Benjamins.
- De Smet, Isabeau & Freek Van de Velde. 2020. A corpus-based quantitative analysis of twelve centuries of preterite and past participle morphology in Dutch. *Language Variation and Change* 32:2.1–25. <https://doi.org/10.1017/S0954394520000101>.
- De Swart, Peter & Helen de Hoop. 2018. Shifting animacy. *Theoretical Linguistics* 44:1–2.1–23. <https://doi.org/10.1515/tl-2018-0001>.
- Euskaltzaindia. 1991. *Euskal gramatika. Lehen urratsak-I*. Bilbao: Euskaltzaindia.
- Euskaltzaindia. 2019. *Orotariko Euskal Hiztegia*. Bilbao: Euskaltzaindia.
- Euskara Institutua. 2013. *Euskal Klasikoen Corpora (EKC)*. University of the Basque Country UPV/EHU. <https://www.ehu.eus/ehg/kc/>.
- Fox, John & Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Thousand Oaks CA: Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html>.
- Geleyn, Tim. 2017. Syntactic variation and diachrony. The case of the Dutch dative alternation. *Corpus Linguistics and Linguistic Theory* 13:1.65–96. <https://doi.org/10.1515/cllt-2015-0062>.
- Gorrochategui, Joaquín, Iván Igartua & Joseba A. Lakarra, eds. 2018. *Historia del euskera*. Vitoria-Gasteiz: Eusko Jaurlaritzza.
- Grafmiller, Jason. 2019. JGmermod: custom functions for mixed-effects regression models. <https://rdr.io/github/jasongraf1/JGmermod/>.
- Gries, Stefan Thomas. 2013. *Statistics for linguistics with R: a practical introduction*. Berlin: Mouton de Gruyter.
- Gries, Stefan Th. 2015a. The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10:1.95–125. <https://doi.org/10.3366/cor.2015.0068>.
- Gries, Stefan Th. 2015b. Some current quantitative problems in corpus linguistics and a sketch of some solutions. *Language and Linguistics* 16:1.93–117. <https://doi.org/10.1177%2F1606822X14556606>.
- Gries, Stefan Th. & Martin Hilpert. 2010. Modeling diachronic change in the third person singular: a multifactorial, verb- and author-specific exploratory approach. *English Language and Linguistics* 14:3.293–320. <https://doi.org/10.1017/S1360674310000092>.

- Harrison, Annette & William J. Ashby. 2003. Remodelling the house: The grammaticalisation of Latin *casa* to French *chez*. *Forum for Modern Language Studies* 39:4.386–399. <https://doi.org/10.1093/fmls/39.4.386>.
- Haspelmath, Martin. 2019. Differential place marking and differential object marking. *STUF - Language Typology and Universals* 72:3.313–334. <https://doi.org/10.1515/stuf-2019-0013>.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56:2.251–299.
- Ibarretxe-Antuñano, Iraide. 2004. Polysemy in Basque locational cases. *Belgian Journal of Linguistics* 18:1.271–298. <https://doi.org/10.1075/bjl.18.14iba>.
- Igartua, Iván & Ekaitz Santazilia. 2018a. How animacy and natural gender constrain morphological complexity: evidence from diachrony. *Open Linguistics* 4:1.438–452. <https://doi.org/10.1515/opli-2018-0022>.
- Igartua, Iván & Ekaitz Santazilia. 2018b. Asimetrías gramaticales asociadas a la animacidad en la lengua vasca: una perspectiva tipológica. *Anuario del Seminario de Filología Vasca Julio de Urquijo* 52:1–2.381–395. <https://dx.doi.org/10.1387/asju.20209>.
- Kittilä, Seppo, Katja Västi & Jussi Ylikoski, eds. 2011. *Case, animacy, and semantic roles*. Amsterdam / Philadelphia: John Benjamins.
- Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316479704>.
- Lafitte, Pierre. 1991 [1944]. *Grammaire basque*. Donostia: Elkar.
- Lakarra, Joseba A. 1984. Bizkaiera zaharreko ablatiboaz. *Anuario del Seminario de Filología Vasca Julio de Urquijo* 18:1.161–193.
- Lakarra, Joseba A. 1997. Euskararen historia eta filologia: arazo zahar, bide berri. *Anuario del Seminario de Filología Vasca Julio de Urquijo* 31:2.447–535.
- Lakarra, Joseba A. 2005. Prolegómenos a la reconstrucción de segundo grado y al análisis del cambio tipológico en (proto) vasco. *Palaeohispanica* 5.407–470.
- Langacker, Ronald W. 1991. *Foundations of cognitive grammar. Vol. 2. Descriptive application*. Stanford: Stanford University Press.
- Le Nouveau Testament*. 1563. *Le Nouveau Testament, c'est à dire, la Nouvelle Alliance de Nostre Seigneur Jesus Christ*. Genève: Jean Bonne-Foy. <https://gallica.bnf.fr/ark:/12148/bpt6k314602j>.
- Levshina, Natalia. 2015. *How to do linguistics with R: data exploration and statistical analysis*. Amsterdam / Philadelphia: John Benjamins.

- Longobardi, Giuseppe. 2001. Formal syntax, diachronic minimalism, and etymology: The history of French *chez*. *Linguistic Inquiry* 32:2.275–302. <https://doi.org/10.1162/00243890152001771>.
- Lüdecke, Daniel. 2018. ggeffects: tidy data frames of marginal effects from regression models. *Journal of Open Source Software* 3:26.772. <https://doi.org/10.21105/joss.00772>.
- Lüdecke, Daniel. 2020. *sjPlot: data visualization for statistics in social science*. <https://CRAN.R-project.org/package=sjPlot>.
- Martínez-Areta, Mikel. 2013. Demonstrative and personal pronouns. *Basque and proto-Basque. Language-internal and typological approaches to linguistic reconstruction* ed. by Mikel Martínez-Areta, 283–321. Frankfurt am Main: Peter Lang.
- Mitxelena, Koldo. 2011 [1970]. Nombre y verbo en la etimología vasca. *Obras completas* ed. by Joseba A. Lakarra & Iñigo Ruiz Arzalluz, vol. VII, 235–267. Bilbao: University of the Basque Country UPV/EHU.
- Nakagawa, Shinichi, Paul C. D. Johnson & Holger Schielzeth. 2017. The coefficient of determination R² and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface* 14:134.20170213. <https://doi.org/10.1098/rsif.2017.0213>.
- Nakagawa, Shinichi & Holger Schielzeth. 2013. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4:2.133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>.
- Oyharçabal, Beñat, Irantzu Epelde & Jasone Salaberria. 2009. *Norantz proiektua*. <http://www.norantz.org/>.
- Padilla-Moyano, Manuel. 2015. Oloroeko katixima: edizioa eta azterketa. Unpublished manuscript.
- Padilla-Moyano, Manuel. 2017. Zuberoako euskararen azterketa diakronikoa: XVI-XIX. mendeak / Analyse diachronique du dialecte souletin : XVI-XIX siècles. Unpublished thesis, University of the Basque Country UPV/EHU & Université Bordeaux Montaigne. <http://hdl.handle.net/10810/26535>.
- R Core Team. 2020. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- de Rijk, Rudolf Pieter Gerardus. 2008. *Standard Basque: a progressive grammar*. Cambridge, Massachusetts: MIT Press.
- Santazilia, Ekaitz. 2013. Noun morphology. *Basque and proto-Basque. Language-internal and typological approaches to linguistic reconstruction* ed. by Mikel Martínez-Areta, 223–281. Frankfurt am Main: Peter Lang.

- Santazilia, Ekaitz. 2019. Animacy effects in inflectional morphology: a typological survey. Unpublished thesis, University of the Basque Country UPV/EHU. <https://addi.ehu.es/handle/10810/37384>.
- Santazilia, Ekaitz. 2020. The two faces of animacy. *Studies in Language* 44:4.812–830. <https://doi.org/10.1075/sl.19089.san>.
- Timberlake, Alan. 1975. Hierarchies in the genitive of negation. *The Slavic and East European Journal* 19:2.123–138. <https://doi.org/10.2307/306765>.
- Trask, Robert L. 1997. *The history of Basque*. London/New York: Routledge.
- Ulibarri, Koldo. 2010. “Viva Jesus” doctrina: edizioa eta azterketa. *Anuario del Seminario de Filología Vasca Julio de Urquijo* 44:2.41–154.
- Wickham, Hadley. 2016. *ggplot2: elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Winter, Bodo. 2020. *Statistics for linguists: an introduction using R*. New York: Routledge.
- Wolk, Christoph, Joan Bresnan, Anette Rosenbach & Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructural variation and change. *Diachronica* 30:3.382–419. <https://doi.org/10.1075/dia.30.3.04wol>.
- Yamamoto, Mutsumi. 1999. *Animacy and reference: a cognitive approach to corpus linguistics*. Amsterdam / Philadelphia: John Benjamins.
- Zuloaga, Eneko. 2020. Liburu virgina santissimien errosario santuena: edizioa eta hizkuntza-oharrak. Unpublished manuscript.
- Zuur, Alain F., Elena N. Ieno, Neil Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. New York, NY: Springer New York.