

José Luis Blas Arroyo*

Same people, different outcomes: the sociolinguistic profile of three language changes in the history of Spanish. A corpus-based approach

<https://doi.org/10.1515/jhsl-2021-0025>

Received November 12, 2021; accepted August 31, 2022

Abstract: Within the framework of Historical Sociolinguistics and using a corpus of ego-documents written by Spaniards from different social backgrounds, this study analyses the sociolinguistic profiles of three phenomena of variation and change that took place in two critical periods in the history of Spanish: the Golden Age and the Early Modern Spanish. The study focuses on three standard variants that would end up displacing several vernacular forms whose use was much more widespread in Golden Age Spanish: (a) the use of the complementiser *que* in doxastic predicates depending on the verb *creer* [‘believe, think’], to the detriment of the variant *creer* + \emptyset ; (b) the analogical pronoun *quienes* in relative clauses with an explicit human antecedent (‘estos son los niños a quienes me dirigí’ [‘these are the children I spoke to’]), as opposed to the traditional relative *quien*; (c) the diffusion of the demonstrative pronoun *allí* [‘there’] at the expense of *allá*. Despite the success of the standard variants in the eighteenth century, the three cases of variation show different sociolinguistic conditioning, which in turn is closely related to several parameters, such as the speed and robustness of the respective changes, the typology of the variables and the linguistic constraints at work in each case.

Keywords: corpus linguistics; egodocuments; historical sociolinguistics; language variation and change; old Spanish.

1 Introduction

The research conducted over the decades since the birth of sociolinguistics has allowed scholars to establish several generalisations about usual correlations between language variation and social structure. Thus, it is well known that the employ of vernacular variants usually appears more in the speech of the lower

*Corresponding author: José Luis Blas Arroyo, Universitat Jaume I, Castelló de la Plana, Spain, E-mail: blas@fil.uji.es

sociolects and more informal styles, unlike the standard forms, which become more frequent in the speech of the higher sociolects and more formal registers (Labov 1972, 2001; Preston 1991; Tagliamonte 2012). Similarly, in the description of language change in progress, several types have been identified, each with a different background in the social structure. This explains the specific metaphors (*from above* and *from below*) that are frequently employed to characterise them (Ash 2013; Labov 2001; Tagliamonte 2012).

Sociolinguistics has also noted that both the spread of variants and sociolectal constraints can fluctuate considerably from one language variety to another. For example, in Peninsular Spanish, a representative example of this is the pluralisation of presentational *haber* (*habían* cambios vs. the prescriptive *había* cambios ‘there were changes’). While the pluralised form (*habían* cambios) is widespread in several Catalan-speaking communities, such as Catalonia and the Valencian Community, its diffusion is still incipient in other Peninsular varieties (Blas-Arroyo 2011, 2018; Bouzouita and Pato 2019). Moreover, a linguistic variable may also present significant changes in terms of diachrony. Hence, a variant that once enjoyed sociolinguistic prestige in a particular speech community later gave way to another, with the corresponding change in the socio-stylistic conditioning. This is what happened, for example, in Spanish with the imperfect and pluperfect of the subjunctive ending, where the traditional prestige form for centuries was *-se* (*cantase* ‘sing’), a role that was transferred to *-ra* (*cantara* ‘sing’) as of the 19th century (Martínez 2001).

In the present study, we aim to delve deeper into this last line of research by performing a comparative analysis of the sociolinguistic profiles exhibited by three linguistic variables at two critical moments in the history of the language: the Golden Age¹ period and Early Modern Spanish (18th century). In addition to dealing with cases of variation that affect several grammatical categories of Spanish, the selection of these three variables is explained because, even though they have not traditionally been analysed as potential sociohistorical variables in the literature, some recent studies have verified the existence of significant extralinguistic correlations in different periods of the history of Spanish. However, although the transition from Golden Age Spanish to Early Modern Spanish involved the firm consolidation of one of the competing variants at the expense of others -relegated to much more restricted

¹ Anglophone Hispanists generally use the concept of *Early Modern Spanish* to refer to Classical Spanish, while Spanish philologists consider the beginnings of *Modern Spanish* to lie in the 18th century. To avoid confusion, the term Golden Age Spanish will be used instead of Classical Spanish. Without entering into a critical assessment of this term, which is also present in the periodisation of other European languages (Rutten 2016), Golden Age is a cultural label widely used since the end of the 18th century in Spain to refer to the extraordinary flourishing of Spanish arts and literature from the late 15th century to the last decades of the 17th century.

uses since then—their diffusion was significantly different in each case. Some of them displayed this privileged condition from the outset, while others only succeeded over time (sometimes, after a long time). Besides, the sociolinguistic outcomes were also heterogeneous, and several factors could have played a determining role, such as the rate at which the changes took place or their linguistic typology.

Based on a corpus consisting of texts close to the pole of communicative immediacy (Oesterreicher 2004), and, presumably closer to the vernacular of past times, the three variables have been the separate targets of several variationist studies carried out as part of a research project in historical sociolinguistics conducted by the author of these pages.² The main objective of this article is to compare the sociolinguistic profiles exhibited by these variables in the two historical periods mentioned above,³ to shed some light on the potential reasons for these different trajectories.

This article is structured as follows. In Section 2, we review the three linguistic variables that are the object of this study, focusing on their genesis and evolution and the most relevant linguistic aspects. Although the empirical analysis also considers these structural aspects, given the aims of this study, in Section 3, we limit the coding description to explain how several socio-stylistic factor groups have been established in all cases. The most important methodological aspects, such as the corpus used or the quantitative analysis carried out, will be the subject of attention in Section 4. The main results are summarised in the different subsections of 5, which will serve as a preamble for discussing the most relevant results arising from the comparative analysis (Section 6) and the synthesis of the main conclusions (Section 7).

2 Origin and evolution of the linguistic variables

2.1 *Creer (que) ‘to think/believe (that)’ + complement clause*

Example (1), included in a letter sent in 1564 by a Spaniard by the name of Martín Delgado to an acquaintance of his living in Mexico City, shows a linguistic variation that continued for a long time in Spanish. In the excerpt, it can be seen how the speaker makes use of the same expression using the complementiser *que*, firstly, and then immediately afterwards without it:

2 The partial data on which this comparative study is based have appeared in previous studies published by members of the research group led by the author of these pages at the Jaume I University (Castellón, Spain), namely: Blas Arroyo (2020), Blas Arroyo and González (2022), Blas Arroyo and Velando (2019). The co-authors of the latter two papers agree that the data will be used again in this article for comparative purposes.

3 For the implications of this regarding the specific periodisation of each linguistic variable, see Section 4 below.

- (1) Después que de esta tierra v.m.d. salió no he visto letra suya, ni aun *creo que* se ha acordado de mí [...] Aquí *creo* \emptyset nos ha engañado un Agustín López, hijo de un conquistador, vecino de Méjico, que se dice Martín López
 ‘Since you left this land, I have not seen any letter of yours, nor do I *think that* have remembered me [...] Here I *believe (that)* we have been deceived by a certain Agustín López, son of a *conquistador*, a resident of Mexico, who calls himself Martín López.’
 (*El hilo que une*, 1564)

The root of these syntactic expressions is to be found in the passage from Latin to Romance when constructions with a complementiser were used to mark subordination in clauses with a verb in a finite form, thus replacing the schemes typical of classical Latin, where this subordination was generally presented in the infinitive and with a subject in the accusative (AcI). However, even in the classical period, this syntactic scheme competed with other complements introduced by *quod* or *ut*, albeit initially in different contexts (Väänänen 1968). However, as time went by, the boundaries between these uses started to become hazy, and this led to a progressive replacement of the AcI structures with those introduced by a conjunction. Although in Romance, the AcI constructions were still to be maintained occasionally in some causative structures and with verbs of perception (Cano 1977), the completive clauses with language and understanding verbs went on to generalise the structure with the subordinating *que*.

Although this became the variant par excellence, at various times in Spanish history, it has coexisted with another: the elision of the complementiser. Examples of this elision appear as early as the *Cantar de Mio Çid* at the beginning of the 13th century (“Digades al conde \emptyset non lo tenga a mal” ‘Tell the count \emptyset not to take it badly’). According to Folgar (1997: 350), in literary texts: “the conjunction can be elided on certain occasions (especially if the governing verb is placed immediately after the governed one and if there is no coreferentiality in the subjects of these two verbs), without this implying any modification in the syntactic structure” (our translation). The omission of the complementiser, however, could also be due to the analogy with other complementary structures with subjunctive verbs governed by volitional verbs (“*Suplicando* a sus Alteza \emptyset mande rreuocar la merçed” ‘Begging his Highness \emptyset to revoke the grace’; apud Serradilla 1997: 209). In these cases, the post-positioning of the subordinate clause and the very presence of the subjunctive could be sufficient to ensure the relationship of syntactic dependency. Hence, as an added mark of subordination, the conjunction could be interpreted as redundant (Raible 1983: 279).

Nevertheless, elision has also been characterised as a stylistic resource to prevent the accumulation of complementisers, mainly when the sentence consists of

two or more coordinated clauses (García Cornejo 2006: 234–235). Additionally, it appears to have been consolidated as a significant syntactic resource in the works of some classical authors (we find testimonies in Cervantes, Calderón, Quevedo or Tirso), but especially in Saint Teresa de Jesús. In this respect, Folgar (1997: 380) defended that “elision must be interpreted as a stylistic resource that implies a deviation deliberately sought by the writer” (our translation), and in no way as an imitation of common speech, since the omission is infrequent in the writings of other authors such as Lope de Rueda, Diego Hurtado de Mendoza or Fray Luis de León (in the same line, see Keniston 1937: 676; Pountain 2016: 18–19). Octavio de Toledo y Huerta (2011: 264 ff.) recalled that the Saint Teresa’s preference for elisions was also advocated by other authors such as the humanist Juan de Valdés, who was also for a simpler syntactic scheme.

However, a review of several grammars of the time suggests that the most prestigious form in the written language involved the use of *que*. This perception is reinforced in some 17th-century works, such as the *Gramática* by Correas (1626/2001), who explicitly defended the variant with the subordinate complementiser and described its elision as a “novelty” or “slip”. Similarly, Jiménez Patón (1614/2001) and Villar (1651/2001) insisted that certain verbs must be accompanied by phrases introduced by *que*.

2.2 The alternation of *quienes/quien* ‘who’ in relative subordinate clauses with a plural antecedent

For the construction of relative subordinates with plural animate antecedents, Spanish has different resources, one of which involves using the pronoun *quien*.⁴ However, until the almost definitive generalisation of the plural analogical form (*quienes*), this variant coexisted in history with the more traditional and invariable pronoun, *quien*, which served to express both singular and plural referents. Examples (2) and (3) show this alternation in letters written by Spanish emigrants in America during the last third of the 16th century:

- (2) Y en dandosela, se puso delante, sin ver el dicho Fonseca quién le avia dado, por aver sido por detras y estar él ocupado con los demás, a *quienes* dezia que se tuviesen ...

4 Another resource is the combination of the article and the pronoun *que*: “... y en todo muestra a *los que* te trataren la virtud y cordura *ques* razon” [and in everything he shows those who have had dealings with you the virtue and good sense that is due] (*Cartas de particulares en Indias del siglo XVI*, 1590).

‘And on giving it to them, he stood before them, not seeing the aforementioned Fonseca *who* had given it to him because he was behind him and was busy with the others, whom he said they had ...’
(*Documentos lingüísticos de la Nueva España*, 1576).

- (3) ... le da espuelas para q me le inbie con brebedad pues para ellos sera lo mejor pues no tengo hijos a *quien* lo dejar
‘... he urges him to send it to me soon because it will be better for them as I have no children (*who*) to leave it to.’
(*Cartas de particulares en Indias del siglo XVI*, 1576).

According to Corominas and Pascual (1980–1991; s.v. *què*), *quien* originates in the accusative *quēm* of the Latin interrogative pronoun (*quis, quae, quid*) and during the Middle Ages was invariable for the singular and the plural. Things started to change as of the 16th century, when the form *quien* developed a specific variant for the plural, *quienes* (Lapesa 1980: 397; Menéndez Pidal 1904/1941: §101). Nevertheless, the opinions are far from unanimous about what the timeline of this change might have been. For example, López Ruano (2011: 82) concluded that the use of *quienes* “did not spread until well into the 17th century”. Girón Alconchel (2004: 871) dated this innovative variant back to the end of the 16th century, although he noted that the change was slow. Iglesias (1996: 513) insisted on this idea and recalled that “the first testimonies of analogical plurals appeared in the first half of the 16th century, but they were still met with great resistance throughout the 17th century”. However, Lope Blanch (1997: 290–291) suggested that the plural *quienes* probably began to be used in the spoken language as of the first half of the 16th century; it later moved into literary language and was not to become generalised until the last decades of the next century, which accounts for its late discovery in written texts.

There are few references to these relative pronouns in the grammars and dictionaries of the 16th century, although, with some exceptions, the preference for the invariable use of *quien* prevails, either explicitly or in the examples provided in them. Things changed, however, in the 17th century, when *quienes* was already considered for plural antecedents. Even so, still in this period, prominent grammarians (Correas, Jiménez Patón, Doergangk, etc.) regarded the invariable use of *quien* as usual, and some of them even qualified *quienes* as “not very elegant” (Salazar 1627: 397) or “unnecessary” (Villar 1651: 8–9) (*apud* Lope Blanch 1997: 285–288). From the 18th century onwards, however, the tendency to include the plural form *quienes* became so intense that people barely noticed the old use of *quien* in this context (Martínez Gómez-Gayoso 1769: 103–104; Sobrino 1738: 45).

Nevertheless, even in this period, some authors still accepted the invariable etymological form (Benito de San Pedro 1769).

2.3 The variation *allí/allá* ‘there’

Examples (4) and (5) show the alternating use of *allí* and *allá* in the same semantic setting. Note how in both cases, the adverbial form is associated with a precise spatial reference, the city of Seville, which two emigrants in the Indies refer to in their letters to their relatives in Spain:

- (4) ... se vien[e?] fasta sevilla en vn carro o dos y *alli* se aliñan de todo p[ar]a pasar adelante
 ‘... they come to Seville in a carriage or two and *there* they pick up supplies with which to be able to continue on their way’
 (*Cartas de particulares en Indias del siglo XVI*, 1587)
- (5) ... yo ubiera umbiado tres u quatro barras de plata que valen alla mil y quinientos ducados. aca nos dan tan malas nuevas que *alla* en sevilla la toman i toda para el rey
 ‘... I would have sent three or four ingots of silver that are worth one thousand five hundred ducats there. Here the bad news they give us is that, *there*, in Seville, they take it all for the king’
 (*Cartas de particulares en Indias del siglo XVI*, 1577)

The expression **ad illic* has been proposed as the origin of the adverb *allí* (Corominas 1980–1991: s.v. *allí*; Sánchez Lancis 1990). Since the first written examples (back in the early 12th century), *allí* appeared in texts with different locative values such as direction, location, origin and trajectory. Later, from the 15th century onwards, the first temporal values also appeared as a metaphorical translation of the spatial coordinates (Company and Espinosa 2014: 161). On the other hand, *allá* would have derived from **ad illac*, an expression in which the second element already indicated transition in Latin (‘por allá’ [over there]). It appeared in the *Cantar de Mio Çid* and, in the opinion of some authors, would initially have a different meaning to *allí*. For example, according to Corominas (1980–1991: s.v. *allá*) “[*allá*] has always been distinguished from *allí* because it refers to a broader [...] or more vaguely located place”. In passing, this would explain why it accepts increases such as quantification (más allá ‘beyond’) to a greater extent than *allí* (? más allí) (Company and Espinosa 2014: 129–246). Sedano (1996) also suggested that the primitive meaning of adverbs ending in *-á* (*lugar por donde* ‘place through which’) may have extended overtime to the coding of broad, ill-defined spaces. In contrast, *allí* (like *aquí*) would delimit

spaces and times in a more precise way, which would allow them to be interpreted preferably as specific points and not as generic areas (in the same sense, see RAE/ASALE (2009). Other authors, however, have drawn attention to the fact that these differences are more likely to be due to the speakers' subjective perception than to truly objective distinctions (Pérez Saldanya and Rigau 2011; Terrado 1990).

Furthermore, it has also been proposed that *allá* denotes dynamic circumstances, unlike *allí*, which is more closely linked to static values (Bello 1847/1972; RAE/ASALE 2009). The Latin case system allowed for the distinction between the place in which and directionality, which could explain why such duality has been maintained in Spanish. Likewise, the adverbial opposition has also been related to other linguistic constraints such as contrast (Terrado 1990)⁵ or idiomaticity (Company and Espinosa 2014; Sedano 1996).⁶

The fact that both forms are possible in most contexts explains the publication of several dialectological and sociolinguistic studies on this variation in contemporary Spanish-speaking communities (Carbonero 1979; Gallucci 2007; Sedano 1996, 2000; etc.). Things change, however, when it comes to diachrony. In fact, we do not know of any studies that have exhaustively examined the conditioning of this variation throughout the Spanish language history.

3 Socio-stylistic coding

All the empirical analyses on which the current study is based address the potential influence of linguistic, stylistic and social factors, considered together in a mixed-effects logistic regression analysis (for more details, see Section 4 below). However, considering that our primary interest is focused on the socio-stylistic profile shown by the linguistic variables, in this section, we will concentrate on how these factor groups have been coded.⁷

To define the stylistic factor group (*tenor* in the coding), we combined the domain in which communication occurs and the type of relationship established between the interlocutors, particularly in epistolary correspondence (see Section 4). Thus, in the *close* sphere, we place all those occurrences of the linguistic variables

5 According to this hypothesis, the forms ending in *-í* are used to take account of absolute circumstances, which do not need to be set against others – just the opposite to the adverbs ending in *-á*, which are used in contrastive settings.

6 It has been said that *allá* appears more frequently in the formation of phraseological units (encontramos eso *aquí* y *allá* ‘we find that here and there’), which could result from the above-mentioned greater semantic imprecision.

7 Moreover, unlike the linguistic factor groups, specific to each variable, such social and stylistic factors are common to all three phenomena.

taken from letters in which close family and personal relationships are observed, such as those between spouses, parents and children, siblings, close friends, lovers, etc. Similarly, a significant part of the autobiographical texts is included in this section. In the absence of a direct interlocutor, in those diaries and memoirs, the writer also instils a high dose of subjectivity in narrating the vicissitudes of his or her life. Conversely, we consider the epistolary communication between people separated on the axes of power and solidarity as being representative of a more *distant* character, whether they are from subordinates to superiors (or vice versa) or between distant relatives, which is well reflected, for instance, in the use of considerably more formal forms of address, as well as in periphrastic sequences such as greetings and leave-takings. Consequently, we test the hypothesis that as the variants analysed became the prestige forms in their respective variables, they were used increasingly more frequently in distant contexts, whereas close contexts hindered their expansion.

At the social level, we consider first the social *status* or *rank*, which are concepts that, to avoid unfortunate anachronisms, we prefer to that of *social class*. In the three cases investigated here, we initially established a tripartite classification with representatives located at three stages, even though the affiliation criteria may vary from one period to another. In all periods, what we have called the social elites included representatives of the nobility and high clergy and some cultural elites such as writers, intellectuals, humanists, scientists, grammarians, etc. However, in the 18th century, some of the commercial and financial bourgeoisie joined this group, as individuals whose rise in society placed them in a privileged position (Fernández Díaz 2004). The *middle* groups were made up of several intermediate clergies (friars, priests, chaplains, deans, etc.), civil professionals (graduates, lawyers, doctors, teachers, traders, shopkeepers, etc.), as well as middle-ranking officials in different branches of state or municipal administration (mayors, army officers, etc.). Lastly, the *low* stratum comprised individuals in manual trades (turners, miners, tanners, carpenters, shoemakers, tailors, barbers, shop assistants, servants, etc.) and domestic work (mainly in the case of women). However, given the proximity of the results obtained in the middle and low groups, which were in all cases a significant distance away from those found among the elites, for the regression analysis, we compared these latter with the rest of society. By doing so, we intended to facilitate the statistical analysis and one of the main objectives of the research, namely, to check how these elites behaved as regards the three variation phenomena studied and which (we hypothesise) should always be aligned with the more prestigious variants.

In contrast to the relative balance of the samples obtained in this social stratification, the imbalance between men and women prevails in the corpus. Only 10–15% of the writers in the Golden Age corpus were women, a rate that decreases further

still among 18th-century texts. This fact undoubtedly reflects a strongly patriarchal society in which women's access to education and writing was very restricted until recent times (Castillo 2006). Unfortunately, this is a significant limitation, which may adversely affect the results. Still, we tested the hypothesis according to which women championed the use of prestige variants to a greater extent than men, which has been a recurring outcome in many contemporary western speech communities (Tagliamonte 2012: 61–62) but perhaps also in earlier periods of history according to some recent findings in historical sociolinguistics (Nevalainen and Raumolin-Brunberg 2017 [2003], Nevalainen 2018).

Lastly, we have also considered the *context* in which the texts were written. In this regard, we distinguished between documents written by Spaniards in *Spain* and those written by Spanish emigrants in different territories of *America*, which had been conquered by Spain as of the late 15th century and where individuals from very diverse dialectal backgrounds came together, with potential consequences on linguistic variation and change (Penny 2000). So far from their places of birth and in intense contact with migrants arriving from many different places, it could be assumed that those who wrote from America would show patterns of variation significantly different from those of their fellow citizens who had remained in Spain.

4 Corpus and methodology

As said above, this work is part of a research project in historical sociolinguistics to study language variation and change at different points in the history of Spanish (see note 2). In this case, such moments coincide with two significant milestones in the periodisation of the language, namely, the Golden Age (16 and 17th centuries) and Early Modern Spanish (18th century), respectively. Table 1 shows some dimensions of the corpus of immediacy communicative texts written by Spanish authors or by individuals who, although born elsewhere, lived most of their lifetime in Spain. The volume of words available in each century is relatively balanced, which is not the case for the number of writers. The two best represented centuries are

Table 1: Some dimensions of the corpus by century.

	Words	Writers
16th	2.407.913	1.129
17th	2.205.103	685
18th	2.334.749	1.737
Total	6.947.765	3.551

the sixteenth and the eighteenth, largely thanks to the meritorious task of several philologists and social historians in editing these texts in recent years. In contrast, the seventeenth century shows the worst figures regarding the number of writers considered. However, this circumstance is offset by the greater length of its texts.

Nevertheless, since the data used in this research come from previous variationist studies (see note 2), with specific targets and interests in each case, the exact periodisation of each variable differs. These are the precise stages analysed in each case:

- (a) Use of *creer que* (vs. *creer* Ø) + complement clause: 1st period (16th, 17th centuries); 2nd period (18th century)
- (b) Extension of *quienes* (vs. *quien*) in relative clauses with a plural antecedent: 4 periods (1500–1580; 1581–1660; 1661–1740 and 1741–1820)
- (c) Diffusion of the demonstrative adverb *allí* (vs. *allá*): 1st period (16th century); 2nd period (18th century)

One of the main methodological innovations introduced in this historical sociolinguistics research project is the use of texts that, although written in nature, are conceptually closer to the pole of communicative immediacy (Oesterreicher 2004). Approximately 90% of these documents are private correspondence written by Spanish people. There is also a tiny sample of autobiographical works, such as diaries and memoirs – genres that were not very frequent among Spaniards until more recent times, but of which we also have some representative samples from between the 16th and 18th centuries. Taken from different sources (public and private archives, private collections, etc.), all the texts included in the corpus come from editions made from original documents, not later copies that could alter the language previously used. Still, in the final selection of texts, we had taken special care to employ only those in which the editors had explicitly stated to respect the original in transcribing the manuscripts, with no modernisations except for some slight changes, if any, to modernise accentuation and punctuation in order to facilitate reading.⁸ Most of the texts are autographic, although some letters, especially those sent by the lower ranks in the 16th century – a period when illiteracy rates were very high among these people (Castillo 2006) – were probably dictated to scribes who, in exchange for payment, set them down on paper. Although dictated letters make any approach to phonic and graphic variation extremely difficult, this need not necessarily be the case in the study of grammatical and discursive variation (see Bergs 2005 for an overview).

⁸ The reader can find the complete list of these works, which are the same in the analysis of the three variables, in the following link: https://www.researchgate.net/publication/368336094_PRIMARY_SOURCES#fullTextFileContent.

As has been stressed in recent times, these texts are conceptually closer to oral speech than other more formal discursive traditions (Oesterreicher 2004), making them particularly suitable for studying the vernacular in earlier times when there were no oral testimonies available. Among the different “ideal” communicative contexts favouring the production of spoken–written testimonies outlined by Oesterreicher (2004), we have given preference to texts that this author characterised as having “written competence with an oral component”. The latter is particularly true with epistolary correspondence, especially when (as in this case) we are dealing with private letters whose writers are distributed across the whole society (Elspaß 2012; Schneider 2013). The periods analysed also represent relevant milestones for the genre, as they match decisive moments in America’s colonisation by Spaniards when letters were continuously being sent from one side of the Atlantic to the other (Fernández Alcaide 2009; Stangl 2012). Moreover, the corpus also offers various styles, as the topics and the relations of power and solidarity between the interlocutors are diverse (see Section 3 above). Besides, private letters provide valuable testimonies for sociolinguistic analysis as they reveal biographical details for measuring specific extralinguistic parameters (Okulska 2010; Raumolin-Brunberg 2005). Furthermore, another reason that justifies a particular association of private correspondence with the vernacular usages of past times is that, at least initially, the letters were not written with the intention of them ever being published (Palander-Collin and Nevala 2005).

The variable that is found most frequently in the texts is the alternation between the demonstrative adverbs *alli/allá*. The very nature of the letters can explain this fact since the senders often mention referents within the space-time reference framework of those participating in the communicative act. It is followed by the variation among doxastic predicates (*creer Ø/que*), which is also frequent in letters and autobiographical texts. Conversely, the use of *quien* or *quienes* in relative clauses with a plural antecedent occurs far less frequently in discourse.

The use of a non-tagged corpus makes it necessary to employ a concordancer (Wordsmith 6.0) to retrieve the instances of the linguistic variables. Following the principle of accountability (Labov 1972), all tokens of the reference variants are located, besides those of the alternative forms that could appear in their place. Each token is then coded according to several linguistic, stylistic and social factor groups, which allows us to assess different hypotheses about the conditioning of variation. Given the specific interests in this study (see Section 1), we will focus mainly on stylistic and social issues, with occasional references to linguistic conditioning when the occasion requires it.

For the quantitative analysis, we used *Rbrul* (Johnson 2009). This computer program performs a mixed-effects logistic regression statistical model in a relatively intuitive and straightforward way. It can be used to discover the relevance of the

independent variables considered and their potential interactions.⁹ Rbrul also evaluates the consistency of the results against the background of the idiolectal variation, strengthening the statistical significance of those factor groups whose effect is sufficiently robust to be able to resist the potential bias of specific individual preferences.

Together with the absolute frequencies and percentages, which allow for an initial exploration of the results, the multivariate analysis provides different elements of proof that can be used to interpret the variation in each period (Tagliamonte 2012: 122ss.):

- (a) the list of independent variables selected and not selected by the regression model
- (b) the robustness of these variables
- (c) the significance of the different factors within each factor group, offered in the form of statistical weights
- (d) the explanatory hierarchy, or direction of effect, from the most favourable to the most unfavourable constraints

Moreover, the comparative method adopted in this study involves investigating the connection between the patterns of variation presented by similar samples, as occurs in this case with parallel corpora of immediacy texts in different moments in Spanish history. The idea underlying this approach to analysing language change is that its characterisation can be extracted from the contrastive examination of the quantitative results reported in each period (Poplack 2011; Tagliamonte 2012).

5 Results

In this section, we offer the results obtained in the empirical examination of each variable. As already mentioned, our interest is centred on the socio-stylistic conditioning, given our main objectives in the study. However, this does not prevent us from summarising some of the most relevant structural aspects involved in each case, as some may also have implications for specific profiles of language change, as we will see later (see Section 6).

⁹ In these analyses, all linguistic and extra-linguistic predictors were run together in Rbrul, including the identity of the writers, which is analysed as a random factor. On the other hand, although the focus of attention in this article is how the same socio-stylistic factor groups condition the three linguistic variables examined, in each section, we also review -albeit more succinctly for the sake of brevity- the main linguistic constraints, which, not surprisingly, are largely specific to each variation.

5.1 Creer que/Ø + complement clause

Table 2 shows the distribution of both variants in the two periods considered (see Section 4). As can be observed, entry into modern Spanish (18th century) led to a significant increase in the use of the *que* variant (70%), compared to the averages obtained in Golden Age Spanish (57%).

However, Figure 1 shows that the progression of this variant throughout the 18th century was very gradual and sustained over time, as opposed to the more abrupt nature of changes in the classical period. Indeed, the figure shows how, in Golden Age Spanish, *creer que* even saw its primacy over the elided form jeopardised, especially during the second half of the 16th century, when the latter peaked.

In the Golden Age period and the 18th century, the variation was constrained linguistically by the degree of adjacency between the verbs and the person of *creer*. In the first case, the data show a more significant presence of the complementiser the

Table 2: Distribution of the variable in two periods.

	Golden age		18th century	
	<i>N</i>	%	<i>N</i>	%
Creer Ø	340	43	258	30
Creer que	450	57	599	70
Σ	790		857	

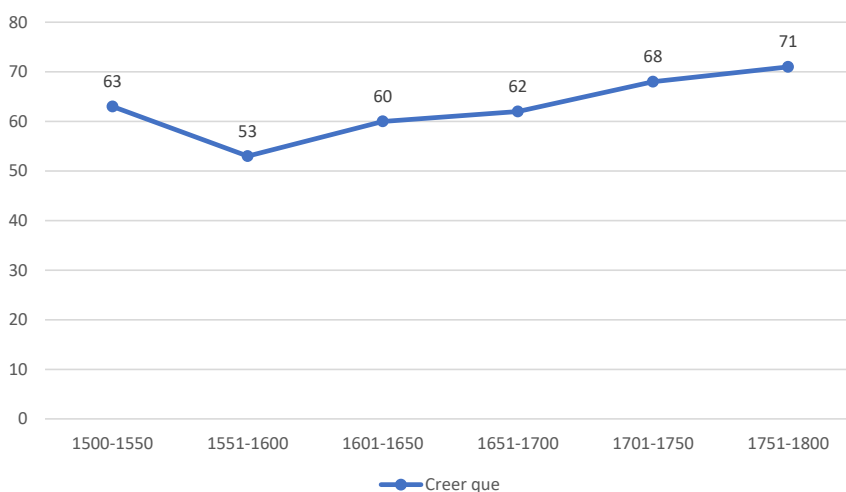


Figure 1: Distribution of the uses of the variant *creer que* in periods of fifty years (%).

further *creer* and the verb of the subordinate clause are from each other. Cognitively, this distance seems to warrant the presence of the subordinate link to a greater extent. The results also show that the highest rates of *creer que* are produced in contexts where the first person does not intervene (*yo creo* ‘I think/believe’), which is the opposite to what happens with elisions throughout history. Along with these signs of structural persistence, however, a few changes in grammar can also be seen over time. Thus, in the 18th century (but not in the Golden Age period), the presence of a *que* in the previous context produces a “horror aequi” effect that discourages the selection of this variant, perhaps to avoid the cacophonous effect that would be created by the simultaneous presence of two (or more) *que*. At the same time, while in Golden Age Spanish, the future in the complement clause appears as one of the most favourable verbal contexts for elisions (*creo* \emptyset *no vendrá* ‘I don’t think \emptyset he will come’), in the 18th century, this factor group underwent a semantic restructuring. Elisions are now favoured in all tenses characterised by high rates of virtuality and low factuality (this is the case again of the future, but also the conditional or the subjunctive tenses), which is just the opposite to the more factual paradigms (the present and past indicative tenses), among which the use of the complementiser increased significantly.

However, despite the latter’s relevance, the main changes in the configuration of the variable took place at the socio-stylistic level. Indeed, the results exhibited in Table 3 show some notable changes in the transition from one period to another. Hence, in Golden Age Spanish, the variation is sensitive to both the communicative tenor and the writers’ social status, while these factor groups faded away in the Enlightenment period.

Table 3 shows the selection of *creer que* is favoured in the more formal and distant contexts (0.59; 72%) during the classical period, while the *closer* ones disfavour it to a similar extent (0.41; 49%). Nevertheless, the influence of this factor group is diluted among the 18th-century data when the proportions of the *que* variant are much closer in the two communicative settings. Moreover, the same happens on a sociolectal level: in Golden Age Spanish, the variation shows a break between the elites and the rest of society, with the former acting as the main spearhead of *creer que* (0.62; 66%). Given these results, the linguistic variable could be interpreted as a sociolinguistic marker during the Golden Age period, given its particular sensitivity to sociolectal and stylistic variation (Labov 1972: 346–347; Tagliamonte 2012: 27–30). This characterisation can be better appreciated in Table 4, which offers the results of a re-analysis where the two independent variables, tenor and social status, are cross-tabulated.

As can be seen, there is indeed a close correlation between the two factors, so that as we go down the social pyramid and the scale of formality, the uses of the complementiser decrease. Hence, both in the social elites and in the rest of society,

Table 3: Statistical contribution of the extralinguistic predictors to the selection of *creer que* in the two periods analysed.

	Golden age				18th century			
	<i>N</i>	%	Σ	FW	<i>N</i>	%	Σ	FW
TENOR								
Close	283	49	579	0.41	160	60	268	–
Distant	105	72	147	0.59	224	65	342	–
GENDER								
Men	327	53	616	–	465	62	587	0.31
Women	59	55	108	–	19	83	23	0.69
STATUS								
Elites	98	66	149	0.62	100	63	159	–
Rest	290	50	577	0.46	167	63	471	–
CONTEXT								
Europe	127	59	217	–	207	74	280	0.63
America	261	52	509	–	177	54	330	0.37

Golden Age: Intercept: -0.983 ; log.likelihood: -391.839 ; AIC: 801.93 . Dxy: 0.84 . Input 0.57 . 18th century: Intercept: -0.47 ; log.likelihood: -280.07 ; AIC: 574.327 . Dxy: 0.88 . Input 0.70 . Along with the absolute frequencies, the percentages and the factor weight of each selected constraint, these measures contain several data derived from the mixed-effects regression analysis implemented with random intercepts. Thus, the AIC indicator is related to the predictive power of the statistical model, calculated from the log.likelihood, as a goodness-of-fit criterion, and the number of parameters as a measure of complexity. The Dxy statistic is also a goodness-of-fit measure whose figures closer to 1 indicate higher predictive power.

Table 4: Comparative results of a new regression analysis after crossing the communicative tenor and the social status predictors in the two periods.

	Golden age				18th century			
	<i>N</i>	%	Σ	FW	<i>N</i>	%	Σ	FW
TENOR + STATUS								
Distant – high	61	78	78	0.79	75	68	110	–
Distant – rest	44	64	69	0.55	105	69	153	–
Close – high	45	57	79	0.41	25	52	49	–
Close – rest	238	48	500	0.36	179	60	298	–

creer que increases in the more formal and distant communicative spheres and decreases in the closer and more personal ones. Still in both stylistic domains, the higher numbers are always for the elites. However, little of this remained in the 18th century: the proportions of this variant were still somewhat greater in the distant sphere, but the differences were minor, while the sociolectal correlations differed from those noted previously.

Conversely, during the Enlightenment period, some social constraints arose that had not worked before. One of them was the *gender* of the writers, which was selected by the regression and revealed the greater diffusion of *creer que* in female speech (0.69; 83%), at some distance from that of their male counterparts (0.31; 62%). In this regard, our results are in line with one of the usual outcomes of genderlect variation, in which women champion the use of the prestige variants, either in the contexts of change that point towards the prestige norms or in stable variation, as seems to be the case with this linguistic variable in the 18th century. Nevertheless, the highly unbalanced sample between men and women in the corpus (see Section 3) forces us to be cautious in interpreting this result.

These cautions disappear with the context in which the texts were written. We have distinguished between letters and other ego-texts written in Spain and those written in America under migratory conditions in which interdialectal contact must have been the norm, with potential consequences for variation and language change. Far more balanced than gender in terms of sample size, the results of this factor group show a considerable distance between the two periods examined. Thus, in Golden Age Spanish, the differences in frequency between the two groups of writers were of minor importance and were ruled out by the regression analysis. In contrast, the frequency distances became broader in the 18th century, with a notably greater vitality of *creer que* in the European varieties (0.63; 74%) than in the American ones (0.37; 54%). Could elisions be thought of as an archaism still deeply rooted in America in the 18th century while mainly receding in Spain? Did this variation change on the two sides of the Atlantic in line with the notable differences in frequency?

We carried out a new regression analysis to answer these questions, comparing American and European data from the 18th century after cross-tabulating them with the most relevant linguistic factor groups reported at the beginning of this section. The results of this comparative analysis are shown in Table 5.

The data show that, despite the differences in frequency in nearly all contexts, striking parallelisms are observed in the conditioning on both sides. The independent variables selected as significant and non-significant in America were the same as those in Europe. The only exception was the person/tense of the verb *creer*, which was not selected in European varieties, although the differences in frequency pointed in the same direction as in America.

5.2 *Quienes/quien* with a plural antecedent

The results obtained in this variable differed from those observed in the previous section. Thus, as shown in Table 6, the variant that later became almost categorical in relative clauses with animate antecedents (*quienes*) was very much in the minority

Table 5: Two independent regression analyses for the American and European contexts (structural predictors, 18th century).

	America				Europe			
	<i>N</i>	%	Σ	FW	<i>N</i>	%	Σ	FW
VERB ADJACENCY								
Adjacent	34	29	119	0.16	73	58	126	0.23
Intermediate distance	49	45	108	0.26	45	73	62	0.33
Distant	94	91	103	0.94	89	97	92	0.88
TENSE (COMPL. CLAUSE)								
Virtual tenses	101	46	219	0.34	99	66	149	0.41
Factual tenses	76	69	111	0.63	108	82	131	0.60
PERSON-TENSE (CREER)								
<i>(Yo) creo</i>	94	47	202	0.39	100	68	147	–
Rest	83	65	128	0.62	107	80	133	–
PRIMING								
<i>Que</i>	18	36	50	0.29	13	45	29	0.32
Rest	159	57	280	0.71	194	77	251	0.68

America: Intercept: -0.2 ; log.likelihood: -156.612 ; AIC: 327.571 . Dxy: 0.91 ; Input: 0.54 . Europe: Intercept: -0.988 ; log.likelihood: -118.655 ; AIC: 249.617 . Dxy: 0.81 ; Input: 0.74

during a large part of the Golden Age period. It was only from the second half of the 17th century onwards that *quienes* began to take off, in an abrupt change that took place in a short time, and which continued, although at a more moderate pace, in the following period.

From 1660 onwards, when the change began to accelerate in society, the variation was linguistically constrained by the type of relative clause, so that the innovative form *quienes* found a preferential environment in non-defining clauses (“*estos son mis familiares, de quienes te hablé hace una semana*” ‘these are my relatives, whom I spoke to you about a week ago’), rather than in defining clauses (“*estos son los familiares de quienes te hablé*” ‘these are the relatives I spoke to you about’). In contrast,

Table 6: Distribution of the relative pronouns *quien/quienes* in four periods.

	<i>Quien</i>		<i>Quienes</i>		Σ
	<i>N</i>	%	<i>N</i>	%	<i>N</i>
1500–1580	209	97	7	3	216
1581–1660	311	87	46	13	357
1661–1740	74	40	111	60	185
1741–1820	54	13	386	87	440
Total	648		550		1198

in the initial period of the change (1581–1660),¹⁰ the variation was conditioned by the syntactic complexity of the antecedents and by their semantic category. Thus, the coordination of antecedents favoured *quienes* (“mi tío y mi padre, a *quienes* no veo desde hace tiempo” ‘my uncle and my father, whom I have not seen for some time’), and especially those involving a proper noun (“mi padre y Pedro, a *quienes* no veo desde hace tiempo” ‘my father and Pedro, whom I have not seen for some time’). However, the significance of both factor groups weakened in the following stages, coinciding with the phase of abrupt diffusion of the innovative variant.

On the extralinguistic level (see Table 7), we begin by highlighting the irrelevance this time of the context factor group, with very similar distributions in the European and American varieties. Neither does the gender of the writers show any significant differences between men and women in this variation. The apparent distances in the period 1661–1740 are based again on a significant imbalance in the

Table 7: Contribution of the extralinguistic factors to the selection of *quienes* in three periods.

	1581–1660			1661–1740			1741–1820		
	<i>N</i>	%	FW	<i>N</i>	%	FW	<i>N</i>	%	FW
Tenor									
Close	17	12	–	94	63	–	276	86	–
Distant	16	14	–	19	50	–	119	93	–
Gender									
Men	35	12	–	104	62	–	389	88	–
Women	11	17	–	6	38	–	9	89	–
Status									
Elites	33	14	–	12	27	0.33	135	90	–
Rest	14	12	–	96	68	0.67	261	87	–
Context									
Europe	32	13	–	48	63	–	147	88	–
America	15	13	–	63	58	–	249	88	–
1581–1660:	<i>n</i>	intercept	overall proportion	log.likelihood	AICc	Dxy.total			
	357	1.7	0.13	–123.066	256.302	0.592			
1661–1740:	<i>n</i>	Intercept	Overall proportion	log.likelihood	AICc	Dxy.total			
	185	0.334	0.60	–108.11	224.443	0.747			
1741–1820:	<i>n</i>	Intercept	Overall proportion	log.likelihood	AICc	Dxy.total			
	440	2.323	0.88	–133.767	275.627	0.96			

¹⁰ Given the virtual absence of variation during the first stage considered (1500–1580), with barely 3% of the occurrences of *quienes*, for the regression analysis, we only examine the following three periods: initial (1581–1660), intermediate (1661–1740) and final (1741–1820).

samples, with only 16 female tokens (8% of the total). These rules out the possibility of measuring any possible genderlect differences with precision, which are not observed in the other periods either.

The *status* of the writers is an entirely different case. This factor group shows some changes in the direction of effect that suggest diverse sociolectal forces at different times. For example, it is not the case in the initial stage of change (1581–1660), when both the *elites* and the *rest* of society displayed similar figures (14% vs. 12%). However, things are pretty different in the next phase, in which the change begins to expand. At that moment, the highest rank does not favour the new uses of *quienes* (0.33; 27%), unlike the rest of society (0.67; 68%). It seems, in short, that during this critical stage in the evolution of the relative pronouns, the representatives of the highest levels of society still exhibited a conservative behaviour, opposed to a change from below, such as the one that seemed to be triggered in this period.

Another piece of evidence favouring the interpretation of *quienes* spreading as a change from below can be obtained after cross-tabulating the social *status* and the *tenor* of the interactions. Although the latter factor group is not selected in isolation by the regression analysis, the period 1661–1740 is the only one with some apparent differences between close (63%) and distant contexts (50%). Initially, this contrast fits well with the preferential diffusion of changes from below in more spontaneous registers (Labov 2001: 274). However, these distances become even more significant when connected with the writers' social status, as shown in Table 8.

The table shows that the greatest obstacle to the diffusion of *quienes* with a plural antecedent takes place in the most distant communicative settings by the social elites (0.38; 30%), with their more personal communications in a neutral position (0.49; 42%). In contrast, the rest of society appears as much more favourable towards using the innovative variant (0.72; 68%).

On the other hand, the conversion of this variant into the definitive new prestige form was already very close in time, as revealed by the data in Table 7 for the following period (1741–1820). At that moment, the stylistic and sociolectal differences noted in the previous phase of sharp expansion had wholly disappeared, reflecting the almost universal adoption of the plural pronoun.

Table 8: Cross-tabulation of the social status and the tenor predictors in the period of abrupt expansion of *quienes* (1661–1740).

Status – tenor	N	%	FW
Elites – distant	6	30	0.38
Elites – close	11	42	0.49
Rest	96	68	0.72

5.3 *Allí* (vs. *allá*)

Table 9 reflects the distribution of the adverbs *allí* and *allá* in the two periods now under consideration (16th and 18th centuries), showing a notable change in the preferences for each form over time. Thus, while in early Golden Age Spanish, the unmarked variant was *allá*, in the 18th century this role was played by *allí*, with averages already close to the figures found in contemporary speech communities.

Table 9: Distribution of the adverbs *allí/allá* in the two periods.

	16th century		18th century	
	<i>N</i>	%	<i>N</i>	%
Allí	271	22	834	64
Allá	962	78	462	36
Σ	1233		1296	

Moreover, the following figures provide information about another relevant fact: the change favouring the variant *allí* is visible along the whole-time axis. Nevertheless, in the 16th century (Figure 2), this change only seems to have been activated in the last three decades, with figures that double the far more moderate ones in the rest of the century.

The 18th century also displayed a favourable evolution of this adverb, with significant differences between the first part of the century and the rest. As Figure 3 reveals, there was a significant leap forward in selecting *allí* from the middle of the century onwards, even though, from then on, frequencies level out.

That said, we have also found that the linguistic conditioning of the variation is identical at all times. Thus, in both periods, *allí* is favoured in the same contexts: (a) precise circumstantial coordinates; (b) states and movements that point to the origin; (c) non-contrastive contexts; (d) the presence of adverbs of the same series in the previous context (forms in *-í*, such as *allí* or *aquí*); and (e) the positions before the governing element (for more details on each of them, see Section 2.3). Despite this structural persistence in linguistic conditioning, the strength of these factors weakened over time so that the distance between the most and least favourable contexts was considerably reduced in the 18th century.

Notwithstanding the (weakened) maintenance of these linguistic constraints, this variation shows some notable changes in the socio-stylistic conditioning. As shown in Table 10, the extralinguistic factor groups display a very different profile at each stage.

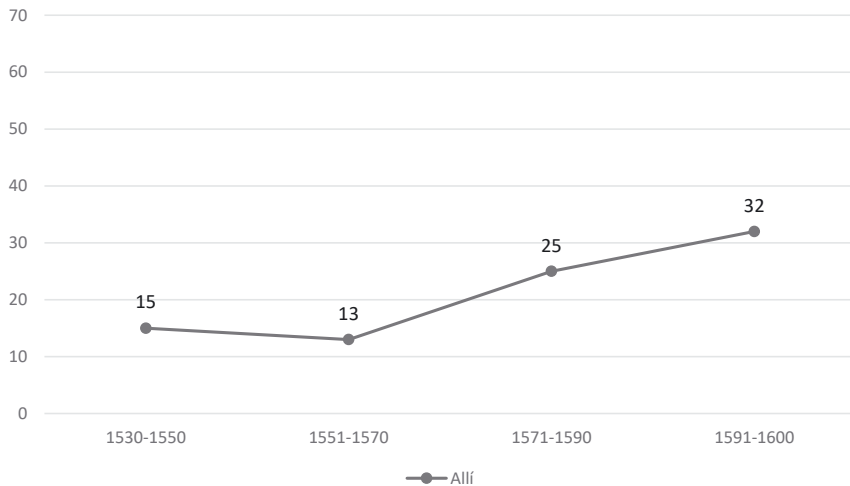


Figure 2: Distribution of the uses of *allí* in different subperiods of the 16th century (%). In the corpus, there are no data from before 1528.

The summary of what happened in the 16th century is simple, as none of the independent variables considered showed any notable (and far less significant) differences. It is perhaps the correlate of a still incipient moment in the spreading of the change, which was only clearly activated in the last three decades of the century (see Figure 2). It would not be surprising if the linguistic variable failed to show any

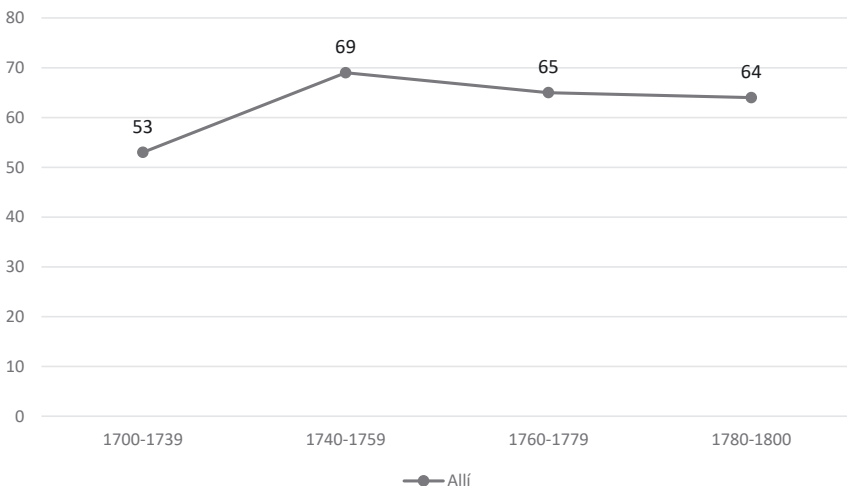


Figure 3: Distribution of the uses of *allí* in different subperiods of the 18th century (%).

Table 10: Contribution of the socio-stylistic predictors to the selection of *allí* in the 16th and 18th centuries.

	16th century				18th century			
	<i>N</i>	%	Σ	FW	<i>N</i>	%	Σ	FW
Tenor								
Close	188	23	794	–	303	50	604	0.40
Distant	83	19	439	–	531	77	692	0.60
Gender								
Men	248	23	1094	–	824	65	1262	–
Women	23	16	139	–	10	32	34	–
Status								
Elites	79	18	444	–	531	76	700	0.65
Rest	190	24	789	–	303	51	596	0.35
Context								
Europe	27	19	140	–	554	71	780	– ^a
America	244	22	1093	–	280	54	516	

16th century: $n = 1233$; intercept: -42.697 ; overall proportion: 0.22; log.likelihood: -429.308 ; AICc: 880.832; Dxy.total: 0.90. 18th century: $n = 1296$; intercept: -24.993 ; overall proportion: 0.64; log.likelihood: -595.347 ; AICc: 1219.021; Dxy.total: 0.77. ^aSee Table 12 for cross-tabulation between this factor group and the social status.

clear sociolinguistic correlations in these circumstances while the variation ran preferably along structural paths. The case was different, however, in the 18th century. Leaving aside the gender differences in this last period, also burdened here by a considerable imbalance in the samples, the data reveal the significant incidence of some extralinguistic factors that, either in isolation or in interaction, highlight a sociolinguistic profile that is far more advanced and runs parallel with the notable frequency changes mentioned earlier.

One of those factors is the *tenor* of the interactions. The results of the regression confirm the incidence of this factor group and suggest that the more distant texts are those that favour the selection of *allí* the most (0.60; 77%), as opposed to the closer and more personal documents, in which the adverb still competes on equal terms with *allá* (0.40; 50%). The influence of this factor also interacts with the writers' social *status*, the independent relevance of which is also guaranteed by the regression, with social elites as the leading promoters of the expanding variant (0.65; 76%), as opposed to the more conservative behaviour of the rest of society (0.35; 51%).

Moreover, cross-tabulation between these two factor groups, whose proportions are summarised in Figure 4, shows some noteworthy trends. As can be observed, the elites always outnumber the rest of society in the use of *allí*. Nevertheless, the difference between the two groups in the more distant texts is considerably lower than in the closer ones, where the use of *allí* by the higher rank doubles (67%) the

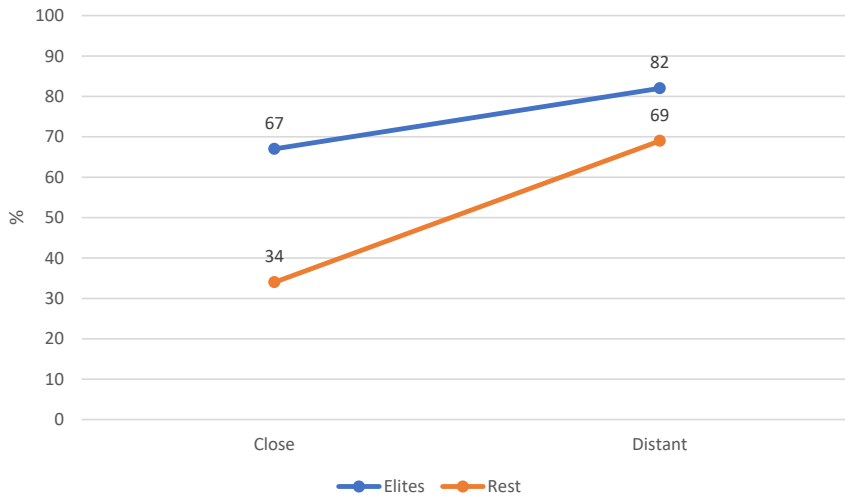


Figure 4: Cross-tabulation between the social status and the tenor factor groups in the selection of *allí* in the 18th century corpus (%).

averages found in the others (34%). A re-analysis in which both factor groups were considered together allowed us to discover a perfect gradation between three levels (see Table 11): at one extreme, the distant texts (regardless of social status) are those that favour *allí* the most (0.66; 77%), followed by close communication among the elites (0.55; 67%); finally, at the other end of this continuum, the most personal and intimate documents written by the rest of society exert an unfavourable effect on the selection of this adverb (0.30; 34%).

A similar interaction operates with the *context* in which the texts were written. The data in Table 10 show differences in the uses of *allí* in Europe (71%) and America (54%) but the regression did not select the factor group. Why not? The answer to this question will probably be found again by cross-tabulating it against status (see Table 12). This crossing shows that the preference for *allí* existed indiscriminately among the elites settled both in America (79%) and in Europe (75%) (0.68; 76%). Ultimately, the

Table 11: Contributions to the selection of *allí* after the cross-tabulation between the social status and tenor (18th century corpus).

	<i>N</i>	%	Σ	<i>FW</i>
Distant	531	77	692	0.66
Elites – close	195	67	289	0.55
Rest – close	108	34	302	0.30

Table 12: Contributions to the selection of *allí* after the cross-tabulation between the social status and the origin of the texts (18th century corpus).

	<i>N</i>	%	Σ	<i>FW</i>
Elites	531	76	700	0.68
Rest – Europe	174	64	271	0.49
Rest – America	129	40	325	0.39

differences were limited to the rest of society, with averages significantly more favourable towards *allí* in Spain (0.49; 64%) than in America (0.39; 40%).

6 Discussion

Table 13 summarises the incidence of the extralinguistic factors considered in analysing three linguistic variables at different moments between the Golden Age period and early modern Spanish.

The data in the table reveal some notable differences between the three variables. As we have seen, *creer que* + *complement clause* was always the majority form in Spanish syntax, although, during the Golden Age period, a considerable degree of variation occurred in alternating with elisions (*creer Ø*). In fact, in the 16th and 17th centuries, a significant social and stylistic correlation was observed. The instances of *creer que* in the most distant contexts by the elites always exceeded those of other people, and the same sociolectal distribution is observed in the closer and more personal settings, with the difference that the uses of this variant were now significantly lower in all sociolects. However, this socio-stylistic profile disappeared

Table 13: Incidence of the socio-stylistic predictors in the three phenomena of variation and change considered in the study.

	<i>Creer que</i>		<i>Quienes</i>			<i>Allí</i>	
	Golden age	18th century	1580–1660	1660–1740	1740–1820	16th century	18th century
Tenor	+	–	–	(+) ^a	–	–	+
Status	+	–	–	+	–	–	+
Gender	–	+	–	–	–	–	–
Context	–	+	–	–	–	–	(+)

^aThe brackets reflect those factor groups that are not selected in isolation by the mixed-effect regression analysis, but whose statistical relevance is undisputed in combination with others.

in the 18th century, when the *zero* variant became weaker. The old correlations left only a few traces in the genderlect variation, with women as the leaders in the use of the prestige variant, although the imbalances in the corpus between male and female speech samples mean that we have to be cautious regarding the true incidence of this factor. In contrast, in the century of the Enlightenment, a new extralinguistic constraint seems to take over: the European varieties took the lead in the generalisation of the prestige variant, while in America, the elisions were still notable, even though the grammar remained essentially the same.

In contrast to the uses of *creer que*, the relative pronoun *quienes* was very much in the minority until well into the 17th century, without any signs of socio-stylistic variation until then. As of the second half of that century, however, an abrupt change boosted the use of this innovative form rather than the invariable and traditional *quien*. This change was first driven by the lower ranks but faced resistance from the social elites. However, from the mid-18th century onwards, in line with the almost categorical spread of the new variant, these sociolectal differences were neutralised entirely. In short, we seem to deal with a quick change, with significant sociolectal and stylistic implications at that point, which were nevertheless soon to be abandoned once the innovative variant became (almost) definitive.

Although not in the same proportion as *quienes*, the demonstrative adverb *allí* was also the minority form throughout the 16th century, with no signs of socio-stylistic variation. However, things changed radically in the 18th century, with a significant increase in use. Likewise, *allí* displayed social and stylistic variation in this period, but with a different sense to what is observed in the other two variables. Hence, if in the stage of expansion of *quienes* the diffusion of the relative pronoun was produced above by the lower-middle sectors of society and in close communicative settings, the generalisation of *allí* pointed in the opposite direction, that is, towards a change driven by the elites, and with a particular emphasis on the more distant registers. These differences are also evident in comparison with *creer + complement clause*, since, in this case, the consolidation of the *que* variant in the 18th century diluted the social and stylistic constraints that were prevalent in the Golden Age period. Even so, both *creer que* and *allí* showed parallelism in their dialectal diffusion during early modern Spanish: they expanded more rapidly in the European varieties than in the American ones.¹¹

Of the extralinguistic factor groups considered in this study, the social status appears to be the most crucial since, at some point, it independently conditioned the

¹¹ The conservativeness of the American varieties in two of the three linguistic changes seems to support the “colonial lag” hypothesis, that is, that the colonial varieties of a language change less than the variety spoken in the mother country (Nevalainen 2006). However, much more data of this kind would be needed to confirm the hypothesis more precisely in the history of Spanish.

three phenomena analysed. Nevertheless, we have seen that the direction of effect of this conditioning varied substantially depending on several parameters, such as the speed and robustness of changes, their typology, the vitality of the alternative variants or the role played by the structural constraints governing the variation. Thus, in the paradigm of relative pronouns, we witness an abrupt increase of the novel variant, *quienes*, by the middle of the 17th century. Nevertheless, this new form was soon to become almost categorical, relegating the traditional and invariable *quien* to sporadic uses. In the first stages, the leading role of the lower sociolects can be seen in this change, which the elites resisted for some time. Nonetheless, as the new variant became firmly established by the mid-18th century, these sociolectal differences disappeared.

What is so special about a linguistic change like this taking place in a few decades and with its back turned, initially, to the dominant classes? In our view, this fact may be related to the analogical nature of the change, which led to the creation of a functional variant similar to those observed in other paradigms of grammar, where the singular and the plural have different forms (*alguno* vs. *algunos* ‘one vs. some’; *esta* vs. *estas* ‘this vs. these’), etc. These analogical changes were akin to vernacular speech, so it is not surprising that they may have been initially promoted by the lower-middle classes, although they were later joined by the elites, given the new variant’s high functional value.

However, the picture was different from other forms that, even if they ended up imposing themselves on their alternative variants, did not do so at the same rate or with the same strength, and they did not respond to the typology of analogical change we have just outlined. This occurred with the progress of *allí* in the paradigm of the demonstrative pronouns of distance in early modern Spanish. In this case, in which the alternative variant (*allá*) still showed signs of a certain vitality (especially in the American varieties), the force driving the changes was the elites, who left the rest of society behind, as opposed to the absence of sociolectal correlations in Golden Age Spanish, when *allí* was in the minority. What occurred with *creer (que) + complement clause* was also different. Although we already find evidence of the alternation between the full and the elided forms in the mediaeval period, *creer que* was always in the majority, even if its primacy was seriously affected by the impetus of the elisions in Golden Age Spanish. In this period, however, the elites acted as a retaining wall against these zero forms (with notable exceptions in some idiolects, as revealed by the letters of Saint Teresa de Jesús), preventing the consolidation of a change from below that, by the 18th century, had been completely deactivated.

The different profiles displayed by the three variables in the transition from classical to early modern Spanish also show some correlations with the structural forces governing the variation in each case. Thus, the particular rate at which the analogical pronoun *quienes* spread runs parallel to an apparent anarchic internal

variation conditioned by different constraints in each period, with no continuity between them. It is just the opposite to the expansion of *allí* over *allá*, which, despite the notable differences in frequency between Golden Age and early modern Spanish, always respond to similar internal conditioning. Finally, on this scale, *creer (que) + complement clause* holds an intermediate position, keeping some structural constraints over time and restructuring others.

7 Conclusions

In the preceding pages, we have reviewed three examples of variation and change in two critical moments in Spanish language history: the Golden Age period and early modern Spanish during the 18th century. Some of them were to become definitively consolidated in the language, as occurred with the categorical substitution of the invariable *quien* by the analogical *quienes* among the relative clauses with an animate antecedent. The same can be said of the (almost) definitive success of the syntactic variant *creer que + complement clause*, to the detriment of the forms with no complementiser (*creer Ø*), which are currently restricted to specific formal registers and sporadic uses. However, it is not the case of the expansion of *allí* at the expense of *allá*, since the latter is still widespread in Spanish, especially in many American varieties.

Using the tools of variationism and with the aid of a corpus composed entirely of egodocuments, the quantitative analysis reveals a unique sociolinguistic profile of each variable in the transition period analysed. We have also verified that this profile is related to some parameters such as the rate and intensity of the change, its typology or the linguistic constraints that operate in each case. Hence, the spread of analogical *quienes* took place as a consequence of an abrupt change from below, in which all the structural constraints were blown aside and, as a result, the new variant became firmly installed. The opposite occurred with the expansion of *allí* versus *allá*, which gained particular momentum in the 18th century, this time at the behest of the social elites, although with the same structural constraints as in previous times. Finally, *creer que + complement clause* was to corner its competitor (*creer Ø*) more and more in modern Spanish, with some changes in the internal grammar and the neutralisation of the sociolectal differences that governed it in the Golden Age period.

In short, both *creer que* and *quienes* or *allí* were to end up becoming the prestige variants – and in some cases, almost the only ones – in their respective paradigms. However, the processes that led to the generalisation of these forms could not have been more different.

From a methodological point of view, we believe that the present study also has some exciting implications for Spanish historical corpus linguistics. On the one hand, in line with what has been observed in the diachronic research on English and other European languages in recent times, it confirms the usefulness of corpora composed of texts close to the pole of communicative immediacy, and especially of private letters, as a means of approaching the Spanish vernacular of past times. For the sake of argument, some of the distributions found in the analysis of these variables differ from previous accounts based on more formal texts (literary, historiographic, legal, etc.) on which the diachronic study of Spanish has traditionally relied. For example, our data on the diffusion of the elided forms of *creer* \emptyset in the Golden Age period contrast with the more modest figures found in these other discursive traditions, thus contradicting the minority condition pointed out to date in the literature. On the other hand, applying the principles and methods of linguistic variationism and the comparative perspective also proves to be a handy tool for unravelling the factors that have constrained linguistic variation and change in the past.

Acknowledgements: The Author would like to thank Brittani Cortés and Miriam Salvador for their collaboration in the Sociolinguistic Lab on various tasks related to the project. Author would also like to thank two anonymous referees and the editors for their insightful and constructive comments. Any remaining errors are my sole responsibility.

Research funding: The present study is part of the research project “Socio-stylistic, idiolectal and discursive factors in Spanish variation and change: contributions from Historical sociolinguistics”, funded by the Spanish Ministry of Science and Innovation and ERDF funds (Ref. PID2021-122597NB-I00).

References

- Ash, Sharon. 2013. Social class. In Jack K. Chambers & N. Schilling-Estes (eds.), *The handbook of language variation and change*, 350–367. Malden, MA: John Wiley & Sons.
- Bello, Andrés. 1847/1972. *Gramática*. Caracas: Ministerio de Educación.
- Bergs, Alexander. 2005. *Social network analysis and historical sociolinguistics*. Berlin & New York: Mouton de Gruyter.
- Blas Arroyo, José Luis. 2020. Entre el español clásico y el primer español moderno: La evolución de las completivas dependientes de un predicado doxástico (1500-1800). *Bulletin of Spanish Studies* 98(1), 1–30.
- Blas Arroyo, José Luis. 2011. Spanish in contact with Catalan. In Manuel Díaz-Campos (ed.), *Handbook of hispanic sociolinguistics*, 374–394. Oxford: Blackwell.
- Blas Arroyo, José Luis. 2018. At the boundaries of linguistic convergence: variation in presentational *haber* / *haber*-hi. A sociolinguistic comparative analysis of Spanish and Catalan grammars. *Languages in Contrast* 18(2), 35–68.

- Blas Arroyo, José Luis & Juan González. 2022. Debilitamiento estructural y cambio social: La evolución de los adverbios demostrativos en el español antiguo (siglos XVI–XVIII). *Boletín de la Real Academia Española* 52(326). 435–470.
- Blas Arroyo, José Luis & Mónica Velando. 2019. Fases y condicionantes en un cambio gramatical abrupto: Origen y expansión del relativo *quienes* en el español medio (1500–1740). *Revue de Linguistique Romane* 331–332. 407–441.
- Bouzouita, Miriam & Enrique Pato. 2019. Antes había(n) pozos en el pueblo: La pluralización del verbo haber existencial en español rural europeo. *Revue de Linguistique Romane* 83(329). 137–165.
- Cano, Rafael. 1977. Las construcciones causativas en español. *Boletín de la Real Academia Española* 57. 221–258 y 323–351.
- Carbonero Cano, Pedro. 1979. *Deixis espacial y temporal en el sistema lingüístico*. Sevilla: Publicaciones de la Universidad de Sevilla.
- Castillo, Antonio. 2006. *Entre la pluma y la pared. Una historia social de la escritura en los Siglos de Oro*. Madrid: Akal.
- Company, Concepción & Rosa Espinosa. 2014. Adverbios demostrativos de lugar. In Concepción Company (dir.), *Sintaxis histórica de la lengua española. Tercera parte: Adverbios, preposiciones y conjunciones. Relaciones interoracionales*, 129–246. México: Fondo de Cultura Económica y Universidad Nacional Autónoma de México.
- Corominas, Joan & José Antonio Pascual. 1980–1991. *Diccionario Crítico Etimológico Castellano e Hispánico*. Madrid: Gredos.
- Correas, Gonzalo de. 1626. Arte de la Lengua Española Castellana. In José J. Gómez Asencio (ed.), *Antiguas gramáticas del castellano*. Madrid: Fundación Histórica Tavera: Digibis, DL. (CD_ROM edition).
- Elsaß, Stephan. 2012. The use of private letters and diaries in sociolinguistic investigation. In Juan M. Hernández-Campoy & Juan C. Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*, 156–169. Malden, MA & Oxford: Wiley-Blackwell.
- Fernández Alcaide, Marta. 2009. *Cartas de particulares en Indias del siglo XVI. Edición y estudio discursivo*. Frankfurt: Vervuert.
- Fernández Díaz, Roberto. 2004. *Historia de España. La sociedad del siglo XVIII*. Madrid: Austral.
- Folgar, Carlos. 1997. *Decir* + clausula completiva objeto en español. Algunos aspectos de su diacronía. *Moenia* 3. 377–410.
- Gallucci, María José. 2007. Variación entre los adverbios demostrativos aquí-acá en el habla de Mérida. *Lingua Americana* 20. 54–66.
- García Cornejo, Rosalía. 2006. *Morfología y sintaxis de que en la Edad Media*. Sevilla: Universidad de Sevilla.
- Girón Alconchel, José Luis. 2004. Cambios gramaticales en los Siglos de oro. In Rafael Cano (dir.), *Historia de la Lengua Española*, 859–893. Barcelona: Ariel.
- Iglesias, Isabel. 1996. *Los relativos en la prosa renacentista castellana*. Oviedo: Departamento de Filología Española.
- Johnson, Daniel Ezra. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1). 359–383.
- Jiménez Patón, Bartolomé. 1614/2001. *Breves Instituciones de la Gramática Española*. In José J. Gómez Asencio (ed.), *Antiguas gramáticas del castellano*. Madrid: Fundación Histórica Tavera: Digibis, DL. (CD_ROM edition).
- Keniston, Hayward. 1937. *The syntax of the Castilian Prose*. Chicago: The University of Chicago Press.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. 2001. *Principles of linguistic change. Vol. 2. Social factors*. Oxford: Blackwell.
- Lapesa, Rafael. 1980. *Historia de la Lengua Española*. Madrid: Gredos.
- Lope Blanch, Juan Miguel. 1997. Sobre la forma plural *quienes*. *Anuario de Letras* 25. 283–291.

- López Ruano, Raquel E. 2011. Los relativos: Usos y valores a partir de un texto del siglo XVI. *Anuario de Estudios Filológicos* 34. 77–95.
- Martinez, Glenn A. 2001. Política lingüística y contacto social en el español méxico-tejano: La oposición *-ra* y *-se* en Tejas durante el siglo XIX. *Hispania* 84(1). 114–124.
- Martínez Gómez-Gayoso, Benito. 1769. *Gramática de la lengua castellana reducida a breves reglas y fácil método para instrucción de la juventud*. Madrid: Imprenta de D. Gabriel Ramírez.
- Menéndez Pidal, Ramón. 1904/1941. *Manual de gramática histórica española*. Madrid: Espasa-Calpe.
- Nevalainen, Terttu. 2006. *An introduction to early modern English*. Oxford: Oxford University Press.
- Nevalainen, Terttu. 2018. A wider sociolinguistic perspective. In Terttu Nevalainen, Minna Palander-Collin & Tanja Säily (eds.), *Patterns of change in 18th-century English: A sociolinguistic approach*, 256–270. Amsterdam: John Benjamins.
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2017 [2003]. *Historical sociolinguistics: Language change in Tudor and Stuart England*, 2nd edn. New York: Routledge.
- Octavio de Toledo y Álvaro S. Huerta. 2011. Santa Teresa y la mano visible: Sobre las variantes sintácticas del Camino de perfección. In Mónica Castillo & Lola Pons (eds.), *‘Así se van las lenguas variando’*. *Nuevas tendencias en la investigación del cambio lingüístico en español*, 241–304. Peter Lang: Bern.
- Oesterreicher, Wulf. 2004. Textos entre inmediatez y distancia comunicativas. El problema de lo hablado escrito en el Siglo de Oro. In Rafael Cano Aguilar (ed.), *Historia de la lengua española, 729–769*. Barcelona: Ariel.
- Okulska, Urszula. 2010. Performing the world of politics through the discourse of institutional correspondence in Late Middle and Early Modern England. In Urszula Okulska & Piort Cap (eds.), *Perspectives in politics and discourse*, 173–197. Amsterdam: John Benjamins.
- Palander-Collin, Minna & Minna Nevala. 2005. Letters and letter writing: Introduction. *European Journal of English Studies* 9(1). 1–77.
- Penny, Ralph. 2000. *Variation and Change in Spanish*. Cambridge: Cambridge University Press.
- Pérez Saldanya, Manuel & Gemma Rigau Oliver. 2011. *Totz hic viuen, ací, e yo sol no yc puc viure!*: Una aproximació històrica als adverbis demostratius. In Maria R. Lloret & C. Clàudia Pons (eds.), *Noves aproximacions a la fonologia i a la morfologia del català*, 459–488. Alacant: Institut Interuniversitari de Filologia Valenciana.
- Poplack, Shana. 2011. Grammaticalisation and linguistic variation. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalisation*, 209–224. Oxford: Oxford University Press.
- Pountain, Christopher J. 2016. Tradiciones de discurso y Santa Teresa. *Scriptum Digital* 5. 5–23.
- Preston, Dennis. 1991. Style, status and change: Three sociolinguistic axioms. In Francis Byrne & Thom Hubner (eds.), *Development and structures of creole languages. Essays in honor of Derek Bickerton*, 43–59. Amsterdam: John Benjamins.
- Raible, Wolfgang. 1983. Knowing and believing and syntax. In Herman Parret (ed.), *On believing. De la croyance. Epistemological and semiotic approaches*, 274–291. Berlin: W. de Gruyter.
- Raumolin-Brunberg, Helena. 2005. Language change in adulthood: Historical letters as evidence. *European Journal of English Studies* 9(1). 37–51.
- Real Academia Española (RAE)/Asociación de las Academias de la Lengua Española (ASALE). 2009. *Nueva gramática de la lengua española (NGRAE)*. Madrid: Espasa Libros.
- Rutten, Gijsbert. 2016. The ‘Golden Age Myth’: The construction of Dutch as a national language. In Gijsbert Rutten & Kristine Horner (eds.), *Metalinguistic perspectives on Germanic languages*, 13–34. Oxford: Lang.
- Salazar, Ambrosio de. 1627. *Espejo general de la gramática en diálogos, para saber la natural y perfecta pronunciación de la lengua castellana*. Ruan: Louys Loudet.
- San Pedro, Benito de. 1769. *Arte del romance castellano*. Valencia: Imprenta de Benito Monfort.

- Sánchez Lancis, Carlos. 1990. *Estudio de los adverbios de espacio y tiempo en el español medieval*. Barcelona: Universitat Autònoma de Barcelona dissertation.
- Schneider, Edgar. 2013. Investigating variation and change in written documents. In Jack K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*, 67–96. Oxford: Blackwell.
- Sedano, Mercedes. 1996. *Los adverbios demostrativos en el español de Caracas y de otras ciudades hispanohablantes*. Madrid: UNED.
- Sedano, Mercedes. 2000. Los adverbios demostrativos en diez ciudades hispanohablantes: Resultados de una encuesta. In Carbonero Pedro, Manuel Casado & Pilar Gómez (eds.), *Lengua y discurso: Estudios dedicados al profesor Vidal Lamíquiz*, 917–934. Madrid: Arco Libros.
- Serradilla, Ana María. 1997. *El régimen de los verbos de entendimiento y lengua en español medieval*. Madrid: Universidad Autónoma de Madrid.
- Sobrino, Francisco. 1738. *Gramatica nueva española y francesa*. Bruselas: Francisco Foppens.
- Stangl, Werner. 2012. *Zwischen Authentizität und Fiktion. Die Korrespondenz spanischer Emigranten aus Amerika, 1492–1824*. Wien: Weimar.
- Tagliamonte, Sali A. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Malden/Oxford: Wiley-Blackwell.
- Terrado Pablo, Xavier. 1990. Sobre la forma de contenido de los adverbios de lugar. Cuestiones de diacronía. *Sintagma* 2. 55–66.
- Väänänen, Veikko. 1968. *Introducción al latín vulgar*. Madrid: Gredos.
- Villar, Juan del. 1651/2001. Arte de la Gramática Española. In José J. Gómez Asencio (ed.), *Antiguas gramáticas del castellano*. Madrid: Fundación Histórica Tavera: Digibis, DL. (CD_ROM edition).