

José Luis Blas Arroyo* and Javier Vellón Lahoz
**On the trail of grammaticalization in progress:
has *el que* become a compound relative pronoun
in the history of Spanish prepositional relative
clauses?**

<https://doi.org/10.1515/probus-2017-0010>

Abstract: In this paper, we examine the grammatical status of *el que* in Spanish prepositional relative clauses (el lugar *en el que* vivo) [the place where I live] from a variationist perspective of the theory of grammaticalization. At least from the nineteenth century onwards, several authors have defended the nature of *el que* as a compound relative pronoun, even if these forms continue alternating today with others without the article [el lugar *en que* vivo], in contrast to *el cual*, a fully grammaticalized relative since the late fifteenth century.

Based on a 3,200,000 word corpus of immediacy text (mainly private letters), we test the hypothesis of *el que* being a case of grammaticalization in progress from a variationist point of view, examining in depth what happens inside the grammar and the socio-stylistic matrix in different periods of history, from 1700 to 1960. The idea underlying this approach is that the structure of changes as well as the grammaticalization in progress can be inferred from the comparative analysis between different quantitative magnitudes of functionally similar variants. To do so, we have performed three independent mixed-effects regression analyses (*Rbrul*), one for each century.

The results of these comparative analyses confirm the progression of *el que* in prepositional subordinate clauses between the early eighteenth century and the first half of the twentieth. Yet, this progression has taken place at a slow pace and, objectively, can only be described as moderate. Moreover, a number of elements of continuity in history are revealed, such as several conditioning factors that are systematically selected and with the same explanatory direction in all periods. This casts serious doubt on the existence of a grammaticalization in progress in the case of *el que*, and sees the evolution as not essentially different from other morphosyntactic changes that have taken place in the history of Spanish.

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

Javier Vellón Lahoz, Universitat Jaume I, Castellón de la Plana, Spain, E-mail: vellon@fil.uji.es

Keywords: Grammaticalization in progress, historical sociolinguistics, syntactic variation, linguistic change, relative clauses, communicative immediacy, variationist method, Spanish

1 Introduction

The concept of grammaticalization has been used over and over again in recent times to explain very different evolutionary processes, some of which have little in common except for the incorporation of novelties at some level of grammar (Elvira 2015: 93). Despite its age, this has led to some critics to question the actual notion itself, which they have described as too fuzzy and with imprecise limits as compared to other changes at the grammatical level (Joseph 2011).

In the Romance languages, the concept has been applied to the study of change phenomena that are already firmly established or in very advanced phases. This is the case, for example, of the evolution of the demonstrative ILLE and the appearance of the category of the article in different languages, or the transformation of the periphrases with HABEO from the Latin to the Romance morphological future, and the later grammaticalization of the periphrasis *ir a* + infinitive with identical temporal values, together with the progressive reclusion of the morphological value within the modal sphere, which is very advanced in some varieties of Spanish (Pérez Saldanya 2003; Company 2004; Garachana 2016). Yet, we do not know so much about other evolutionary processes, whose progression has been much slower, but which, in essence, could reflect structurally similar outcomes.

Here, we will be addressing one of these processes: that which affects the variants of the relative pronoun *que* with an article (*el que*) in prepositional relative clauses with an antecedent ('la casa en la que vivimos'... [the house in which we lived]), and which, for centuries, have competed with the more traditional ones, without an article ('la casa en que vivimos'), to make a space for themselves in the grammar of these sentences. If by grammaticalization we understand not only the evolution that leads a lexical item to take on some grammatical or discursive functions, but also an already grammatical unit developing new roles in the grammar (Bybee 2006: 719–721; Elvira 2015: 93), it might be supposed that the extension of the variants with the article at the expense of those that do not carry it fits well into this heuristic paradigm. This is, of course, what has already happened in the past with another relative pronoun, *cual*, which has been completely grammaticalized since the late fifteenth century, to the point where, since then, it only appears in combination with an article (*el cual*, *la cual*, etc.). In the case of *el cual*, we

are therefore undoubtedly dealing with the existence of a compound relative pronoun, and that is how it has been even ‘grammatized’ (Auroux 1994) in the normative texts. But, what can be said of *el que*? Is it also a compound relative pronoun, as held by a number of authors, at least since the observations made by Bello in the nineteenth century (1988 [1847]: 106, § 325)? And how do we prove that status?

In any event, if we initially start from that premise – i.e. the grammaticalization of the compound relative pronoun *el que* – we would have to acknowledge, as did Girón (2006a: 1485), that the process we are dealing with here can only be characterized as partial, or, to use a metaphor commonly employed in socio-linguistic literature, *in progress*, since the evolution followed until reaching the status of being a compound relative pronoun is still a long way from completion, as shown by the prolonged competition among both forms – with and without an article – in prepositional relative clauses throughout history. The following are some representative examples taken from a corpus of epistolary texts that were used as the basis for this research. As will be seen, the linguistic context in all of them is very similar, as is also the antecedent to which the relative clauses are linked, the noun *ciudad* (city/town):

- (1) ...Si lo tienes a bien puedes proporcionar embarque para esta **ziud[a]d** *en la que* te proporcionaré acomodo (*Die private Korrespondenz spanischer Emigranten aus Amerika*, 1789)
[... If you see fit to do so, you can board ship to this city in which I will provide you with accommodation]
- (2) ...Ni en una **ciudad** tan grande como esta *en que* he tenido los empleos que son públicos... (*Correspondencia de emigrantes guipuzcoanos y navarros*, 1732)
[... Not even in a **city** as large as this one *in which* I have had public employment...]
- (3) ...Participo a Vds he llegado con toda felicidad a esta **ciu[da]d** *en la que* el s[eñ]or provisor me ha hospedado en casa de su sobrino (*Die private Korrespondenz spanischer Emigranten aus Amerika*, 1808)
[... I hereby notify you that I have arrived happily in this **city** *in which* His Worship the Vicar-General has offered me accommodation at his nephew’s house]
- (4) ...deví en México partticulares atenciones no sólo del s[eño]r virrey [...] sino a todas las demás personas de primera clase y desttinción de aquella

hermosa **ciudad**, *en que* recibí obsequio y aprecio singular (*Die private Korrespondenz spanischer Emigranten aus Amerika*, 1801)

[... in Mexico I was in debt for particular attentions not only from the Viceroy ... but to all the other people of first class and distinction in that beautiful **city**, *in which* I received gifts and singular kindness]

Now, what are the arguments that allow us to think that *el que* is, indeed, a compound relative pronoun and that its extension is due to a case of grammaticalization in progress? To respond to this question, in this paper, we will undertake a study under the variationist analysis paradigm, which in recent years has provided several new theoretical and methodological insights that are relevant to studies of linguistic change in general and, more particularly, to those on grammaticalization. Unlike other heuristic models, the variationist approach goes beyond research into change and grammaticalization by the simple analysis of frequencies to examine in greater depth what happens inside grammar at different times in history during which different variants compete for the same functional area (Poplack and Tagliamonte 2001; Poplack and Malvar 2007; Poplack and Dion 2009; Poplack 2011; Torres Cacoullós 2012). The idea underlying this approach is that the structure of such changes can be inferred from the comparative analysis between different quantitative magnitudes of functionally similar variants, in the contexts in which they are used and that make up their envelope of variation (Poplack 2011: 176).

The aims of this work are to provide answers to the following questions: what are the main routes of penetration of the variant *el que* in prepositional relative clauses with an antecedent in three different historical periods (eighteenth, nineteenth and twentieth centuries), in both the linguistic and the social matrix?¹ Is there any evidence of continuity in that evolutionary process or, in contrast, does the change come about abruptly between different eras? And, above all, is there any proof of a path of grammaticalization in progress that justifies the status of *el que* as a relative compound in the two and a half centuries analysed?

¹ These prepositional relative clauses were chosen because, in the opinion of some authors (see Girón 2006a: 1568), they represent the intermediate link between “the phrase *el que*, made up of an antecedent article-pronoun and *que*, and the compound relative pronoun (CR) *el que*,” while also allowing the variation in wider contexts (e.g. restrictive clauses, which are prohibited in other syntactic settings). Moreover, within these, we chose those headed by the preposition *en*, due to their being by far the most numerous in discourse and, therefore, the ones in which the evolutionary tendencies within this process of secular linguistic change and variation can be analysed with the greatest clarity.

2 The linguistic variable

The current coexistence of the forms with and without the article in prepositional relative clauses with an explicit antecedent is, in the opinion of Brucart (1999: 496), “one of the most complex phenomena in Spanish grammar” and is subject “to a considerable margin of diatopic and even idiolectal variation,” despite the fact that, from the normative point of view, some authors – see Gómez Torrego (1989: 125–136) – recommend the use today of the construction with the article, especially in the spoken language.

Lapesa (2000, 388ss.) considers that the origin of the construction formed by the article and the relative pronoun lies in the correlation of the Latin anaphoric pronoun *IS* + relative, and its later replacement by *ILLE QUI*, with the consequent transformation of the paradigm of the demonstratives in Vulgar Latin. The problem arises when it comes to determining the category of the element that precedes the relative pronoun: is it a demonstrative pronoun or an article with an updating function with respect to the relative? The issue is relevant because it has implications regarding the nature of the resulting construction, which in turn enables us to interpret some of the phenomena analysed in this study.

On the one hand, some authors take the opinion of the academic grammar endorsed by the Real Academia Española (RAE) (1931) (see Gili Gaya 1970: 304, § 231, and Girón 2006a: 1495) and point out that these distributions of the article conserve the etymological pronominal value. Thus, Bosque and Moreno (1990: 44), after analysing the case of *lo*, consider the remaining articles as being “pronominal variants.” In similar terms, Briz and Prunyonosa (1987: 104–105) hold that the article in these constructions acts as “both the nucleus and, at the same time, the antecedent of the relative clause” which proves “their concurrence with the demonstrative pronouns.”

For most scholars, however, this is a clear case of grammaticalization in which “the original pronoun has become a new article and, consequently, has lost its tonicity” (Hernández 1986: 479). We are thus before a “process of fossilisation” (Marcos Marín 1980: 237) of the form *article* + *relative pronoun* to the point where it is converted into a compound relative pronoun (CR), perhaps imitating the form *cual* “...whose mandatory article has also gradually reached in modern Spanish the form *que*” (Lope Blanch 1998: 575). In this evolution, before it occurred with the relative pronoun *que*, the history of the language shows combinations with other relative pronouns – *el cuyo*, *el por quien*, *el donde*, etc. – and sequences – *article* + *preposition* + *relative pronoun* – documented up until the twentieth century, which, in accordance with those in favour of the grammaticalization of the CR, represent a midway stage,

“a demonstrative pronoun that has begun its shift towards the article, but which continues to be a pronoun” (Girón 2006b: 767).² Yet, such usages having now been lost, and being limited to the combinations of the article with *que*, we are witnessing the birth of “a new relative pronoun, in which the article, having lost its status as an antecedent, was incorporated within the nexus as a simple marker of gender and number” (Eberenz 2000: 368). Consequently, the new CR, in a similar manner to the aforementioned *el cual*, possesses a mark of morphological variation, with the article converted into a mere “agreement affix” (Girón 2006b: 768). This opinion has recently been endorsed by the new *Grammar* of the RAE (2009), despite considering that this CR does not form a morphological paradigm, but instead that it would be included within “the syntactically constructed pronominal groups” (2009: 3294, § 44.1 i).³

Now, if this is so, why, after centuries of fierce competition, has not the process been completed, and the solutions with and without the article continue to coexist today, unlike what happened with *el cual*? Before attempting to put forward an answer to this question by means of a variationist study, we should first consider whether both structures are semantically equivalent or, in contrast, some differences can be found between them. For Alarcos (1994: 108, § 146), for example, there is no doubt that they are synonymous constructions, since “the article adds no significant value; it serves only to prevent the confusion that could arise between the relative pronoun and the conjunction *que* when they are preceded by a preposition.” And a similar opinion is voiced by other authors, who see the alternation as a purely syntactic phenomenon that does not affect the structure of the sentence in the slightest (cf. García García 1990: 56, Trujillo 1990: 27–28).

In any case, the literature contains a number of authors who have seen the decisive influence of several conditioning factors in the alternation, which would justify the progressive extension of the forms with the article at the expense of the simple forms; or vice versa: the still relative strength of the latter

2 Advocates of the grammaticalization thesis consider that the disappearance of the *article + preposition + relative pronoun* sequence demonstrates that another phase has been completed in the transformation of an old demonstrative pronoun to a simple concordance affix (Girón 2006b: 772–773). In our corpus, we also find one example of this structure in the eighteenth century:

“Ginovi me a escrito en respuesta a *la en que* le avise se hallava [...] (*Al recibo de esta*, 1790) [Ginovi wrote to me in reply to the one in which I warned him that he was [...]]

3 Since in non-prepositional relative clauses (“está el padre, *el que* te dije el otro día”), the morphosyntactic values of the pronoun and the article seem to coexist in a more manifest manner, some authors restrict the consideration of *el que* as a compound relative clause to just prepositional relative clauses (Brucart 1999: § 7.5.1).

in present-day Spanish. Hence, Brucart (1999: 494–495) pointed out that the fact that the form without the article (*en que*) has been preserved in current language usage is favoured by definite antecedents, although he acknowledges the possibility that this constraint “has not been generalised in all periods of history.” At the same time, he states that the variation does not reach the negative polarity (in this same line, see also Santana 2004).⁴ Likewise, several authors have related this variation with other additional factors, such as the semantic nature of the antecedent (human/non-human) (Porto Dapena 1997: 24–25), the type of relative (López García 1994: I, 440–442), the function of the relative pronoun (Martínez 1989: 154), the degree of determinacy of the antecedent (Santana Marrero 2004) or of the whole NP (Balbachán 2011), or due to stylistic reasons (Rivero 1991: 73–75) (for other conditioning factors from a cognitive grammar, see also Delbecque 2011).

For the purposes of our study, the research conducted on the evolution of the phenomenon in diachrony is of particular interest. In this regard, some linguists have pointed out that the extension of the relative pronoun preceded by the article in prepositional relative clauses did not come about in an apparent manner until the eighteenth century, although some report scattered testimonies with the monosyllabic prepositions (*a, con, de, en, por*) in previous centuries. Eberenz (2000: 368), for example, offers a small sample from the fifteenth century and acknowledges that “the phenomenon is still a long way from the grammaticalisation later achieved” (for further similar examples in the sixteenth and seventeenth centuries, see also Keniston (1937: § 15.158, § 15.228; Lapesa 2000: 398 and Girón 2006a: 1525ss).

In American texts written in the eighteenth century, Guzmán Riverón (2012: 184–187) found 79 occurrences of *en el que*, which account for 9% of the total, versus the remaining 91% ($n = 775$) consisting of the simple form. Likewise, this author mentions “some particularities” of the uses of the article, such as the relatively greater association with certain discursive traditions (geographical relations), the locative value of the antecedent and the distancing of the relative pronoun with respect to the antecedent.

All of these factors were also analysed by Girón (2006a: 1484 and ss), although based on a far more comprehensive basic corpus of texts of peninsular Spanish, between the mid-seventeenth and the present century. After pointing out that “the frequency of *el que* has increased notably since the 18th and nineteenth centuries,” this author highlights the factors that make the greatest

⁴ However, as Girón (2006a: 1499) recalls, it is not difficult to find examples of the opposite: “Llega un momento en la vida en que nada tiene importancia... [A point in life is reached in which nothing has any value].”

contribution to the gradual use of this variant in prepositional relative clauses: the type of relative clause (appositive), the function of the relative pronoun (non-adverbial) and certain syntactic-discursive properties of the antecedent (non-immediate position, syntactic expansion and non-defined nature).

Now, in spite of the relevance of these approaches to the syntactic issue we are dealing with here, it must be stressed that the vast majority of them are based on a statistic that is too simple and descriptive, which goes little further than merely counting the absolute frequencies and percentages. Moreover, these calculations can lead to errors without a minimum inferential apparatus to uphold them, especially if the samples are not sufficiently comprehensive and representative. Furthermore, claiming that certain factors are associated with a particular form does not necessarily mean that in that context that form is the predominant one; in fact, quite the opposite may occur. As an example, in the analysis of texts from the eighteenth century performed by Girón (2006a: 1529), this author claims that “The type of relative clause makes it clear that *‘el que’* is widely preferred in appositive relative clauses (81%), whereas the preferred form in restrictive clauses is *‘que’* (82%), with the same distance of 63 percentage points between the two relative pronouns in both schemes” (our italics). Yet, a review of the frequencies reveals that, while the second statement is true, the first is not. Indeed, the number of appositive relative clauses constructed with *que* (81) is quite a lot higher than those with *el que* (26), regardless of the fact that the relative proportions of each variant change drastically in that context. Last, this method bases its conclusions on a univariate statistic, in which each factor is analysed in isolation, thus making it impossible to know the real significance of each one when studied in combination with the others, and thereby precluding any chance of interactions or relations of collinearity among them.

An alternative method, based moreover on specific theoretical principles such as those of the linguistic variationism, will be used for the interpretation of data in this study.

3 The variationist paradigm in the study of change and grammaticalization

In this study, we shall analyse the functioning of the alternating forms at the beginning of prepositional relative clauses with an express antecedent between the eighteenth century and 1960. As we shall see, from this analysis, it can be inferred that such forms have never acted as manifestations of free or

idiosyncratic variation, but, to the contrary, they represent the consequence of a variability regularly determined by a number of structural and non-structural factors. With this in mind, the research has focused on the most relevant principles that variationism has developed in recent times.

As our starting point, we assume the principle of functional neutralization in discourse, that is, the existence of forms that, over and beyond the potential differences in the linguistic system, can be used by speakers to express similar referential or functional contents (Sankoff 1988). In the theory of grammaticalization, this inherent variability corresponds in broad terms to the concept of *layering*, that is to say, the capacity of different constructions to represent “similar or even identical functions” (Hopper 1991: 22–24), in a process by means of which *new layers* emerge without displacing the older ones within the same functional area. This implies that the grammaticalization of one form cannot be isolated from what happens with those that compete with it for that linguistic area. As highlighted by Poplack (2011: 176): “...grammaticalisation is usually construed – and studied – as the set of changes involved in the association of *one* form with a new (presumably more grammatical) meaning or function, *downplaying, or even ignoring, the role of other layers coexisting in that context*” (our italics).

The foregoing leads us directly to the second fundamental principle of variationism, that of *accountability* (Labov 1972), according to which research must focus not only on the forms under study, but also on those that could have a place in their reference domain. Besides, the systematic study of variation requires the examination of not only the alternating forms, but also the contexts in which these could be neutralized in discourse. Thus, the analysis does not take the linguistic forms exactly as its point of departure, but rather the *envelope of variation*, that is, the wider reference domain in which they can alternate (Torres-Cacoullous 2011: 151). By reviewing the distribution of the variants in the different contexts, we will be in a position to know not only which forms are preferably used in different settings, but also their nature, interpreted heuristically in the form of conditioning factors. In short, the fundamental idea is not to account for just the grammaticalized form, but to compare the context in which it operates with that of the alternating variants, and to do so on a diachronic axis that is as extensive as possible.

This delimitation of the envelope of variation obliges us to specify in detail the syntactic–semantic spheres in which the variation takes place, thereby ruling out other contexts in which it is either non-existent or very limited. In the case we are dealing with here, this is the reference field of prepositional relative clauses with a nominal antecedent – or functional equivalent – due to the reasons outlined above (see note 2).

Based on these principles, the variationist paradigm considers that it is possible to diagnose both the existence and the extent to which grammaticalization has taken place between two points on a diachronic axis. To do so, it is necessary to compare several quantitative magnitudes derived from the distribution of the variants in the different contexts in which they coexist. In the case we are dealing with, we performed three independent multivariate analyses, one for each century (eighteenth, nineteenth and twentieth), with the aim of evaluating the way in which the above-mentioned predictors have conditioned the variation in each period.

In addition to comparing the frequencies and percentages, in the analysis of the results, we draw on four means of proof: (a) the list of factor groups selected and not selected by the logistic regression statistic; (b) the explanatory relevance of each factor group, derived from the corresponding *range*, which is a non-statistical measure obtained after subtracting the significances of the most favourable and inhibiting contexts, respectively; (c) the statistical significance of the different contexts within each factor group, measured by the corresponding probabilistic weights or *P* values (see Tagliamonte 2012: 173); and (d) the explanatory hierarchy, or direction of the effect, which runs from the most favourable environment of a variant to the most inhibiting.

On comparing these indices, we obtain a sample that sheds light on the role played by each variant – and in our case, especially the form *el que* – as the change has taken place, as well as the influence of the envelope of variation in each period.

4 Coding and methodology

In order to locate all the occurrences of the linguistic variable in the corpus, we used the concordance software application *Wordsmith* v. 6.0. In the logistic regression analysis, the occurrences of the majority variant (prepositional relative clauses introduced by *en que*) and those of the far more minority variant (headed by the compound relative pronoun *en el cual*) were considered jointly, and they were compared with the occurrences with an article (*en el que*), considered in this work as being the *application value*.⁵

⁵ In certain contexts, these variants may alternate with others, such as *quien*. Yet, this pronoun appears with animated and preferably human antecedents (Brucart 1999: 502 and ss., § 7.5.3), a context which in our corpus occurs only two or three times in each century, and was therefore excluded from the analysis. More possibilities of alternation arise with the adverbial relative *donde*, but again these are limited essentially to locative contexts, and are far more anecdotal in

Once the examples have been located, and every exclusions have been carried out, they have been coded on the basis of a series of different linguistic and extralinguistic factor groups constituting the broad environment in which the variants covary. Thus, in the variationist approach to grammaticalization, this is not only related to an increase in the overall frequencies of the grammaticalized variants, but also – and above all – with the existence of significant changes over time in its conditioning environment. In this way, if the grammaticalization process actually operates, it would be expected that the main contextual factors that constraint a particular variant in the past will gradually weaken or (in the most advanced cases) disappear. Through the variable rule analysis performed at different points on the time axis, and the statistical lines of evidence described above (see Section 3), we obtain a snapshot of the way in which that variable context affected the variants at each moment. From there, the use of a comparative methodology allows, in the words of Poplack (2011: 215): “[to] trace not only the rise and fall of variant forms but also their entry points into the system the trajectory of their functions, which is of particular interest to gramticalization theorists. In its capacity to transcend frequencies to reveal the *patterns* of variability and change, this is perhaps where VT has the most to offer”.

In the following, we offer a list of these conditioning factors, although a more detailed explanation and exemplification of those that were found to be significant will be undertaken in a later section (see § 6).

With regard to the linguistic factors, these can be classified into two large groups: those that directly involve the antecedent and those that affect the relative clause itself.

The first group includes the following:⁶

- *Lemma* of the antecedent
- *Distance*: Adjacent/Intermediate/Distant
- *Presence of the definite article*: Yes/No
- *Degree of determinacy*: Non-determined/Indefinite/Definite
- *Gender*: Feminine/Masculine
- *Number*: Singular/Plural

other semantic values. Thus, they are also precluded from the envelope of variation to which the variation is restricted in this study.

⁶ In Tables 3 to 8 (see Section 6), some of the groups mentioned here are grouped to facilitate statistical analysis, especially when similar distributional trends are detected, or the number of tokens are too small, which could jeopardize the analysis of regression. On other occasions, we also present the combined effect of two factor groups, when the explanatory power of these combinations is more significant than those reached by each factor group separately.

- *Function*: Adverbial/Others
- *Category*: Nominal/Others
- *Type of referent*: Concrete/Abstract

The second type of linguistic factors affect the relative clause, and come down to the following contexts:

- *Degree of coreferentiality of the subjects* (ruling clause/relative clause): Coreferential/Non-coreferential
- *Sentence modality*: Affirmative declarative/Negative declarative/Others
- *Syntax of the subject*: Explicit/Omitted
- *Sentential syntax*: Independent and main/Subordinate/Coordinate
- *Verb tense/mode*: Present indicative/Past (perfect and imperfect) indicative/Future indicative/Subjunctive/Others
- *Meaning of the relative clause*: Locative/Others
- *Function of the relative pronoun*: Circumstantial/Others
- *Degree of agentivity*: Impersonal and passive/Others
- *Type of relative clause*: Appositive/Restrictive

As can be noted, these factor groups are categorically diverse. Most are *grammatical* (gender, number, function and category of the antecedent, type of relative clause, function of the relative pronoun, etc.), but some of them are also of *semantic* nature (type of referent of the antecedent, meaning of the relative clause, sentence modality, degree of agentivity) and even *discursive* (distance from the antecedent, presence of the definite article in the antecedent). Some of them have already been considered in previous (non-variationist) analysis; however, to the best of our knowledge, it is the first time that such a number of factor groups are considered and, more importantly, that a joint interpretation of their combined effects is carried out. In fact, and as we shall see later (Section 6), there are occasions when the incidence of a given factor may be accelerated, countered or, simply, neutralized, by other factors. In those cases, only a multivariate analysis can unravel its true explanatory power.

On the other hand, and in line with the above, the diagnosis about the existence (or not) of grammaticalization (or the degree reached by it between two points in the temporal axis) will be related to a significant increase in the overall frequencies of the (grammaticalized) variant, but also, and specially, to a weakening on the constraints found in the past and/or its replacement by new ones. For example, if grammaticalization has turned the *el que* variant into a new relative compound, in the same way as *el cual*, it would be expected that those factors constraint mainly the variation in the eighteenth century (i. e. type and content of the relative clause or the presence of a definite article in the

antecedent; see 6.1) have significantly reduced, or even cancelled, its explanatory relevance two centuries later. This would be objective evidence, beyond the intuitions of the analyst, that the internal grammar of the grammaticalized variant has changed significantly and that the discursive functions performed by it are now qualitatively different.

Last, whenever it has been possible to obtain the corresponding information, we have also taken the following extralinguistic factors into account:

- *Gender* (Males/Females)
- *Status*: Despite the different configuration of the Spanish society in each period, for the purposes of the comparative analysis, we have classified the writers whose biographical data were available into three different social groups, in a hierarchy based on a combination of different parameters, such as power, the social class they belonged to, their economic status, profession, the degree of social and geographic mobility, etc.: High/Medium/Low.
- *Degree of relationship* (only in the epistolary genre): Distant/Close or familial
- *Origin of the author*: there are writers from 16 different regions of Spain. In order to make the analysis easier, these were grouped into two broad dialectal areas: Northern regions/Southern regions.
- *Migratory context*: Texts written from America/from Spain
- *Decade*
- *Writer*

Of these, the time axis (*decade*) is interpreted as a continuous factor, while the factors linked to the antecedent (*lemma*) and the identity of the *writer* are considered to be random intercepts. All the others are categorical factors, given that coding is established on the basis of fixed options: appositive/restrictive; explicit/omitted, man/woman, etc.

To account for the way in which variation and change have acted within the sphere of prepositional relative clauses throughout history, we carried out three independent mixed-effects regression analyses, one for each century, with the aim of evaluating the way in which the above-mentioned factors have conditioned variability within each period. With this aim in mind, we used the *Rbrul* toolkit (Johnson 2009), which enables us to obtain a list of factor groups selected as being significant, when all of them are analysed together, as well as the potential interactions among them.⁷

⁷ Because, initially, we did not know whether the two random factors could be collinear with respect to each other (e.g. whether the value of the factor *antecedent* would predict that of the variable *writer*), two parallel analyses were performed, each with a random variable. The results of these analyses are shown in Tables 3 to 8.

5 The corpus

As pointed out by Labov (1994: 11), data representing the early stages of language, especially if the aim is to find samples of vernacular speech, have generally been incomplete and fragmented. This has frequently meant that they are unsuitable for tracing the evolution of the alternating forms throughout history. Furthermore, there is also the need to have access to corpora that are sufficiently comprehensive and representative, which until recently has been more of a wish than a reality.

In order to reduce the effect of this *bad data problem*, sociolinguists have recently experimented with different genres that, while not exactly samples of oral speech, do however share many of the features that characterize orality, which makes them good candidates for studying variation and linguistic change (Poplack and Malvar 2007; Elspass 2012).

Within this framework, and as part of a currently ongoing research project in historical sociolinguistics (see note 1), for this study we compiled a corpus made up of texts close to communicative immediacy (Oesterreicher 2004). Today, such texts have become a valuable source for diachronic research (Oesterreicher 2004; Palander-Collin and Nevala 2005; Schneider 2013). For this paper, we have selected, for the most part, a series of private letters written by Spaniards from different social, dialectal and stylistic backgrounds.⁸

Due to their structural characteristics, letters are highly interesting testimonies for linguistic variation studies: they offer autobiographical data that reveal the relationships of power and solidarity between senders and receivers, in addition to their social status (Nevala 2009; Okulska 2010); likewise, they offer ethnographic details about social life at that time (Raumolin-Brunberg 2005; Dossena and Tieken-Boon van Ostade 2008); and they become samples that are reasonably close to the vernacular speech of past eras, among other reasons because they were not originally conceived with the intention of later being edited (Palander-Collin and Nevala 2005; Elspass 2012). Although letters display a certain amount of rigidity in their structure (above all in their opening and closing sequences) (Cano 1996), both the lower sociocultural level of many of their authors and the feelings that are expressed in them help make the language used in these texts of great interest for studies into the history of language. As remarked by García Mouton (1999: 264), “literary texts can imitate

⁸ Likewise, the corpus contains a smaller selection of other discursive genres that are close to communicative immediacy, such as autobiographies, diaries and the memoirs of persons from a range of social types.

reality; letters, even though subjected to all kinds of shaping, are a part of it.” At the same time, the conquest of America by Spain gave rise to an unprecedented exchange of communications, through the letters sent to and from both sides of the Atlantic, either with an official and administrative content or of a private and familiar nature (in the latter case, between emigrants and their families). Thanks to all this, today we now have an invaluable documentary corpus for studies into social history as well as for the type of linguistic research we are dealing with here.

Whenever possible, we have used the transcriptions undertaken by philologists and linguists, although those recently carried out by different social historians have also been included in the corpus, provided that they have respected the literal sense of the documents, perhaps only modernizing the spelling and the punctuation in order to make the text easier to understand by the reader. Although this kind of modernization invalidates the documents as material with which to study graphic and phonic variation, it is not necessarily the case in the research of syntactic and discursive variation, such as that being undertaken here (in the same vein, see Bergs 2012).

This has provided us with a corpus containing over three million words written by 2518 different individuals, although the distribution is somewhat unbalanced across the centuries, as can be seen in Table 1.⁹

Table 1: Number of words and writers by centuries.

| | Words | Writers |
|--------------------|-----------|---------|
| Eighteenth century | 1,242,588 | 1345 |
| Nineteenth century | 1,013,610 | 763 |
| Twentieth century | 1,037,849 | 410 |
| TOTAL | 3,294,047 | 2518 |

6 Results and analysis

After coding all valid tokens and excluding the cases that failed to meet the above-mentioned envelope of variation (see § 3), we found 2134 occurrences of the variable, whose distribution by centuries can be seen in Table 2.

⁹ The data for the twentieth century, however, only extend as far as 1960. The reader can find the full list of books that make up the corpus on the following link to the project: <http://sociolinguisticawe.wix.com/sociolinguisticauji>.

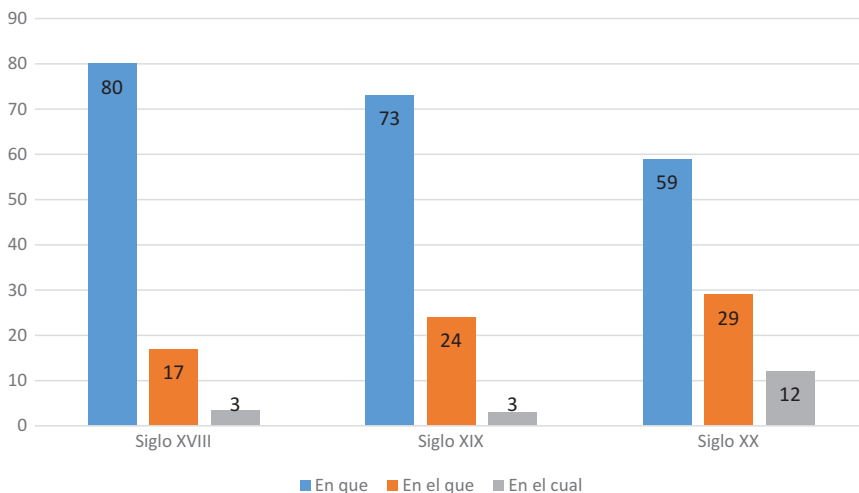
Table 2: Distribution of the variants by centuries.

| Variants | Eighteenth century | | Nineteenth century | | Twentieth century | |
|--------------------|--------------------|----|--------------------|----|-------------------|----|
| | <i>N</i> | % | <i>N</i> | % | <i>N</i> | % |
| En que | 601 | 80 | 539 | 73 | 379 | 59 |
| En el/la... que | 127 | 17 | 178 | 24 | 186 | 29 |
| En el/la... cuales | 26 | 3 | 17 | 3 | 72 | 12 |
| Total | 754 | | 743 | | 637 | |

As can be observed, in the 260 years covered in this study, the variant *el que* has gained ground at the expense of the form without the article. Additionally, a lesser advance can be observed for the variant with the compound relative (*el cual*), rising from figures that were practically anecdotal in the eighteenth and nineteenth centuries (3%) to a slightly higher proportion in the first half of the twentieth (12%). Nevertheless, in order to facilitate the multivariate analysis, we collapse these forms with those of the majority variant, and contrasted them with the one we are dealing with here, *el que*.

Despite the progression of *el que*, it can nevertheless be considered moderate, as can be seen more graphically in the Figure 1 that follows:

So much for the global uses of the variants, but what happens inside the grammar and in the socio-stylistic matrix of the variation? To be able to find an

**Figure 1:** Evolution of the use of variants in prepositional relative clauses between the eighteenth and twentieth centuries (%).

answer to this question, we performed three independent multivariate analyses, the results of which are summarized below in subsections 6.1 to 6.3, where we review the data corresponding to the eighteenth, nineteenth and twentieth centuries, respectively.

6.1 Eighteenth century

Table 3 shows the results of the multivariate analysis performed with the data from the eighteenth century, where the form *el que* reached overall

Table 3: The effect of linguistic factors on the selection of *el que* in prepositional relative clauses in the eighteenth century.

| | Total <i>N</i> | % | P. (antecedent) | P. (writer) |
|--------------------------------|----------------|----|-----------------|-------------|
| Type of relative | | | | |
| Non-defining | 551 | 21 | 0.64 | 0.65 |
| Defining | 203 | 4 | 0.36 | 0.35 |
| <i>Range</i> | | | 28 | 30 |
| Content of the relative | | | | |
| Locative | 400 | 26 | 0.64 | 0.65 |
| Other | 354 | 7 | 0.36 | 0.34 |
| <i>Range</i> | | | 28 | 31 |
| Distance (antecedent) | | | | |
| Distant | 60 | 37 | 0.63 | 0.65 |
| Adjacent | 258 | 6 | 0.33 | 0.32 |
| Other | 436 | 20 | 0.54 | 0.54 |
| <i>Range</i> | | | 30 | 32 |
| Degree of determinacy | | | | |
| Determinate | 632 | 18 | 0.58 | 0.59 |
| Non-determinate | 122 | 11 | 0.42 | 0.41 |
| <i>Range</i> | | | 16 | 18 |
| Verb tense/mode | | | | |
| Future ind. | 31 | 35 | 0.67 | 0.68 |
| Present ind. | 395 | 18 | 0.50 | 0.51 |
| Past ind. | 249 | 17 | 0.46 | 0.46 |
| Other | 194 | 11 | 0.37 | 0.35 |
| <i>Range</i> | | | 30 | 33 |

With the antecedent: deviance –576.904; AIC (Akaike’s information criterion): 596.904; input: 0.17.

With the Writers: deviance: –460.454; AIC (Akaike’s information criterion): 596.086; input: 0.17.

percentages of 17%.¹⁰ For the sake of clarity, only the results for linguistic factors are shown, those of an extralinguistic nature being offered in a later subsection (see section 6.1.2). It should be noted, however, that all data come from the same run of the logistic regression analysis.

In the following, we examine these conditioning factor groups, as well as the way in which they were coded in this research, by means of several representative examples.

6.1.1 Linguistic factors

The variation was strongly conditioned in the eighteenth century by the *type of relative clause*. Thus, the variant *el que* is clearly favoured in appositive relative clauses (21%; 0.64), such as those in (5), which form their own phonic group, since their function does not lie in delimiting the semantic extension of the antecedent, but instead only in providing additional information about it (Brucart 1999: 408–417; de Vries 2002: 181–195). This does not mean – as can sometimes be intuited in previous literature – that the variant is predominant in this context; rather, we observed quite the opposite: 79% of the appositive clauses are still constructed without *el que*. What the logistic regression analysis offers, then, is a relative interpretation of the factor: within the minority condition occupied by the *el que* variant in this century, it is especially favoured in the syntactic context of appositive relative clauses, a long way from the group of restrictive clauses, such as those in (6), of which only 4% (0.36) are constructed with this form¹¹:

- (5) ... paseo grandemente por las tardes, saliendo al anochecer [...] andando hora y media o dos horas por las calles, *en las que* siempre encuentro muchas señoras llenas de misericordia... (*Correspondencia extraoficial de Ignacio de Heredia con Manuel de Roda*, 1773)

[... I take a long walk in the afternoon, going out a dusk ... walking for an

¹⁰ These figures are a little higher than those reported by Girón (2006a) (10%) and Guzmán Riverón (2012) (9%) in (mostly) Spanish texts in the first case and American ones in the second. Girón, however, takes the combinations with all the prepositions into account in these calculations, although he points out that there are no differences between them (p. 1528).

¹¹ For the sake of clarity and owing to space restraints, from here on we will only give the P. values obtained for each factor when the *Antecedent* has been taken as a random variable. We will only refer to those of the variable *Writers* when it is necessary to complete the profile in some particular factor. It should be noted, however, that the values of both are nearly always very similar.

hour and a half or two hours through the streets, *in which* I always find many ladies full of compassion...]

- (6) ... Sea como fuere estos hermanos de Ángel Carvallo apreciarían sobre manera que vm. procure liquidar quanto antes lo que le corresponde en lo que dejó su hermano y remitirlo a Cádiz a don Eduardo y a don Jacobo Gough de aquel comercio *en el que* son matriculados... (*Correspondencia canario-americana*, 1795)

[However, it may be these brothers of Ángel Carvallo would be greatly appreciative if you made efforts to pay off as soon as possible the part corresponding to you of that left by your brother and send it to Cadiz to Don Eduardo and to Don Jacobo Gough from that business *in which* they are registered...]

A second strongly conditioning parameter is the *semantic nature* expressed by the relative clause. Within the corpus, relative clauses present several semantic values, which we have divided into three groups. Hence, fragments such as those in (7) display a clear temporal sense, unlike (8), where this meaning is clearly locative. Last, we have created a third group with utterances that exhibit other more scattered meanings, such as modal, causal, consecutive, etc. Example (9) is a representative sample of this last group.

- (7) ... fuimos al navío, donde nos habíamos de embarcar, en donde estuvimos tres días, *en los que* yo no comí tan sólo un bocado (*Vivir y morir en México*, 1748)

[... we went to the ship, where we were to embark, where we stayed for three days, *in which* I did not eat even a morsel]

- (8) Si llegava a salvamento una fragata Americana *en la que* tengo echo embarque de algunos frutos... (*Die private Korrespondenz spanischer Emigranten aus Amerika*, 1791)

[If to our rescue came an American frigate *in which* I have arranged for shipment of some fruits]

- (9) ... cuio exemplo deve consolar a Vm y unir su conformidad con aquella para tolerar con paciencia los naturales sentimientos de la pérdida de un padre *en la que* acompañamos a Vm (*Die private Korrespondenz spanischer Emigranten aus Amerika*, 1775)

[... whose example should comfort you and unite your conformity with that of tolerating with patience the natural feelings on the loss of a father *for which* we ask you to accept our deepest condolences]

A preliminary analysis of frequencies indicates that there is a clear over-representation of the locative values ($N = 400$) in comparison to the others. In addition, the proportions of *el que* in those other semantic contexts are very similar to each other (close to 6–7%). These percentages are clearly below the overall average, and a considerable distance behind the locative values, which are above it (26%). In the regression analysis, the contrast between the latter and the other two (taken together for the multivariate analysis under the title of ‘Others’) is selected as significant: *el que* has the locative semantic contexts as one of its main points of penetration into the system (0.64) unlike the other meanings (0.36).

The presence of the *definite* article¹² in the antecedent and the *distance*¹³ from it with respect to the relative pronoun also appear as conditioning factor groups, but they act in tandem rather than independently. With regard to the *distance*, Bello (1988 [1847]: § 1076) noted that the distancing of the antecedent favours the presence of what he interpreted as a compound relative to mark the agreement in a clearer manner. Girón (2006a: 1530), moreover, also claimed that in the eighteenth century, distant positions favour the presence of the CR, as well as other factors like the expansion of the antecedent, which, in practice, means insisting on the distancing, since it implies the spacing out of the placement of the relative pronoun with respect to its antecedent (in the same line, see Cortés Rodríguez 1986: 19; Porto Dapena, 1997: 24–25; Santana Marrero 2004: 71; Guzmán Riverón 2012: 187).

Our data, however, suggest an interaction between this factor and the presence of the article in the antecedent.¹⁴ Hence, the adjacency of the antecedent is significant (and negative) when it appears with the definite article. This context, as exemplified in (10), especially disfavours the variant *el que* (6%; 0.33). Conversely, the presence of the article in distant antecedents, such as in (11), raises the number of occurrences of the variant to an extent that is

12 Through this factor group, we evaluated a potential cognitive constraint, the priming effect exerted by the presence of the same article in the antecedent (*el* barco en (el) que fuimos [the boat in which we were ...]). As a preliminary hypothesis, it could be assumed that such presence will favor the selection of the application value (on the psycholinguistic effect of this structural priming, see Pickering and Ferreira 2008).

13 Initially, we encode all tokens of the variable in three contexts. The *Adjacent* ones consist of those cases in which the antecedent is placed next to the relative (*la carta en (la) que me dices...* [the letter in (the) that you tell me that ...]). The *Intermediate* contexts are those in which an adjectival or a prepositional phrase is interposed between both categories (*la carta del banco en (la) que...* [the letter from the bank in which ...]). Finally, in the group of *Distant* contexts, we include the remaining tokens, in which the categories of the antecedent and the relative pronoun are separated by an even greater linguistic material.

14 The combination of these two factor groups results in the following three contexts: (1) Distant + article; (2) Adjacent + article; (3) All the remaining combinations (“Other” in coding).

unusual in this period (37%; 0.63). The other combinations stand somewhere in between – an example can be found in (12) – which exerts an influence that is close to neutrality (20%; 0.54).

- (10) ... había hecho una gran ramada toda colgada y cubierta de damascos donde tenía puestas **las mesas** *en las que* se sirvió una abundante y esquisita comida (*Un epistolario virreinal*, 1790)
[... had made a large shelter from overhanging apricot tree branches where they had set **the tables** *on which* they served a copious and exquisite meal]
- (11) ... acabó con mucho lucimiento **el curso** de artes en la ciudad de Querétaro, *en el que* sacó el primer lugar (Al recibo de esta, 1771)
[... finished outstandingly well **the course** in arts in the city of Querétaro, *in which* he came first]
- (12) ... con el motiuo de auer venido de dicha Hauana estos dos **nauíos** de guerra, nombrados *El Fuerte* y *La Galga*, *en los que* remito quinientos pesos a entregar a D. Miguel de Lauaqui (*Correspondencia de emigrantes guipuzcoanos y navarros*, 1749)
[... on the occasion of yesterday's arrival from Havana these two **warships**, called *El Fuerte* and *La Galga*, *in which* I send five hundred pesos to be handed over to Don Miguel de Lauaqui]

In short, it appears that copying the article preferably takes place when the antecedent is sufficiently far away so that, cognitively, the speaker feels the need to retrieve it. In contrast, when the antecedent and the relative pronoun are together, that pressure not only does not exist, but, on the contrary, becomes an important inhibiting constraint.

Another of the factors selected by the regression analysis concerns the degree of determinacy of the antecedent. Thus, in examples like the one in (13), we can see how the noun *cartas* [letters] appears without any kind of determination whatsoever, which is quite the opposite to *aviso* [notification] in (14), where the noun is preceded by an indefinite determiner, and *compañía* [company] in (15), which is preceded by a definite determiner:

- (13) ...En este mismo correo he recibido **cartas** de Espafia *en las que* me solicittan para la defensa de vn maiorazgo y barrios vínculos (*Cartas desde la otra orilla*, 1790)
[... In this same post I have received **letters** from Spain *in which* I am requested for the defence of an entailed estate and related districts]

- (14) Ahora tengo noticia que a llegado **un aviso** a Veracruz en el que é rreziuido varias cartas (*Correspondencia de emigrantes guipuzcoanos y navarros*, 1732)

[I now have news that **a notification** has arrived in Veracruz *in which* I have received several letters]

- (15) ... mientras se berifiquen las cobranzas de sus haveres, que son de alguna consideración, podrá estar bien hallado en **mi compañía**, *en la que* creo sacará algún partido decente (*Die private Korrespondenz spanischer Emigranten aus Amerika*, 1788)

[... while the amounts received for your belongings, which are quite considerable, are verified, you would be welcome to partake of my **company**, *in which* I think you will enjoy some benefit]

Indeed, the factor is seen to be significant, although with one of the lowest explanatory ranges, and in a manner slightly different from what was initially presumed. In any case, the two regression analyses coincide in highlighting the fact that the presence of *el que* is favoured in determinate settings (18%; 0.58) – independently of their character, being definite (18%) and indefinite contexts (17%) – and, conversely, disfavoured in non-determinate ones (11%; 0.42).

Finally, a significant correlation can also be observed with the verb tense in the relative clause. As we have seen in other phenomena of variation and change in the history of Spanish (Blas Arroyo and Vellón 2015), a restricted context like the one represented by the future indicative represents a point of entry for a novel variant such as *el que* (35%, 0.67), in contrast to what happens with the subjunctive forms and, in general, the less frequent conjugation paradigms, which have a clearly more conservative influence (11%; 0.37). On the other hand, the majority contexts, represented by the present indicative (18%; 0.50) and the past tense forms (17%; 0.46), stand in positions close to neutrality.

6.1.2 Extralinguistic factors

Of all the extralinguistic factors taken into consideration, only the diachronic axis is selected as significant independently (see Table 4).¹⁵

¹⁵ As can be observed in Table 4, in some of the extralinguistic factors considered in this table, there are fewer occurrences of the variable. This is due to the lack of biographical information about some of the writers, a fact that is especially apparent in variables such as their dialectal origin and age.

Table 4: The effect of extralinguistic factors on the selection of *el que* in prepositional relative clauses in the eighteenth century.

| | <i>N</i> | % | P. (A) | P. (E) |
|--------------------------|----------|----|-----------|-----------|
| Sex | | | | |
| Males | 725 | 17 | [] | [] |
| Females | 29 | 7 | [] | [] |
| <i>Range</i> | | | | |
| Social status | | | | |
| High | 232 | 12 | [] | [] |
| Middle | 281 | 17 | [] | [] |
| Low | 207 | 21 | [] | [] |
| <i>Range</i> | | | | |
| Tenor | | | | |
| Distant | 316 | 16 | [] | [] |
| Familial/private | 350 | 19 | [] | [] |
| <i>Range</i> | | | | |
| Migration context | | | | |
| Yes | 564 | 18 | [] | [] |
| No | 190 | 12 | [] | [] |
| <i>Range</i> | | | | |
| Age | | | | |
| Youngers | 169 | 22 | [] | [] |
| Adults | 353 | 15 | [] | [] |
| <i>Range</i> | | | | |
| Region | | | | |
| Northern | 397 | 19 | [] | [] |
| Southern | 192 | 19 | [] | [] |
| <i>Range</i> | | | | |
| Decade | | | | |
| continuous logodds | | | + 1 0.118 | + 1 0.138 |

With regard to this continuous factor group, an in-depth examination of the frequencies, grouped by decades, like the one shown in Figure 2, enables us to observe the presence of different moments in the diffusion of the innovative variant. At the beginning of the century, the variant is in very limited usage (3%), in proportions that are similar to those observed in the seventeenth century (1,3%).¹⁶ In the second and third decades, however, an upward movement can be

¹⁶ These data come from the counting of tokens made on two corpus of communicative immediacy similar to the one handled in these pages for the sixteenth (0,5%) and seventeenth (1,3%) centuries (see Blas Arroyo and Vellón 2017).

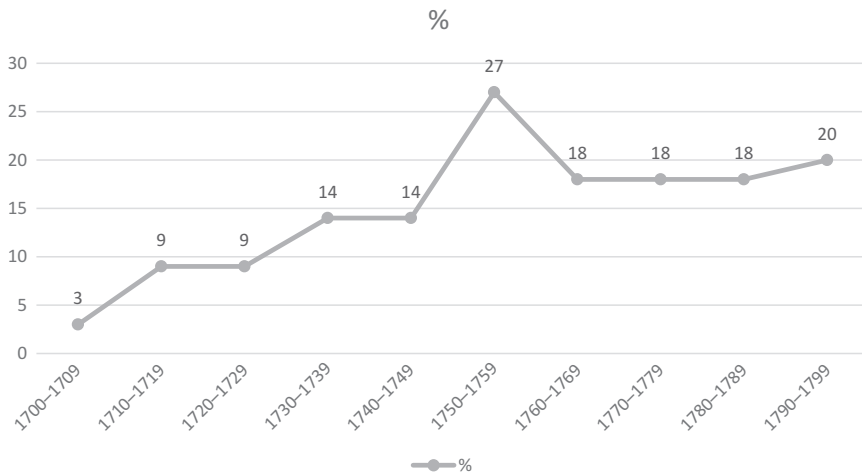


Figure 2: Evolution of the use of *el que* by decades during the eighteenth century (%).

seen (9%), which was to increase at the beginning of the second third of the century (14%), and rose sharply at the end of that period (27%).¹⁷ Yet, from that moment on, the advance of the phenomenon slowed down, with more stable distributions at the end of the century (around 20%).

Real-time studies of historical sociolinguistics help to evaluate the degree of progression and diffusion of linguistic changes. In our case, Figure 2 shows that after the early stages of the century, in which we seem to witness an *incipient* phase of change (below 15%), this is accelerated sharply around the year 1750, when its frequencies were increased twofold with respect to the figures from the previous decade. In this second phase, we reach the already characteristic *S-shaped curve of the new and vigorous changes* described in the sociolinguistic literature, and which spread forcefully within the social matrix until reaching quite considerable figures (between 15% and 35%) (Labov 1994: 79–83; Nevalainen and Raumolin-Brunberg 1996: 55). Somewhat unsurprisingly, then, the meagre figures of the beginning of the century rise to almost 30% in a space of barely 50–60 years. Yet, unlike other processes of change of this same kind reported in diachronic linguistics, and which have eventually become firmly established in a relatively short period of time (Martínez 2001),

¹⁷ Girón (2006a: 1528) also draws attention to an increase in *el que* towards the middle of the century.

in our case, the novel variant seems to lose momentum in the late eighteenth century, although with figures that clearly exceed those of the beginning of the century.¹⁸

No differences are found between the different groups in terms of sex,¹⁹ migratory context or the writers' birthplaces, either independently or interacting with others. And the same is true of the tenor of the writings, measured on the basis of the distance between the interlocutors and/or the topic addressed in the texts.

Of greater interest, however, are two other social constraints, namely status and age, among other reasons because their still-timid differences in terms of frequency point in a direction that was to become firmly established in later centuries. In the case of age, and as shown by Figure 3, the behaviour of the two groups under consideration is quite stable throughout the century: in all the periods into which the time axis was divided, except one, the younger generations are ahead of the older adults. Besides, since these latter usually react later

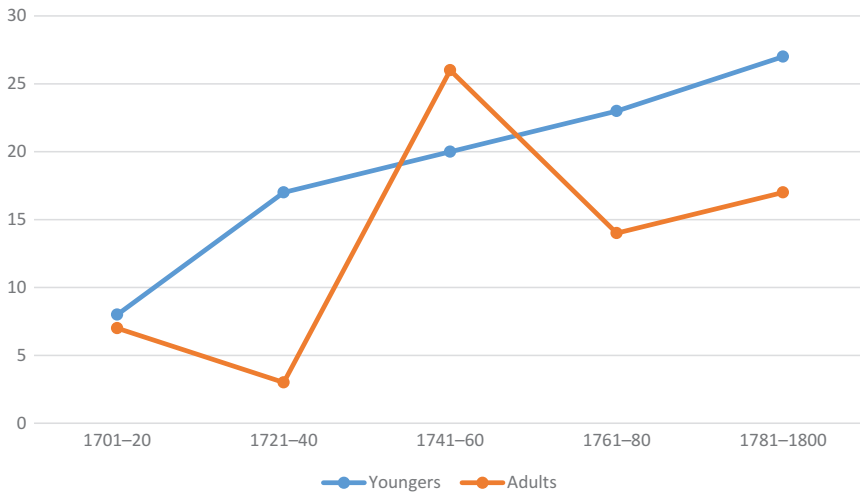


Figure 3: Distribution of *el que* by age groups in the eighteenth century (data grouped by periods of two decades (%)).

¹⁸ This fact has also been highlighted by Girón (2006a: 1572), who speaks of a possible reaction by the academicist, on the basis of classical texts from the sixteenth century, upon request by the recently created Real Academia Española.

¹⁹ Nevertheless, the extraordinary irregularity of the samples in terms of male and female speech does not allow us to go any further into our conclusions.

to changes affecting linguistic variables (Cukor-Avila and Bailey 2013; Tagliamonte 2012), we could suppose that there were already some differences in the sociolinguistic prestige of the variants in that century.

The sociolectal distribution that is observed also contributes to this impression. Although the differences among groups are not strong enough to be validated statistically, in the distribution of the novel variant there is a characteristic linear distribution, in which its usage appears to be associated to the writers' status: low (21%), intermediate (17%) and high (12%).

Moreover, crossing with age reveals some interesting results that suggest a possible change from below (Labov 2007: 346),²⁰ encouraged by the lower classes and, within these, especially by the young. As can be seen in Figure 4 the vernacular variant is led by the subordinate classes, and more particularly by the young (26%). In addition, this change seems to be accepted by the younger members of the upper and middle classes (18%) sooner than by older individuals (12%).

In sum, these results seem to indicate that, in the late eighteenth century, there may have been pressure opposing the variant with the article, a vernacular form that had awoken from a long dormant state in the first half of the century among the less privileged sectors and the younger generations. Hence, this

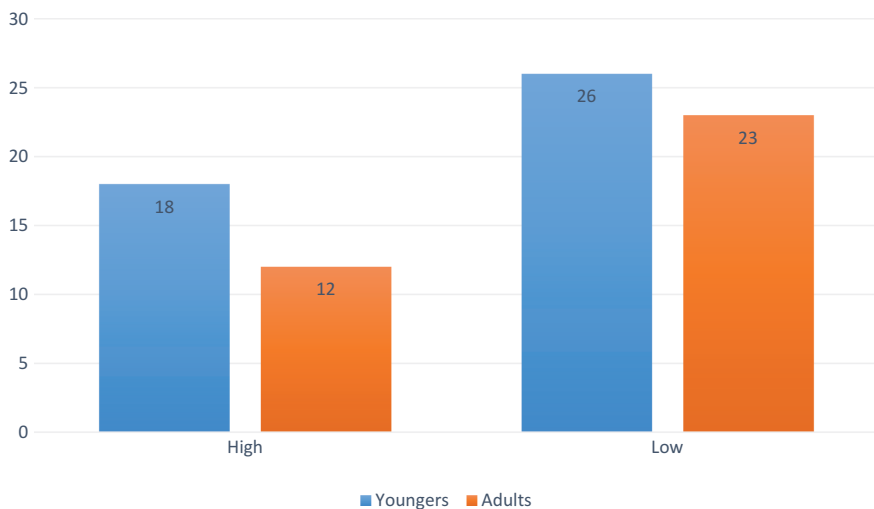


Figure 4: Distribution of *el que* by status and age groups in the eighteenth century (%).

²⁰ In order to facilitate the calculations, here we group the values of the upper and middle strata and we contrast them with those of the lower class.

change from below was to be halted later on by an opposing movement – from above – that favoured a return to the old orthodoxy of the pronoun without an article. And like many changes of this type, it seems to have been led by the upper classes (Labov 2007: 347; Tagliamonte 2012: 55ss.), in a pendulous movement that would end up affecting the whole social spectrum, but to a lesser extent among the more popular classes, whose members (above all the youngest) were to remain connected to an innovative variant that they had helped to disseminate.

6.2 Nineteenth century

6.2.1 Linguistic factors

In the nineteenth century (see Table 5), the difference between *appositive* and *restrictive* subordinate clauses was still the most relevant factor group of all those considered in the analysis and, furthermore, with the same explanatory direction as that seen in the previous period. Thus, the sentences that encourage the pronominal variants most are, by far, the appositive ones (43%; 0.77). In contrast, restrictive relative clauses continue to hinder the advance of the variant *el que* (3%; 23). On comparing the magnitudes of each period, it can be seen how the relevance of the factor increases with respect to the eighteenth century. Accordingly, appositive clauses have twice the frequency of the variant (eighteenth century: 21%; nineteenth century: 43%), whereas restrictive clauses hardly change (eighteenth century: 4%; nineteenth century: 3%), which take the factor weights values to the extreme (Non-defining: 0.77; Defining: 23) with respect to the century of Enlightenment (Non-defining: 0.64; Defining: 0.36).

This is not the case in relation to the semantics of the antecedent. Although the factor group continues to be selected as one of the most relevant to explain variability, its range is now almost identical to that of the previous period. As at that time, in the nineteenth century, the locative sense of the subordinate clause favours the appearance of *el que* (36%; 0.64), in contrast to the other values (modal, temporal, etc.), which clearly discourage this form (9%; 0.36).

Another element of continuity with respect to the past can be seen in the particular association of the variant with certain verbal paradigms. Thus, the future indicative in the subordinate clause continues to be the context that is most closely linked to the presence of *el que* (47%; 0.73), whereas the subjunctive and other less frequent forms in discourse inhibit it to the same extent (11%;

Table 5: The effect of linguistic factors on the selection of *el que* in prepositional relative clauses in the nineteenth century.

| | <i>N</i> | % | <i>P. (A.)</i> | <i>P. (E)</i> |
|--------------------------------|----------|----|----------------|---------------|
| Type of relative | | | | |
| Non-defining | 393 | 43 | 0.77 | 0.80 |
| Defining | 341 | 3 | 0.23 | 0.20 |
| <i>Range</i> | | | 54 | 60 |
| Content of the relative | | | | |
| Locative | 414 | 36 | 0.64 | 0.69 |
| Other | 320 | 9 | 0.36 | 0.31 |
| <i>Range</i> | | | 28 | 28 |
| Distance + Article | | | | |
| Adjacent-Yes | 295 | 7 | 0.34 | 0.35 |
| Adjacent-No | 205 | 23 | 0.51 | 0.49 |
| Other | 234 | 46 | 0.65 | 0.65 |
| <i>Range</i> | | | 31 | 30 |
| Verb tense/mode | | | | |
| Future ind. | 19 | 47 | 0.73 | [] |
| Present ind. | 350 | 22 | 0.41 | [] |
| Past ind. | 278 | 29 | 0.52 | [] |
| Other | 87 | 11 | 0.33 | [] |
| <i>Range</i> | | | | |
| Sujeto | | | | |
| Null | 488 | 28 | 0.59 | 0.60 |
| Explicit | 246 | 16 | 0.41 | 0.40 |
| <i>Range</i> | | | 18 | 20 |

With the Antecedent: deviance: -263.21; AIC (Akaike's information criterion): 552.42; input: 0.24.

With the Writers: deviance: -254.418; AIC (Akaike's information criterion): 528.835; input: 0.24.

0.33). Finally, the most recurring tenses, such as the present (22%; 0.41) and the past indicative forms (29%; 0.52), again stand in intermediate positions.

The interaction between the distance from the antecedent and the presence of the definite article within it is also selected as explanatory, but now submitted to a partial restructuring of the contexts. The adjacency of an antecedent with an article continues to exert an important disfavoured effect (7%; 0.34), unlike those same adjacent settings without an article, whose contribution to the variant offers figures close to neutrality (23%; 0.51). At the opposite extreme, however, there is no longer any difference between the distant contexts according to whether the antecedent is preceded or not by

the article: in both cases, the factor raises the number of appearances of the application value to levels that are much higher than the average for the period (46 %; 0.65).

The two main novelties with respect to the eighteenth century lie in the loss of significance of the type of determination – although the differences in frequency point in a direction that will become firmly established in the twentieth century (see 6.3) – and the appearance of a new conditioning factor group that was not operative in the past. This novel factor is the syntax of the subject of the subordinate clause, where there are two possibilities: elision, the predominant option in a null-subject language like Spanish ($n = 488$), as in (16), and expression ($n = 246$), as in (17).

- (16) ... estuve anoche en casa de Pedro Gil donde me dieron una carta tuya, *en la que* Ø me dices estás buena y en Buñol (*Cartas de Sorolla a su mujer*, 1895) [... last night I was in Pedro Gil's house where they gave me a letter of yours, *in which* you tell me that you are well and in Buñol]
- (17) No quiero detenerme en decirte el sentimiento que nos causó su temprana muerte, y mucho más en un tiempo *en el que* florecían **todos sus negocios** (*Cartas de emigrantes escritas desde Cuba*, 1817)
[I do not wish to dwell on telling you how we felt on hearing about his early demise, and even more so at a time *in which* **all his business undertakings** were flourishing]

Whether due to the predominant nature of the former or to the greater cognitive simplicity of elided contexts, the truth is that, in the nineteenth century, the omission of the subject is favourably associated with the forms of the relative with the article (28 %; 0.59), in contrast to the greater restrictions imposed by the alternative context (16 %; 0.41). Nevertheless, it is the least relevant of all the factors considered in this period, as highlighted by its scarce range (18), together with the fact that its days were numbered, as it was to disappear from the chart of explanatory factors in the next century.

6.2.2 Extralinguistic factors

As occurred in the previous period, the data from the nineteenth century show the time axis as being significant, which is proof that the change was following its course. Nevertheless, its distribution throughout the century is very irregular,

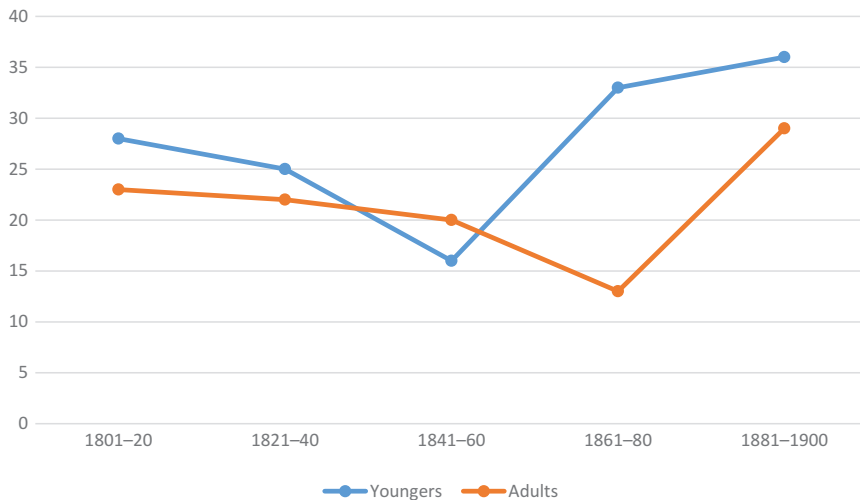


Figure 5: Distribution of *el que* by age groups in the nineteenth century (data grouped by periods of two decades (%)).

and only movements that are decidedly favourable to the increased use of the innovative variant are observed in the last two decades, when it was to reach figures above the average (31%).

Crossing it with the generational factor, as shown in Figure 5, allows us to observe other interesting facts. One of them is the curvilinear distribution that can be seen in the cut-off points for age into which we have classified the members of the sample. This distribution accounts for a non-linear extension of the variant, perhaps reflecting the existence of oscillations in its sociolinguistic prestige with respect to the traditional *en que*. Moreover, it can also be seen again how, except for one time cut-off point, in all the others the younger cohorts exceed the more adult ones in selecting the novel variant, just as we observed in the previous century. Nevertheless, it is also telling that, in the last period, the lines of both groups point in the same direction.

Meanwhile, the sociolectal factor not only continues the tendencies of use that were already intuited in the past, but also consolidates them to a notable degree. As can be seen in Table 6, the variable is now selected as significant by the mixed-effect regression analysis, and again shows a linear distribution that is clearly associated to the status of the writers. Hence, members of the social elite seem to be the most reluctant to use the novel variant (14%; 0.27), followed

Table 6: The effect of extralinguistic factors on the selection of *el que* in oblique relative clauses in the nineteenth century.

| | <i>N</i> | % | P. (A) | P. (E) |
|--------------------------|----------|----|-----------|-----------|
| Sex | | | | |
| Males | 598 | 26 | [] | [] |
| Females | 136 | 16 | [] | [] |
| <i>Range</i> | | | | |
| Social status | | | | |
| High | 213 | 14 | 0.27 | 0.20 |
| Middle | 408 | 25 | 0.55 | 0.57 |
| Low | 113 | 39 | 0.69 | 0.75 |
| <i>Range</i> | | | 38 | 55 |
| Tenor | | | | |
| Distant | 204 | 16 | [] | [] |
| Familial/Private | 262 | 24 | [] | [] |
| <i>Range</i> | | | | |
| Migration context | | | | |
| Yes | 162 | 29 | [] | [] |
| No | 572 | 23 | [] | [] |
| <i>Range</i> | | | | |
| Age | | | | |
| Youngers | 271 | 28 | [] | [] |
| Adult | 426 | 23 | [] | [] |
| <i>Range</i> | | | | |
| Region | | | | |
| Northern | 407 | 27 | [] | [] |
| Southern | 235 | 23 | [] | [] |
| <i>Range</i> | | | | |
| Decade | | | | |
| continuous logodds | | | + 1 0.175 | + 1 0.177 |

by the intermediate strata (25%; 0.55). In contrast, *el que* spreads predominantly among the subordinate classes (39%; 0.69), a fact that was already noted in the previous century, but now in a much clearer and stronger way. At the same time, crossing it with age, as illustrated in the Figure 6, shows how the movement favouring the novel variant is especially intense among the younger members of the popular classes (45%).

Finally, attention must be drawn to the non-selection of the other factors, which were already discarded in the eighteenth century.

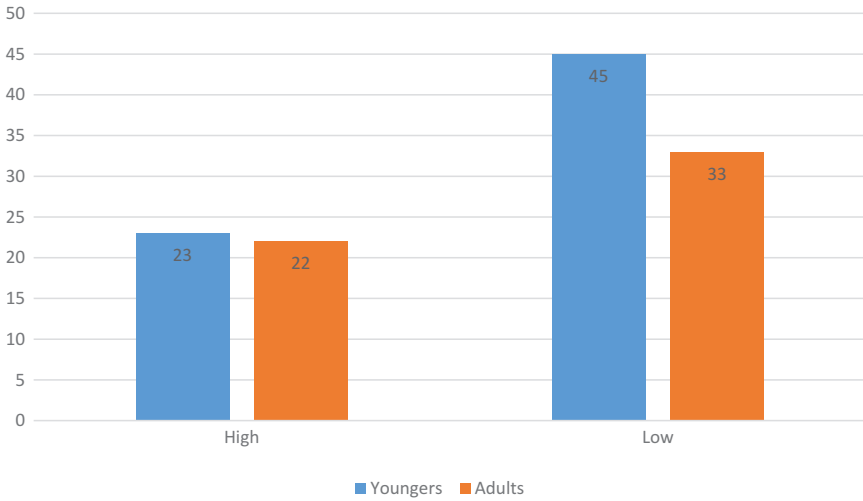


Figure 6: Distribution of *el que* by status and age groups in the nineteenth century (%).

6.3 Twentieth century

In the first six decades of the twentieth century, some conditioning factors show signs of being very strong and continue to determine the variation and, therefore, the possibilities of change. Yet, some things are starting to change. On exploring many of the contexts under consideration, the variable acquires new constraints that were not operative up until that moment (at least not with enough strength), while others disappear or lose force.

6.3.1 Linguistic factors

Some of the factors that in the past played a significant role in explaining the variation continue to do so now, and in the same explanatory direction (see Table 7). Hence, there is still a clear advance of *el que* among appositive relative clauses, which continue to gain popularity until they finally account for practically half the choices of the variable (46%). Nevertheless, the probabilistic value of this context drops with respect to the previous century (0.68) as the variant advances, albeit only moderately, among restrictive relative clauses. Although the frequency indices in this case are still unremarkable in absolute terms (10%; 0.32), they increase significantly with respect to previous centuries, in which

Table 7: The effect of linguistic factors on the selection of *el que* in prepositional relative clauses in the twentieth century.

| | <i>N</i> | % | P. (A.) | P. (E) |
|--------------------------------|----------|----|---------|--------|
| Type of relative | | | | |
| Non-defining | 338 | 46 | 0.68 | 0.69 |
| Defining | 299 | 10 | 0.32 | 0.31 |
| <i>Range</i> | | | 36 | 38 |
| Content of the relative | | | | |
| Locative | 334 | 45 | 0.66 | 0.67 |
| Other | 303 | 12 | 0.34 | 0.33 |
| <i>Range</i> | | | 32 | 34 |
| Article | | | | |
| Yes | 299 | 15 | 0.38 | 0.39 |
| No | 338 | 41 | 0.62 | 0.61 |
| <i>Range</i> | | | 24 | 22 |
| Degree of determinacy | | | | |
| Non-determinate | 71 | 25 | 0.49 | 0.46 |
| Indefinite | 164 | 47 | 0.62 | 0.64 |
| Definite | 402 | 22 | 0.38 | 0.40 |
| <i>Range</i> | | | 24 | 24 |
| Gender | | | | |
| Feminine | 326 | 40 | 0.58 | 0.59 |
| Masculine | 311 | 18 | 0.42 | 0.41 |
| <i>Range</i> | | | 16 | 18 |
| Function (antecedent) | | | | |
| CC | 180 | 16 | 0.40 | [] |
| Other | 457 | 34 | 0.59 | [] |
| <i>Range</i> | | | 19 | |
| Function (relative) | | | | |
| CC | 620 | 29 | 0.33 | 0.31 |
| Other | 17 | 47 | 0.67 | 0.68 |
| <i>Range</i> | | | 34 | 37 |

With the Antecedent: deviance: -285.697; AIC (Akaike's information criterion): 587.623; input: 0.29. With the Writers: deviance: -284.377; AIC (Akaike's information criterion): 584.754; input: 0.29

their presence was practically anecdotal (around 3%). Likewise, there is also a persistence over time with respect to the semantics of the relative clause, and thus the novel variant continues to progress at a good pace among locative sentences (45%; 0.66), at the expense of the other meanings (12%; 0.34), where it advances to a lesser extent.

Together with these signs of continuity and persistence, however, the variable starts to encounter new structural constraints in its slow flow across the time axis. Some of them are the result of a restructuring of others that worked in the past. Thus, the influence of the article with the antecedent becomes firmly established as an independent factor for the first time in this century, without any interaction with others. If in the previous centuries the presence of the article only restricted to any significant extent the selection of *el que* in the adjacent contexts, it now does so in every position. As a result, the appearance of the article becomes firmly established as an unfavourable setting for its repetition in the relative pronoun (15%; 0.38). In contrast, in the contexts in which the article does not appear in the antecedent – and regardless of the position of the latter – there is a clear progression of the variant (41%; 0.62).

In other cases, the new conditioning factors consolidate differences that had already been noted in the past, but were without sufficient strength, at least in the corpus handled in this study. For instance, in the nineteenth century, a visible difference began to stand out amongst the definite and indefinite contexts in relation to the antecedent. Whereas the innovative variant hardly increased in the former contexts, it did so considerably among the latter ones. This movement within Spanish grammar therefore grew stronger at the turn of the twentieth century, to the point where it reveals itself as a new conditioning factor (definite: 22%; 0.38; indefinite: 47%; 0.62).²¹

In the same line, we could speak of the *gender* and the *function* of the antecedent, factors that add themselves in this century to the list of those that significantly constraint the variation. With regard to the first, in previous centuries we have already observed the particular affinity of the combinations of *el que* for feminine antecedents, but only now do those differences (40%; 0.58) reach sufficient explanatory power with respect to the masculine ones (18%; 0.42). This does not change even when we remove from the analysis the antecedent *carta*, which is by far the most frequent in the corpus. This reanalysis yields practically identical results, thus confirming the relevance of gender as an independent factor.

Regarding the function of the antecedent in its own clause, in the nineteenth century, we observed a small difference in favour of the solutions with *el que* among the non-adverbial functions, such as those in (18), although this difference now becomes considerably larger (34%; 0.59). Conversely, the function as an adverbial, as in (19), continues to include a number of pitfalls hindering the

²¹ The non-determined ones, in minority in the corpus with respect to the other two ($n = 71$), maintain positions close to neutrality (25%; 0.49).

advance of the variant, with levels similar to those of two centuries before (16%; 0.40).

- (18) Y otra vez **un bombardeo terrible** *en el que* he pasado más miedo que en el primero (*Epistolario amoroso*, 1939)
[And again **a terrible bombardment** *in which* I felt even more afraid than in the first]
- (19) Estamos **en unas horas difíciles** *en las que* necesito no perder (*Una pasión imposible*, 1940)
[We find ourselves **in difficult times** *in which* I need not to lose]

Finally, the same happens with the function played by the relative in its own clause: the adverbial function, as in (20) is hostile towards the innovative solutions (29%; 0.33), unlike other syntactic functions, as in (21) (47%; 0.67).

- (20) ... en la primera ocasión *en la que* nos vimos todos los compañeros les di lectura de tus líneas (*Correspondencia de Eugenio Granell con sus camaradas del POUM*, 1945)
[On the first occasion in which all the comrades saw each other, I read your letter to them]
- (21) El otro es un amigo de París *en el que* no tengo tanta confianza (*Correspondencia de Eugenio Granell con sus camaradas del POUM*, 1939)
[The other one is a friend from Paris who I do not trust so much]

3.3.2 Extralinguistic factors

At least in the period analysed in this paper, which only covers the first six decades of the twentieth century, the variable seems to have entered into a stage of certain stability, as can be seen from different results shown in Table 8.

As can be observed, this time none of the extralinguistic factors are selected as significant by the regression program. In this respect, it is telling that, in contrast to what occurred in the previous centuries, in this period the time axis does not leave any mark at all. The same can be said of the complete neutralization of the generational differences, which disappear completely in the twentieth century.

Table 8: The effect of extralinguistic factors on the selection of *el que* in prepositional relative clauses in the twentieth century.

| | <i>N</i> | % | P. (A) | P.(E) |
|--------------------------|----------|----|--------|-------|
| Sex | | | | |
| Males | 583 | 29 | [] | [] |
| Females | 54 | 30 | [] | [] |
| <i>Range</i> | | | | |
| Social status | | | | |
| High | 294 | 26 | [] | [] |
| Middle | 201 | 28 | [] | [] |
| Low | 142 | 38 | [] | [] |
| <i>Range</i> | | | | |
| Tenor | | | | |
| Distant | 360 | 30 | [] | [] |
| Familial/Private | 257 | 29 | [] | [] |
| <i>Range</i> | | | | |
| Migration context | | | | |
| Yes | – | – | – | – |
| No | – | – | – | – |
| <i>Range</i> | | | | |
| Age | | | | |
| Youngers | 272 | 29 | [] | [] |
| Adults | 357 | 29 | [] | [] |
| <i>Range</i> | | | | |
| Region | | | | |
| Northern | 553 | 29 | [] | [] |
| Southern | 69 | 27 | [] | [] |
| <i>Range</i> | | | | |
| Decade | | | | |
| | | | [] | [] |

The distribution of frequencies on sociolectal terms also shows some marked differences in comparison with previous eras. While it is true that the novel variant appears more often in the lowest strata (38%), the gap between these and the middle (28%) and upper ones (26%) becomes notably narrower with respect to the previous century. Consequently, it appears that the variant extends across the whole social spectrum, including the upper classes, which were until then the most reluctant towards the extension of the form *el que*.

Finally, there are still no differences in the factors sex, communicative distance (*tenor*) and the writers' place of birth.²²

7 Discussion

In the foregoing, we have been able to confirm the advance of the forms *el que* in prepositional subordinate clauses between the early eighteenth century and the first half of the twentieth. Yet, this progression has taken place at a slow pace and, objectively, can only be described as moderate. Thus, at the beginning of the second half of the twentieth century, the forms of the relative pronoun without an article were still the majority variant, a situation that appears to have changed dramatically from then on, judging by the far higher frequencies found in other surveys that include more recent periods (Girón 2006: 1539 § 14.5.2.2; Blas Arroyo and Vellón 2017; Vellón and Moya 2017). In short, it seems that the change has become far more abrupt in recent decades, as has occurred with other syntactic variation phenomena, such as the extension of the modal periphrasis *tener que* at the expense of the traditional *haber de*, which we have analysed elsewhere (Blas Arroyo and Vellón 2015).

In any case, the advance of *el que* has not taken place only in global terms, but also in most of the contexts taken into account. Nevertheless, there are several significant exceptions to this tendency with some of the contexts that have been seen to be the most reluctant throughout history. Hence, both the adverbial function of the antecedent (eighteenth century: 16%; nineteenth century: 20%; twentieth century: 16%), its presence in an adjacent position (eighteenth century: 6%; nineteenth century: 7%; twentieth century: 9%), or with the definite article (eighteenth century: 16%; nineteenth century: 20%; twentieth century: 15%) have inhibited the selection of *el que* in very similar proportions in the three periods.

Together with these, the three independent mixed-effect regression analyses that were performed reveal a number of elements of continuity in history. On the one hand, there are a significant number of structural factors (e. g. number, category and type of referent to the antecedent, degree of agentivity, sentence modality, etc.) and several of the non-structural ones (e. g. sex, region, tenor) that are not selected in any of the centuries.

²² The variable *Migratory context* is not taken into account this time owing to the scarcity of samples of the variable produced by Spanish emigrants in America, a situation that contrasts with that observed in the previous centuries.

More telling are, however, a series of conditioning factors that are systematically selected and with the same explanatory direction in all periods. Of these, the most important is the type of relative clause. Thus, the extension of *el que* has taken place above all within appositive clauses. Between the eighteenth (21%) and the nineteenth centuries (43%), the frequency of the variant in these clauses increases twofold, a progression that continues in the first decades of the twentieth century (46%). In contrast, restrictive clauses continue to represent hostile territory for the variant, despite the timid advance made in the last century (eighteenth century: 4%; nineteenth century: 3%; twentieth century: 10%).

Variation and change are also related to the semantic content of the relative clauses, and thus *el que* is benefited in the three centuries by locative contexts (eighteenth century: 26%; nineteenth century: 36%; twentieth century: 45%), whereas in non-locative ones (time, modal, etc.) the increase is far more moderate (eighteenth century: 7%; nineteenth century: 9%; twentieth century: 12%).

Throughout long periods of time, some factors acted in tandem, as is the case of the distance of the antecedent with respect to the relative pronoun and the presence of the definite article within it. Neither of them are seen to make an independent contribution during the eighteenth and nineteenth centuries, but they do when the two interact. The result of that interaction oscillates, however, between different eras. Thus, if at all times the appearance of the article in an adjacent position represents one of the most adverse points for the innovative variant, this is not the case when it is in more distant positions. In the nineteenth century, the distancing of the antecedent favoured *el que* regardless of whether the article was present or not, but this was not so in the twentieth century, where the absence of the determiner led the frequencies of the variant to unusual extremes.

Likewise, as the use of *el que* expanded, new conditioning factors appeared in the grammar in the twentieth century. Yet, in most cases what occurred was a consolidation of differences that had already appeared in a more timid fashion in the past. This is the case, for example, of determinacy. So we see how the antecedents that favour the innovative variant to a greater extent in the twentieth century – but also in the nineteenth, although below the threshold of statistical significance – are the indefinite ones, with proportions (47%) that contrast dramatically with the definite ones (22%), especially the article, as we have seen. And the fact that these latter are far more frequent in discourse than the former (at least in the corpus analysed here) has surely contributed to the fact that the occurrences of the variant *el que* have not reached higher levels.

Furthermore, a secular attraction of this variant for feminine antecedents becomes firmly established as a significant factor in the twentieth century, when it doubles the number of cases with respect to the masculine antecedents (40%/18%). And the same happens in the functional plane: both in the antecedent and in the relative pronoun itself, *el que* is favoured in the non-adverbial functions, while the adverbial ones favour the variant without the article to a greater extent.

In contrast to the foregoing, two factors that were operative in the past are now no longer so. One of them is the tense of the verb in the relative clause. In both the eighteenth (35%) and nineteenth centuries (47%) *el que* is especially inclined to appear in relative clauses constructed with the verb in the future indicative. In contrast, at the other extreme, we find the subjunctive forms, that is, as an especially unfavourable setting for the variant (11%). Between the two points, we find the verb tenses most commonly used in discourse, such as the present and past indicative, with contributions that are closer to neutrality in both periods. In the twentieth century, however, the differences between these verbal contexts become considerably smaller: the future loses all the importance it had in the past, while the frequencies of the subjunctive forms catch up with the rest to the point where they become almost equal.

With regard to the social factors, we have already pointed out that some show no differences – let alone any significant ones – in any of the centuries. Neither sex (although here the scant presence of the feminine may have introduced some disturbance into the data) nor the dialectal origin of the writers nor the fact of having written from Spain or from America generate any visible differences among the corresponding groups. More interesting results are obtained, however, between the two age groups into which we divided the samples based on an emic and experiential criterion. Despite not being significant, the greater inclination towards *el que* among the younger cohorts throughout most of the eighteenth and nineteenth centuries is striking and, together with the significance of the time axis in both periods, leads us to consider the possible existence of a slow but sustained generational change in relation to the variable. Hence, in each period considered, young people were driving the diffusion of *el que* beyond the figures of the more adult age groups.

The selection of social status as a significant factor in some periods also contributed to this sociolinguistic profile. Although it was only properly so in the nineteenth century, the same explanatory direction can be seen in all the others: the variant *el que* is systematically more often associated with the more popular strata, while the form without the article appears mostly in the social elite and, to a lesser extent, in the middle classes. Nevertheless, these differences begin to cancel each other out in the twentieth century – especially the distance between

the two latter – thus confirming a greater extension of the variant across the whole social spectrum, and a certain stabilization of the linguistic variable in this century.

8 To conclude: has *el que* become a (grammaticalized) compound relative pronoun?

As we have seen in previous sections, several authors defend the nature of *el que* as a compound relative pronoun, in the same way that the academic grammar has recognized *el cual* for several centuries. Nonetheless, those who think thus consider that this status is a result of a process of grammaticalization in progress, as yet incomplete, as shown by the secular alternation, which continues today, between the forms of *que* with and without the article. Furthermore, this grammaticalization would be the last milestone in a long process that, with its origins in Latin, would have led to the transformation of some demonstratives into articles and, later on, within the context we are addressing here, to their evolution as a mere inflected agreement affix.

By applying the variationist method, and comparing the patterns of variation over the last three centuries, in this study, we have tried to test this hypothesis and determine whether, in the period under study, there is enough evidence to support it.

As grammaticalization theorists recall, an increase in the absolute and relative frequencies (in relation to the other variants) is an indispensable factor in it (Bybee 2003). The variant *el que* fulfils that requirement, and today we know that it represents a majority solution, at least in peninsular Spanish. However, throughout the 260 years we have analysed in these pages, that increase was far more moderate over time, with tenuous increases from one century to another. On the other hand, the increases in frequency of the form subjected to the process of grammaticalization are also accompanied by several significant changes in the underlying grammar. These are related to several principles, two of which are of particular interest to us: the principle of retention, or persistence, and that of generalization (Hopper 1991). The first implies that the grammaticalized expression retains old features from the construction in that process. In our case, we have confirmed how the variant *el que* has been strongly associated at all times to diverse syntactic and semantic contexts, such as appositive sentences and locative contents, the two main conditioning factors in every process of change. In the same way, the restrictive relative sentences

and non-locative values have persisted over time as strongly inhibiting factors, with scarce variations from one period to another.

However, this very fact casts some doubts on the second principle, that of generalization, which is even more important than the first to explain cases of grammaticalization. According to this principle, in grammaticalized expressions the old features that operated in the past gradually lose importance, and as a result their explanatory relevance becomes weaker. Yet, in our case this weakening does not occur, as revealed by the explanatory ranges reached by them in each period. This means that in 1960, the type of relative clause and its content continued to exert a strong conditioning effect on the variation.

It is true that some conditioning factors that operated in the past lose strength until they eventually disappear, as happened with the verbal paradigm of the subordinate clause, which was operative in the eighteenth and nineteenth centuries, but not in the twentieth, or the presence/absence of the subject in the relative clause, which was among the significant factors in the nineteenth century. Nevertheless, this generalization in the first case affects very restricted forms, such as the future (on the favouring side) or the subjunctive mode (on the inhibiting side), which were in strong minority at all times with respect to those of the present and past indicative. And in the second case, its relevance is not only moderate (the most moderate of the whole century) but also short-lived, since it was not present before or after the nineteenth century.

What did occur in the twentieth century is that, with the growth of the uses of *el que*, within the envelope of variation appeared a series of new conditioning factors that, in some cases, consolidated tendencies that were already timidly on the rise in the past, like the particular association of *el que* with feminine or indefinite antecedents or those with non-adverbial functions, which were already visible, albeit with less force, in the previous periods.

Therefore, at least in the long period analysed in these pages, the process of grammaticalization of *el que* as a relative compound, if it is the case, is not at a stage as developed as some authors claim (see § 2). While it is true that the article has completed the transformation that has led it to empty its pronominal content into a mere affix with morphological value, we are still far from a fossilized structure such as *el cual*, which in this case is, no doubt, a fully compound relative pronoun. Besides, some contemporary corpus data show that there are still many contexts in which the relative *que* and *el que* continue to alternate and – more importantly – are conditioned by similar constraints (degree of adjacency, degree of determination, type of relative clause, etc.). The main difference is that, in our days, the variant with the article (*el que*) is clearly more frequent than the bare one (*que*) (Vellón and Moya 2017).

In short, it cannot be concluded that the forms of the relative pronoun *que* preceded by an article are a grammatical unit resulting from a process of grammaticalization in progress, as is the case of *el cual*. In the long period analysed here at least (260 years), the most likely interpretation in view of the data examined is that of the existence of two variants of *que*, with and without the article, which would have been struggling for the same functional space over a period of 260 years. In this regard, the now confirmed change in progress is not essentially different from other morphosyntactic changes.

Acknowledgement: This study is part of the research project entitled “Structural, social and idiolectal dimensions of linguistic change: new contributions from historical sociolinguistics to the study of Spanish,” funded by the Universitat Jaume I (Ref. UJI-B2017-01) and the Spanish Ministry of Economy and Competitiveness (Ref. FFI2017-86194-P). We would like to thank Carme Barberà, Susana Martínez, Maria Chiara Marullo and Jordi Ayza for their collaboration in various tasks related to the project. We also thank three anonymous reviewers for their insightful and constructive comments. Any remaining errors are our sole responsibility.

References

- Alarcos Llorach, Emilio. 1994. *Gramática de la lengua española*. Madrid: Espasa-Calpe.
- Blas Arroyo, José Luis & Vellón Javier. 2015. The refuge of a dying variant within the grammar: Patterns of change and continuity in the Spanish verbal periphrasis *haber de* + infinitive over the past two centuries. *Language Variation and Change* 27. 89–116.
- Blas Arroyo, José Luis & Vellón Javier. 2017. En los albores de un cambio lingüístico: Factores condicionantes y fases en la inserción del artículo en las relativas oblicuas del siglo XVIII. *Zeitschrift für romanische Philologie* 133(2). 492–529.
- Auroux, Sylvain. 1994. *La Révolution technologique de la grammatisation*. Liège: Mardaga.
- Balbachán, Fernando. 2011. Asimetrías en el uso del artículo definido con cláusulas relativas prepositivas u oblicuas; un análisis semántico-pragmático. *Boletín de Lingüística XXIII* 35–36. 31–56.
- Bello, Andrés. 1988 [1847]. *Gramática de la lengua castellana destinada al uso de los americanos*. Madrid: Arco Libros.
- Bergs, Alexander. 2012. The uniformitarian principle and the risk of anachronisms in language and social history. In Juan Manuel Hernández-Campoy & Juan Conde-Silvestre (eds.), *The Handbook of Historical Sociolinguistics*, 80–98. Malden: Wiley-Blackwell.
- Bosque, Ignacio & Juan Carlos Moreno. 1990. Las construcciones con *lo* y la denotación del neutro. *Lingüística* 2. 5–50.

- Briz, Antonio & Manuel Prunyonosa. 1987. *Sintaxi i semántica de l'article*. Valencia: Universitat de València.
- Brucart, José María. 1999. La estructura del sintagma nominal: Las oraciones de relativo. In Ignacio Bosque & Violeta Delmonte (eds.), *Gramática descriptiva de la lengua española*, 395–522. Madrid: Espasa-Calpe.
- Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.
- Camarero, Garachana. Mar 2016. Restricciones léxicas en la gramaticalización de las perífrasis verbales. *Rilce* 32(1). 134–158.
- Cano Aguilar, Rafael. 1996. Lenguaje 'espontáneo' y retórica epistolar en cartas de emigrantes españoles a Indias. In Thomas Kotschi, Wulf Oesterreicher & Klaus Zimmermann (eds.), *El español hablado y la cultura oral en España e Hispanoamérica*, 375–404. Frankfurt: Vervuert.
- Company, Concepción. 2004. ¿Gramaticalización o desgramaticalización? Reanálisis y subjetivización de verbos como marcadores discursivos en la historia del español. *Revista de Filología Española* 84(1). 29–66.
- Cortés Rodríguez, Luis. 1986. Alternancia de los relativos *donde*: Que/el que: El cual en el español hablado. *Revista Española de Lingüística Aplicada* 2. 9–22.
- Cukor-Avila, Patricia & Guy Bailey. 2013. Real time and apparent time. In Jack Chambers & Natalie Schilling (eds.), *The Handbook of Language Variation and Change*, 239–262. Malden/Oxford: Wiley-Blackwell.
- Dapena, Porto & Álvaro. José. 1997. *Relativos e interrogativos*. Madrid: Arco Libros.
- Delbecque, Nicole. 2011. Alternancia de los pronombres relativos *el que* y *que* tras preposición: Especificación vs. apoyo discursivo. In Isabel Molina Ana M^a Cestero & Florentino Paredes (eds.), *Congreso Internacional de la Asociación de Lingüística y Filología de América Latina. La lengua, lugar de encuentro*, 253–263. Alcalá de Henares: Universidad de Alcalá.
- Dossena, Marina & Ingrid Tiekens-Boom (eds.), 2008. *Studies in Late Modern English Correspondence: Methodology and Data*. Bern: Peter Lang.
- Eberenz, Rolf. 2000. *El español en el otoño de la Edad Media. Sobre el artículo y los pronombres*. Madrid: Gredos.
- Elspass, Stephan. 2012. Between linguistic creativity and formulaic restriction: Cross-linguistic perspectives on nineteenth-century lower class writers' private letters. In Marina Dossena & Gabriella Del Lungo (eds.), *Letter Writing in Late Modern Europe*, 45–64. Amsterdam: John Benjamins Publishing Company (Pragmatics & Beyond New Series (218)).
- Elvira, Javier. 2015. *Lingüística histórica y cambio gramatical*. Madrid: Síntesis.
- García García, Serafina. 1990. *Los transcritores oracionales en la obra histórica alfonsí. Estudio de sintaxis funcional*. Oviedo: Universidad de Oviedo.
- García Mouton, Pilar. 1999. Tratamientos en las cartas de Indias. In Eduardo Forastari, Humberto López Morales, Julia Cardona & Amparo Morales (coords.), *Estudios de lingüística hispánica. Homenaje a María Vaquero*, 263–276. Puerto Rico: Universidad de Puerto Rico.
- Gili Gaya, Samuel. 1970. *Curso superior de sintaxis española*. Barcelona: Vox.
- Girón, José Luis. 2006a. Las oraciones de relativo II. Evolución del relativo compuesto EL QUE, LA QUE, LO QUE. In Concepción Company (dir.), *Sintaxis histórica de la lengua española. Segunda parte: La frase nominal*, vol. 2. 1477–1590. México: UNAM/FCE.
- Girón, José Luis. 2006b. La historia del artículo como antecedente de relativas oblicuas desde el punto de vista de la gramaticalización. In José Luis Girón & José Jesús Bustos Tovar

- (coords.), *Actas del VI Congreso Internacional de Historia de la Lengua Española*. Madrid, 2013, vol. 1. 765–776. Madrid: Arco Libros.
- Gómez Torrego, Leonardo. 1989. *Manual de español correcto*. Madrid: Arco Libros.
- Guzmán Riverón, Martha. 2012. El artículo en las relativas oblicuas [prep. + (art. definido) + que] en textos americanos del siglo XVIII. *Cuadernos dieciochistas* 13. 175–208.
- Hernández Alonso, César. 1986. *Gramática funcional del español*. Madrid: Gredos.
- Hopper, Paul J. 1991. On some principles of grammaticalization. In Elizabeth Traugott & Bernd Heine (eds), *Approaches to Grammaticalization*, vol. 1. 17–35. Amsterdam: John Benjamins.
- 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.
- Joseph, Brian D. 2011. Grammaticalization: A general critique. In Bernd Heine & Narrog Heiko (eds.), *The Oxford Handbook of Grammaticalization*, 193–205. Oxford: Oxford University Press.
- Keniston, Hayward. 1937. *The syntax of Castilian prose. The Sixteenth century*. Chicago: The University of Chicago Press.
- Labov, William. 1972. *Sociolinguistics Patterns*. Philadelphia: Pennsylvania University Press.
- Labov, William. 1994. *Principles of linguistic change*. Oxford: Blackwell.
- Labov, William. 2007. Transmission and Diffusion. *Language* 83. 344–387.
- Lapesa, Rafael. 2000. El, la, lo como antecedente de relativo en español. In Rafael Cano & M^a Teresa Echenique (eds.), *Estudios de morfosintaxis histórica del español*, 388–401. Madrid: Gredos.
- Lope Blanch, Juan M. 1998. Los relativos en la segunda carta de Hernán Cortés. In Claudio García, Fabián González & José Javier Mangado (coords.), *Actas del IV Congreso Internacional de Historia de la Lengua Española. La Rioja, 1-5 de abril de 1997*, vol. I. 573–580. Logroño: Universidad de la Rioja.
- López García, Ángel. 1994. *Gramática del español. I La oración compuesta*. Madrid: Arco-Libros.
- Marcos Marín, Eugenio. 1980. *Curso de gramática española*. Madrid: Cincel.
- Martínez, Glenn. 2001. Política lingüística y contacto social en el español México-tejano: La oposición -ra y -se en Tejas. *Hispania* 84(1). 114–124.
- Martínez, José Antonio. 1989. *El pronombre II: Numerales, indefinidos y relativos*. Madrid: Arco Libros.
- Nevala, Minna. 2009. Altering distance and defining authority: Person reference in Late Modern English. *Journal of Historical Pragmatics* 10(2). 238–259.
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 1996. *Sociolinguistics and Language History. Studies Based on the Corpus of Early English Correspondence*. Amsterdam: Atlanta.
- Oesterreicher, Wulf. 2004. Textos entre inmediatez y distancia comunicativas. El problema de lo hablado escrito en el Siglo de Oro. In Rafael Cano (coord.), *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 an Early Modern England. In Urszula Okulska & Piotr Cap (eds.), *Perspectives in Politics and Discourse*, 137–197. Amsterdam: John Benjamins.
- Palander-Collin, Minna & Minna Nevala. 2005. Letters and Letter Writing: Introduction. *European Journal of English Studies* 9(1). 1–77.
- Pérez Saldanya, Manuel. 2003. La gramaticalización del verbo *ir* en construcciones narrativas del español. *Medievalia* 35. 62–89.
- Pickering, Martin J. & Victor S. Ferreira. 2008. Structural priming: A critical review. *Psychological Bulletin* 134(3). 427–459.

- Poplack, Shana. 2011. Grammaticalization and linguistic variation. In Bernd Heine & Narrog Heiko (eds.), *The Oxford Handbook of Grammaticalization*, 209–224. Oxford: Oxford University Press.
- Poplack, Shana & Nathalie Dion. 2009. Prescription vs. praxis: The evolution of future temporal reference in French. *Language* 85(3). 557–587.
- Poplack, Shana & Elisabete Malvar. 2007. Elucidating the transition period in linguistic change: The expression of the future in Brazilian Portuguese. *Probus* 19. 121–169.
- Poplack, Shana & Sali Tagliamonte. 2001. *African American English in the diaspora: Tense and aspect*. Oxford: Blackwell.
- Raumolin-Brunberg, Helena. 2005. Language change in adulthood: Historical letters as evidence. *European Journal of English Studies* 9(1). 37–51.
- Real Academia Española. 1931. *Gramática de la lengua española*. Bilbao-Madrid-Barcelona: Espasa-Calpe.
- Real Academia Española. 2009. *Nueva gramática de la lengua española*. Madrid: Espasa-Calpe.
- Rivero, María Luisa. 1991. *Las construcciones de relativo*. Madrid: Taurus.
- Sankoff, David. 1988. Variable Rules. In Ulrich Ammon, Norbert Dittmar & Klaus Mattneier (eds.), *Sociolinguistics*, vol. 2. 984–997. Berlin: Walter de Gruyter.
- Santana Marrero, Juana. 2004. Preposición + (artículo) + que relativo: Análisis en la norma lingüística culta panhispánica. *Boletín de Lingüística* 21. 66–91.
- Schneider, Edgar W. 2013. Investigating Variation and Change in Written Documents. In *New Perspectives*, Jack Chambers & Natalie Schilling *The Handbook of Language Variation and Change*, Second, 57–81. Malden/Oxford: Wiley-Blackwell.
- Tagliamonte, Sali. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Oxford: Wiley-Blackwell.
- Torres Cacoullous, Rena. 2011. Variation and Grammaticalization. In Manuel Díaz-Campos (ed.), *The Handbook of Hispanic Sociolinguistics*, 148–166. Oxford: Wiley-Blackwell.
- Torres Cacoullous, Rena. 2012. Grammaticalization through inherent variability: The development of a progressive in Spanish. *Studies in Language* 36(1). 73–122.
- Trujillo, Ramón. 1990. Sobre la supuesta despronominalización del relativo. *Estudios de Lingüística de la Universidad de Alicante* 6. 23–46.
- Vellón, Javier & Rosana Moya. 2017. Pervivencia de las relativas oblicuas sin artículo: Factores y contextos condicionants. *Spanish in Context* 14(3). 464–486.
- Vries, Mark de. 2002. *The Syntax of Relativization*. Utrecht: LOT.