José Luis Blas Arroyo*

# Comparative variationism for the study of language change: five centuries of competition amongst Spanish deontic periphrases

**Abstract:** Based on a corpus composed entirely of texts close to the pole of communicative immediacy, mainly private letters from the sixteenth, eighteenth and twentieth centuries (c. 1960), this paper analyses the results of a variationist study on the historical evolution undergone by the Spanish modal periphrases with three distinct auxiliary verbs (*haber, tener, deber*). Using the heuristic tools of the comparative method, the data show that variation has been constrained by a handful of common factor groups over almost five centuries. Nonetheless, with the odd exception, these factors have conditioned each verb in a different way. Moreover, the sense of this variation changes as time goes by, with especially relevant reorganisation in the first part of the twentieth century. Furthermore, there is a notable association between these constraints and the degree of markedness and the frequency of the conditioning contexts, giving support to a usage-based approach to language change in which cognitive processes such as entrenchment play a decisive role. These data also allow a particular profile to be traced for each modal verb in the history of Spanish, in which *tener* and *haber* finally undergo a complementary distribution, whereas *deber* follows a different pattern. After several centuries of stagnation, *tener* becomes the star in the deontic firmament of spontaneous communication, diffusing abruptly as a change from below in the twentieth century, and replacing *haber*, which had been the unmarked variant for centuries.

**Keywords:** variationist linguistics, comparative method, deontic periphrases, Spanish, communicative proximity (*Nähesprache*), epistolary corpus, frequency effects, cognitive entrenchment

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

# 1 Introduction

In this study we offer a thorough analysis of the evolution undergone by the Spanish modal periphrases with the infinitive within the deontic sphere over a period of five centuries. The following examples illustrate this variation at three points along that time axis: the sixteenth century (examples 1 to 3), the eighteenth century (4 to 6) and the twentieth century (7 to 9). All the examples, taken from a corpus of private correspondence, show a clear notion of obligation or necessity, as well as a similar referential activity, coded by the same main verb in all the periphrases (*hacer* ['to do']). Conversely, the periphrases contain three different modal verbs, (*haber, tener* and *deber*) as auxiliaries, which have been vying for this semantic category all the way back to mediaeval Spanish (Beardsley 1921; Yllera 1980; Olbertz 1998; Gómez Torrego 1999; Fernández De Castro 1999; Hernández Díaz 2006; García Fernández 2006, 2013; NGRALE (Real Academia Española y Asociación de Academias de la Lengua) 2009; Blas Arroyo and Vellón 2015; Blas Arroyo and Schulte 2017; among others):

(1)     ... avnque yo fuera de vn corazon de hierro me paresçe que me avian de mober espeçialmente siendo hr$^{no}$ de vm y que de justa just$^{a}$ *avia de haçer* lo que vm me pedia (*Cartas de particulares en Indias*, 1582).
        [... even though I had a heart of iron I think they ought to move me especially being your brother and to be fair I should do what you asked of me]

(2)     ... el uirrey mi señor esta con mucha salud y *tengome de hazer* mucha merced por muchas cosas que me a significado (*Cartas de particulares en Indias*, 1577).
        [... my lord the viceroy is in very good health and I have to thank him very much for many things he has told me about]

(3)     ...lo qual siento en el alma cada dia en no aver podido regalalle ante[s] de su muerte como yo lo *devia hazer* (*Cartas de particulares en Indias*, 1585).
        [... which I feel deeply sorry for each day for not having been able to please him before his death as I ought to have done]

(4)     Hija, en cuanto a mis cuidados, doite aviso de lo que *has de hacer* con Pedro Salvador (*Cartas desde América*, 1728).
        [Daughter, regarding my affairs, I am giving you notice of what you must do, with Pedro Salvador]

(5)   Le buscarás en casa de Don José Nolasco, que allí asiste, y si no ha venido de Vizcaya aguardarás que venga, que te dirá lo que *tienes que hacer* (*Cartas desde América*, 1787).
      [Look for him in Don José Nolasco's house, where he usually is, and if he hasn't returned from Biscay, wait for him to come, and he will tell you what you must do]

(6)   ... te encargo que consultes con nuestro rector o el doctor Cathalano [...], pues tengo hecha la súplica a dicho señor para que te dirija lo que *debes hacer* a mi favor (*El hilo que une*, 1771).
      [... I put you in charge of asking our rector, Dr Cathalano [...] for I have requested this gentleman to let you know what you must do for me]

(7)   Yo he de radiar por la noche en Tarrasa una impresión y *he de hacer* otra para «L'Acció», diario de la Esquerra en Tarrasa (*Un catalanófilo de Madrid*, 1936).
      [I have to broadcast a print in the evening in Tarrasa and I have to do another one for "L'Acció", a left-wing newspaper in Tarrasa]

(8)   Voy a bajar de aquí y lo primero que *tengo que hacer* es lavarme bien (*Once cartas de mi padre*, 1936).
      [I am going to get down from here and the first thing I must do is have a good wash.]

(9)   Por consiguiente te advierto que, una vez admitido, lo primero que *debes hacer* es obedecer (*Francia no nos llamó*, 1939).
      [I therefore warn you that, once accepted, the first thing you must do is obey]

In order to analyse the route followed by the changes that have taken place within these periphrases, we use tools from linguistic variationism to explore in greater depth not only the evolution of frequencies of use from one period to another, but also – and above all – to examine what happened *within* the grammar at diverse times in which different variants competed with one another for a particular functional area (Poplack and Tagliamonte 2001; Torres Cacoullos 2009; Poplack 2011). Initially, this means starting out from the principle of *neutralisation in discourse* (Sankoff 1988), by which the potential differences between two or more forms in the language system are sometimes not relevant to the speakers, and so their use in discourse can be alternated to cover the same referential area. Moreover, this neutralisation, which for Sankoff (1988: 153) represents "[the]

fundamental discursive mechanism of variation and change", must be analysed in accordance with the *principle of accountability* (Labov 1972), one of the seminal differences with respect to other approaches to the study of language change. According to this principle, to be able to account for the changes that take place within the grammar accurately it is necessary to analyse not only the way in which the variable context conditions a particular linguistic form, but also that of the one (or ones) that could appear in its place. Otherwise, the association of certain variants with one function or another could give rise to partial, and even incorrect, interpretations (Sankoff 1990).

Likewise, examining the changes requires a precise delimitation of the envelope of variation, without taking into account other spheres in which it is far more limited. In the case we are dealing with here, for example, it is known that some modal periphrases also work well in the area of epistemic modality, as well as in expressing temporal meanings that are not strictly modal (Yllera 1980; Lapesa 2000; Sinner 2003; Hernández Díaz 2006; NGRALE (Real Academia Española y Asociación de Academias de la Lengua) 2009; Blas Arroyo and Porcar 2014; Blas Arroyo and Schulte 2017), as in examples (10) to (11) and (12) to (13), respectively.

(10)  … jamás nadie con quien he tratado se ha quejado de mi ni aun tampoco con razón *tiene de estarlo* el dicho señor Treviño (*El hilo que une*, 1585).
[… nobody I have ever had dealings with has ever complained about me and this Mr Treviño has no reason to do so either]

(11)  … por alli diz que aporto perdida que se cree que alguno dellos *deve ser* nicuesa capitan quel catolico Rey don fernando de gloriosa memoria… (*Textos del Caribe*, 1519).
[… they say that he ended up getting lost, and it is believed that one of them must be Captain Nicuesa, who King Ferdinand the Catholic, of glorious memory…]

(12)  …beros a vos y a mis hijos q os quiero mas q a mi vida yo espero [en] dios q los *tengo de ber* [ = veré] muy breue plazdo a nro sor (*Cartas de particulares en Indias*, 1583).
[… see both you and my children who I love more than life itself I wish to God that I will see you soon]

(13)  … en sebilla esta pº de moya mercader qe *a de benir* [ = vendrá] en la flota con ese puede ynbiar vmd los recados… (*Cartas de particulares en Indias*, 1596).
[…Pedro de Moya, a merchant who will come on the fleet, is in Seville and I can send you messages through him]

However, the distribution of these meanings among the periphrases considered in the article is far more irregular and in some cases almost anecdotal. Thus, the verb *deber* is hardly used in the temporal meanings, and the same could be said for *tener* from the eighteenth century onwards. Moreover, historically this latter verb has been far less commonly used to express conjectural contents than *haber* and *deber*.[1] This study is therefore limited to the alternation among these three modal verbs within the field of deontic modality, which is the only one in which variation takes place in truly significant proportions throughout history.

Finally, the comparative method employed in this study involves exploring the connection between the patterns of variation presented by similar samples (Tagliamonte 2012: 162), whether we are dealing with different languages or speech communities or, as in the present case, different moments in the history of a language. The idea underlying this approach to language change is that their structure can be derived from the contrastive analysis of different quantitative magnitudes extracted from a multivariate statistical analysis (Poplack 2011: 212). To do so we perform different independent logistic regression analyses in which, in each case, a modal verb and a particular historical period are taken as the reference.

In a previous study we carried out a partial exploration in this direction by analysing the evolution of modal periphrases in two key moments in the evolution of the Spanish language: the first stage of Classical Spanish (sixteenth century) and the earliest Modern Spanish (eighteenth century) (Blas Arroyo and Schulte 2017). On that occasion we focused our attention on the evolution of *haber de*, which was the most common periphrasis from the Middle Ages until recent times. The empirical analysis performed there shows that, first of all, there was a marked decrease in the frequencies of use of *haber de* to express deontic contents as one period gave way to the next (sixteenth century: 74 % / eighteenth century: 53 %), these figures contrasting with the far more generous values that had been obtained in earlier studies (Martínez Díaz 2003; López Izquierdo 2008).[2] Nevertheless, in the Age of Enlightenment *haber de* continued to be the majority variant, not only in

---

**1** In fact, in the corpus of the eighteenth century we did not find any tokens with this value.
**2** These differences may have been influenced, on the one hand, by the different discursive traditions analysed in each case (hence, the importance of working with texts that are as close to orality as possible), but also by the insufficient delimitation of the envelope of variation, as a result of including all the potential meanings of the periphrases (and not only the deontic ones), regardless of whether the variation occurs in practice or not.

general terms, but also within a number of specific contexts. At the same time, the loss of vitality between one period and another was not incompatible with the existence of similar conditioning factors and, furthermore, with the same explanatory direction. Another significant piece of data about that continuity was the fact that the unmarked variant (*haber de*) was systematically favoured in contexts that were in turn unmarked – both in general terms or in the type of discourse considered (see Section 5.3) – and therefore often more frequent in discourse. On the other hand, the results of this analysis also showed several areas of the grammar that no longer operated in the eighteenth century as they had in the past.

One of objectives of the present study is to find out whether if this entrenchment is maintained in more advanced phases of that evolution, or, conversely, whether some abrupt changes are expected within the grammar beyond a significant change in the uses of these periphrases. Hence, in the following we analyse the way in which the envelope of variation affects each of the modal verbs at three points on the temporal axis with enough separation between them to allow a reliable comparison: the sixteenth, eighteenth and twentieth centuries. Besides, in this last period (up until around 1960), we analyse both structural and stylistic as well as social factors.[3]

The results of this study can also be of interest for the theory of grammaticalisation, since at the end of the period analysed, by the second half of the twentieth century, *haber* had not only lost the prevalence it enjoyed throughout the history of Spanish, but had also started to display some signs of fossilisation, as can be deduced from the outcomes shown in several of the factor groups analysed in this study.

Before explaining these outcomes (Section 5) and the main implications deriving from them (Section 6), in the next section we devote some attention to the origin of the infinitive periphrases and their different formal variants (Section 2). We then go on to outline the main features of the corpus, consisting in texts that are close to the pole of communicative immediacy (Oesterreicher 1996, 2004), which was compiled as part of a historical sociolinguistic project aimed at the diachronic study of Spanish (Section 3). Finally, in Section 4, we will describe the essential aspects of the encoding process and the methodology that was used.

---

**3** The data of this latter type in previous centuries are, in some cases, too irregular and incomplete, and so it was decided they should be left out.

# 2 A brief overview of the origin and evolution of the modal periphrases with *haber, tener* and *deber* as the auxiliary verbs

In Latin, the periphrases with *habere* (originally, 'to have something') were mostly used without any element linking them to the main verb. Nevertheless, from Late Latin onwards, as well as at the dawn of the Romance languages, this construction appears with the alternate use of the prepositions *a* and *de*, which are particles that are added to specify the modal sense and to differentiate it from the merely prospective values, which have also been present in these periphrases since ancient times (Yllera 1980: 92). Consequently, in mediaeval Spanish we have these three variants, *aver Ø /aver a / aver de*, which are used in all kinds of texts. *Aver Ø* was the least frequently used construction of the three, and we have only a small number of examples in the twelfth and thirteenth centuries and even fewer from the fourteenth century onwards. Moreover, its use as a modal periphrasis declined as it became grammaticalised in the construction of the synthetic future. Both *aver a* and *aver de* then reached a higher frequency of use. The first is predominant until the fourteenth century, although it practically disappears in the fifteenth, except in some markedly dialectal texts (Stengaard 2003: 1151). From then on, *aver de* was to go on alone to become the modal periphrasis par excellence for much of the history of the Spanish language. A fourth variant, *aver que* + infinitive, also appears in documents dating from the fourteenth and sixteenth centuries, although it was never a serious competitor for the others. In later centuries it was to gradually take on a restrictive value, which it still has and which makes it a wholly valid construction today. Thus, the form *hay que* is now the impersonal modal periphrasis of necessity-obligation par excellence (*hay que trabajar mucho para salir adelante* – one has to work hard to get ahead in life). Precisely because of this value, and therefore the fact that it cannot be alternated in the same contexts with the other personal periphrases, we decided not to include it in this study.

Between the thirteenth and fifteenth centuries the obligational periphrases with *tener* ( < Lat. *tenere* 'dominate, retain') as an auxiliary make their way into the language, as part of a long process of substitution of the predicative contexts originally reserved for *haber* (Yllera 1980: 110). Garachana and Rosemeyer (2011: 39) have underlined the fact that these constructions are a clear example of how changes in the grammar can arise from a mere substitution in the vocabulary. One consequence of this process was the appearance of the variant *tener de*, a few examples of which already occur in the thirteenth century. The new modal

verb selects this preposition in its earliest appearances, probably influenced by *aver de*. In the fifteenth century the use of *tener de* consolidated itself and went on to become more firmly established during the following century. *Tener a*, on the other hand, appears in texts from the fourteenth century and, always with a very limited usage, finally declined a century later together with its analogous form *aver a*. Also by analogy, as of the fifteenth century, examples have been recorded of *tener* with an infinitive in a non-prepositional construction, either without the linking element or with the conjunction *que*. The first variant barely took root and was considered as a rather literary formula (Yllera 1980: 117). In contrast, the second (*tener que*) began its upward progression as of the second half of the seventeenth century, when it started to replace *tener de*, which is today only found in very restricted dialectal contexts (Yllera 1980; Olbertz 1998; Fernández De Castro 1999; López Izquierdo 2008; García Fernández 2013; Blas Arroyo and González 2014).

Finally, as regards the third modal verb, it must be pointed out that in Latin the original meaning of *deber* was 'to be the debtor of'. Yet, very soon it was also to cover the senses of obligation or necessity which are seen today in the formation of deontic modal periphrases in practically all the Romance languages (Yllera 1980: 92). In the earliest Spanish texts, constructions with *deber* appear without any preposition whatsoever, although the analogy with other obligational periphrases (above all, *haber de*) would justify that, in the thirteenth century, we can see the first examples of alternation between the two constructions: *deber* + infinitive and *deber de* + infinitive (Beardsley 1921; apud Yllera 1980: 128). Although, given its lower cognitive load, the non-prepositional variant has been clearly predominant in all the stages of Spanish, *deber de* nevertheless presented an acceptable level of vitality during the classical period, with a significant increase in its usages as of the second half of the sixteenth century (Blas Arroyo and Porcar 2016). From the eighteenth century onwards, however, a period of marked decline began, partly affected perhaps by the normative caveats issued by the Real Academia Española regarding the uses of the periphrasis in deontic (but not epistemic) contexts (Blas Arroyo 2014).

# 3 The corpus

The corpus used in the present research, which was compiled as part of a historical sociolinguistics project for the study of Spanish, is made up entirely of texts that lie close to the pole of communicative immediacy (Oesterreicher 2004) (see footnote 1). Most of the texts are letters of a private nature, although

there is also a smaller sample of autobiographical works, such as diaries or court testimonies from trials. They are, therefore, texts which, despite being written, can give us a more faithful idea of the vernacular varieties used in periods for which there are no audio recordings. Moreover, the selected writings are representative of diverse geographical, stylistic and social varieties, as is to be expected of the writings of thousands of individuals of different social and dialectal conditions.

As different authors have recently observed, private letters are an especially fertile ground for diachronic studies. Thus, Oesterreicher (1996: 325) situated these texts within the category of written works with oral features, since, originally, most of them were not conceived to be published, and therefore it can be expected that the pressure on the author to eliminate vernacular features in them was considerably lower than in more formal written texts. Likewise, in most cases private letters lack any kind of prior planning, which also explains their being closer to orality (Danilova 2012). Other authors add important extra-linguistic reasons for working with collections of private letters, such as the fact that this correspondence includes many biographical details which, in spite of their limitations, make it much easier to undertake a sociolinguistic study (Raumolin-Brunberg 2005; Okulska 2010; Elspass 2012). In the same way, by using letters it also becomes possible to discover other relevant data about the communicative act, such as the relations of power and solidarity between the sender and his or her addressees.

Whenever possible, we have used the editions undertaken recently by philologists and linguists, with transcription criteria that are closer to the original texts than those that were usual in the past. Nevertheless, we also include several editions carried out in recent years by some social historians, who are playing a crucial role in unveiling these kinds of documents, which had been forgotten for centuries. However, for inclusion in the corpus, the editor must respect the literalness of the texts, with the only potential exception of modernising the spelling or the punctuation, in order to facilitate understanding. Although this kind of modernisation invalidates the documents as a material for studying graphic and phonic variation, it is not necessarily the case in the research of syntactic and discursive variation, as is the case here.[4]

The sixteenth century subcorpus consists of 1,935 letters, as well as a small sample of the oral testimonies obtained in trials carried out by the Inquisition (Eberenz and De La Torre 2002). Furthermore, we have included some diaries and chronicles written by individuals who had only a limited formal education

---

**4** Readers can find a complete list of the works used in each period on the following website: http://sociolinguisticawe.wix.com/sociolinguisticauji.

(Stoll 2002; Stoll and Vázquez 2011). Taken as a whole, this subcorpus represents the voices of 700 Spaniards who wrote their texts either in Spain or (the majority) from the recently conquered territories in America. Altogether the texts amount to 842,658 words. The corpus from the eighteenth century contains a sample of 1,263 letters written by over 500 different authors, together with two diaries and an accounts book, all of which adds up to a total of 624,456 words. Lastly, the twentieth century subcorpus consists of 2,045 letters and two autobiographical works, with a total of 695,090 words written by nearly 400 different authors.

Many of the letters were written (or dictated, in the sixteenth century)[5] by individuals from different social classes, ranging from manual workers and craftsmen at the bottom of the social pyramid to the highest aristocracy at the other extreme. Furthermore, in this corpus we also find different styles, which range from examples of the utmost spontaneity and intimacy between the interlocutors to more hierarchical and formal relationships that depend on the differences of power between the interlocutors (more details about the stylistic differences will be given below in Section 5.2.7).

# 4 Methodology

The variationist method seeks to identify the contexts that favour or, conversely, inhibit the choice of a particular linguistic form from among others in order to express the same referential or functional content. To do so, it tests a series of hypotheses about the incidence of certain settings (linguistic, stylistic and/or social) that it interprets as potential conditioning factors in a multivariate statistical analysis. A concordance program (*Wordsmith 6.0*) was used to extract all the tokens of the variable in the corpus (and to discard those that exceeded the envelope of variation), so that they could then be encoded according to a selection of these contexts. Moreover, for the purposes of the multivariate

---

**5** As several studies have recently shown, direct dictation could have exerted a much less decisive effect on the vernacular than expected, at least as regards grammatical variation. Thus, in his study on the private letters of the Paston family in the sixteenth and seventeenth centuries, Bergs (2005) has proven that the grammar of different authors differed significantly from one to another, despite the fact that those letters were dictated to a single scribe. For my part, in an analysis on the variation between *deber* and *deber de* + infinitive in letters written by private citizens in the sixteenth century (Blas Arroyo 2016), I noticed that the constraints of this variation are similar when both texts – dictated and written directly by the authors – are taken into account.

analysis, the formal variants of each modal verb were gathered within the same group. Thus, we took together the cases of both *tener de* and *tener que* (examples 14 and 15) and the prepositional (*deber de*) and non-prepositional variants (*deber*) of this verb, as in examples (16) and (17).

(14)   ... considerando lo que tengo dcho [...] yo *tengo de enbiar* la sangre del brazo si alla fuese de probecho (*Cartas de particulares en Indias*, 1571)
       [... bearing in mind that I have said [...] I would send the blood from my arm if it were needed there]

(15)   Yo estoy en compañía de una señora de allá de mi tierra, muy honrada; lo demás no *tengo que encomendaros* sino que miréis que sois cristiano... (*El hilo que une*, 1574)
       [I am in the company of a very honourable lady from my homeland; apart from that I have nothing else to ask of you except that you ensure you are Christian...]

(16)   Y así no *debes de escribirme* por esta misma mano; avísame cómo está tu salud y todo lo que tuvieras de nuevo *(El hilo que une*, 1761)
       [And so you shouldn't write to me by this same hand; tell me about your state of health and any other news you may have]

(17)   En cuanto a lo que me dices del padre Castrejón *debo decirte* que ya te tengo escrito en otras tres que te mantengas en tu régimen (*El hilo que une*, 1760)
       [With regard to what you told me about Father Castrejón I must tell you that I have already written three other letters to you saying that you should remain as you are]

The linguistic factors, which have already been examined in previous studies (Balasch 2008, 2012; Blas Arroyo and Porcar 2014; Blas Arroyo and Schulte 2017), are of a different nature. In the following we show a list of these factors, whereas those selected as significant by the logistic regression analysis are described and exemplified in more detail in Section 5.

   *Discourse factors (priming):* presence/absence within the immediate previous co-text of another periphrasis with an auxiliary verb from the same group.

   *Phonetic factors:* next phonetic context (first phoneme of the next word, disregarding the linking particle).

   *Syntactic factors:* 1) person and number (first person/others); 2) tense of the conjugation; 3) syntax of the subject (omitted/explicit); 4) type of clause

(subordinate/others); 5) syntax of the main verb (simple/compound); 6) degree of (im)personalisation (active/passive and reflexive impersonal); and 7) polarity (affirmative/negative sentences).

*Semantic-pragmatic factors:* 1) modal values (external/subjective obligation); 2) degree of animacy (human/non-human subjects); 3) mode of action of the main verb (stative/motion/language/others); and 4) contextual modalisation (intensified/neutral settings).

Two criteria are used to evaluate the *stylistic* axis: a) the main topic dealt with in the texts, and b) the closeness of the relationship between the two interlocutors. Crossing the two criteria enables us to distinguish between more private and spontaneous texts (*personal*, in the encoding) and more formal and *distant* texts (see further details in Section 5.2.7).

Lastly, in the data from the twentieth century, we also include the *sociolectal* matrix.[6] In this case we will analyse the potential incidence of some social factors in order to examine the way in which the variants spread throughout the community during this century. These factors are: *sex* (males/females), *social status* (upper, middle and lower classes), *age* (young/adults), *geographical origin* (northern/southern regions) and the *migratory* context (texts written from America/Spain).

Finally, for the quantitative analysis we used tools from the comparative variationist method (Poplack and Tagliamonte 2001; Tagliamonte 2012), consisting in performing different independent variable rules analyses, for each of the periods under study. Comparing the results of these analyses makes it possible not only to trace the fate of each of the alternating forms but also the structural and non-structural path such changes have followed throughout history. To carry out these analyses we used the logistic regression program Rbrul (Johnson 2009), which provides different means of proof, the data of which will be submitted to a contrastive inquiry. These are: a) the number and identification of the factor groups that are selected and not selected as significant in each language according to the corpus available; b) the explanatory hierarchy of those selected as significant, calculated by means of their respective *ranges*;[7] c) the degree of significance of each individual factor, measured in terms of

---

6 Of the three periods analysed in this study, the twentieth century is the only one in which we have an almost completely access to an exhaustive and reliable extralinguistic information. For the other ones (and specially for the sixteenth century), even though we know the names of most of the letters contained in the corpus, in many cases we don't have knowledge of their corresponding biographical information, in such relevant traits for a sociolinguistic analysis as the social status, the age or the geographical origin of the writers.

7 Range is a non-statistical measure based on the difference between the maximum and minimum P. values within each group of factors. The hypothesis is that, the higher that value is, the greater the predictive relevance of some factors over others will be.

probabilities or factor weights (P. values); and d) the explanatory direction, that is to say, the order of the factors on the axis that ranges from the contexts that most favour the variant to the least favouring.

# 5 Results and analysis

## 5.1 General results

Table 1 shows the distribution of absolute frequencies and percentages of use of the modal verbs and their respective formal variants in each century. As far as these latter are concerned, in the case of *tener* we can see the change that favours the variants with the linking conjunction (*tener que*) more than those with a prepositional one (*tener de*) that we referred to earlier. This second form exceeds the first in the sixteenth century, but the distribution is clearly inverted two centuries later and even more so in the twentieth century, when the presence of *tener de* is almost anecdotal. Likewise, Table 1 shows the always minority position of the prepositional variants of *deber* (*deber de*) with respect to the non-prepositional forms (*deber*), although this difference was to grow even larger as time went by (for more details, see Blas Arroyo and Schulte 2017).

**Table 1:** General distribution of the periphrases in the corpus.

| Periphrases | Sixteenth Century | | Eighteenth Century | | Twentieth Century | |
|---|---|---|---|---|---|---|
| | **N** | **%** | **N** | **%** | **N** | **%** |
| *Haber de* + infinitive | 938 | 74 | 420 | 53 | 228 | 16 |
| *Deber* (*de*) + infinitive | 211 | 16 | 310 | 39 | 457 | 32 |
| (*Deber*) | (182) | (14) | (296) | (37) | (431) | (30) |
| (*Deber de*) | (29) | (2) | (14) | (2) | (26) | (2) |
| *Tener de/que* + infinitive | 120 | 9 | 60 | 8 | 742 | 52 |
| (*Tener de*) | (65) | (5) | (13) | (2) | (19) | (1) |
| (*Tener que*) | (55) | (4) | (47) | (6) | (723) | (51) |
| **Total** | **1269** | | **790** | | **1427** | |

With these formal variants now grouped around their respective auxiliary verbs, Figure 1 shows the evolution of the periphrastic uses over the three periods under consideration. As can be seen, the periphrasis with *haber* occupied a dominant position in the sixteenth (74%) and eighteenth centuries (53%),
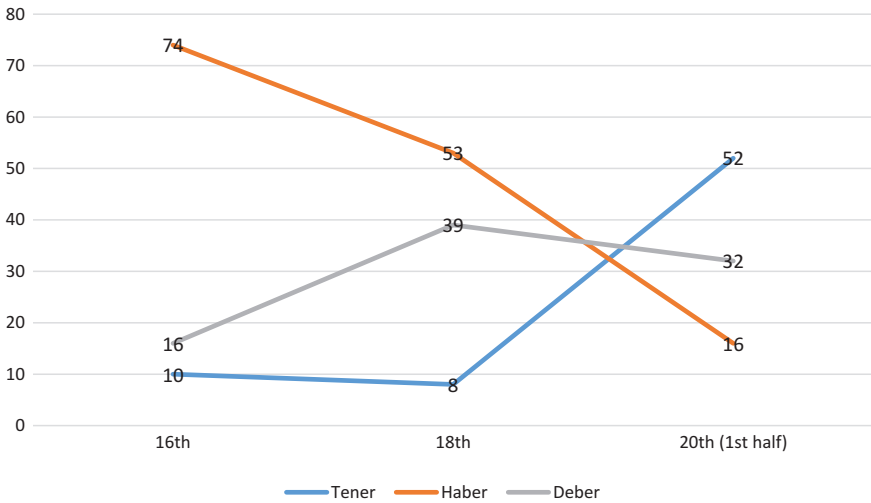
**Figure 1:** Evolution of the frequencies of use between the sixteenth and twentieth centuries (%).

although the decrease between the two centuries is significant. This drop, however, was much sharper in the twentieth century (16 %), when *haber de* became the least commonly used of all the variants. *Tener*, on the other hand, started out as a minority modal verb in early Classical Spanish (9 %), a position that it still maintained two centuries later (8 %). Nevertheless, from that moment on its fate was to change up to the point of becoming the most common form in the sphere of deontic modality during the first half of the twentieth century (52 %).[8] Finally, *deber* displays a distribution that differs from that of its competitors. Like *tener*, this verb is used relatively infrequently in the sphere of obligation in the first period (16 %), but doubled its presence in the Age of Enlightenment (39 %), when it started taking over usages previously belonging to *haber*, which was still the unmarked and majority variant. Unlike *tener*,

---

**8** Elsewhere we have seen how this change in favour of *tener* was already quite apparent in the nineteenth century (Blas Arroyo and Vellón 2015). Even though that study referred to the whole envelope of variation in which both verbs could alternate (and not only to the deontic domain), which prevents direct comparison with the current data, in that article we could already see how *tener* reached relative proportions close to 40 % during the whole century. These are figures considerably higher than those noticed previously for the nineteenth century (see López Izquierdo 2008: 793), a divergence that could be attributed to the different genres analysed and which seems to confirm the relevance of private correspondence to the study of the vernacular in the past.

however, *deber* shows far more constant figures and even slightly decreased its uses in contemporary Spanish (32%).

So far we have been dealing with the evolution of the general uses, but in what spheres of grammar and of the socio-stylistic matrix do these changes operate? Do they do so in a similar way in all contexts? And what is the intensity of such changes? In order to answer these questions, in the following we examine the results from several independent variable rules analyses, one for each variant and period considered in the study. For illustrative purposes, first we show the evolution undergone by each modal verb based on the structural, stylistic (Section 5.2) and social (Section 5.4) factor groups selected as significant by the regression program. We will then go on to examine some common patterns in that evolution, which are related to the degree of markedness and the frequency of both the variants and the contexts in which modal verbs appear (Section 5.3). All this will allow us to obtain a clearer picture of the transformation of the periphrases over the last five centuries, as well as the main points of convergence and contrast among them (Section 5.5).

## 5.2 The evolution of the variable context

Table 2 shows the relative frequencies of use and the factor weights of each variant in the sixteenth, eighteenth and twentieth centuries in relation to the structural and stylistic factors that were selected as significant at some time or another to explain the variation. In the following we address the factors that show an independent relevance and are therefore not altered by interaction with others.

At first glance, the table shows a considerable degree of consistency in this variable conditioning, such that practically the same factor groups come into play in order to explain the variation over the centuries.[9] Some of them condition the evolution of the periphrases all the time, sometimes with the same explanatory direction, but in other periods there are notable changes in that direction. Nevertheless, there is no shortage of those that affect only one of the variants but not the others. In the next sections we offer a more detailed analysis of all these factors.

---

**9** And equally important, the same could be said of those factor groups that were not selected: in the corpus their explanatory force is null, or is influenced by other factors in all the variants and periods under consideration.

**Table 2:** Factor weights and frequencies of use of the modal verbs *tener*, *haber* and *deber* in three different periods: sixteenth, eighteenth, and twentieth centuries (linguistic and stylistic factor groups).

| | | Sixteenth Century | | | | | | | Eighteenth Century | | | | | | | Twentieth Century | | | | | | |
| | | Tener | | Haber | | Deber | | Σ | Tener | | Haber | | Deber | | Σ | Tener | | Haber | | Deber | | Σ |
| | | P. | % | P. | % | P. | % | | P. | % | P. | % | P. | % | | P. | % | P. | % | P. | % | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Subject** | Omitted | 0.67 | 13 | – | [71][a] | – | [16] | 814 | 0.60 | 10 | – | [45] | 0.58 | 45 | 498 | 0.59 | 61 | 0.43 | 12 | 0.43 | 27 | 957 |
| | Explicit | 0.33 | 3 | – | [78] | – | [18] | 455 | 0.40 | 4 | – | [67] | 0.42 | 29 | 292 | 0.41 | 34 | 0.56 | 24 | 0.57 | 42 | 470 |
| **Person** | 1st person | 0.88 | 53 | 0.17 | 37 | 0.37 | 10 | 127 | 0.65 | 15 | 0.31 | 34 | – | [52] | 137 | 0.57 | 63 | – | [16] | – | [21] | 558 |
| | Non-1st person | 0.02 | 2 | 0.57 | 80 | 0.63 | 18 | 1142 | 0.35 | 5 | 0.56 | 60 | – | [35] | 653 | 0.43 | 46 | – | [15] | – | [39] | 869 |
| **Polarity** | Affirmative | 0.31 | 7 | 0.53 | 77 | – | [16] | 1112 | 0.23 | 6 | 0.52 | 54 | – | [40] | 722 | 0.57 | 55 | – | [16] | 0.41 | 29 | 1254 |
| | Negative | 0.69 | 25 | 0.28 | 54 | – | [21] | 157 | 0.77 | 31 | 0.31 | 39 | – | [30] | 68 | 0.43 | 34 | – | [14] | 0.59 | 52 | 173 |
| **Modal meaning** | External obligation | – | [6] | 0.53 | 80 | 0.37 | 14 | 959 | – | [7] | 0.57 | 61 | 0.39 | 32 | 494 | 0.63 | 66 | – | [13] | 0.39 | 21 | 688 |
| | Non-external ob. | – | [19] | 0.39 | 58 | 0.63 | 23 | 310 | – | [8] | 0.37 | 40 | 0.61 | 52 | 296 | 0.37 | 39 | – | [18] | 0.61 | 43 | 739 |
| **Agentivity** | Human | – | [12] | – | [72] | – | [16] | 962 | – | [11] | 0.46 | 48 | – | [41] | 617 | – | [56] | – | [14] | – | [30] | 1130 |
| | Non-human | – | [1] | – | [80] | – | [19] | 307 | – | [0] | 0.64 | 70 | – | [30] | 173 | – | [39] | – | [24] | – | [39] | 297 |

*(continued)*

**Table 2:** (continued)

**Sixteenth Century**

| | Tener P. | Tener % | Haber P. | Haber % | Deber P. | Deber % | Σ |
|---|---|---|---|---|---|---|---|
| **Degree of (im)personality** | | | | | | | |
| Active | – | [10] | – | [76] | – | [14] | 1026 |
| Pas.&Impers | – | [1] | – | [75] | – | [24] | 243 |
| **Tenor** | | | | | | | |
| Personal | – | [12] | 0.62 | 88 | 0.33 | 10 | 673 |
| Distant | – | [7] | 0.38 | 65 | 0.67 | 28 | 596 |
| **Priming[b]** | | | | | | | |
| Same modal verb | 0.79 | 51 | 0.79 | 91 | 0.87 | 69 | |
| No | 0.21 | 8 | 0.50 | 74 | 0.36 | 17 | |
| Other modal verbs | | | 0.8 | 18 | 0.21 | 8 | |

**Eighteenth Century**

| | Tener P. | Tener % | Haber P. | Haber % | Deber P. | Deber % | Σ |
|---|---|---|---|---|---|---|---|
| **Degree of (im)personality** | | | | | | | |
| Active | – | [8] | 0.68 | 72 | – | [20] | 704 |
| Pas.&Impers | – | [0] | 0.46 | 51 | – | [49] | 86 |
| **Tenor** | | | | | | | |
| Personal | – | [8] | 0.63 | 65 | 0.34 | 27 | 490 |
| Distant | – | [7] | 0.27 | 33 | 0.66 | 59 | 300 |
| **Priming[b]** | | | | | | | |
| Same modal verb | 0.74 | 39 | 0.77 | 84 | 0.84 | 80 | |
| No | 0.26 | 7 | 0.52 | 55 | 0.40 | 39 | |
| Other modal verbs | | | 0.11 | 13 | 0.22 | 16 | |

**Twentieth Century**

| | Tener P. | Tener % | Haber P. | Haber % | Deber P. | Deber % | Σ |
|---|---|---|---|---|---|---|---|
| **Degree of (im)personality** | | | | | | | |
| Active | – | [54] | 0.40 | 15 | 0.42 | 31 | 1344 |
| Pas.&Impers | – | [10] | 0.61 | 36 | 0.58 | 54 | 83 |
| **Tenor** | | | | | | | |
| Personal | 0.60 | 60 | 0.33 | 9 | – | [31] | 1085 |
| Distant | 0.40 | 26 | 0.67 | 38 | – | [36] | 342 |
| **Priming[b]** | | | | | | | |
| Same modal verb | 0.58 | 71 | 0.63 | 31 | 0.64 | 44 | 153 |
| No | 0.42 | 49 | 0.37 | 14 | 0.62 | 39 | 1008 |
| Other modal verbs | | | | | 0.25 | 11 | 266 |

[a] The figures in brackets refer to the factor groups not found significant by the variable rule analysis.

[b] In contrast to the remaining factor groups, the raw numbers for *Priming* are unevenly distributed across modal verbs. These are the numbers for every verb and period: *Tener* (16th): Same modal verb (39), Other modal verbs (26), No (1064); *haber* (16th): Same modal verb (140), Other modal verbs (65), No (1064); *deber* (16th): Same modal verbs (179), No (1064). *Tener* (18th): Same modal verb (18), Other modal verbs (102), No (670); *haber* (18th): Same modal verb (61), Other modal verbs (59), No (670); *deber* (18th): Same modal verb (41), Other modal verbs (79), No (670). *Tener* (20th): Same modal verb (216), Other modal verbs (203), No (1008); *haber* (20th): Same modal verb (153), Other modal verbs (266), No (1008); *deber* (20th): Same modal verb (153), Other modal verbs (266), No (1008).

### 5.2.1 Degree of agentivity and impersonality

The first thing that must be highlighted in this respect is the existence of several spheres of grammar that are an almost categorical conditioning factor for certain modal verbs in some periods. This is the case, for example, of the semantic factor group of *agentivity*, by which we distinguish between human and non-human subjects. As can be seen in Table 2, its role was absolutely essential in the early centuries in the case of *tener*, which does not appear with non-human subjects (only one token out of 1,269 in the sixteenth century and none in the eighteenth century), unlike *haber* and *deber*. Yet, things change drastically in the twentieth century, when *tener que* starts to be used in these contexts (39 %), although still some way behind the human ones (56 %).

The same can be said now for a syntactic factor like the *degree of impersonality*: passive and impersonal sentences, in which the subject is camouflaged or impossible to identify, were forbidden territory for periphrases with *tener* for centuries. Even in the twentieth century, when the change favouring this verb was beyond all doubt, its presence in passive or impersonal sentences was very scarce (10 %; n.s.), unlike the cases of *haber* (36 %; 0.61) and *deber* (54 %, 0.58), which are verbs for which such sentences were a privileged context. In this sense, the evolution of *haber* is revealing because, as an unmarked variant throughout much of the history of Spanish, it had been predominant in active sentences, especially in the eighteenth century (0.68).

### 5.2.2 Modal values

The remaining factor groups selected by the multivariate analysis condition the variation of all the periphrases, but they nearly always do so in different ways according to the periphrasis.

One of the most relevant is the semantic value acquired by the notions of obligation or necessity, which are closely associated with deontic modality. Initially, we divide this axis into two spheres, which we call *external obligation* and *non-external or subjective obligation*, respectively.[10] In the first case, the

---

**10** For other attempts to delimit this semantic axis in Spanish, see Yllera 1980; Olbertz 1998; Gómez Torrego 1999; Fernández De Castro 1999; García Fernández 2006; López Izquierdo 2008, among others. Furthermore, we understand the obligational modality in a broad sense, that is, "the expression of the subjectivity of the (enunciated) statement, meaning that the subject of the enunciation may or may not be the same as that of the enunciated statement" (Martínez

obligation always comes from an entity other than the subject, which, in cases like those exemplified in (18) and (19), involves a commitment imposed by external circumstances that have nothing to do with the subject, and which determine the involuntary and almost inevitable need for compliance (at least, from his or her point of view):

(18)   El q a de coger forzoso *a de senbrar* (*Documentos para la historia lingüística de Hispanoamérica* (1581)
       [He who has to harvest has to sow]

(19)   ...porque ya saue V.M. los trabaxos que a los señores mis tíos an benido, y como *tengo de quedar* vn día destos sola y sin tener a quien boluer la cabeça. (*Cosmovisión de una mujer salmantina*, 1569)
       [... because you know the hardship that has fallen upon my uncles and how one of these days I will be left alone and without anyone to turn my head to...]

At other times, however, the external agent is more clearly defined in the obligation. Now the contents of obligation, unavoidable necessity or imperative and coercive convenience are of a nature that is external to the agent of the action described by the main verb, that is to say, the obligation results from his or her being submitted to the will of another, whether it is as a consequence of: a) regulations, agreements, instructions, social conventions, etc., as in (20), regardless of whether the external agent is explicit in the text or not; or b) commands, orders or similar speech acts, generally issued by the subject of the utterance, in order to oblige certain actions to be carried out, as in (21):

(20)   Don Miguel dará a vm. la instrucción de cómo se *a de azer* porque la enbío a él (*Vida y fortuna del emigrante navarro*, 1563)
       [Don Miguel will give you instructions on how to do it because I am sending it to him]

(21)   *Debe Vd. suprimir* de raíz su proemio. Por muy aficionado que sea Vd. a la paradoja no tiene Vd. derecho a protestar de que la lógica pueda ser rectora de la política (*Un catalafónilo de Madrid*, 1927)

---

Díaz 2008: 1285). Otherwise, for some categories we would be forced to consider only those sentences related to the first person singular.

[You should completely remove your preface. However keen you might be on paradox, you have no right to protest that logic may be the guide of politics]

In addition to these obligations, we find others whose origin does not lie outside the subject, but which follow his or her own introspection. A number of senses can be seen among these *subjective* obligations. Thus, in examples such as those in (22), the subject is the one who imposes or refuses him/herself to carry out something. In these cases, the inner conviction rests upon moral, ideological or any other type of values, which lead him or her to the need to act. In other fragments, where the subjects of the enunciation and of the enunciated statement are not the same, it is the first that advises or (morally) obliges the second to fulfil what the main verb refers to, as can be seen in (23):

(22)    ... yo siempre *he de cumplir* con mi ocupación, pues mi mayor deseo es darte gusto en todo para que conozcas lo mucho que te estimo y venero (*Cartas desde América*, 1717)
        [... I must always fulfil my duty, for my greatest wish is to indulge you in everything, so that you understand how much I love and cherish you]

(23)    ... y tanbién se os acuerda de la carta que la reyna, nuestra señora, me enbió, la qual os mostré, asy que todo os lo pongo delante. Vos verés lo que *devés hazer* (*Epistolario del Conde de Tendilla*, 1504)
        [... and I also remind you of the letter that Our Highness, the Queen, sent me, which I will show you and put before all of you]

On other occasions, however, what predominates is not so much the obligation or duty to fulfil something but instead the idea of *necessity* or *advisability*, also considered subjectively, such that the subject controls or decides to impose (upon others or upon him/herself) a certain obligation not so much because of ethical or ideological questions, as in the previous case, but due to reasons purely related to convenience or because of the benefits that could derive from it. The following are a few illustrative examples:

(24)    ... no *tiene que procurar* si no es dos barriles de biscocho no *tiene que gastar* otra cossa y el muchacho lo trayga consigo yo rehusso de yr alla (*Cartas de particulares en Indias*, 1570)
        [... he has only to obtain two barrels of sea biscuits, he doesn't need to buy anything else and the boy must bring it with him – I refuse to go there]

(25)  Confío en su buena acción teniendo en cuenta que son para América y que por esta razón la caja *ha de ser* sólida (*Darío de Regoyos: sus cartas inéditas*, 1911)
[I trust in his good actions bearing in mind that they are for America and for this reason the chest has to be solid]

Finally, we also include within this group several nuances in which the same idea of obligation is less prototypically visible (encoded as "others"). This is the case of utterances where it is difficult to distinguish the purely obligatory from other values such as desire, intention, or recrimination, as in (26), or even, as in (27), a *phatic* value that the speaker uses as an elocutionary aid in order to "enter into communication" with the interlocutor (Gómez Manzano 1992; Gómez Torrego 1999):

(26)  Si pedido yo os mudáis a otra parte, no sé para qué me *tengo de mover*. Si acaso hubiérades enviado orden para que yo vaya a esa tierra... (*El hilo que une*, 1589)
[If I ask you to move somewhere else, I don't know why I have to move. Maybe if you had sent an order for me to go to that land...]

(27)  Si te *he de decir* la verdad ya desconfiaba de que me contestases, creía que ya te habías olvidado definitivamente de mí (*Madrina de guerra¸* 1938)
[To tell you the truth I didn't expect you to answer, I thought you had completely forgotten about me forever]

Table 3 shows the distribution of the verbs for expressing the different meanings considered here.

The above-mentioned general opposition between external and non-external obligations acts as a conditioning factor for some periphrases in certain periods. The case that has lasted the longest in time is that of *deber*, which appears especially favoured in contexts involving non-external obligation in all the periods analysed, and moreover with very similar factor weights in the three centuries (see Table 2: sixteenth century: 0.63; eighteenth century: 0.61; twentieth century: 0.61). Conversely, obligations of an external origin disfavoured this verb to the same extent.

An altogether different case is that of *tener*. In the sixteenth century, the factor did not condition the variation despite the apparent advantage of non-external obligations (19 %) over the external ones (6 %) in terms of frequency. In practice, however, it is not independent from others with which it interacts, and

**Table 3:** General distribution of the main modal meanings among the verbs.

| Modal meanings | Sixteenth Century | | | | Eighteenth Century | | | | Twentieth Century | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Tener % | Haber % | Deber % | N(Σ) | Tener % | Haber % | Deber % | N(Σ) | Tener % | Haber % | Deber % | N(Σ) |
| Inevitability | 6 | 81 | 13 | 321 | 17 | 60 | 23 | 122 | 76 | 11 | 13 | 285 |
| External oriented obligation | 7 | 79 | 14 | 638 | 4 | 61 | 35 | 372 | 59 | 15 | 26 | 403 |
| Necessity/advisability | 31 | 53 | 16 | 112 | 12 | 47 | 41 | 138 | 46 | 9 | 45 | 337 |
| Internal oriented obligation | 14 | 61 | 25 | 182 | 5 | 34 | 61 | 121 | 43 | 14 | 43 | 279 |
| Others | 25 | 7 | 68 | 16 | 6 | 35 | 59 | 37 | 14 | 51 | 34 | 123 |

whose explanatory force is more relevant, as occurs with polarity.[11] Neither was this the case in the eighteenth century, although what did happen as of that moment was a restructuring of the factor group in a new order that will be consolidated in contemporary Spanish (see Table 4). There, *tener* becomes specialised as a verb associated with unavoidable obligations, followed in the twentieth century (but not yet in the eighteenth) by agent-oriented external obligations. It is clear that the difference lies in the important differences in frequency between the two periods: whereas only 17 % of the unavoidable senses employed the verb *tener* in the eighteenth century, these proportions increased fourfold in the twentieth century (76 %).

**Table 4:** Realignment of the deontic sphere in the history of *tener*.

|  | Sixteenth Century | | Eighteenth Century | | Twentieth Century | |
|---|---|---|---|---|---|---|
|  | P. | % | P. | % | P. | % |
| Inevitability | – | [6] | 0.69 | 17 | 0.66 | 76 |
| External oriented obligation | – | [7] | 0.39 | 4 | 0.51 | 58 |
| Others | – | [19] | 0.41 | 8 | 0.32 | 39 |

Moreover, it must be remembered that, in this latter period, the verb also becomes firmly established in practically all the contexts considered, including those involving internal obligation and subjective necessity. Nevertheless, in these cases it struggles against *deber*, a verb that, as we have noted, has always been favoured in such settings.

On the other hand, the presence of *tener* among the values we have classified as less prototypically obligational (others) (14 %) is clearly lower than that of *deber* (34 %), but above all smaller than *haber* (51 %), which in the twentieth century becomes specialised as a means to express these peripheral meanings. After being the unmarked variant in most contexts throughout the history of the Spanish language, the direction of the variability changed completely during its period of obsolescence. Now the old categories are no longer operational for *haber* and instead they are replaced by another in which the more prototypical obligations clearly disfavoured the periphrasis (0.27; 13 %), in contrast to those noted as being less prototypical, whose values still encouraged its use throughout the first half of the twentieth century (0.75; 51 %).

---

**11** In practice, the only advantage for external modality in this century takes place within negative sentences, which is, in any case, a very minority context.

### 5.2.3 Syntax of the subject

As is well known, Spanish is a null-subject language, and so the unmarked expression of this function normally takes place in utterances such as those in (28), whereas its explicit presence, as in (29), is associated with a complex network of syntactic, semantic, pragmatic and even dialectal factors (Luján 1999; Travis 2007; Carvalho et al. 2015).

(28)  … te encargo que consultes con nuestro rector o el doctor Cathalano – el que nos casó, que por entonces era vicario –, pues tengo hecha la súplica a dicho señor para que te dirija lo que *debes hacer* a mi favor (*El hilo que une*, 1771)
[… I put you in charge of asking our rector, Dr Cathalano – who married us, since back then he was a vicar – for I have requested this gentleman to let you know what you must do for me]

(29)  Yo en vista que V. no *tenía que hazer* este desembolzo que tal vez le sería perjudicial […] accedí a ello (*Al recibo de esta*, 1790)
[As I realised that you did not have to spend this sum, which may have been unfavourable to you […] I decided to agree to it]

The influence of this factor appears as one of the most consistent of all those considered in the study, although it comes across in different ways in each periphrasis. In the case of *tener*, a sustained incidence is observed in all the centuries. Hence, and regardless of the frequencies of use in each period, the selection of the periphrasis is always encouraged preferentially in the periphrases that have an omitted subject (sixteenth century: 0.67; eighteenth century: 0.60; twentieth century: 0.59). This is not, however, the case for *haber*, for which the factor is not significant in the earlier two periods, but it is in the twentieth century, when it combines above all with explicit subjects (0.56). Finally, the case of *deber* is apparently the most erratic: the factor is not selected in the sixteenth century, but it is in both the eighteenth and the twentieth century, except that, in the latter, it shows an opposite explanatory direction to that of the Age of Enlightenment (for the theoretical implications of these differences, see Section 6 below).

### 5.2.4 Person

As in the previous case, and leaving aside the general frequencies of use at each time, the data analysis shows a consistent influence of the first person on the

selection of *tener* throughout history (sixteenth century: 0.88; eighteenth century: 0.65; twentieth century: 0.57). This is just the opposite of what can be seen for the remaining persons, whose percentages of use are similar and which we have therefore grouped together for the regression analysis (sixteenth century:0.02; eighteenth century: 0.35; twentieth century: 0.43). Nevertheless, a comparison of these figures shows how the influence of this grammatical factor group diminishes as time goes by. Moreover, in the twentieth century, and unlike the previous ones, where the constraint only affects the first person singular, it now expands to the whole sphere of the first person, including the plural.

Conversely, *haber* displayed a distribution that was largely complementary to the one we have just noted, at least in the sixteenth and eighteenth centuries: if *tener* is favoured in the first person in the two periods, *haber* does the same with the remaining ones. Nevertheless, the factor group was no longer operational in the twentieth century, when the periphrasis falls into the utmost decadence.

Lastly, the case of *deber* is again qualitatively different: the persons other than the first favoured the verb in the sixteenth century, but the influence of this factor group wanes in the following centuries.

### 5.2.5 Polarity

The regression analysis reveals a new example of complementary distribution between *haber* and *tener*. This latter verb is favoured in the two first periods in negative sentences (sixteenth century: 0.69; eighteenth century: 0.77), thus contradicting the idea that negative polarity favours more conservative solutions (cf. Givón 1979; Poplack and Dion 2009; Tagliamonte et al. 2014; Torres Cacoullos and Walker 2009). In contrast, in the twentieth century the polarity factor equally constrains the variation, but now with an important change in the explanatory direction, as affirmative sentences are now the ones that favour these periphrases to a greater extent (0.57 versus 0.43 for negative ones).

This distributional pattern is exactly the opposite of the one we observe with *haber*: during the period in which they enjoy their greatest vitality, modal periphrases with this verb are encouraged more in affirmative utterances (sixteenth century: 0.53; eighteenth century: 0.52) than in negative ones (sixteenth century: 0.28; eighteenth century: 0.31). Nonetheless, when these entered a period of clearly identifiable obsolescence, during the twentieth century, the factor no longer conditions the variation.

Again, the example of *deber* is a particular case: it is not affected by this factor group in the sixteenth and eighteenth centuries, but it is in the twentieth, when it becomes more common in negative utterances (0.59).

### 5.2.6 Priming

In spite of the pronounced imbalance between the samples under consideration here, we find ourselves before one of the most consistent factor groups. It is always selected as significant by the regression analysis and, moreover, with the same explanatory direction. This shows that the presence of the same verb in the immediate previous co-text, as in (30) and (31), encourages its being replicated in all the periods:[12]

(30)  ... y como es jornada que todos **hemos de hazer** *emonos de conformar* con la voluntad devina (*Cartas de particulares en Indias*, 1565)
      [... and as it is a task that we must all do, we must accept God's will]

(31)  ... toda la noche de servicio, **tenia que estar** a las 7 en Vigo con un temporal de la gran 7 (el que sobreviva a este invierno *tiene que contar*) llego a Puenteareas y no había un coche... (*Una familia y un océano de por medio*, 1962)
      [... on duty all night, I had to be in Vigo at 7 with a big force 7 storm (whoever survives this winter will have a story to tell) I got to Puenteareas and there was no car...]

In the other cases, however, we find some telling differences between the verbs. Thus, with both *deber* and *haber*,[13] the presence of the other verbs in this previous setting, as in (32), is always the most disfavouring context, whereas the remaining cases ("no" in the encoding) stands in an intermediate position:

(32)  Y me parece que también se le **debe hacer** cargo del crédito de este dinero que injustamente ha retenido. Si la guerra permanece *habré de remitir* los

---

**12** In order to measure this factor in the most objective way possible, we counted the potential influence exerted by the previous context situated at a distance equal to or less than 25 words away. We believe that, by so doing, the arbitrariness of the limit imposed is offset by its objectivity. Moreover, the distance considered is short enough to exert a foreseeable discursive parallelism, unlike other approaches to this factor which take into account much larger distances (Pickering and Ferreira 2008: 447), although, as we see it, they are more problematic.
**13** Except in the twentieth century, where this difference is not significant for *haber*.

reales asegurados (*Al recibo de esta*, 1795)
[And I believe that he should also be charged for this credit that he has retained without justification. If the war goes on I shall have to pay the fixed amount of money]

With the verb *tener*, however, the presence or absence of these alternative modal verbs does not make any difference, and so for the statistical analysis the two contexts have been grouped together.

### 5.2.7 Tenor

The combination of different stylistic parameters, which we referred to earlier (Section 4), allows us to divide the tokens of the variable into two large groups. On the one hand, we have the more *personal* correspondence, of a private or intimate nature, with close ties between sender and addressee. In the majority of cases, these are members of the same family, but letters exchanged between close friends or lovers also falls into this category. The case of (33) is a representative example of this kind of letters, in this case sent from Havana in the late eighteenth century by a man who was writing to his wife in Spain asking for the family to be reunited. On the other hand, we have letters dealing with non-personal matters and those with a clear distance along the axes of familiarity and solidarity, such as letters sent by commoners to their superiors (or vice versa), to members of the clergy or nobility (or vice versa), or, in general, to social superiors (or vice versa). This correspondence (encoded as *distant*) is assumed to be further away from the colloquial register than the previous one. Fragment (34), taken from the correspondence between the Count of Tendilla and one of his subordinates at the beginning of the sixteenth century, illustrates well this second stylistic context:[14]

---

**14** Considering the content of the corpus, it is understandable that the *personal* pole is represented better than the *distant* one, especially in the twentieth century (with 76 % of the tokens), although these differences are smaller in the eighteenth (62 %) and the sixteenth centuries (53 %). Nevertheless, none of these figures represent a serious problem of evenness or imbalance in the data. On the other hand, with a few exceptions, most of the correspondence analysed in the corpus was written by an author to an addressee with whom he or she kept a balanced position on the axes of power and intimacy. In many cases, this addressee is the same person all the time, as occurs, for instance, in the correspondence between a father (or mother) and his or her beloved son (or daughter), who immigrated to America (or vice versa) in search of a better life. Even when we find writers who send letters to several addresses, they are normally

(33)  Luego no digas que tu marido es un pícaro, y que no cumplo con lo que
      Dios manda. No *tienes que traer* ningunos sino tu cama y algunos trastos
      para aderesar tus paredes (*La emigración en tinta y papel*, 1786)
      [But then do not say your husband is a scoundrel and I fail to comply with
      God's will. You do not have to bring anything except your bed and a few
      things to adorn your walls]

(34)  Dezildes a esa buena gente que huelguen y aya plazer que el rey y la
      reyna, nuestros señores, los conocen y saben su buena voluntad [...] *tengo
      de creer* dellos syno todo bien (*Epistolario del conde de Tendilla*, 1504)
      [Tell those good people to rest and to take pleasure from the fact that the
      king and queen, our lord and lady, meet them and know of their good will
      [...] I cannot but think well of them]

The stylistic factor also reveals differences between the modal periphrases. After
not being significant in the first two periods considered in the study, it makes
itself felt strongly for *tener* in the twentieth century, when the verb is preferred
in the more spontaneous contexts (0.60). These results, in accordance with
others from the sociolectal matrix that we will discuss later (Section 5.4), suggest
that at that time we are dealing with the favourite variant in more spontaneous
communication. A different stylistic pressure is perceived, however, with *haber*:
while it is the unmarked periphrasis in the deontic sphere during the sixteenth
and eighteenth centuries, its presence is also especially favoured in more
informal and spontaneous contexts (sixteenth century: 0.62; eighteenth century:
0.63). Yet, after reaching its decline in the first half of the twentieth century,
*haber* loses ground as an active variant, reducing the contexts in which it can be
found to the more formal and distant ones (0.67 versus 0.33 in more personal
settings). Lastly, the periphrases with *deber* again show that they are addressing
different forces, since they are encouraged in formal contexts in the first cen-
turies, but in the twentieth century the opposition is no longer operational.

## 5.3 Distributional patterns: markedness and frequency

A review of what has been noted up until now allows us to see that the majority
variants in each period are favoured preferably in unmarked contexts, which at
the same time are more frequent in discourse. Furthermore, although with the

---

situated at the same pole in the above-mentioned axes. This fact prevents the possibility of
gauging potential stylistic differences derived from a more varied type of recipients.

odd exception, the opposite occurs with the minority variants: they either do not respond to the incidence of a factor group or they are preferably selected in marked contexts, which are less common in discourse.

When we speak of high or low frequency of contexts, we do so in relative terms, comparing them with possible alternative settings. Moreover, these differences in frequency may be due to general properties of the discourse itself or might depend on the genre analysed. In the first case, for example, we know that, generally speaking, there are always many more affirmative sentences than negative ones and in our corpus this is overwhelmingly true: whereas negative sentences account for barely 10 % of the total number, affirmative ones represent the remaining 90 %. The same can also be said of the opposition between active and passive/impersonal sentences. In contrast, the external modality meanings do not necessarily have to be objectively more frequent than those of subjective modality, but they are in our corpus, given its structural characteristics. As the reader will recall, most of the texts are of an epistolary nature, in which a sender writes to one or several addressees on whom he or she often wishes to exert an influence by means of his or her ideas, reflections or mandates in a variety of circumstances (emigration, exile, etc.). The same occurs with the stylistic axis: the fact that most of the letters in the study deal with private affairs and between people with a close relationship results in the *personal* sample being considerably larger in the corpus than the *distant* one.

There are a number of different examples of the mentioned rule, by which the more frequent forms are in turn selected preferably in the more recurrent settings. Hence, we have seen that external modality favours the periphrasis with *haber*, which was predominant in the sixteenth and eighteenth centuries, but also those that replace the unmarked periphrasis in contemporary Spanish, as it is the case of *tener que*. Conversely, the opposite senses, of a non-external nature, are the areas in which either certain alternative variants find a space (the case of *deber* in the three centuries), or those that have reached an important degree of obsolescence remain confined. The latter is fundamentally the case of *haber de* in the twentieth century, a time when the periphrasis becomes specialised in contexts that are less prototypically obligational and considerably less frequent in discourse.

The same correlation is found among other factors where the distinction between contexts that are more frequent than others is operational. Thus, we have seen that, from the very start, first person subjects clearly favour a novel variant such as *tener*, which back in the sixteenth century began to vie for this small area of the grammar with the all-powerful *haber*, which was far more favoured in the remaining personal paradigms. It is true that the influence of the first person upon *tener* is maintained even in the twentieth century, when it

becomes the most frequent variant, but, as we have seen, in this moment we witness two important novelties: on the one hand, the positive influence is not limited to just the first person singular, but covers all the referential contexts in which the 'I' of the utterance is included; on the other hand, the influence of the factor also becomes weaker, so that the distance between the explanatory weights of the two contexts (first person / Others) is now considerably lower than in the previous centuries.

As regards polarity, we have also seen that the overwhelmingly predominant affirmative sentences are fertile ground for *haber* during its long period of dominance (sixteenth and eighteenth centuries), while in its decline, in the twentieth century, the factor is no longer operational. In contrast, *tener* starts to spread especially in the minority negative contexts, both in early Classical Spanish and in the Age of Enlightenment, but it eloquently changed the direction of this influence in the twentieth century, when it has appropriated many of the domains of the deontic modality. From this point on, the affirmative polarity will be the most favourable setting.

Finally, we find new proof of this correlation on the stylistic axis: *haber* is predominant in more spontaneous and personal contexts in the sixteenth and eighteenth centuries, when it is the unmarked variant, while it completely changes that direction in the moment of its decline (twentieth century), when it takes refuge in the more formal sphere. The opposite occurs with *tener*, which goes from being a variant that is not affected by the stylistic axis in the early centuries to being the preferred variant in more intimate and spontaneous settings in contemporary Spanish. The third variant, *deber* – which is always in a minority, either with respect to *haber* in the Classical and Enlightenment periods or with respect to *tener* in current Spanish – shows two stylistic profiles: either it is favoured in distant contexts in the sixteenth and eighteenth centuries, or it does not react to this factor groups in the twentieth century.

Certainly, there are some exceptions to these variationist patterns, as occurs with the influence of structural priming. The attraction resulting from the recycling and reutilisation of recently used linguistic material or, as a complement to this, the dissimilatory effect of the alternative forms in the previous cotext is a sufficiently relevant cognitive factor as to revert the above-mentioned rule. This explains why both contexts, which are very restricted in the corpus, are the ones that now clearly favour and disfavour, respectively, each and every one of the variants analysed.

Another exception is that represented by factors such as agentivity and the degree of impersonality in the specific case of *tener*. As we have seen, for a very long time modal periphrases with non-human subjects were practically "off-limits" for this verb. They account for a lower proportion in the corpus (20–25 %)

with respect to the human ones, which comes as no surprise if we bear in mind that obligation is an entity that is prototypically associated with human activity and can only be attributed to non-human entities by extension. The same can be said of the degree of impersonality: *tener* appears exclusively in active sentences in the sixteenth and eighteenth centuries, and even in the twentieth century, when its periphrases have clearly imposed themselves, its presence in impersonal contexts is very scarce (10 %), as these are the favourite context of *haber* (36 %) and, even more so, *deber* (54 %).

## 5.4 Social diffusion of the periphrases in the twentieth century

As noted above (Section 4), in the analysis of the twentieth century we also include sociolectal information, the results for which are presented in Table 5 and summarised in the following.

From the sample we have available, the social factors with the highest explanatory power are social status and age. In the first case, the factor shows a linear stratification of *tener* that is intimately associated with the writers' sociolect.[15] Thus, the periphrasis is preferably driven by lower class individuals (0.62; 68 %), followed by those from the middle class (0.47; 50 %). In contrast, the members of the more cultured classes show significantly lower indexes of the variant (0.40; 34 %). This is exactly the opposite pattern to that seen in the case of *haber*, the classic variant, which is clearly in decline during the first half of the twentieth century, as can be seen graphically in the figure 2.

---

**15** For the regrouping of the individuals in the sample according to their social status, I follow a classical Weberian classification of the social classes based on the interaction among variables such as wealth, power and prestige. Moreover, in the Spain of the first half of the twentieth century, the limits between these social classes were far more marked than today. Thus, the representatives of the High group corresponded to the social-economic and cultural elites of the country, that is, people who, in many cases, inherited this condition at birth. In the opposite pole, we find the bottom of society (Low), a large portion of the population composed of people who carried out manual, non-prestigious and poorly paid jobs. In fact, many of these impoverished people were still compelled to migrate to America in search of a better life, as in the previous centuries. Finally, between the two extremes mentioned we find an incipient, mostly urban, middle class made up of skilled occupations such as white collar workers, small traders and liberal professions.

**Table 5:** Distribution of the modal verbs *tener, haber* y *deber* in three social factor groups (twentieth century).

| | | Tener | | Haber | | Deber | |
|---|---|---|---|---|---|---|---|
| | | FW. | % | FW. | % | FW. | % |
| **Social status** | High | 0.40 | 34 | 0.67 | 26 | – | [40] |
| | Medium | 0.47 | 50 | 0.57 | 21 | – | [29] |
| | Low | 0.62 | 68 | 0.27 | 6 | – | [26] |
| **Age** | Youths | 0.61 | 69 | – | [12] | 0.37 | 19 |
| | Adults | 0.38 | 38 | – | [19] | 0.63 | 41 |
| **Gender** | Men | – | [51] | – | [16] | – | [33] |
| | Women | – | [63] | – | [11] | – | [26] |
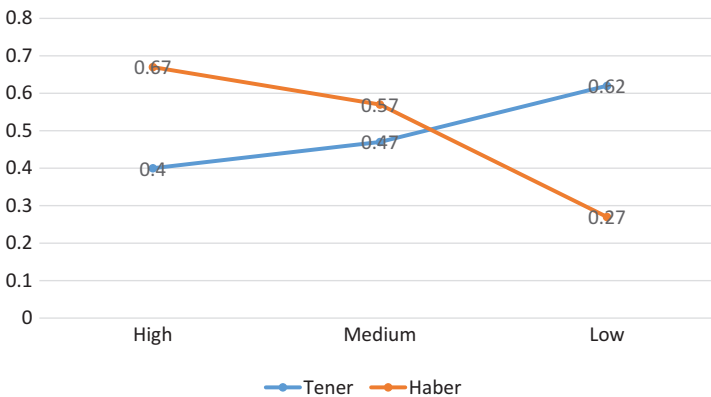


**Figure 2:** Factor weights associated with the selection of the modal periphrases with *tener* and *haber* in the twentieth century.

In contrast, the differences found in *deber* are not significant, despite the fact that the indexes of usage among the uppermost class (40 %) are some way ahead of the other groups (middle: 29 %, low: 26 %).

In fact, the reason for this lack of significance is to be found in the interaction with age,[16] which in return is now really significant to explain the variation of *deber*. Indeed, the preference for this verb is significantly apparent among

[16] This factor was configured on the basis of an emic – not always chronological – interpretation, given the difficulties in determining the precise age of some writers, whose biographical data are unknown to us. Nonetheless, even in those cases, we can often deduce the period of

adult age groups (0.63; 41 %), at a considerable distance from the younger individuals (0.37; 19 %), regardless of their status. This is just the opposite of what occurs with *tener*: in this case, it is the younger individuals who lead in the use of the variant in expansion (0.61; 69 %) versus the greater conservatism of the groups made up of older individuals (0.38; 38 %). In contrast, the relative differences in age within the *haber* variation are much smaller and not at all significant.

Specific mention should be made of gender, which is not initially selected as an independent factor, although this may well be related to the extraordinary imbalance between men ($n = 1130$; 90 %) and women ($n = 117$; 10 %) tokens in the sample.[17] In any case, with the data available, it is worthwhile highlighting the scarce weight of the differences found for the verbs *haber* and *deber*, in contrast to *tener*, a verb that was encouraged more by women (63 %) than by men (51 %). Although these differences are not significant as an independent factor according to the regression analysis, it must be remarked that in the cross-tabulation with social status and with age, the women in each subgroup always exceeded their male counterparts. Particularly striking, in this respect, are the results among lower class (77 %) and younger (76 %) women, with very advanced values that are clearly above the mean and far from those of the remaining subgroups.[18] These data, in addition to the preference for the variant noted earlier among the low strata, the younger age groups and the more intimate and spontaneous contexts, could suggest a pattern of change from below in the social diffusion of *tener que* (further implications on this type of change will be discussed later in subsequent sections).

---

their lifetime in which they found themselves at the moment of writing – reduced here to either youthfulness or maturity – on the basis of the details given in the letters. It goes without saying that, in the absence of either direct or indirect information, the corresponding tokens have been discarded from the statistical analysis of this factor group, treating them as missing data.

**17** Neither the *geographical origin* of the writers nor the *migratory* context in which the texts were written (America/Spain) have been found to be decisive for promoting any modal verb to a significant extent. In the case of the former factor group, this lack of significance remains the same when the analysis is performed across both the North-South and the East-West axes. I thank one of the reviewers for pointing out this last possibility, which, nevertheless, has not been found to be relevant with these data.

**18** These are the percentages after cross-tabulation between *Age* and *Gender*: Younger-Men (68 %), Younger-Women (76 %); Adult-Men (38 %); Adult-Women (56 %). And these are the results of cross-tabulation with the *Social status*: High-Men (33 %), High-Women (38 %), Medium-Men (51 %), Medium-Women (67 %), Low-Men (69 %), Low-Women (78 %).

## 5.5 The history of three modal verbs

A thorough analysis of deontic modal periphrases at different points on the time axis allows us to obtain an accurate overview of the development of each of them over the almost five centuries that separate early Classical Spanish (sixteenth century) and contemporary Spanish (first half of the twentieth century).

If we start with *haber*, the modal verb par excellence throughout most of the history of Spanish, we can see how its position is one of absolute dominance in the sixteenth century, when its frequencies of use are far higher than those of the other two verbs, both in general terms (74 %) and in the vast majority of the linguistic and stylistic contexts considered. At that time, its selection is either unrelated to most of the factors considered or, if this is not the case, its presence is favoured further still in the more unmarked and recurrent contexts, whether we are dealing with external obligation, most of the grammatical persons (except for the first person singular), affirmative sentences or the most spontaneous communication that lies closest to the colloquial register. Nevertheless, some restricted contexts already act like small fissures that will gradually erode that dominant position in the coming centuries. In the sixteenth century this occurs in some minority contexts, such as the first person singular (37 %), negative sentences (54 %) or non-external obligation (56 %), whose figures, although objectively high, are clearly below the mean.

These chinks were to progressively weaken the verb in the eighteenth century, where the competition against the other periphrases – especially *deber* – increased. Thus, although *haber* is still the most frequent periphrasis in the deontic sphere in the Age of Enlightenment, it has already lost a good proportion of its past strength. Its overall frequency of use diminished significantly (53 %) and the same happened in each and every one of the contexts considered. Nevertheless, the internal grammar shows a considerable degree of stability, so that, with the odd exception, most of the significant factors that operated in the past continue to do so in this period and, furthermore, in the same explanatory direction.

This panorama, however, was to change dramatically in the early twentieth century. At that time, the verb *haber* had lost a large part of the power it had previously enjoyed, while also displaying some signs of grammaticalisation, as can be deduced from: a) its taking refuge in the less prototypically obligational modal nuances (phatic nuances, desire, intention, etc.); b) the total loss of explanatory relevance of certain factors that acted in the past (person, polarity, external/internal modality opposition, agentivity); c) inversion of the explanatory direction of the few that were left standing (taking refuge in the more formal spheres of communication, in contrast to the past); and d) the appearance of

new factor groups that were not involved in the past (preference for passive and impersonal sentences). Moreover, the preference for more formal registers is complemented by the incidence of social status. Thus, we have seen that the lower classes use *haber* less often, even in more distant communication.

The evolution of *tener* is largely the opposite and complementary to that observed with *haber*. In the sixteenth and eighteenth centuries it is no rival at all for this last verb, as shown by the clearly minority frequencies of use (around 10 %). Furthermore, in addition to this apparent stabilisation over two hundred years, another no less telling one also takes place within the internal grammar. The quantitative analysis reveals that the variation of *tener* is conditioned by almost the same factors and with the same explanatory direction. The variant is therefore significantly more favoured in both centuries within contexts involving the presence of the same verb in the previous co-text (priming), omitted subjects, the first person singular and negative sentences. Nonetheless, the modality begins to show certain relevant changes with respect to the past, which will go on to become firmly established later on. Thus, we have observed how, in the eighteenth century, and despite displaying rather modest frequencies of use, the periphrasis with *tener* begins to show a special preference for the expression of unavoidable obligations.

This specialisation will become firmly established in the twentieth century, when no less than 76 % of the utterances of these senses found in the corpus are already encoded by means of this auxiliary verb. Moreover, this consolidation coincides with the period of greatest splendour of the periphrasis, which becomes very common in most of the contexts that had previously been the privileged territory of *haber*. Hence, *tener que* becomes not only the main conveyor of unavoidable duties, but also of the obligations caused by an external agent, which had been preferentially associated with *haber de* in the past. Likewise, some factors appear for the first time or change their explanatory direction to adapt to the new times and to contexts that are more in keeping with the triumphant variant. Thus, while in the past *tener* was especially encouraged in the minority negative sentences, that place was now taken up by the majority affirmative ones. In the same way, the periphrasis becomes firmly established as the unmarked variant in the more spontaneous communication associated with the more personal and intimate letters. Yet, unlike the process undergone by *haber*, whose original conditioning patterns change almost radically five centuries later, the periphrasis *tener que* remains constant in some points of its internal grammar. Particularly striking, in this regard, is the consolidation in this long period of the preference for omitted subjects and the first person, even though both of them display some signs of weakening in the last period.

Lastly, from the sociolectal point of view, *tener* exhibits a clear stratification linked to social status, age and (provisionally) sex. We have found that the use

of the periphrasis *tener que* is particularly encouraged among members of the lower strata, the young population and women, above all where these groups intersect. All of this, together with the above-mentioned preference for more spontaneous and colloquial contexts, allows us to describe the diffusion of *tener que* in the twentieth century as an example of change from below (Chambers 1995; Labov 2001; Tagliamonte 2012).

In contrast to the very often complementary distribution between the verbs *haber* and *tener* throughout history, the case of *deber* displays a different distributional profile. Over the period from the sixteenth to the eighteenth centuries, the verb undergoes a substantial increase in frequency within the deontic sphere, mainly at the expense of *haber* (sixteenth century: 16 %; eighteenth century: 39 %). Nevertheless, unlike *tener*, which burst onto the scene in the twentieth century, *deber* seems to become stabilised or even to decline slightly in contemporary Spanish (32 %). Over these five centuries, however, *deber* has become specialised as a modal verb associated with subjective obligations, a meaning to which its periphrases are seen to be particularly sensitive in the three periods analysed. Thus, during the first centuries (sixteenth and eighteenth), *deber (de) + inf.* is seen to be a periphrasis that is favoured more in more formal contexts, in contrast to the more spontaneous nature of *haber*, although like this latter – and unlike *tener* – it shows a special inclination for subjects other than the first person. Yet, both conditioning factors cease to be operational in the twentieth century. In exchange, within this last period, *deber* also coincides with *haber*, both of which are now minority, in finding a significant niche in restricted contexts, such as negative sentences, explicit subjects and passive and impersonal sentences. Finally, from the sociolectal point of view, in the twentieth century *deber* does not appear to be sensitive to sex or status but it is to age: unlike *tener*, which is clearly preferred by the youngest age groups, and the absence of genolectal variation of *haber*, it appears above all in the speech of the older speakers.

# 6 Conclusions

Variation within the infinitive modal periphrases has been conditioned by a handful of common factors of a syntactic, semantic and discursive nature over almost five centuries. Nonetheless, with the exception of structural priming, which is active all the time and with the same explanatory direction, the other factors condition each verb in a different way. Furthermore, the sense of this variation sometimes changes as time goes by, with especially relevant

alterations in the twentieth century. Therefore, the external/non-external modal opposition explains the variation of *deber* over five centuries, when this verb becomes specialised as a means to express subjective obligational contents, which is just the opposite to *haber*, a periphrasis preferentially associated with external obligation up until the twentieth century, when it was to take refuge mainly in the more peripheral modal contents. On the other hand, *tener* has become progressively more specialised as a means to express unavoidable obligations and, as of the twentieth century also external obligation, thereby taking over from *haber*. Likewise, in the case of *tener* the constant association with the first person and omitted subjects is significant. Other factors display an incidence that changes over time and which affects the modal verbs in different ways. For example, *haber* and *tener* exhibit complementarity when factors like polarity and style are involved. While *haber* has been the unmarked variant in affirmative sentences and in spontaneous communication throughout much of history, in the twentieth century it takes refuge in the opposite contexts (negative sentences and more formal contexts), as did *deber*, and exactly the contrary to *tener*. Finally, mention should also be made of some strong constraints related to the degree of agentivity and impersonality: during the sixteenth and eighteenth centuries, *tener* was very rarely used to express modal contents with non-human and impersonal subjects.

Furthermore we have detected the existence of a notable association between these constraints and the degree of markedness and the frequency of the conditioning contexts. On the whole, a rule is pervasively fulfilled: with the odd exception, the most frequent variants in a certain period are especially favoured in the most unmarked and frequent contexts, and vice versa: the alternate forms do the same in the minority contexts. It is known that the more frequent forms are more resistant to change due to their higher cognitive entrenchment. As stated by Bybee and Thompson (2000: 380): "the more a form is used, the more its representation is strengthened, making it easier to access the next time". This preserving effect has been demonstrated on different levels of analysis, regardless of whether we are dealing with lexical (Bybee 1985; Langacker 1987) or syntactic material (Givón 1979; Croft 2000; Bybee and Hopper 2001). What this study shows is that this same effect also operates at the level of the morphosyntactic, semantic, pragmatic and stylistic settings in which the variants are used in discourse. From a theoretical perspective, these data give support of a usage-based approach to language change in which cognitive processes such as entrenchment have a decisive role (Schmid 2012). In essence this is understandable if we bear in mind that the most frequent contexts "allow specific information about instances of use to be retained in representation" (Bybee 2006: 717). Consequently, this would explain how factors

like external modality, affirmative polarity, active sentences, persons other than the first person or more intimate and spontaneous communication have a positive influence on the selection of *haber* throughout much of history, as well as on that of *tener* from the twentieth century onwards, when this modal verb takes over as the modal verb par excellence. In contrast, both *tener* in the sixteenth and eighteenth centuries and *deber* throughout the period under analysis are preferentially associated with the alternative contexts, as the minority variants they are. And the same occurs with *haber* when it reaches its decline in the twentieth century.

The foregoing allows us to trace a particular profile for each of the periphrases in the history of Spanish, in which *tener* and *haber* undergo what is largely a complementary distribution, as regards both the frequencies of use and their variable conditioning.[19] After this latter having been the most characteristic verb in the deontic sphere for centuries, it went into severe decline in the twentieth century, when it was confined to minority contexts that were the opposite of those that encouraged it in the past, such as negative polarity, passive and impersonal sentences, less prototypically obligational modal contexts or more formal communication. In contrast, the case of *tener* is unique because, after centuries of stabilisation in very moderate figures, it appears in the twentieth century, emerging as the most frequent variant in most contexts, including some that were out of bounds to it in the past. Moreover, from the sociolectal point of view it can be seen how this favourable evolution has been diffused as a change from below, especially encouraged on the lower social levels, in the younger age groups and among women. The fact that these latter make a decisive contribution to the change, exceeding men at all the intersections, may mean that, as in the first half of the twentieth century, *tener que* could already be seen as the new triumphant variant in ordinary communication, in line with what is observed regarding the role of women in similar processes of change (Labov 2001). This is exactly the opposite of what happens with *haber*, which is relegated to the most formal communication and very seldom heard in the speech of the lower social groups, and now with very few differences in terms of age and gender.

Lastly, the third verb, *deber*, displays some specific features that draw it away from the recently observed complementary distribution. Throughout the

---

**19** In this sense, the replacement of *haber* by *tener* seems to be part of a more general change in the history of Spanish, not limited to deontic periphrases, but also present in the domain of possession (Pountain 1985), as an auxiliary with the past participle (e. g. Doroga 2013) and as a temporal modifier (in this case, replaced by *hacer, hace tres años que vivo aquí* vs. ha tres años que vivo aquí (e. g. Howe and Ranson 2010). I thank one of the reviewers for this reminder.

whole of the period analysed it is seen as a verb preferably associated with internal obligation, while it shows certain signs of weakening from the eighteenth century onwards. This would explain why its figures remain almost constant in the last two centuries, as well as the fact that its selection is encouraged, as the minority variant it is, in certain marked contexts (negative sentences, explicit subjects and passive and impersonal sentences).

This ends our account of the historical evolution of the infinitive modal periphrases over almost five centuries. But that history continues and the question remains as to what has happened in the deontic sphere over the last fifty or sixty years, from the end of the period analysed in this study to the present day, and for which, fortunately, we have oral corpora available to us. Has the process of retrocession of *haber de* and the consolidation of *tener que* as the unmarked periphrasis of obligation continued in Spanish? What role is played in all this process by what has been shown to be the singular variant, *deber*? To what extent can having access to oral and not written texts influence the variation, however close these latter come to the pole of communicative immediacy, as is the case of private letters? These are undoubtedly interesting questions, but unfortunately cannot be addressed here and must be left for a future study.

# References

Balasch, Sonia. 2008. *Debe (de) ser*: Evolución de la variación. In Maurice Westmoreland & Juan A Thomas (eds.), *Selected proceedings of the 4th workshop on Spanish Sociolinguistics*, 109–119. Somerville, MA: Cascadilla Proceedings Project.

Balasch, Sonia. 2012. Revisión amplia de la variación *deber (de)* + infinitivo en la diacronía del español. In Emilio Montero (ed.), *Actas del VIII Congreso Internacional de Historia de la Lengua Española*, 607–620. Madrid: Arco Libros.

Beardsley, Wilfred A. 1921. *Infinitive constructions in old Spanish*. New York: Columbia University Press.

Bergs, Alexander. 2005. *Social network analysis and historical sociolinguistics*. Berlin: Mouton de Gruyter.

Blas Arroyo, José Luis. 2014. La alternancia *deber/deber de* + infinitivo en la tradición grama- tical española: Entre la distinción funcional y la variación libre. In María Luisa Calero, Alfonso Zamorano, Francisco J. Perea, María Del Carmen GarcíaManga & María Martínez Atienza (eds.), *Métodos y resultados actuales en Historiografía de la Lingüística*, 378–388. Münster: Nodus.

Blas Arroyo, José Luis. 2016. The rise and fall of a change from bellow in the Spanish syntax of golden age period: Deber & deber de + infinitive in communicative immediacy texts. *Journal of Historical Linguistics* 6(1). 1–31.

Blas Arroyo, José Luis & Juan González. 2014. ¿Qué *tengo que/de hazer*?': Variación y cambio lingüístico en el seno de las perífrasis de infinitivo a partir de textos escritos de impronta oral en el español clásico. *Studies in Hispanic and Lusophone Linguistics* 7(2). 241–274.

Blas Arroyo, José Luis & Margarita Porcar. 2014. De la función a la forma: La influencia del contexto variable en la selección de las perífrasis modales de infinitivo en el primer español clásico. *Anuario de Lingüística Hispánica* 30. 9–49.

Blas Arroyo, José Luis & Margarita Porcar. 2016. Patrones de variación y cambio en la sintaxis del Siglo de Oro. Un estudio variacionista de dos perífrasis modales en textos de inme- diatez comunicativa. *RILCE. Revista de Filología Hispánica* 32(1). 47–81.

Blas Arroyo, José Luis & Kim Schulte. 2017. Competing modal periphrases in Spanish between the sixteenth century and the eighteenth century: A diachronic variationist approach. *Diachronica* 34(1). 1–39.

Blas Arroyo, José Luis & Javier Vellón. 2015. The refuge of a dying variant within the grammar: Patterns of change and continuity in the Spanish verbal periphrasis *haber de* + *infinitive* in the past two centuries. *Language Variation and Change* 27(1). 89–116.

Bybee, Joan. 1985. *Morphology: A study between the relation of meaning and form*. Amsterdam: John Benjamins.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.

Bybee, Joan & Paul Hopper (eds.). 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Bybee, Joan & Sandra Thompson. 2000. Three frequency effects in syntax. In Matthew L Juge & Jeri L Moxley (eds.), *Proceedings of the 23rd annual meeting of the Berkeley Linguistics Society: General session and parasession on pragmatics and grammatical structure (1997)*, 378–388. Berkeley: Berkeley Linguistics Society.

Carvalho, Ana María, Rafael Orozco & Naomi L Shin. 2015. *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington: Georgetown University Press.

Chambers, Jack K. 1995. *Sociolinguistic theory*. Oxford: Blackwell.

Croft, William. 2000. *Explaining language change: An evolutionary approach*. London: Longman.

Danilova, Oxana. 2012. La hipérbole como recurso expresivo en las cartas privadas del siglo XVI. In Victoria Béguelin-Argimón & Gabriela Cordone (eds.), *En pos de la palabra viva: Huellas de la oralidad en textos antiguos. Estudios en honor al profesor Rolf Eberenz*, 195– 215. Bern: Peter Lang.

Doroga, Jason P. 2013. Late attestations of participle agreement in Spanish perfective constructions. In Chad Howe, Sarah E. Blackwell & Margaret Lubbers Quesada (eds.), *Selected proceedings of the 15th hispanic linguistics symposium*, 17–28. Somerville, MA: Cascadilla Proceedings Project.

Eberenz, Rolf & Mariela De La Torre. 2002. *Conversaciones estrechamente vigiladas: Interacción coloquial y español oral en las actas inquisitoriales de los siglos XV a XVII*. Lausanne: Centro de Estudios Hispánicos.

Elspass, Stephan. 2012. The use of private letters and diaries in sociolinguistic investigation. In Juan M Hernández-Campoy & Juan C Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*, 156–169. Malden: MA & Oxford: Wiley-Blackwell.

Fernández De Castro, Félix. 1999. *Las perífrasis verbales en el español actual*. Madrid: Gredos.

Garachana, Mar & Malte Rosemeyer. 2011. Rutinas léxicas en el cambio gramatical: El caso de las perífrasis deónticas e iterativas. *Revista de Historia de la Lengua Española* 6. 35–60.

García Fernández, Luis (ed.). 2006. *Diccionario de perífrasis verbales*. Madrid: Gredos.

García Fernández, Luis. 2013. *Las perífrasis verbales en español*. Madrid: Castalia.

Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.

Gómez Manzano, Pilar. 1992. *Perífrasis verbales con infinitivo (valores y usos en la lengua hablada)*. Madrid: UNED.

Gómez Torrego, Leonardo. 1999. Los verbos auxiliares. Las perífrasis verbales de infinitivo. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, 3223–3390. Madrid: Espasa-Calpe.

Hernández Díaz, Axel. 2006. Posesión y existencia. La competencia de haber y tener y haber existencial. In Concepción Company Company (ed.), *Sintaxis histórica de la lengua española 1*, 1055–1164. Mexico DF: Fondo de Cultura Económica de España.

Howe, Chad & Diana L Ranson. 2010. The evolution of clausal temporal modifiers in Spanish and French. *Romance Philology* 64(2). 197–207.

Johnson, D. Ezra. 2009. Getting off the goldVarb standard: Introducing rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1). 359–383.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 2001. *Principles of linguistic change: Social factors*. Oxford: Blackwell.

Langacker, Ronald. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford: Stanford University Press.

Lapesa, Rafael. 2000. *Estudios de morfosintaxis histórica del español*. Madrid: Gredos.

López Izquierdo, Marta. 2008. Las perífrasis modales de necesidad: Emergencia y renovación. In Concepción Company Company, José G. Moreno De Alba (ed.), *Actas del VII Congreso Internacional de Historia de la Lengua Española*, 789–806. Madrid: Arco Libros.

Luján, Marta. 1999. Expresión y omisión del pronombre personal. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, 1275–1315. Madrid: Espasa-Calpe.

Martínez Díaz, Eva. 2003. La frecuencia de uso de haber y tener en las estructuras perifrásticas de obligación: Algún fenómeno de variación en el español de Cataluña. *Interlingüística* 14. 681–694.

Martínez Díaz, Eva. 2008. La expresión de la modalidad de obligación en *el Corpus del español conversacional de Barcelona y su área metropolitana (2001)*. In Antonio Moreno Sandoval (ed.), *El valor de la diversidad (meta)lingüística: Actas del VIII Congreso de Lingüística General*. http://elvira.lllf.uam.es/clg8/actas/pdf/paperCLG73.pdf (accessed 2 March 2017).

NGRALE (Real Academia Española y Asociación de Academias de la Lengua). 2009. *Nueva gramática de la lengua española*. Madrid: Espasa.

Oesterreicher, Wulf. 1996. Lo hablado en lo escrito: Reflexiones metodológicas y aproximación a una tipología. In Thomas Kotsch, Wulf Oesterreicher & Klaus Zimmermann (eds.), *El español hablado y la cultura oral en España e Iberoamérica*, 317–340. Madrid & Frankfurt: Iberoamericana/Vervuert.

Oesterreicher, Wulf. 2004. Textos entre inmediatez y distancia comunicativas: El problema de lo hablado escrito en el Siglo de Oro. In Rafael Cano Aguilar (ed.), *Historia de la lengua española*, 729–769. Barcelona: Ariel.

Okulska, Urszula. 2010. Performing the world of politics through the discourse of institutional correspondence in late middle and early modern England. In Urszula Okulska & Piort Cap (eds.), *Perspectives in politics and discourse*, 173–197. Amsterdam: John Benjamins.

Olbertz, Hella. 1998. *Verbal periphrases in a functional grammar of Spanish*. Berlin: Mouton de Gruyter.

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 Heiko Narrog & Bernd Heine (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 & Sali Tagliamonte. 2001. *African American English in the diaspora: Tense and aspect*. Oxford: Blackwell.

Pountain, Christopher J. 1985. Copulas, verbs of possession and auxiliaries in old Spanish: The evidence for structurally interdependent changes. *Bulletin of Hispanic Studies* 62(4). 337–355.

Raumolin-Brunberg, Helena. 2005. Language change in adulthood: Historical letters as evidence. *European Journal of English Studies* 9(1). 37–51.

Sankoff, David. 1988. Sociolinguistics and syntactic variation. In Frederick Newmeyer (ed.), *Linguistics: The Cambridge survey*, 140–161. Cambridge: Cambridge University Press.

Sankoff, Gillian. 1990. The grammaticalization of tense and aspect in Tok pisin and sranan. *Language Variation and Change* 2(3). 295–312.

Schmid, Hans-Jörg. 2012. Entrenchment, salience and basic levels. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The oxford handbook of Cognitive Linguistics*. Oxford: Oxford University Press.

Sinner, Carsten. 2003. Valores y usos de < *haber (de)* + infinitivo > y < *tener que* + infinitivo > en diferentes variedades del castellano. In Claus D Pusch & Andreas Wesch (eds.), *Verbalperiphrasen in den (ibero-)romanischen Sprachen*, 193–206. Hamburg: Helmut Buske.

Stengaard, Birte. 2003. Función y evolución de la construcción aver + preposición + infinitivo. In José L Girón & José J De Bustos (eds.), *Actas del VI Congreso Internacional de Historia de la Lengua Española 1*, 1147–1152. Madrid: Arco Libros.

Stoll, Eva. 2002. *La memoria de Juan Ruiz de Arce*. Madrid & Frankfurt: Iberoamericana/Vervuert.

Stoll, Eva & María N Vázquez. 2011. *Alonso Borregán: La conquista del Perú*. Madrid & Frankfurt: Iberoamericana/Vervuert.

Tagliamonte, Sali A. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Malden, MA & Oxford: Wiley-Blackwell.

Tagliamonte, Sali A., Mercedes Durham & Jennifer Smith. 2014. Grammaticalization at an early stage: Future *be going to* in conservative British dialects. *English Language and Linguistics* 18(1). 75–108.

Torres Cacoullos, Rena. 2009. Variation and grammaticisation: The emergence of an aspectual opposition. In Stavroula Tsiplakou, Marilena Karyolemou & Pavlos Pavlou (eds.), *Studies in language variation: European perspectives 2*, 215–224. Amsterdam: John Benjamins.

Torres Cacoullos, Rena & James A. Walker. 2009. The present of the English future: Grammatical variation and collocations in discourse. *Language* 85. 321–354.

Travis, Catherine. 2007. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19(2). 101–135.

Yllera, Alicia. 1980. *Sintaxis histórica del verbo español: Las perífrasis medievales*. Zaragoza: Universidad de Zaragoza.