

## Competing modal periphrases in Spanish between the 16th and the 18th century: A diachronic variationist approach

José Luis Blas Arroyo & Kim Schulte  
Universitat Jaume I (Castellón)

The history of Spanish modal constructions has been widely discussed in the literature, focusing primarily on the semantic differences between the available alternatives. This paper offers an innovative analysis of the evolution of these constructions by adopting a diachronic variationist approach that takes into account a wider range of semantic, syntactic, morphological and stylistic factors that influence the choice between the competing modal periphrases during two key stages in the evolution of Spanish. The data is drawn from a diachronic corpus of personal correspondence, reflecting actual language usage during the respective periods as closely as possible. Particular attention is paid to the question of whether the influence of different factor groups remains stable over time or not, and it is shown that the most frequent form–context pairings are particularly resistant to innovation, which can be explained by cognitive entrenchment of the respective variant in specific linguistic environments.

**Keywords:** variationist linguistics; modal constructions; Spanish; epistolary corpus; communicative proximity (*Nähesprache*); frequency effects; cognitive entrenchment

### 1. Introduction\*

It is generally accepted that most linguistic changes proceed gradually and over an extensive period of time, during which different forms compete with each other, vying for dominance and spreading to different areas of the grammatical system at different rates and to a different extent.

A better understanding of the details of this type of process can be gained by means of a variationist approach which makes it possible to obtain genuine insights into what is going on within the grammatical system at different stages during the period in which two or more variants compete for dominance in a specific area of grammar (Poplack & Tagliamonte 2001; Poplack 2011; Torres Cacoullos 2009). This approach is based on the idea that the patterns and principles underlying such changes can be identified by means of a quantitative analysis of the competing variants in the different environments that, together, constitute the VARIABLE CONTEXT of these forms (Poplack 2011: 212).

While the variationist approach has become indispensable for the analysis of recent and ongoing language change, it has so far only made fairly small inroads into traditional historical linguistics, though its methodological innovations are equally important for the study of changes that lie further in the past. In particular, the PRINCIPLE OF ACCOUNTABILITY, which requires us to establish not only how the variable context conditions the use of a particular linguistic variant, but also how it conditions the alternative forms within a given subsystem of grammar (cf. Labov 1972: 72), is essential for a complete analysis, as without taking all available alternatives into account, an apparent link between a particular variant and a particular function might cause us to draw incomplete or even incorrect conclusions.

---

\*We would like to thank Carme Barberà, Susana Martínez, Maria Chiara Marullo and Jordi Ayza for their contribution to this project. We also thank the editor and the anonymous reviewers for their insightful and constructive comments. Any errors are the sole responsibility of the authors.

The application of this principle is a key element of the study presented in this article, which offers a comparative analysis of the usage of the modal periphrasis [*haber de* + infinitive] in the 16th and 18th centuries, which alternates with two other constructions in the same semantic domain: [*deber (de)* + infinitive] and [*tener de/que* + infinitive]. While these competing constructions are not entirely identical in semantic terms, there is a great deal of overlap, and they have, over the centuries, been used as alternative variants to express similar modal meanings, though their usage frequencies have changed over time. In line with Sankoff's (1988) hypothesis of the neutralisation of potential semantic differences between different forms in discourse, we can identify the periphrases exemplified in (1)–(3) as variants of the same syntactic variable, used to express verbal modality:

- (1) *Hija, en cuanto a mis cuidados, doite aviso, de lo que has de hacer, con Pedro Salvador, ya va informado de lo que has de hacer. (Cartas desde América, 1728)*

“Daughter, regarding my affairs, I am giving you notice of what you must do, with Pedro Salvador, [who] has already been informed of what you must do.”

- (2) *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 ienes 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.”

- (3) *... 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, though back then he was a vicar – for I have requested this gentleman to let you know what you must do for me.”

It can be seen that in all three sentences, taken from a corpus of 18th-century epistolary texts, the modal periphrases (underlined) have a lot in common: in all cases, (a) the main verb is *hacer* “to do”; (b) the type of modality they express is deontic, describing an obligation imposed by someone else (command/order); (c) the grammatical person and number (2SG) is the same; (d) the tense and mood (present indicative) is also the same; (e) the clauses containing the periphrasis are affirmative in terms of polarity and (f) active in terms of voice; and (g) the relationship between writer and addressee, as well as the subject matter of the letter, can be classified as personal, since we are dealing with correspondence between family members.

Despite the fact that there is much functional similarity and overlap between these three periphrases in discourse, existing studies examining their diachronic development tend to leave some important questions unanswered. Most quantitative analyses tend to be limited to the variation between the periphrases with *haber* and *tener* (e.g., Martínez Díaz 2003; López Izquierdo 2008), overlooking that the construction with *deber* is a further competitor in the area of verbal modality. As a result, the observations and conclusions regarding the changes in the usage of modal constructions provide an incomplete picture, according to which [*haber de* + infinitive] is practically the only variant used until the beginning of the 20th century. However, as will be shown in this study, quite a different picture emerges if [*deber de* + infinitive] is included in the analysis; though [*haber de* + infinitive] is still the numerically dominant option in the 18th century, there is evidence of strong competition from the alternative variants, especially in certain structural and stylistic contexts.

Furthermore, the existing studies generally provide a straightforward frequency analysis that does not take into account the variable context of these constructions. For instance, [*haber de* + infinitive] appears both in modal and temporal (future) contexts, while the use of the alternative periphrases is limited almost entirely to modal contexts, as we shall see below. As far as we are aware, there are no detailed grammatical studies of the diffusional pathway of this change beyond the simple analysis of global usage frequencies.

One of the objectives of this paper is, thus, to compare the patterns of variation in the two periods, the 16th and the 18th centuries, in order to identify language-internal and socio-stylistic factors that either favour or disfavour the progressive replacement of [*haber de* + infinitive] by its competitors, and to examine the explanatory hierarchy among the most relevant factor groups, as well as the direction of the effect within these factor groups. It is important to establish to what extent the underlying grammar changes between these two periods, which are generally considered to be key stages in the evolution of the Spanish language: the 16th century represents the beginning of Classical or Golden Age Spanish, with significant differences to the medieval language, while the 18th century marks the transition from Classical to Modern Spanish.<sup>1</sup>

As will be shown in §6, contrary to what is suggested in some descriptive studies (e.g., López Izquierdo 2008), the periphrasis [*haber de* + infinitive], which is the dominant modal construction during earlier stages of the language, has already lost considerable ground to the variants [*deber (de)* + infinitive] and, to a lesser extent, [*tener que* + infinitive]<sup>2</sup> in 18th-century Spanish.

From a more theoretical perspective, this paper provides data in support of a usage-based approach to language change in which cognitive processes such as entrenchment have a decisive role (cf. Schmid 2012; Croft 2001: 28); this is particularly apparent from the fact that, in many cases, the most frequent contexts tend to favour the use of the older, more frequent and more established variant [*haber de* + infinitive], whereas low-frequency contexts are typically the point of entry and expansion for the newly emerging alternative.

Before presenting the data, a brief historical overview of the origin and evolution of the Spanish auxiliary constructions with the verbs *haber*, *deber* and *tener* is provided in §2, followed by a description of the corpus in §3. §4 examines the general distribution of the three periphrases in the corpus, which is followed by a discussion of some important methodological issues in §5. After the presentation and analysis of the data in §6, the most important conclusions of this study are summarised and some theoretical implications discussed in §7.

## **2. The origin and evolution of the periphrases with *haber*, *tener* and *deber* as auxiliary verb**

In Latin, [*HABERE* + infinitive] is already used to express various types of deontic modality in the first half of the 1st century, for example in the works of Seneca the Elder (Hertzenberg 2012). While there is, initially, no linking particle between the auxiliary and the main verb of this construction, as early as the Late Latin and Early

---

<sup>1</sup> Note that the term ‘Early Modern Spanish’ is generally used to refer to Classical Spanish by Anglophone hispanists, while Spanish philologists consider the beginnings of ‘Modern Spanish’ to lie in the 18th century. To avoid confusion, the term will not be used in this article.

<sup>2</sup> In the 18th-century corpus used for this study, the presence of the now extinct variant [*tener de* + infinitive], which was occasionally used as an alternative to [*haber de* + inf.] in Medieval and Classical Spanish, is already minimal.

Romance periods the prepositions *a* and *de* appear in this periphrasis. In medieval Spanish, the three variants [*aver* + infinitive], [*aver a* + infinitive] and [*aver de* + infinitive]<sup>3</sup> can be found in all text types and registers; the meaning of all three constructions ranges from clearly temporal (future) to clearly deontic (obligation), though in some cases such a clear distinction is not possible, given the strong semantic link between obligation and future reference.

The construction without a prepositional linker, [*aver* + infinitive], was infrequent even in medieval times, used rarely in the 12th and 13th centuries and even less from the 14th century onwards. Its demise appears to be linked to the emergence of the synthetic future, the outcome of the grammaticalization of [infinitive + *aver*], at a time when the prepositional variants [*aver a* + infinitive] and [*aver de* + infinitive] were extremely common. Of these, [*aver a* + infinitive] is used more frequently until the 14th century, after which [*aver de* + infinitive] gradually takes over as the dominant variant. By the 15th century, the variant with the prepositional linker *a* had all but disappeared, except in some markedly dialectal texts (Stengaard 2003: 1151).

A further set of obligational periphrases, with *tener* “to have, to hold” as their auxiliary verb, emerges between the 13th and the 15th centuries, as part of a gradual process in which *tener* replaces *aver* in an increasing number of contexts. In a recent study, Garachana and Rosemeyer (2011) show that this is a clear example of how a grammatical change can have its origin in an initially purely lexical substitution, arguing that the rise of [*tener de/a* + infinitive]<sup>4</sup> “is based on a process of conceptual identification, in which speakers do not distinguish between the lexical and the grammatical level once the equivalence between the verbs has been established” (Garachana & Rosemeyer 2011: 39, our translation). In the same vein, Yllera (1980: 110) observes that “in the 13th century *tener* can be found in a variety of contexts that previously only admitted *aver*, both as an independent verb indicating possession and when used with an adjective or participle. In addition, though with a certain delay and less frequently, *tener* begins to substitute *haber* in the modal periphrases” (our translation). A further variant of the modal construction with *tener*, [*tener que* + infinitive], begins to gain ground from the 16th and 17th centuries onwards, at the expense of [*haber de* + infinitive] and [*tener de* + infinitive], which have all but disappeared from the present-day language (Blas Arroyo & González, 2014).

Finally, the Latin verb *DEBERE* “to owe”, takes on the meanings of obligation and necessity at an early stage, leading to the emergence of deontic modal periphrases in almost all Romance languages (Yllera 1980: 92), including Spanish. According to Beardsley (1921: 150), [*deber* + infinitive], without a linking preposition, is by far the most common variant of the *deber*-periphrasis in early Medieval Spanish, with a few sporadic cases of [*deber a* + infinitive] also attested in the 13th century, whereas “[t]here seems to be no trace of *deuer*<sup>5</sup> *de* in the early texts” (Beardsley 1921: 31). While the prepositionless construction is the more frequent variant throughout the entire

<sup>3</sup> A fourth variant, [*aver que* + infinitive] is also occasionally documented, approximately between the 14th and the 16th centuries, but it does not turn into a serious competitor. Over the subsequent centuries, it acquires the specialised function that it has to the present day: the default impersonal modal periphrasis of necessity/obligation [*hay que* + infinitive] (‘one must, it is necessary to’ + inf.).

<sup>4</sup> The variant [*tener de* + infinitive] is the first to appear, with a few cases documented as early as the 13th century; the short-lived variant [*tener a* + infinitive] appears in the 14th century but always remains rare and disappears together with its counterpart [*aver a* + infinitive] before the end of the 15th century, whilst the frequency of [*tener de* + infinitive] rises continually until the 16th century (Blas Arroyo & González 2014).

<sup>5</sup> *Deuer*, as well as *dever*, are common historical spellings of modern-day *deber*.

history of Spanish, [*deber de* + infinitive] is also relatively common during the classical period from the second half of the 16th century onwards (Beardsley 1921: 31; Blas Arroyo & Porcar, 2016).

### 3. The corpus

The corpus used in this study, compiled as part of a wider project on diachronic variation in Spanish,<sup>6</sup> consists entirely of documents of a personal nature, primarily personal letters and private diaries. Such documents, characterised by a high degree of communicative proximity (*Nähesprache*; cf. Koch & Oesterreicher 1985; Oesterreicher 2004), tend to reflect natural, spoken language more closely than more formal texts (literary, legal, etc.) that have traditionally been used in diachronic linguistic studies. The documents selected to be included in the corpus represent different regional varieties of Spanish as well as different registers and degrees of formality, ranging from correspondence between close family members to letters sent by private individuals to the authorities.

The private nature of many of these texts, together with the relatively basic level of formal education of many of their writers, makes them a valuable resource for historical linguistics, as language change almost always has its origin in the spoken language, and the more formal or official a text is, the less likely it is to reflect the way in which speakers really used the language in everyday situations during the respective period. As observed by Oesterreicher (1996: 325), personal letters are a fertile ground for textual production within the category of written documents with oral features; as they were not written with the intention of their ever being published, there is little reason for the authors to avoid vernacular features.

Elspass (2012) notes that there are also other reasons why diachronic linguistic studies using data from private correspondence have become more numerous in recent years; one advantage is that these collections of letters tend to contain information about the relationships of power and solidarity between senders and addressees, as well as their social status (Okulska 2010) and their geographical origin, which allows the linguist to draw detailed conclusions about the impact of diastatic and diatopic factors on linguistic choices. Furthermore, the fact that personal letters frequently lack a pre-planned structure and are often emotionally charged means that they are likely to employ linguistic strategies aimed at increasing their expressive force (Danilova 2012), allowing us to identify potential conditioning environments (favouring one variable over another) such as emphasis, intensification or attenuation.

The 16th-century corpus used in this study contains 1,935 letters, a number of official statements recorded, in direct speech, by officials of the Inquisition (Eberenz & de la Torre 2002), as well as several diaries and chronicles authored by individuals with limited formal education (Stoll 2002; Stoll & Vázquez 2011); all in all, the corpus consists of texts by more than 700 Spanish speakers from a variety of social and regional backgrounds, totalling 842,658 words. The 18th-century corpus, meanwhile, contains 1,263 letters by approximately 500 different individuals, as well as two diaries and one account book, with a total word count of 624,456. The majority of the letters

---

<sup>6</sup> This study is part of the project “Linguistic variation and change through texts of communicative proximity: a historical sociolinguistic research project” (2014-2016), funded by the University Jaume I (ref. P1-1B2013-01) and the Spanish Ministry of Economy and Competitiveness (ref. FFI2013-44614-P). Please consult <http://sociolinguisticawe.wixsite.com/sociolinguisticaui> for full details on this and related projects.

are written (or dictated in some cases in the 16th-century corpus)<sup>7</sup> by individuals from a variety of social strata, ranging from farm labourers and craftsmen to members of the aristocracy who had emigrated to the Spanish New World colonies, and the topics dealt with in their correspondence (as well as the relationship between writer and addressee) range from intimate or familiar on one end of the spectrum to formal on the other end.<sup>8</sup>

#### 4. Distribution of the periphrases in the corpus

##### 4.1. General observations

Table 1 shows the overall distribution of the periphrases examined in this study during the two periods analysed. The most frequent construction in both periods is [*haber de* + infinitive], but its proportion decreases from 76% in the 16th century to 61.3% two hundred years later. As seen in Figure 1, this decrease becomes more pronounced as the 18th century progresses, from 64.7% in the first half of the century to 43.8% between 1750 and 1800 ( $p > 0.05$ ); by the end of the 18th century, [*haber de* + infinitive] is, for the first time, no longer the most frequently used modal periphrasis.

The proportion of periphrases with *deber*, on the other hand, more than doubles from 15.5% in the 16th century to 32.2% in the 18th century. This development becomes even clearer in the second half of the 18th century, when the variants with *deber* account for almost half of all modal periphrases,<sup>9</sup> used with approximately the same frequency as [*haber de* + infinitive]. This stands in stark contrast to the periphrases with the auxiliary verb *tener*, which remain relatively infrequent, though a certain increase can be observed towards the end of the 18th century, as seen in Figure 1.<sup>10</sup>

**Table 1.** Overall distribution of the periphrases examined

Periphrasis	16th century		18th century	
	N	%	N	%
<i>Haber de</i> + infinitive	1584	76.2	664	61.3
<i>Deber (de)</i> + infinitive	326	15.7	349	32.2
<i>Deber</i>	235	11.3	335	30.9
<i>Deber de</i>	91	4.4	14	1.3
<i>Tener de/que</i> + infinitive	168	8.1	71	6.5
<i>Tener de</i>	107	3.1	13	1.2
<i>Tener que</i>	61	1.7	58	5.4
<b>Total</b>	<b>2078</b>	<b>100</b>	<b>1084</b>	<b>100</b>

<sup>7</sup> Though letters dictated to scribes are almost certainly subject to some degree of orthographic levelling, thus obscuring certain patterns of phonological variation, there is little evidence of such a levelling effect regarding the choice of syntactic structures (cf. Bergs 2005: 79-80).

<sup>8</sup> A complete list of the sources of the documents contained in the corpus is available at the following URL: <http://sociolinguisticawe.wix.com/sociolinguisticauij#!blank/tntpi>

<sup>9</sup> It should also be noted that, in contrast with the 16th century, the variant [*deber* + inf.] without the linking preposition *de* has supplanted [*deber de* + inf.] almost completely in the 18th century; for a more detailed account of this alternation, see Blas Arroyo & Vellón (2014).

<sup>10</sup> However, a move from [*tener de* + inf.] towards [*tener que* + inf.] is evident, showing that this is the key stage in the shift from the variant with *de*, which was more common in the medieval and classical language, to the variant with *que*, which has, in the modern language, entirely supplanted the former, except in a few regional varieties (cf. Blas Arroyo & González 2014).

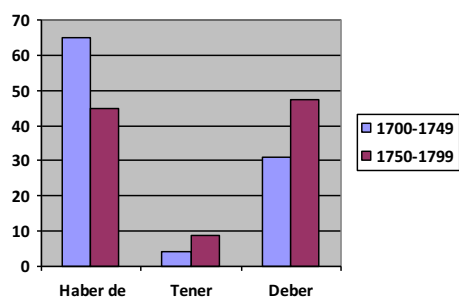


Figure 1. Distribution of modal periphrases in the first and second half of the 18th century

#### 4.2. Semantic values

As mentioned above, the periphrases examined here can express a range of modal notions, among them those exemplified in sentences (4)-(9):

- (4) ... y así pienso gastar poco porque bale tanto allá el bestir y calçar que dizen a de trabajar para sólo esto (*Vida y fortuna del emigrante navarro*, 1596)

“... and thus I intend to spend little, as clothes and shoes are so expensive there that it is said that one has to work just for that.”

- (5) ... aunque estoy con la cabeza como un cántaro por lo mucho que he tenido que trabajar a causa de estar sólo y haverse juntado tantas cosas de una vez... (*Correspondencia extraoficial de Ignacio de Heredia con Manuel de Roda*, 1774)

“... though my head is spinning as I’ve had to work so hard because I’m alone and so many things have happened at the same time...”

- (6) ...se cree que alguno dellos deve ser nicuesa capitan quel catolico Rey don fernando de gloriosa memoria mando yr a tierra fyrme (*Textos del Caribe*, 1519)

“... it is believed that one of them must be Captain Nicuesa, whom King Ferdinand the Catholic, of glorious memory, sent to the Province of Tierra Firme.”

- (7) En el día se halla este renglón hasta 10 pesos pero esperamos suba en junio, agosto y setiembre, por la mucha escasez de granos que ha de aver este año por las malas cosechas (*Al recibo de esta*, 1793).

“At present, the price is around 10 pesos, but we expect it to rise in June, August and September because of the shortage of cereals that will probably come about this year due to the poor harvests.”

- (8) Y que le respondyo la dicha Françisca: ¿Como lo tengo de yr a dezir que lo vido Juan Xymenes y negalo y no tengo con quien provarlo? (*Conversaciones estrechamente vigiladas*, 1515)

“And Francisca answered him: Why should I go and say that Juan Jiménez saw it, and deny it, without having anyone to confirm it?”

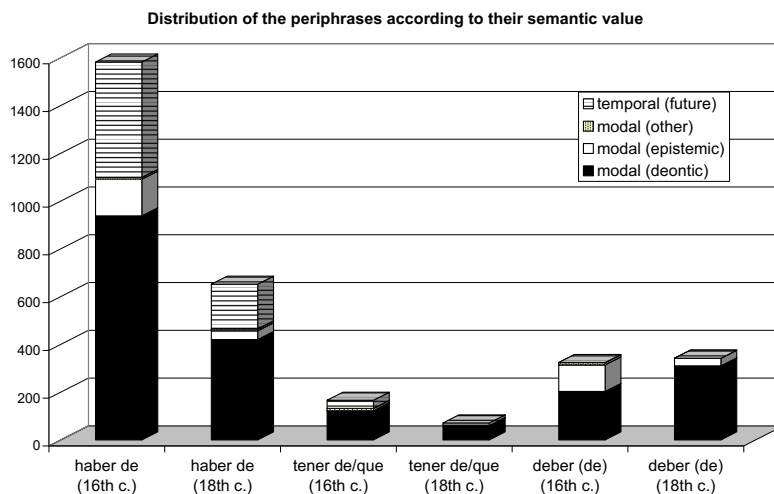
- (9) ... siendo mi fin buscar la vida para nuestra vejez, que es mi ánimo éste y lo ha sido, hallarme en lo mejor de mi edad para poderlo trabajar, y si este tiempo lo pierdo, después qué había de ser de nosotros (*Cartas desde América*, 1723)

“... as my objective is to prepare for our old age, which is and has been my motivation, to be of the best age to be able to work on it, and if I lose this time, what shall become of us later?”

Examples (4) and (5) are typical cases of deontic modality, expressing obligation or necessity, as the majority of tokens in our corpus: in the 16th century, 61% of all modal periphrases are deontic, a figure that rises to 74.1% in the 18th century. (6) and (7), on the other hand, are examples of epistemic modality in which the speaker refers to probable, presumed or approximate events or states of affairs. There are far fewer cases of epistemic modality in the corpus, and their proportion further diminishes between the 16th and the 18th century, from 13.2% to 6.5%.

**Table 2.** Distribution of the periphrases according to their semantic value

	<i>Haber de</i>		<i>Tener de/que</i>		<i>Deber (de)</i>		$\Sigma$	
	16th	18th	16th	18th	16th	18th	16th	18th
<b>Modal (deontic)</b>	74.4% (938)	53.2% (420)	9.4% (119)	7.6% (60)	16.1% (203)	39.2% (310)	61% (1260)	74.1% (790)
<b>Modal (epistemic)</b>	57.5% (154)	53.6% (37)	1.1% (3)	0% (0)	41.4% (111)	46.4% (32)	13.2% (268)	6.5% (69)
<b>Modal (other)</b>	25.9% (7)	100% (8)	37% (10)	0% (0)	37% (10)	0% (0)	0.2% (27)	0.7% (8)
<b>Temporal (future)</b>	93.4% (483)	94.4% (187)	6.6% (34)	5.6% (11)	0% (0)	0% (0)	25% (517)	18.5% (198)



**Figure 2.** Distribution of the periphrases according to their semantic value

Examples (8) and (9), finally, are instances of much rarer modal usages with an expressive force implying surprise, indignation, reproach, etc. (Gómez Torrego 1999: 3356). Added together, these modal values account for a large percentage of the tokens found in the corpus, rising from 75% in the 16th to 81.5% in the 18th century.

This increase in the proportion of modal values is, in fact, primarily due to a decrease of the non-modal use of [*haber de* + infinitive] and [*tener de/que* + infinitive] expressing purely temporal future reference, a function already documented in medieval times (Yllera 1980; Lapesa 2000; Hernández Díaz 2006). It should be kept in mind that the modal usages of these periphrases frequently have an implicit prospective value; this is due to the fact that especially deontic modality often coincides with future reference, as the notion of obligation implies futurity (Sinner 2003: 200). In a considerable number of tokens, this implicit futurity has been reanalysed as the main meaning of the



periphrasis, and the respective construction is used to refer to events that are predicted to take place in the future, without any hint or shade of modality; in these cases, the periphrases are in direct competition with other future constructions such as the synthetic and the periphrastic future tense, as exemplified in (10) and (11), where the writers alternate between the synthetic future and the periphrastic constructions.

- (10) *quiero dezir ansí que pienso que tarde a de ser mi venida, por eso ésta será primera y postrera que tengo de escribir...* (*Vida y fortuna del emigrante navarro*, 1557)

“So I want to say that my arrival will be late, which is why this [letter] will be the first and last that I will write.”

- (11) *Espero que toparé la orden de Vms. para entregar lo que les devo en España. Vms. [...] por quien son, espero que me han de mirar con caridad* (*Al recibo de esta*, 1776)

“I hope that I will receive your order to send what I owe you in Spain.... because of who you are, I hope you will consider me with charity.”

While a quarter of all tokens of these periphrases have non-modal future values in the 16th century, this proportion decreases to 18.5% by the 18th century.

#### 4.3. Distribution of the periphrases according to their semantic values

As shown in Table 2, the periphrasis *haber de* remains the most frequently used variant until the 18th century, not only in terms of overall frequency, but also with each and every one of the semantic modal values examined. Nevertheless, some significant shifts in the effect of the different modal values on the choice between the available variants are also apparent. For instance, for the expression of epistemic modality, *deber* (46.4%) comes close to drawing equal with *haber de* (53.6%) in the 18th century, up from (41.4%) in the 16th century. A much stronger shift can be observed in the area of deontic modality, where the proportion of *haber de* decreases from 74.4% in the 16th century to 53.2% two centuries later. The reduction in frequency of [*haber de* + infinitive] is primarily due to the rise of its main competitor, [*deber* (de) + infinitive], which more than doubles its proportion in deontic contexts, rising from 16.1% in the 16th to 39.2% in the 18th century; [*tener de/que* + infinitive], on the other hand, remains rare, and its percentage in obligatory contexts hardly changes between the two centuries examined.

In contrast, the percentages of non-modal future values hardly differ between the 16th and the 18th century: *haber de* accounts for 93.4% of all purely temporal future values in the 16th and for 94.4% in the 18th century. The only (limited) competition comes from the periphrases with *tener*, whilst [*deber* + inf.] never occurs in such non-modal contexts.<sup>11</sup>

In deontic modal contexts, on the other hand, the three periphrases examined here are in real competition, which is why the ENVELOPE OF VARIATION for the present multivariate analysis has been limited to these contexts.

#### 5. Coding and methodology

The variationist approach aims to establish how certain contexts favour or disfavour the choice of one linguistic form or structure over alternative forms or structures that have the same referential meaning or function. To achieve this, it is necessary to test a series

<sup>11</sup> The fact that [*deber* + infinitive] cannot be used to express non-modal futurity during this period contrasts with its use in medieval times, when it occasionally “weaken[ed] to an approximation of the future use” (Beardsley 1921: 22).

of hypotheses regarding the influence of certain constraints, which are the potential conditioning factors in a multivariate statistical analysis.

In a first step, all tokens of the variable to be examined, i.e., all instances of modal periphrases, were extracted from the corpus using the concordancer Wordsmith 6.0. Subsequently, it was determined for each token whether specific linguistic and extralinguistic factors were present in the particular instance or not, and this information was encoded for the statistical analysis. The factor groups taken into account in this study, listed below, are those that have been shown to be significant in previous work on modal periphrases (Balasch 2008, 2012; Blas Arroyo & Porcar 2014; Blas Arroyo et al. 2013); examples illustrating each of the (potential) factors are given in Table 3.<sup>12</sup> More specific details about the factors found to be statistically significant in this study will be presented in the corresponding parts of §6.

#### Linguistic factors:

- DISCOURSE CONTEXT: presence/absence of another modal periphrasis in the co-text immediately preceding the token,
- PHONEMIC ENVIRONMENT: the first phoneme following the auxiliary verb, disregarding the linking particle (if present),
- MORPHO-SYNTACTIC FACTORS: (1) person and number, (2) tense and mood, (3) presence/absence of an overt subject (4) clause type (subordinate clause, non-subordinate clause), (5) syntax of the main verb (simple or complex), (6) degree of (im)personality (active, passive or impersonal reflexive), (7) clausal polarity (affirmative or negative),
- SEMANTIC/PRAGMATIC FACTORS: (1) modal meaning (internal obligation, external obligation, necessity/advisability, others), (2) animacy of the subject (human/animate or non-human), (3) semantic category of the main verb (stative or dynamic lexical aspect, motion verb, speech verb), (4) degree of assertiveness (neutral, assertive/intensified).

#### Extralinguistic factors:

For the classification of letters along the stylistic continuum, two basic criteria are taken into account: (a) the main topic, and (b) the closeness of the relationship between writer and addressee (see §6.1.5. for further details).

**Table 3.** Examples of the linguistic and extralinguistic factors and factor groups considered in the multivariate analysis

Factor	Example
<b>Priming</b>	
Same periphrasis	... y como es jornada que todos <b>hemos de hazer emonos de conformar</b> con la voluntad devina ( <i>Cartas de particulares en Indias del siglo XVI</i> , 1565) “... and as it is a task that <b>we must all do</b> , we must accept God’s will”
Other modal periphrasis	<i>Y me parece que también se le <b>debe hacer</b> cargo del crédito de este dinero que injustamente ha retenido. Si la guerra permanece <b>habré de remitir</b> los reales asegurados</i> ( <i>Al recibo de esta</i> , 1795) “And I believe that <b>he should also be charged</b> for this credit that he has retained without justification. If the war goes on <b>I shall have to pay</b> the fixed amount of money”
None	<i>Digo Pepa que lo mejor es el correo y <b>han de venir</b> a Marcos de Estrada [...] es tienda muy antigua y muy conocida</i> ( <i>Cartas desde América</i> , 1722)

<sup>12</sup> For the sake of brevity, only examples of [*haber de* + infinitive] are given.

	"I tell you, Pepa, that it's best by mail, and <u>it must be delivered</u> to Marcos de Estrada; that's a very old and very well-known shop"
<b>Phonemic context</b> <sup>13</sup>	
Vowel	... y de los trescientos pesos ha de entregarte 100, luego que llegue. ( <i>Cartas desde América</i> , 1721) "... and of the three hundred pesos <u>he must pay you</u> 100, as soon as [the money] arrives."
Consonant	<i>En el escribir de las cartas a las Yndias se a de recatar que no me traten de mujer sino que digan ermana...</i> ( <i>Vida y fortuna de emigrantes navarros</i> , 1596) "When writing the letters to the New World, <u>care should be taken</u> that I am not addressed as a wife but as a sister..."
Cacophonous	<i>De Antonia no te digo nada porque ya le digo a mi padre lo que <u>había de decirle</u> a él así le escribiera, cosa que nunca haré.</i> ( <i>Cartas desde América</i> , 1716) "I won't say anything about Antonia, as I'd tell my father what <u>I had to say to him</u> if I were to write to him, which I'll never do."
<b>Person/number</b>	
non-1SG	<i>A de venyrse en esta flota, creo la a de traer vn sor hrº suyo</i> ( <i>Cartas de particulares en Indias del siglo XVI</i> , 1571) " <u>She is bound to arrive</u> with this fleet; I think a brother of yours will bring her."
1SG	..., que tocante a el viaje corre por cuenta del dicho padre Don Juan, a quien yo <u>he de satisfacer</u> . ( <i>Cartas desde América</i> , 1721) "... , that regarding the journey, the expenses will be covered by said Father John, whom <u>I shall repay</u> ."
<b>Tense/mood</b>	
Present indicative	<i>Vmd. misma lo ha de confesar.</i> ( <i>Al recibo de esta</i> , 1792) "You yourself <u>must admit it</u> ."
Imperfect indicative	..., él llevó dos mil y novecientos pesos con los cuatrocientos que <u>había de dar</u> a Baltasar de Mendoza y quinientos para la capilla. ( <i>Cartas y memoriales</i> , 1526) "... he took 2900 pesos, as well as the 400 that <u>he had to give</u> to Baltasar de Mendoza, and 500 for the chapel."
Other	... y lo que <u>abreys de traer</u> es lo siguiente: vn manto de tafetan con su ribete de terciopelo ( <i>Cartas de particulares en Indias</i> , 1583) "... and what <u>you will have to bring</u> is the following: a taffetan cape with a velvet trim"
<b>Syntax of subject</b>	
Omitted	<i>Y ha de venir conprado de tres escribanos y que sea por el primer correo</i> ( <i>Cartas de emigrantes navarros y guipuzcoanos</i> , 1791) " <u>And it must be sent</u> certified by three scribes, by first post."
Explicit	... y esto <u>se a de esforçar</u> muy de veras porque de otra suerte no quedará nada ( <i>Vida y fortuna del emigrante navarro</i> , 1596) "... and this <u>must be pursued</u> very seriously because otherwise there will be nothing left"
<b>Clause type</b>	
Subordinate	... esto se hara con el parecer de vn letrado que el sabra mejor de la manera que <u>a de ser</u> . ( <i>Cartas privadas de españoles en el Nuevo Mundo (S. XVI)</i> , 1591)

<sup>13</sup> In this factor group we consider the potential disavouring effect of the phoneme following the preposition *de*. It stands to reason that there may be a tendency to avoid verbs starting with /d-/ after the preposition *de*, which would make these verbs (e.g., *deber*, *dar*, *dejar*) inhibiting context for [*haber de* + infinitive].

	“... this must be done with the approval of a lawyer, as he will know more about the way it has to be.”
Non-subordinate	¿Cómo <u>he de vestir</u> a estos niños? ¿Cómo les <u>he de tener</u> a la escuela? ( <i>El hilo que une</i> , 1729) “How shall I dress these children? How shall I keep them in school?”
<b>Syntax of the main verb</b>	
Simple	<i>Los otros cincuenta los remito yo, para que de ellos se saquen los costos de una diligencia que <u>ha de hacer</u> el cura del Realejo de Abajo</i> ( <i>Al recibo de esta</i> , 1783) “The remaining 50 I’ll send myself, to cover the expenses for a formality that the priest of Realejo de Abajo <b>must attend to</b> .”
Compound	<i>Mas a de ser que luego a de tornar a <u>benir</u> aca.</i> ( <i>Cartas privadas de españoles en el Nuevo Mundo</i> (S. XVI), 1587) “But it is necessary that he must come back here afterwards.”
<b>Degree of (im)personality</b>	
Active	<i>Creo que con el deseo que tengo de ayudar a vm con ese <u>me a de faboreser</u> dios</i> ( <i>Cartas de particulares en Indias</i> , 1594) “I think that, given my desire to help you in this matter, God <b>must assist me</b> ”
Passive/Impersonal	<i>Que por fin ya tengo aquí muy buenos créditos y un principio de veinte mil pesos, de que doy muchas gracias a Dios, que <u>se ha de contar</u> de pocos los que vinieron esta armada.</i> ( <i>Cartas desde América</i> , 1704) “That I finally have good credit and a startup capital of twenty thousand pesos, for which I am very grateful to God, which <b>can probably be said</b> of very few of those that came with this fleet.”
<b>Clausal polarity</b>	
Affirmative sentences	... y esto se <u>a de esforçar</u> muy de veras porque de otra suerte no quedará nada ( <i>Vida y fortuna de emigrantes navarros</i> , 1596) “... and this <b>must be pursued</b> very seriously because otherwise there will be nothing left”
Negative sentences	<i>Y aunque lexitimamente son para ella, no los aya de reserbar de su marido</i> ( <i>Cartas de emigrantes navarros y guipizcoanos</i> , 1720) “And though they are legally hers, she <b>must not keep them</b> from her husband”
<b>Modal meaning</b>	
External obligation	<i>En la ca[rta] de [...] tu tío escriuo lo que <u>as de hazer</u> para venir aca.</i> ( <i>Cartas privadas de españoles en el Nuevo Mundo</i> (S. XVI), 16th c.) “In the letter from your uncle I write what <b>you must do</b> to come here.”
Non-external obligation	... yo siempre <u>he de cumplir</u> con mi ocupación, pues mi mayor deseo es darte gusto en todo para que conozcas lo mucho que te estimo y venero ( <i>Cartas desde América</i> , 1717) “... I <b>must</b> always <b>fulfil</b> my duty, for my greatest wish is to indulge you in everything, so that you understand how much I love and cherish you”
<b>Anymacy (3rd person)</b>	
Human/animate	<i>Y aunque lexitimamente son para ella, no los aya de reserbar de su marido</i> ( <i>Cartas de emigrantes navarros y guipizcoanos</i> , 1720) “And though they are legally hers, she <b>must not keep them</b> from her husband”
Non-human	... la lizençia a de ser para tierra firme y el peru... ( <i>Cartas privadas de españoles en el Nuevo Mundo</i> (S. XVI), 1578) “... the permit <b>has to be</b> for the Province of Tierra Firme and Peru...”
<b>Lexical aspect (main verb)</b>	
Speech verbs	... y en esta tierra <u>aveys de preguntar</u> por p° de vootillo porque ansi es en estas

	<i>partes mi nombre (Cartas privadas de españoles en el Nuevo Mundo (S. XVI), 1565)</i> “... and here <u>you have to ask</u> for Pedro de Bustillo, as that is my name in these parts”
Stative verbs	<i>Diçeme vm en esta q se le murio la mujer [...] y q [...] bibia enferma, si ello <u>abia de ser con tanto trabajo mejor esta [a]lla</u> (Cartas de particulares en Indias, 1586)</i> “You tell me in your letter that your wife has died and that she had been living with disease; if it <u>had to be</u> with so much suffering, she is better off in heaven”
Motion verbs	<i>Creo de ti que cuando no fuera con todas las conveniencias dichas <u>habias de venir sólo por verme</u> (Cartas desde América, 1706)</i> “I believe that, even without all the convenience I’ve just described, <u>you should come here</u> , just to see me”
Other (dynamic) verbs	<i>Y [...] me <u>ha de hacer</u> vuestra merced favor de que se busque a mi hermano Gerónimo Mansebo (Cartas desde América, 1724)</i> “And <u>you must do</u> me the favour of having my brother Geronimo Mansebo searched for”
<b>Degree of assertiveness</b>	
Intensified	<i>... y esto <u>se a de esforçar muy de veras</u> porque de otra suerte no quedará nada (Vida y fortuna del emigrante navarro, 1596)</i> “... and this <u>must be pursued</u> very seriously because otherwise there will be nothing left”
Non-intensified	<i>Y <u>ha de venir conprado de tres escribanos</u> y que sea por el primer correo (Cartas de emigrantes navarros y guipuzcoanos, 1791)</i> “And it <u>must be sent certified</u> by three scribes, by first post”
<b>Register/style (letters only)</b>	
+Personal	<i>Esposa querida de mi corazón: [...]Y así <u>te has de resignar</u> a sus órdenes. (Cartas desde América 1700-1800, 1706)</i> “My dear, cherished wife,... And so <u>you must follow</u> his orders.”
-Personal	<i>vuestra majestad [...] <u>a de saber</u> que al tiempo que los Castellanos entraron en esta ysla avia muchos millares y avn çientos de mill de yndios en ella (Textos del Caribe, 1518)</i> “Your majesty <u>should know</u> that at the time when the Castilians came to this island there were many thousands or even tens of thousands of indians living on it.”

Applying the principles of the sociolinguistic comparative method (Poplack & Tagliamonte 2001), our quantitative analysis consists of two independent multivariate analyses with identical factor groups, one for each of the historical periods examined. By comparing the results of these two analyses, it is possible to trace the path along which the emerging and the receding variants gradually enter or leave the system, focusing on the trajectory of their functions (Poplack 2011: 215). These analyses were carried out using Goldvarb X, with the periphrasis [*haber de* + infinitive] as the APPLICATION VALUE.<sup>14</sup> While the examination of overall frequencies and percentages can provide certain insights, the multivariate analysis shows not only the differences in

<sup>14</sup> An initial straightforward frequency analysis was used to identify and exclude factor groups associated (almost) categorically with only one of the variants, as well as those with insufficient data in one or more cells. This led to a significant reduction of the number of factors considered in the subsequent logistic regression analysis.

usage frequency of the competing variants for each of the contexts considered, but, more importantly, the degree of statistical significance and the order of relevance of different factor groups when considering all potentially significant conditioning factors simultaneously as well (Tagliamonte 2006: 235-245). The individual conditioning factors are arranged along a probabilistic scale between 0 and 1; favouring factors have factor weights greater than 0.5, while disfavouring ones have factor weights lower than 0.5. The further away from 0.5 this figure is, the greater the weight of the respective factor. The relative strength of each factor group is shown by the RANGE value, which is obtained by calculating the difference between the largest and the smallest factor weight in each factor group (Walker 2010); “[t]he higher this number is, the greater the contribution of that factor to the probability of the form” (Tagliamonte 2012: 127).

## 6. Results and analysis

As mentioned above, during the periods examined here [*haber de*+infinitive] is in competition with both [*deber (de) + infinitive*] and [*tener de/que + infinitive*] only when it is used to express deontic modality, which is why the ENVELOPE OF VARIATION for the present multivariate analysis has been limited to deontic contexts.

As will be seen in this section, despite a general decrease in frequency in the majority of contexts, a number of variation patterns remain the same between the 16th and the 18th century, both in terms of significant factor groups and the direction of the effect of individual factors within these groups (§6.1). On the other hand, there are also some conditioning factors that are not yet significant in the 16th century but gain importance by the 18th, thus revealing new paths along which the emerging variants spread (§6.2).

While the different patterns of variation will be discussed separately, it should be pointed out that the data in the corresponding tables is drawn from the same statistical run of Goldvarb.

Table 4 shows the effect of some of the factors outlined in §5 on the selection of [*haber de + infinitive*] during the two periods examined in this study.

**Table 4.** Statistically significant factor groups favouring the choice of *haber de* in the 16th and the 18th centuries<sup>15</sup>

	16th century			18th century		
	FW	%	Σ	FW	%	Σ
<b>Type of deontic modality</b>						
External obligation	.53	79.7	961	.57	60.8	495
Non-external obligation	.39	58.3	294	.37	40.9	295
<i>Range</i>	14			20		
<b>Person/number</b>						
non-1SG	.57	79.9	1090	.56	60.3	607
1SG	.17	37.2	188	.31	34.1	176
<i>Range</i>	40			25		
<b>Tense/mood</b>						
Imperfect indicative	.71	85	206	.67	67.9	81
Present indicative	.48	73.8	947	.51	56.1	578
Others	.28	53.6	125	.34	36.1	144
<i>Range</i>	43			33		
<b>Clausal polarity</b>						
affirmative	.53	77	1111	.52	54.3	718

<sup>15</sup> Square brackets indicate that the respective context is not statistically significant in that century.

negative	.28	54.3	159	.31	39.4	71
<i>Range</i>	25			21		
<b>Register/style (letters only)</b>						
+ personal	.62	79.9	493	.63	64.7	499
- personal	.38	66.3	537	.27	30.8	276
<i>Range</i>	16			36		
<b>Priming effect</b>						
<i>haber de</i>	.79	91.4	140	.77	83.9	62
none	.50	74.9	1070	.52	54.6	680
others ( <i>deber, tener</i> )	.08	17.6	68	.11	13.1	61
<i>Range</i>	71			66		
<b>Degree of impersonality</b>						
Passive/impersonal	[.51]	78.6	243	.68	72.1	86
Active	[.47]	72.5	1035	.46	51.5	717
<i>Range</i>				22		
<b>Clause type</b>						
Subordinate clause	.54	77.1	811	[.47]	52.5	530
Main clause	.43	67.2	457	[.56]	56	273
<i>Range</i>	11					
<b>Degree of agentivity</b>						
Human subject	[.51]	82	297	.45	60.1	283
Non-human subject	[.49]	80.8	399	.57	70	170
<i>Range</i>				12		
<b>Degree of emphasis/intensification</b>						
Intensified/emphatic	[.48]	70.8	271	.64	66.2	142
Non-intensified/non-emphatic	[.51]	74.2	999	.46	50.8	657
<i>Range</i>				18		

Factor groups not selected as significant in either century: phonemic environment, presence/absence of an overt subject, syntax of the main verb, semantic category of the main verb.

Input:	0.785	0.536
Convergence at Iteration:	11	8
Log likelihood:	-542.616	-412.676
Significance:	0.009	0.003

## 6.1. Patterns of continuity

In this subsection, we discuss those factor groups that are statistically significant, with the same direction of the effect, in both periods examined.

### 6.1.1. Type of deontic modality

As mentioned in 4.2. above, the overwhelming majority of tokens in both centuries express deontic modality (necessity or obligation), which makes it particularly interesting to determine whether there are any significant differences regarding the effect of different semantic subtypes within this modal range. Distinguishing and classifying such different types or ‘shades’ of deontic meaning is a difficult task that has been attempted by numerous scholars (Keniston 1937; Yllera 1980; Olbertz 1998; Gómez Torrego 1999; Fernández de Castro 1999; García Fernández 2006; López Izquierdo 2008; Martínez Díaz 2008). The statistical analysis carried out in this study reveals a clear difference between two basic types: externally imposed obligation on the one hand, and self-imposed as well as subjectively perceived obligation on the other. As seen in Table 4, contexts involving the former type are favourable to the use of *haber*

de, whereas the latter type of deontic modality has a disfavouring effect on the choice of this periphrasis.

The category of external obligation includes obligations imposed by written or unwritten rules, agreements, social conventions, laws, etc. (12), obligations as the result of a command or order (13), obligations imposed by external circumstances (14), and inevitability (15).

- (12) ... y logré el 22 de agosto sacar el documento de sus manos con motivo de cierta rebaja que había de hacerse de 500 pesos... (Al recibo de esta, 1789)

“... and on the 22nd of August I managed to take the document out of his hands because of a certain reduction of 500 pesos that had to be applied...”

- (13) te aviso que no as de venir aca sin traer tu madre y hermanas. (Cartas de particulares en Indias del siglo XVI, 1594)

“I inform you that you must not come here without bringing your mother and sisters.”

- (14) Si la guerra permanece habré de remitir los reales asegurados como V ms. me dicen en su apreciable de 4 de diciembre del año anterior (Al recibo de esta, 1795)

“If the war continues, I will have to send the money insured, as you advise me to do in your much appreciated letter written on the 4th of December last year.”

- (15) Dios los tenga en el cielo que como es camino que todos emos de andar no ay mas que dar gracias a Dios por todo... (Cartas de particulares en Indias del siglo XVI, 1594)

“May God keep them in Heaven, for as it is the path along which we shall all go, there is nothing more to do than to thank God for everything...”

A more fine-grained analysis shows that each of these types of externally imposed deontic modality has a different degree of likelihood to trigger the choice of [*haber de* + infinitive] in both centuries, with inevitability being the most favourable factor and obligation due to external circumstances the least favourable one for the choice of this variant.

**Table 5.** Proportion of *haber de* with different sub-types of externally imposed deontic modality in the 16th and 18th centuries

	16th century		18th century	
	%	Σ	%	Σ
<b>Externally imposed obligation</b>	79.7	961	60.8	495
Inevitability	95.8	143	90	30
Order/command	82.2	321	70.9	182
Rules, agreements etc.	75.9	319	51.8	191
External circumstances	69.3	179	50	92

While the percentage of *haber de* decreases for all sub-types of deontic modality between the two centuries, in line with the general development, Table 5 reveals that there is a greater decline of the preference for this periphrasis in those semantic environments that are already less likely to trigger its use in the 16th century (rules and agreements, external circumstances) than in the environments most closely associated with *haber de* (inevitability, order/command). As will be discussed below in more detail, this is a clear example of more strongly ENTRENCHED form-meaning pairings



(e.g., [*haber de*+inf.]–inevitability) resisting change and replacement for longer than less entrenched ones.

Furthermore, there appears to be a correlation between the degree of inevitability or coercive force and the preference for *haber de*: the more predictable it is that the obligation will be carried out, the more likely it is that *haber de* will be used. Thus, if something must inevitably be done, the subject has no choice whatsoever; an order or command must normally be followed, but there is at least a theoretical possibility of ignoring it; rules and agreements must also be followed but are occasionally flouted, while external circumstances can oblige us to act in a certain way, but there may be an alternative course of action available.

In addition to these types of externally imposed obligation, there are also self-imposed obligations caused by the subject's internal convictions (e.g., religious, ethical or philosophical persuasions, gratitude or respect)<sup>16</sup> (16) and a subjective sense of necessity (17). The obligation or subjective necessity can be perceived either by the subject of the sentence containing the modal construction, or on behalf of that subject by the person uttering/writing the sentence.<sup>17</sup>

(16) ... 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.”

Subjective necessity or advisability perceived by the speaker has a lower degree of coercive force than obligations imposed by external circumstances. In contrast to the external circumstances causing necessity in (14), the modal periphrasis in (17) is used to express that something is subjectively desirable.<sup>18</sup>

(17) ... porq el mas pobre [pobre] mata en su casa cada semana vn carnero y no se contenta sino q a de tener su gallina o pollo con su toçino asado y si no tiene gallina a de tener su quartillo de cabrito o cordero y pan. (*Cartas de particulares en Indias del siglo XVI*, 1583)

“... because even the poorest butchers a sheep at home every week, and that's not enough; he must have his hen or chicken with his roast bacon, and if he doesn't have a hen, he must have his quarter of kid or lamb and bread.”

As seen in Table 4 above, subjective necessity and self-imposed obligation have a disfavoured effect on the choice of *haber de*. Table 6 shows that internal (moral) obligation is already least likely to trigger the use of this periphrasis in the 16th century, and even less so two centuries later.

<sup>16</sup> The fact that the need to meet the obligation is primarily subjective means that this type of deontic modality is, semantically, often very similar to volition (Yllera 1980: 114).

<sup>17</sup> Following Martínez Díaz (2008: 1285), we understand modality as “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” (our translation). Otherwise, only 1SG statements would be eligible for certain categories such as *internal obligation*.

<sup>18</sup> This category also includes cases in which the periphrasis is used as a phatic device (*he de decirle* “I must tell you”, *has de saber* “you must know”, etc.), which Gómez Manzano (1992: 160) describes as a “kind of crutch” and Gómez Torrego (1999: 3354) analyses as a manifestation of the speaker's desire “to enter into communication”; it is particularly common in the epistolary genre that most of the texts in our corpus belong to. Though this is a conventionalised and pragmatized usage, it nevertheless retains the notion of beneficial necessity or advisability perceived by the speaker, as in “it would be beneficial if I told you” or “in my opinion it would be beneficial for you to know”.

**Table 6.** Proportion of *haber de* with different types of non-externally imposed deontic modality in the 16th and 18th centuries

	16th century		18th century	
	%	Σ	%	Σ
<b>Subjective necessity/advisability</b>	63.1	95	43.9	148
<b>Internal (moral) obligation</b>	60.6	183	33.9	121

Subjective necessity/advisability is somewhat less self-imposed than purely internal obligation, as external circumstances frequently contribute to the sense of necessity. Considering the correlation between the degree of external coercion and the use of [*haber de* + inf.] identified in Table 5 above, the slightly higher percentage of *haber de* in contexts of subjective necessity does not come as a complete surprise.

### 6.1.2. Person and number

‘Person and number’ was divided into two factors because of a clear difference, in the 16th century, between *haber de* in 1SG contexts, where only 37.2% of tokens contain this periphrasis, and the other person contexts, in which the occurrence of this periphrasis is more than twice as likely (79.9%), as shown in Table 7. This translates into a statistically significant association of non-1SG contexts with *haber de* (FW .57), while 1SG contexts have a clearly negative effect on the use of this periphrasis (FW .17).

The disfavouring effect of 1st person singular contexts on the selection of *haber de* is also evident in the 18th century (34.1%). The main difference when compared with the 16th century is that this reluctance to use *haber de* has spread from the 1st person singular to the 1st person plural, where the most drastic change between the two periods can be observed, with a decline of *haber de* from 82.9% in the 16th to 23.4% in the 18<sup>th</sup> century. In fact, a separate run of the statistical analysis in which 1st person singular and plural are considered together shows that this context is among the least favourable environments for *haber de* (FW .23) and thus a prime route for the expansion of alternative variants.<sup>19</sup>

**Table 7.** Distribution of *haber de* + infinitive by person and number in the 16th and 18th centuries<sup>20</sup>

	16th century		18th century	
	%	Σ	%	Σ
1SG	37.2	188	34.1	176
2SG	100	14	61.3	106
3SG	81.7	622	67.4	325
1PL	82.9	41	23.4	47
2PL	68.2	173	100	1
3PL	81.4	210	54.7	128

### 6.1.3. Tense/mood

A pattern of continuity between the 16th and the 18th century can also be identified for the factor group tense/mood. In terms of representation in the corpus, the present

<sup>19</sup>Log-likelihood: -412.075; significance: 0.003

<sup>20</sup>The considerable difference between the number of singular and plural 2nd-person contexts can be explained by the differences in the ‘semantics of solidarity’ (Brown & Gilman 1960) between the two centuries: while the 2PL verb forms (and the pronoun *vos*) were commonly used to address a single interlocutor, even in contexts of solidarity, in the 16th century, this had changed by the 18th century, when 2SG verb forms (and the pronoun *tú*) had come into general use in those contexts. It is, indeed, remarkable that there is only a single case of a 2PL modal periphrasis in the 18th-century corpus.

indicative is by far the most frequent in both periods, with figures consistently above 70%. This is followed, at a considerable distance, by the imperfect indicative (16th c.: 16.6%; 18th c.: 11.2%). All other tenses and moods occur at such a low frequency in both subcorpora that they were grouped together.<sup>21</sup>

The results of the logistic regression analysis show that, despite a general decrease in the selection of *haber de* in the 18th century in comparison with two hundred years earlier, the constraint hierarchy remains the same. Thus, the imperfect indicative is the most likely to select this periphrasis in both centuries (16th c.: 85%, FW .71; 18th c.: 67.9%, FW .67), followed by the present indicative (16th c.: 73.6%, FW .48; 18th c.: 56.1%, FW .51), whilst the less commonly occurring tenses are less favourable contexts for *haber de* (16th c.: 53.6%, FW .28; 18th c.: 36.1%, FW .34). The fact that the vast majority of modal constructions appear in the present or imperfect indicative and that *haber de* retains a strong link with these tenses goes some way to explaining why this periphrasis remains the numerically dominant modal construction in the 18th century, making it a clear example of how the most frequent contexts tend to support the continued use of the older, more established variant due to their greater degree of entrenchment (cf. Bybee 2006; Rosemeyer 2015).

#### 6.1.4. Clausal polarity and type

Given the general prevalence of affirmative clauses in the types of documents contained in the corpus, it is not surprising that the proportion of affirmative clauses containing periphrastic modal constructions (approx. 85%) is also far greater than that of negative clauses (approximately 13.5%), while non-declarative clauses (exclamations, exhortations, direct questions) account for a mere 1.7% of all modal periphrases.

Leaving aside the latter because they are heterogeneous and also very infrequent (14 and 8 tokens in our 16th and 18th century corpora, respectively), our multivariate analysis reveals that we are dealing not only with a statistically significant factor group in both centuries, but also with the same direction of the effect: affirmative clauses favour the use of *haber de*, while the opposite is the case for negative clauses. Though the percentage of *haber de* in affirmative clauses decreases considerably, from 77% in the 16th to 54.3% in the 18th century, the fact that this clause type accounts for an extremely large proportion of all clauses containing modal periphrases makes them one of the main pillars supporting the use of this periphrasis, not only in the 16th, but also in the 18th century.

On the other hand, the much less frequent negative clause contexts act as an important structural entry point for the diffusion of the emerging variants (16th c.: FW .28; 18th c.: FW .31), which confirms the important role of negative polarity identified in other processes of grammaticalization and syntactic change (Torres Cacoullos & Walker 2009; Tagliamonte et al. 2014).

#### 6.1.5. Stylistic and register variation

A similar pattern of continuity can be identified when considering the factor group 'style', which in this study is understood to consist of a combination of two parameters: (a) the main topic of epistolary texts, and (b) the closeness of the relationship between

<sup>21</sup> This group contains subjunctive forms (around 5% in both centuries), the synthetic future (16th c.: 1.6%; 18th c.: 6%), the simple past or preterite (16th c.: 0.8%; 18th c.: 2.8%), the present conditional (16th c.: 1.6%; 18th c.: 0.3%) and non-finite forms (grouped together), accounting for 1.9% in both centuries.

the sender and the addressee.<sup>22</sup> Based on these two factors, we can establish a stylistic continuum with the following extremes:

a) 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 correspondence between close friends or lovers also falls into this category.

b) Letters dealing with non-personal matters and those with a clear distance along the axes of familiarity and solidarity (e.g., letters sent by commoners to their superiors, to members of the clergy or nobility, or to other recipients belonging to higher social classes).

The analysis of the 16th-century data reveals that [*haber de* + infinitive] is, during this period, already more commonly used in familiar and informal contexts (80%, FW .62) than in letters of a more formal nature or with a greater social distance between the correspondents (66.3%, FW .38). Two centuries later, this trend has not changed: *haber de* occurs less frequently in global terms, but this development is particularly strong in formal letters, with a decrease of more than 50% (16th c.: 66.3%; 18th c.: 30.8%). It is in these contexts that the diffusion of the emerging variants is most notable. This stands in stark contrast to the present-day use of *haber de* in Peninsular Spanish, which, apart from a few regional varieties, is limited to the most formal registers of the written language (Gómez Torrego 1999; Sinner 2003; Martínez Díaz 2003; NGRALE 2009). On the other hand, the more personal and spontaneous contexts remain the stronghold of the older, established variant [*haber de* + infinitive] in the 18th century; the relatively high frequency of this periphrasis in everyday communication is likely to be the reason for its survival over so many centuries.

#### 6.1.6. Discourse factors: the priming effect

In both periods, the choice of periphrasis is clearly influenced by a ‘priming effect’, i.e., a tendency for speakers to repeat linguistic material they have used in previous stretches of their discourse (Labov 1994; Pereira-Scherre & Naro 1992; Travis 2005). Taking into account the potential relevance of ‘structural priming’ (Pickering & Ferreira 2008) in the selection of *haber de*, one of the factors considered in this study is the presence of another, preceding modal periphrasis. Thus, all tokens in the corpus were assigned to one of three different groups. The first, exemplified in (18), contains the tokens immediately preceded by the same modal periphrasis.<sup>23</sup> The second group, exemplified in (19), contains the tokens preceded by a DIFFERENT modal periphrasis, whilst in the vast majority of cases there is no preceding periphrasis (group three).

(18) ... y como es jornada que todos **hemos de hazer** emonos de conformar con la voluntad divina (Cartas de particulares en Indias del siglo XVI, 1565)

“... and as it is a task that **we must all do**, we must accept God’s will.”

(19) 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 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.”

<sup>22</sup> This factor was only encoded for tokens found in letters that were clearly classifiable as either personal/informal or distant/formal.

<sup>23</sup> In order to measure this factor in a uniform way, the presence or absence of another modal periphrasis within the preceding 15 words was considered, though some studies suggest that priming can have an effect over a far longer distance (Pickering & Ferreira 2008: 447).

Our analysis reveals that this factor is a strong constraint in both centuries, with high RANGE values of 67 and 68 respectively. The explanatory hierarchy within this factor group also remains similar: in the 18th century, [*haber de* + inf.] is especially likely to be used when preceded by another [*haber de* + inf.] construction (83.9%, FW .77), and especially unlikely to be used when preceded by a different modal periphrasis (13.1%, FW .11), while the absence of a preceding periphrasis has practically no effect on the choice of periphrasis (54.6%, FW .52). What is important is that the effect of these three contexts hardly changes when compared to the 16th century (*haber de*: 91.4%, FW .79; others: 17.6%, FW .08; none: 74.9%, FW .50), despite the declining overall usage frequency of the construction.

A reason for this continuity might lie in the type of texts examined here, which are less prone to following the prescriptive stylistic norm, typical of more formal registers, which stigmatises the repetition of identical lexical items in close sequence. While literary, legal, administrative and religious texts, on which much of diachronic research has traditionally been based, tend to avoid repetition if possible, the personal nature and style of the letters and memoirs examined in this study, the limited formal education of many of their writers, as well as the spontaneity and emotion of these documents make them more similar to oral discourse, in which formal stylistic norms such as the avoidance of repetition are not generally followed.

This hypothesis is corroborated by the data presented in Table 8, showing that there is, indeed, interaction between the priming effect and register/style (examined in §6.1.5. above), particularly in the 18th century. The cross tabulation of the two factors confirms that a higher degree of formality reduces the priming effect, whereas a less formal, more spontaneous style appears to favour this cognitive process.

**Table 8.** The combined influence on the proportion of *haber de* of the priming effect and the degree of formality/spontaneity

	16th century				18th century			
	+Personal		-Personal		+Personal		-Personal	
	%	Σ	%	Σ	%	Σ	%	Σ
Two tokens of <i>haber de</i> + inf. in close sequence	91	87	89	53	84	56	70	10
<i>Haber de</i> + inf. preceded by a different modal periphrasis	12	49	14	37	14	36	10	29
No priming	76	706	68	539	66	456	32	262

## 6.2. Diverging patterns of variation

The factor groups discussed in this section were selected as significant in one of the centuries but not in the other. In some cases, these factor groups do appear to have a similar effect on the choice of the modal periphrasis in both centuries, but the effect is only statistically significant in one of them.

### 6.2.1. Syntactic (im)personality

An example of this is the factor group ‘degree of (im)personality’, in which the active voice ((20) and (21)) is contrasted with non-active constructions, i.e., the passive voice and the impersonal reflexive construction ((22) and (23)).

- (20) *Creo que con el deseo que tengo de ayudar a vm con ese me a de faboreser dios (Cartas de particulares en Indias, 1594)*

“I think that, given my desire to help you in this matter, God must assist me.”

- (21) ... *y si no se hacen las paces será más dilatada la ida, porque no me he de poner a riesgo de perder la vida*. (*Cartas desde América*, 1704)
- “... and if they do not make peace, the journey will take longer, as I should not put myself at risk of losing my life.”
- (22) ... *se a de poner toda diligencia y a todo trabaxo umano por salvar el anyma* (*Cartas de particulares en Indias*, 1560)
- “... one must put every diligence and human effort into saving the soul.”
- (23) *Que por fin ya tengo aquí muy buenos créditos y un principio de veinte mil pesos, de que doy muchas gracias a Dios, que se ha de contar de pocos los que vinieron esta armada*. (*Cartas desde América*, 1704)
- “That I finally have good credit and a startup capital of twenty thousand pesos, for which I am very grateful to God, which can probably be said of very few of those that came with this fleet.”

As seen in Table 4 above, the preference for *haber de* in non-active contexts is statistically significant in the 18th century (72.1%, FW .68), whereas the degree of impersonality is not a significant factor group two centuries earlier.

### 6.2.2. Agentivity

Similarly, the factor ‘agentivity’, which distinguishes 3rd person<sup>24</sup> human from non-human subjects, is also only significant in the 18th century. According to some scholars, constructions with the verb *haber* already had a deagentivising effect in Old Spanish; for instance, Stengaard (2003: 1151) observes that “by means of the periphrasis with *aver*, the subject of the action expressed by the infinitive either loses its possible role as subject-agent, or the role of subject-recipient or patient implied by the respective verbal action”. This semantic effect is related to the meaning of non-auxiliary *haber*, which expresses non-agentive or receptive possession in which the subject does not exert control over the possessed item (Seifert 1930).<sup>25</sup>

As seen in Table 4, our data confirms this hypothesis for 3rd person subjects in the 18th but not in the 16th century. Though the differences are not very great and the low RANGE value (12) indicates that we are dealing with a relatively weak factor, the analysis nevertheless confirms that, in the 18th century, non-human subjects favour the use of *haber de* (70%, FW .57) more than human subjects do (60.1%, FW .45).

### 6.2.3. Assertiveness/intensification

The variation observed in the 18th century is also sensitive to the degree of assertiveness or intensification, a factor that has no significant effect two hundred years earlier. Despite the absence of intonation and the typical paralinguistic cues of the spoken language, it is possible to identify tokens with a high degree of assertiveness or emphasis in the written texts examined here; the authors of the letters have a range of

<sup>24</sup> As all non-human subjects require the verb to be in the 3rd person, in a separate run of the regression analysis only 3rd person human subjects were considered for this factor group, in order to make the figures comparable and to avoid overlap between this factor group and that of grammatical person/number.

<sup>25</sup> From a cognitive perspective, Garachana (1997) relates this difference to the prototypicality of possession, suggesting that *haber* was subjected to a process of semantic bleaching (figurative control over the possession) that did not extend to *tener* (cf. Garachana & Rosemeyer 2011).

morphosyntactic and semantic resources at their disposal that allow them to emphasise or intensify a statement, either with regard to the content of the statement itself or to the recipient of the letter (in some cases these two types of emphasis coincide). Possible indicators of an emphatic or intensified environment include a wide range of linguistic elements such as specific prefixes and suffixes (e.g., the superlative suffix), the use of evaluative lexical items and expressions, certain types of modality (e.g., imperative, exhortative), certain clause types (e.g., comparative clauses), as well as rhetorical devices such as repetition, enumeration, metaphors and hyperboles.<sup>26</sup> Examples (24) and (25), taken from the 16th and the 18th century corpus, respectively, are clear cases of tokens with a high degree of assertiveness:<sup>27</sup>

(24) ... y esto *se a de esforçar muy de veras* porque de otra suerte no quedará nada (*Vida y fortuna del emigrante navarro*, 1596)

“... and this must be pursued very seriously because otherwise there will be nothing left”

(25) *A lo que se añade que por no pagar salarios, casa, gastos y otras cosas, [...] el albacea había de vender los géneros en mucho menos de lo que valieren.* (*Cartas desde América*, 1706)

“In addition, so as not to pay for **salaries, accommodation, expenses and other things**, the executor would have to sell the goods **for much less than they are worth.**”

While the degree of assertiveness makes virtually no difference in the 16th century, a statistically significant effect of this factor group can be observed two hundred years later, when *haber de* has become associated with these intensified contexts (66.2%, FW .64), whereas non-intensified contexts seem to have an almost neutral effect on the choice of this periphrasis (50.8%, FW .46).

A more detailed analysis shows that there is an interesting correlation between this factor group and the stylistic tenor of the respective text. As discussed in §6.1.5., in letters dealing with private, personal or intimate matters, *haber de* is used more frequently than in less personal letters in both our corpora, but particularly so in the 18th century. When viewed in combination with the degree of assertiveness (see Table 9), it becomes apparent that this preference for *haber de* increases in intensified contexts in letters of a highly personal nature (72%), while it is disfavoured in those that are less private (54%). An even greater difference is observed in non-intensified contexts, where *haber de* is selected for 62% of modal periphrases in letters of a more intimate nature, while this construction accounts for only 27% of the tokens in less personal correspondence. These environments, i.e., non-intensified contexts in more formal, less spontaneous texts, are thus an important area for the expansion of the newer, less established variants, especially of [*deber (de) + inf.*], which accounts for a considerably higher proportion of tokens (66%).

**Table 9.** The combined influence on the proportion of *haber de* of the degree of emphasis/intensification and the degree of intimacy of the text

		16th century				18th century			
		Intensified		Non-intensified		Intensified		Non-intensified	
		%	Σ	%	Σ	%	Σ	%	Σ

<sup>26</sup> For a more exhaustive review of how these strategies are used in the colloquial language, see Albeda (2007).

<sup>27</sup> The increased assertiveness is, here, achieved by adding emphatic lexical material; in (25), it can furthermore be seen that the enumeration preceding the modal periphrasis adds to the emphatic effect.

+ Personal	74	179	74	655	72	88	62	424
- Personal	63	127	67	497	54	37	27	262

#### 6.2.4. Clause type

Finally, the reverse tendency can be identified with regards to the factor group ‘clause type’, i.e., whether the periphrasis appears in a main or in a subordinate clause. While subordinate clauses have been identified as more resistant to syntactic change in other processes of variation (Tarallo 1989; Matsuda 1993), for the Spanish modal constructions this is only the case in the 16th-century corpus, whereas there is no significant difference between main and subordinate clauses in the 18th century, as seen in Table 4.

### 7. Conclusions

As shown by the data presented here, there is a significant shift in the choice of modal periphrases between the 16th and the 18th centuries. The changes involve a gradual reduction in the use of the most common variant, [*haber de* + inf.], in favour of its competitors, [*deber (de)* + inf.] and [*tener que* + inf.]. While the last of these is still comparatively infrequent during this period, the competition between the other two periphrases is in full swing by the 18th century. Though [*haber de* + inf.] remains numerically dominant, the advances made by [*deber* + inf.], especially in the second half of the century, are considerable.

This general shift is characterised by a significant drop in the frequency of *haber de*, over this 200-year period, in the majority of contexts examined here. However, a variable rule analysis comparing the two periods shows that the change does not occur with the same intensity in all contexts, and, more importantly, it does not progress along a single, uniform pathway; there are, in fact, at least two different patterns of development.

The first of these patterns, continuity through time, in which the same factor groups have a statistically significant impact with the same direction of effect in both centuries, can be identified for the following factor groups: type of deontic modality, person/number, clausal polarity, tense/mode, structural priming, and the stylistic tenor of the respective text. While the overall proportion of *haber de*-tokens drops across the board in the 18th century, there is little difference in the weight or ordering of factors within these factor groups – in both centuries, low-frequency contexts<sup>28</sup> such as negative clauses, 1st person singular subjects, non-external obligation and tenses other than the present and imperfect indicative favour the newly emerging variants, whereas high-frequency contexts such as affirmative clauses, non-1SG subjects, external obligation, as well as the present and imperfect indicative tenses favour the continued use of established *haber de*.

The effect of stylistic tenor (i.e., +/- formal) on the choice of the periphrasis also remains similar in the two periods examined, with *haber de* clearly preferred in less formal contexts. In other words, the replacement of *haber de* is particularly advanced in more formal contexts, whereas the more spontaneous environments that resemble the oral language to a greater extent are more favourable to the older, established variant. Given that our two corpora consist primarily of letters dealing with personal and private

<sup>28</sup> High and low frequency refers to *relative* frequency, i.e., in comparison with the possible alternative contexts in the same factor group. For instance, non-external obligation is considered a low-frequency context because it is less frequent than externally imposed obligation in our corpus.



matters, this is also likely to be one of the reasons for the robustness of [*haber de* + inf.] in the present study.

Contrasting with this predominant pattern of continuity, our comparative analysis has also identified several factor groups that have a significant effect in one century but not in the other. In some cases, they appear to show a pattern of continuity (albeit not statistically significant in both periods), in the sense that an incipient effect in the 16th century turns into a fully-fledged significant factor two centuries later. An example of this is non-active contexts (i.e., passive and impersonal clauses), which, in the 18th century, become a stronghold of *haber de*. Generally, however, these cases simply show that a factor group can, over time, attain or lose significance regarding its impact on variation, which is to be expected in an area of grammar that is going through a period of change.

In conclusion, the prevalence of [*haber de* + infinitive] in the 16th century is weakened by alternative variants, in particular [*deber (de)* + infinitive] two centuries later (see Table 1), and this decline can be observed in the majority of the contexts analysed. Nevertheless, *haber de* generally retains its dominance in the most frequent environments, while the alternative periphrases spread primarily in the less frequent contexts.

It has long been known that high-frequency forms are more resistant to change, due to their greater degree of cognitive entrenchment: “the more a form is used, the more its representation is strengthened, making it easier to access the next time” (Bybee & Thompson 2000: 380). This ‘conserving effect’ of high frequency forms has been demonstrated not only at the lexical level (e.g., Bybee 1985; Langacker 1987), but also with regards to syntactic phenomena (e.g., Givón 1979; Croft 2000; Bybee & Hopper 2001). What the present study has shown is that this entrenchment effect does not only apply to the overall frequency of a form, but crucially also to individual form-context combinations: A variant (i.e., form) that occurs particularly frequently in a specific highly frequent morphosyntactic, semantic, pragmatic or stylistic context tends to be more resistant to being replaced IN THAT CONTEXT, but not necessarily in other, lower-frequency environments, which are more prone to admitting alternative variables at an earlier stage. As pointed out by Bybee (2006: 715), the conserving effect of high token frequency means that the memory representation of specific strings or sequences of morphemes and words is strengthened in the language user’s mind, making them more readily accessible and therefore reinforcing their entrenchment; these ‘exemplar representations’, which form part of clusters of similar settings, furthermore “allow specific information about instances of use to be retained in representation” (Bybee 2006: 717). As constructions are, in this model, understood as the result of a cognitive process in which strings that are both formally and semantically similar are “stored close to one another” (Bybee 2006: 716), the specific information about the use of a particular string that is stored together with individual representation also constitutes part of the information about the construction that these individual representations form part of.

Regarding our study of modal periphrases in Spanish, the relevant specific information associated with the respective (competing) constructions is the context in which their representations, be they semantic, stylistic or morphosyntactic, typically occur. This information is thus an integral part of the construction itself, which to some extent must lead to a perpetuation of its preferential use in these contexts. This does not, however, preclude constructions from expanding to contexts in which they are not initially the most typical choice, as such tokens are merely less prototypical and more

distant from the exemplar representation, but still share a number of formal and semantic features with it.

While exemplar theory predicts a general correlation between high frequency and conservation of established linguistic forms, Poplack (2001) observes that this is not universal in ongoing competition between two variants, and some counterexamples have, indeed, been identified in our analysis, such as passive/impersonal contexts which, despite their comparatively low frequency, favour the use of the older, established variant to a greater extent than the far more frequent active construction does (cf. §6.2.). However, in the vast majority of significant factor groups, the most frequent contexts are also the most conservative ones, thus generally confirming the important role of entrenchment in slowing down syntactic change, even at the micro-level of individual contextual factors.

As a result of this split pattern, *haber de* remains the overall most frequently used variant during the 18th century and beyond; despite the gradually increasing proportion of contexts favouring the variant [*deber (de) + inf.*], its overall dominance only comes to an end at the beginning of the 20th century (cf. Blas Arroyo & Vellón 2014).

Summing up, the end of the 18th century represents a milestone in the evolution of modal periphrases in Spanish, though there is also evidence that certain patterns of variation continue beyond this period. Future variationist studies examining the more recent history of Spanish will be able to reveal whether the variables identified as significant in this analysis continue to play a role in the further decline of [*haber de + inf.*] and the consequent rise of its competitors, or whether other factor groups gain importance and begin to affect the language users' choice between the different available modal periphrases, thereby implicitly contributing to long-term language change.

## References

- Albelda, Marta. 2007. *La intensificación como categoría pragmática: Revisión y propuesta*. Frankfurt am Main: Peter Lang.
- 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 & Juan González. 2014. La alternancia *deber/deber de + infinitivo* en el siglo XVI: Factores condicionantes en un fenómeno de variación sintáctica a partir de un corpus epistolar. *Spanish in Context* 11(1). 76–96.
- 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

Con formato: Inglés (británico)

- textos de inmediatez comunicativa. *RILCE. Revista de Filología Hispánica* 32(1). 47–81.
- Blas Arroyo, José Luis & Javier Vellón. 2014. 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.
- Blas Arroyo, José Luis, Margarita Porcar & Javier Vellón. 2013. Un hito clave en la evolución de las perífrasis modales de infinitivo: Análisis sociolingüístico de la alternancia *haber de/tener que + infinitivo* en textos de inmediatez comunicativa del siglo XIX. *Revista de Historia de la Lengua Española* 8. 29–62.
- Brown, Roger & Albert Gilman. 1960. The pronouns of power and solidarity. In Thomas A. Sebeok (ed.), *Style in language*. 253–276. Cambridge, MA: MIT Press.
- 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.
- Croft, William. 2000. *Explaining language change: An evolutionary approach*. London: Longman.
- Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: OUP.
- Danilova, Oxana. 2012. La hipótesis 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* 6, 195–215. Berne: Peter Lang.
- 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 Camarero, Mar. 1997. Acerca de los condicionamientos cognitivos y lingüísticos de la sustitución de 'aver' por 'tener'. *Anuario Galego de Filloxia* 24. 203–235.
- Garachana Camarero, 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.
- Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.

Con formato: Inglés (británico)

Con formato: Inglés (británico)

- 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 City: Fondo de Cultura Económica de España.
- Hertzenberg, Mari Johanne. 2012. The use and development of *habere* + infinitive in Latin: An LFG approach. In Ans M.C. van Kemenade & Nynke de Haas (eds.), *Historical linguistics 2009: Selected papers from the 19th International Conference on Historical Linguistics, Nijmegen, 10–14 August 2009*, 373–398. Amsterdam: John Benjamins.
- Keniston, Hayward. 1937. Verbal aspect in Spanish. *Hispania* 19. 163–176.
- Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36. 15–43.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. 1994. *Principles of Linguistic Change*. Vol. 1. *Internal Factors*. Oxford: Blackwell.
- Langacker, Ronald. 1987. *Foundations of cognitive grammar*. Vol. 1. *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 (eds.), *Actas del VII Congreso Internacional de Historia de la Lengua Española* 1, 789–806. Madrid: Arco Libros.
- Martínez Díaz, Eva. 2003. La frecuencia de uso de haber y tener en las estructuras perífrásticas 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.llf.uam.es/clg8/actas/pdf/paperCLG73.pdf> (accessed March 2, 2015).
- Matsuda, Kenjiro. 1993. Dissecting analogical leveling quantitatively: The case of the innovative potential suffix in Tokyo Japanese. *Language Variation and Change* 5(1). 213–234.
- 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.

Con formato: Inglés (británico)

Con formato: Español

- Okulska, Urszula. 2010. Performing the world of politics through the discourse of institutional correspondence in Late Middle and Early Modern England. In Urszula Okulska & Piotr 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.
- Pereira-Scherre, Maria M. & Anthony J. Naro. 1992. The serial effect on internal and external variables. *Language Variation and Change* 4(1). 1–13.
- Pickering, Martin J. & Victor S. Ferreira. 2008. Structural priming: A critical review. *Psychological Bulletin* 134(3). 427–459.
- Poplack, Shana. 2001. Variability, frequency, and productivity in the irrealis domain of French. In Joan L. Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 405–428. Amsterdam: John Benjamins.
- 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 & Sali Tagliamonte. 2001. *African American English in the diaspora: Tense and aspect*. Oxford: Blackwell.
- Rosemeyer, Malte. 2015. Entrenchment and discourse traditions in Spanish auxiliary selection. In Rolf Kailuweit & Malte Rosemeyer (eds.), *Auxiliary selection revisited: Gradience and gradualness*. Berlin/New York: De Gruyter.
- Sankoff, David. 1988. Sociolinguistics and syntactic variation. In Frederick Newmeyer (ed.), *Linguistics: The Cambridge survey*, 140–161. Cambridge: Cambridge University Press.
- 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.
- Seifert, Eva. 1930. ‘Haber’ y ‘tener’ como expresiones de la posesión en español. *Revista de Filología Española* 17. 345–389.
- 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. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.
- Tagliamonte, Sali A. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Malden/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.

Con formato: Inglés (británico)

Con formato: Inglés (británico)

Con formato: Inglés (británico)

Con formato: Español

- Tarallo, F. 1989. Inside and outside relative clauses. In Ralph Fasold & Deborah Schiffrin (eds.), *Language change and variation*, 255–274. Washington: Georgetown University Press.
- Torres Cacoullou, 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 Cacoullou, 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.
- Walker, James A. 2010. *Variation in linguistic systems*. New York: Routledge.
- Yllera, Alicia. 1980. *Sintaxis histórica del verbo español: Las perífrasis medievales*. Zaragoza: Universidad de Zaragoza.

Con formato: Inglés (británico)

Con formato: Inglés (británico)

### Résumé

On a examiné en profondeur l'histoire des constructions modales en espagnol, en se concentrant essentiellement sur les différences sémantiques entre les alternatives disponibles. Cette communication propose une analyse innovatrice de l'évolution de ces constructions en adoptant une approche diachronique centrée sur la variation, qui prend en considération un ensemble élargi de facteurs sémantiques, syntaxiques, morphologiques et stylistiques ayant influencé le choix entre les périphrases modales en compétition lors des deux étapes-clés de l'évolution de l'espagnol. Les données de l'analyse sont tirées d'un corpus diachronique de correspondance privée, reflétant aussi précisément que possible l'usage réel et courant de la langue pendant les deux périodes données. On porte une attention particulière à la question de la stabilité de l'influence de ces groupes de facteurs différents, et on démontre que les paires forme-contexte les plus fréquentes sont particulièrement résistantes à toute innovation. Ceci pourrait s'expliquer par la stabilité cognitive de chaque variante dans des environnements linguistiques spécifiques.

### Zusammenfassung

Die Geschichte der spanischen Modalkonstruktionen ist in der Fachliteratur eingehend beschrieben worden, wobei sich gewöhnlich das Augenmerk besonders auf die semantischen Unterschiede zwischen den verfügbaren Alternativen richtet. Der vorliegende Beitrag untersucht die historische Entwicklung dieser Konstruktionen in einer variationslinguistischen Analyse, in der verschiedene semantische, syntaktische, morphologische und stilistische Faktoren identifiziert werden, welche die Wahl zwischen den konkurrierenden Modalperiphrasen während zweier wichtiger Etappen der spanischen Sprachgeschichte beeinflussen. Die Daten der Studie stammen aus einem diachronischen Korpus aus privaten Briefen und ähnlich persönlichen Dokumenten, die den tatsächlichen, alltäglichen Sprachgebrauch so genau wie möglich widerspiegeln. Besondere Aufmerksamkeit gilt der Frage, ob der Einfluss verschiedener Faktorengruppen zeitlich stabil ist. Es zeigt sich, dass die am häufigsten vorkommenden Form-Kontext-Paare besonders veränderungsresistent sind, was sich auf kognitives *entrenchment* der betreffenden Variante in bestimmten sprachlichen Kontexten zurückführen lässt.

*Authors' addresses*

Prof. José Luis Blas Arroyo  
Facultad de Ciencias Humanas y Sociales  
Universitat Jaume I  
12071 CASTELLÓN DE LA PLANA  
Spain

blas@uji.es

Dr. Kim Schulte  
Facultad de Ciencias Humanas y Sociales  
Universitat Jaume I  
12071 CASTELLÓN DE LA PLANA  
Spain

kschulte@uji.es