Title: The refuge of a dying variant within the grammar: patterns of change and continuity in the Spanish verbal periphrasis *haber de + infinitive* over the past two centuries.

Authors:

José Luis Blas Arroyo (Universitat Jaume I. E-mail: blas@fil.uji.es.

Phone number: +34 964 729625).

Javier Vellón Lahoz (Universitat Jaume I. E-mail: vellon@fil.uji.es.

Phone number: +34 964 729862)

Short Title: The refuge of a dying variant within the grammar

THE REFUGE OF A DYING VARIANT WITHIN THE GRAMMAR: PATTERNS OF CHANGE AND CONTINUITY IN THE SPANISH VERBAL PERIPHRASIS *HABER DE* + INFINITIVE OVER THE PAST TWO CENTURIES1

ABSTRACT

Based on a corpus of ego-documents (private letters, diaries, memoirs) from the 19th and the first half of the 20th centuries, this article presents a variationist comparative study to determine the fate of the modal periphrasis *haber de* + infinitive in the history of modern Spanish. Detailed analysis of the envelope of variation enables us to show that, despite an abrupt decline in the selection of *haber de* relative to *tener que*, both 'to have to', grammatical environments that favour its use remain in the mid-20th century: many of the factor groups and the hierarchy of constraints during this period are similar to those that operated in previous periods. Nevertheless, a generalised decrease in the explanatory power of these factor groups, as well as some divergent patterns within several of these groups are also observed, mainly as a result of the fact that *haber de* + infinitive is increasingly relegated to some restricted areas of the grammar and lexicon. Based on these results, some theoretical implications for changing rates and constraints in language change and grammaticalisation are discussed.

Key words: *Haber de/tener que* + infinitive, syntactic variation, language change, grammaticalisation, variationist comparative approach, corpus linguistics, Spanish

INTRODUCTION

In the summer of 1936, the father of the man who was to become a leading figure in the Spanish Socialist Party, Victor Manuel Arbeloa, wrote eleven letters to his wife, Josefina. In these personal missives he shared his exploits on the San Sebastián front,

shortly before being seriously wounded in the siege of this Republican stronghold. On reading those letters, one finds fragments such as the following:[2]

(1) …si tengo tiempo voy un rato a la iglesia a la noche mientras los mozos se van al bar, aunque algún día ya les acompaño a echar una copa, todo no *ha de ser* tampoco estar pensativo y triste. (*Once cartas de mi padre*, 8-25-1936)

'… if I have the time I go to church for a while at night while the lads go to the bar, although some days I go with them to have a drink; one doesn't *have to be* thoughtful and sad all the time, either.'

(2) Ahora que me acuerdo, el domingo, según la Epístola, era el de los ramos y todo aquello. Ahora todos *tenemos que ser* bravos. (*Once cartas de mi padre*, 8-25-1936)

'Now that I remember, that Sunday, according to the Epistle, was Palm Sunday and all that. Now we all *have to be* courageous.'

These examples illustrate the alternating use of two Spanish infinitive periphrases, *haber de* and *tener que* + infinitive, both of which convey deontic content, associated with the notions of obligation or necessity, 'to have to', and the main verb is identical in both examples (*ser* 'to be').

Research has confirmed the coexistence of these periphrases, with a predominance of *haber de* until recent times. However, after a lengthy process of grammaticalisation, *tener que* has become the dominant choice, relegating its competitor to a few formal registers and certain regional varieties both in Spain and in Latin America (e.g., Fernández de Castro, 1999; García Fernández, 2006, 2013; Gómez Torrego, 1988, 1999; Hernández Díaz, 2006; Hernández García, 1998; López Izquierdo, 2008; Martínez Díaz, 2003, 2008; Sinner, 2003). However, most studies have set their sights on issues of a

modal and, to a lesser extent, temporal nature, ignoring other structural constraints that may have played a part in this linguistic change.

Thus, an important aim of this study is to examine the trajectory of *haber de*—once the star in the paradigm of modal periphrases—over the last two centuries, and more specifically in the first half of the 20th century, when it appears to finally give way to its competitor. We will attempt to answer this question by comparing two corpora consisting of the same type of documents, collections of private letters, diaries and memoirs written by Spaniards from different social and dialectal backgrounds.[3] As stated by Poplack (2011:212), the objective of the variationist approach to grammaticalization "is not just to record the grammaticalising form, but to compare the structure of the context hosting it at each stage over as long a time frame as possible." Variation theory provides accountable measures of several key parameters in grammaticalisation theory, such as persistence, bleaching and generalisation, several examples of which will be presented in this study. The variationist comparative method (Poplack & Tagliamonte, 2001) adopted here involves the comparison of two independent variable rule analyses (one for each century). Nevertheless, unlike other studies dealing with grammaticalisation that are primarily interested in the emergent structure, in this paper we will focus on the receding variant. We aim to trace the fate of *haber de*, the prevailing modal periphrasis for many centuries, which in the 20th century undergoes a sudden decline in favour of *tener que,* its previously weaker competitor.

HABER DE / TENER QUE + INFINITIVE IN SPANISH. A BRIEF HISTORICAL REVIEW

The origins of the modal periphrases of obligation/necessity lie in the corresponding Latin constructions formed with the verbs *debeo* 'owe' and *habeo* 'have'. The

periphrases with the verb *tener* 'hold, have' emerged later as a result of its historical semantic-functional analogy with the verb *haber*. In fact, the appearance of *tener* as a periphrastic auxiliary within the domain previously reserved for *haber* is linked with the general process of semantic neutralisation of both verbs in their possessive sense at the end of the medieval period (Garachana & Rosenmeyer, 2011; Gili Gaya, 1970; Hernández Díaz, 2006; Seifert, 1930).

The lexical erosion of *haber* was furthered by its generalised usage as an auxiliary in the formation of compound tenses and its extension as an existential and impersonal verb. Its shift towards modal as well as temporal values also played an important role in its increasing association with the infinitive. According to Yllera (1980:100-101), modal periphrases with *haber* in mediaeval Spanish expressed different types of obligation (attenuated necessity, moral obligation, obligation imposed by law, commands and orders or custom, etc.). Lapesa (2000:882) also comments on the variety of modal meanings of *haber de*, together with its uses with a prospective or future value, or simply as the equivalent of a simple form (the so-called *pleonastic* periphrases). In short, *aver de* (occasionally also *aver a*) + infinitive had three basic usages in the Middle Ages: first, with a modal value of obligation/necessity ('Corrié un rio bono cerca de la mongía, *aviélo de passar* el monge…' 'A big river flowed near the monastery, the monk needed to cross it…'); second, with a temporal meaning of subsequent or future action ('Oráculo: "ca todo *ha* asy *de venir* e non puede otra cosa ser, pues que a los dios plaze"' 'Oracle: "Everything will come that way, it cannot be otherwise, because that pleases the gods"'); and third, as a pleonastic periphrasis ('corrió par el astil ayuso, las manos *se ovo de untar*' 'He ran down with the stick, he smeared his hands…' (examples taken from Yllera (1980:101-108).

At the end of the Middle Ages, the obligative periphrases with *tener* as the auxiliary also became firmly established in Spanish, coexisting with *haber* in the semantic field of modality and, to a lesser extent, in the temporal reference system. This construction gradually became more popular from the 16th and 17th centuries onwards, until it finally became predominant in modern Spanish, at the expense of both *haber de* and *tener de* (Blas Arroyo & González 2014; Gómez Torrego, 1999; NGRALE, 2009). As the Real Academia Española (2009:2146-47) recalls, the use of *haber de* is today commonly relegated to the more formal registers in European Spanish, as well as to some varieties influenced by Catalan in Eastern Spain (see also Blas Arroyo, 2014; Sinner, 2003). Nevertheless, in Latin American Spanish (especially in Mexico, Central America and the Antillean islands), these uses are extended to other levels, especially in the expression of prospective (non-modal) meanings, as a result of an accelerated process of displacement of the morphological future by temporal and obligative periphrases. Apart from the (main) obligative and (to a lesser extent) future uses, the expression also appears nowadays in several idiomatic expressions (i.e. 'agua que no *has de beber*' 'if you're not interested, don't spoil things for me/for other people'), as well as in some expressive ('¿siempre se *ha de sentir* lo que se dice?' 'do you always *have to feel* what you say?' and conjectural meanings, '*hubo de haber leído* el Quijote` 'he must have read the *Quixote*'.

In Table 1, based on López Izquierdo (2008:793), the raw frequencies of different personal periphrases over a period of six centuries (1200s-1900s) are given[4] based on data extracted from the *Corpus del Español* (Davies, 2002), a 100 million words corpus mainly from texts belonging to formal discourse traditions, such as literary works, moral and pious books, administrative and scientific works. As illustrated in Figure 1, the proportions of the periphrastic units indicate a fundamental

change in this grammatical paradigm in the past century, confirming that, in the 18th and 19th centuries, *haber de* was still preferred over *tener que* in all text types (Martínez Díaz, 2003).

TABLE 1: *Raw frequencies of* haber de / tener de/que *+ infinitive by centuries in the* Corpus del Español (Davies, 2002; López Izquierdo, 2008:793)

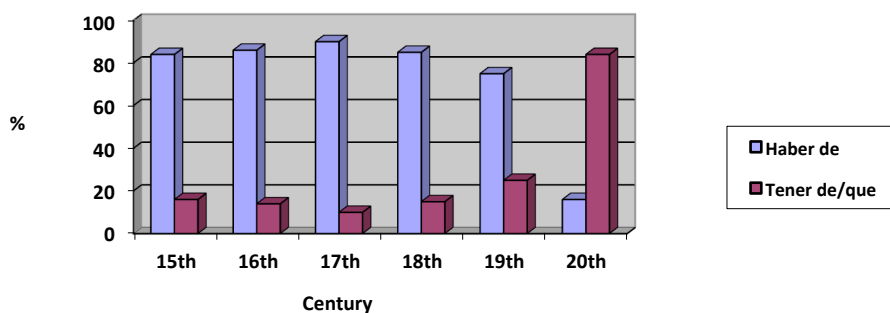|	**15th**	**16th**	**17th**	**18th**	**19th**	**20th**
haber de	2339	17643	17329	4963	8458	1492
tener de	307	1249	1082	284	274	91
tener que	59	666	644	536	2406	7793

FIGURE 1: Distribution of the periphrases made with auxiliary verbs *haber* and *tener* by century (%).[5]

With this in mind, the aims of this study are the following:

a) We shall attempt to establish which factors condition the selection of this periphrasis in the 20th century, considering all potential factors simultaneously and thereby allowing us to compare their explanatory magnitudes and hierarchy; and b) we shall check the (in)consistencies of those factors against the data from the 19th century, with the aim of analysing the process of change that affects these periphrases.

As we shall see, despite confirming the generalised extension of *tener que* to the linguistic contexts that were previously occupied by or shared with *haber de*, the system

still offers some refuge for the latter. Most of the relevant factors are the same as in previous periods, though with some changes in their explanatory power and the hierarchy of constraints.

CORPUS AND METHODOLOGY

Within the framework of a wider research project on historical sociolinguistics (see note 2), we compiled a corpus of written ego-documents for this study. Such materials are considered to be more informal and closer to the vernacular than other, more formal text types (Oesterreicher, 2004). The texts, mainly private letters and (to a lesser extent) several autobiographical works, were written by Spaniards of different social and dialectal origins. Various registers are represented, ranging from documents dealing with more personal matters to others of a less intimate nature.

The private properties of these texts make them attractive for the study of informal language in earlier periods for which no oral testimonials have survived. This is especially true in the case of personal letters (Elpass, 2012). They contain many autobiographical details, which make it possible to determine the relationships of power and solidarity between senders and addressees, as well as their social status (Okulska, 2010). Likewise, they contain ethnographic information that enables researchers to unravel some of the details of social life in bygone times (Raumolin-Brunberg, 2005). Moreover, the letters were not written with the intention of their ever being published, which ensures that the language employed in them is closer to the vernacular than other types of discourse.

For the 20th century, the corpus used here contains 2045 letters and two autobiographical works, amounting to a total of 695,090 words, written by more than 350 different authors. The 19th-century corpus, on the other hand, contains 1389 letters,

two autobiographical texts and one account book, authored by approximately 250 different writers, totalling 490,014 words. In Table 2, the period covered by this diachronic corpus is subdivided further into 33-year sub-periods. For a complete list of the sources and the corresponding time periods, see the Appendix.

Table 2: *Number of texts in the corpus, broken down by time and document type*

Period	Number of texts
1800-1832	215 letters and 1 diary
1833-1866	466 letters, 1 diary and 1 account book
1867-1899	708 letters and 1 memoir
1900-1932	970 letters and 1 memoir
1933-1966	1075 letters and 1 diary

A concordance program (*Wordsmith* v.4) was used to locate all the occurrences of the two variants. This method resulted in a total of 1326 tokens: 282 from the 19th century and 1044 from the 20th century materials.

The tokens were then coded on the basis of 14 factor groups, namely: 1) length of the periphrasis; 2) phonemic context (phoneme following the complementiser); 3) modal shades of meaning; 4) sentential modality; 5) tense and mood; 6) person and number; 7) agentivity; 8) level of semantic (im)personality; 9) lexical aspect of the main verb; 10) clause type; 11) contextual modalisation; 12) syntax of the main verb; 13) subject expression; and 14) lexical priming. In Table 3 we show a representative example for every factor group in which the first periphrasis is the instance found in the corpus (sometimes simplified for the sake of brevity) and the second is the same sentence with the competing periphrasis, provided in order to show the contextual equivalence between them.

Table 3: *Examples of the factors and factor groups considered in the multivariate analysis*

Factors	Example

Length of the verbal group	
4 syllables or less	Esta es la noticia que te *tengo que/he de dar* 'This is the news that *I have to give*'
5 syllables or more	Por muy terrible y dolorosa que sea mi desaparición física *tienes que/has de sobreponerte* 'No matter how painful and terrible my physical disappearance would be, *you must get over it*'
Phonemic context	
Vowel	Te lo *habría de agradecer* muchísimo 'I would have greatly appreciated it'
Consonant	"Y *habían de ser* cocidas" dije yo '"And *they should be boiled*", I said'
Cacophonous	Si te *he de decir* la verdad, ya desconfiaba de que me contestase 'If *I am to tell* you the truth, I already doubted that he would answer me'
Modal shades of meaning	
Internal obligation	He pensado que me la *tenía que/había de tomar* yo también 'I've been thinking that *I should have taken* it, too'
External obligation	En unos minutos *hube de/tuve que terminar* la carta 'In just a few minutes *I had to finish* the letter'
Necessity/Advisability	*He de/tengo que completar* el libro con algo más de lectura '*I must finish* the book with a little more reading'
Other	Quien *ha de/tiene que saber* bien eso es el marido de Olivia 'The one *who must be well aware* of this is Olivia's husband'
Sentential modality	
Affirmative sentences	Todos *tenemos que/hemos de ser* bravos 'We all have to be brave'
Negative sentences	La política de ahora no *ha de/tiene que ser* de engaños 'The politics of today must not be about deceiving'
Other	¡Cómo no *tengo que/he de tenerte* cariño! 'How *can I not be fond* of you!'
Tense/mood	
Present indicative	*He de/tengo que entregar* el libro otra vez '*I have to return* the book'
Imperfect indicative	Yo no perdía la confianza de que *tenía que/había estar* allí 'I remained confident that *it was bound to be* there'
Other	*Habría de/tendría que copiarse* el libro entero 'The whole book *would have to be copied*'
Person/number	
3rd person	Esa mujer forzosamente *ha de/tiene que ser* sevillana 'This woman *must necessarily be* from Seville'
Other	*He de/tengo que decirles* que he tenido a José María muy presente en la Santa Misa '*I must tell you* that I bore José María very much in mind in the Holy Mass'
Agentivity (3rd person)	
Human	Siempre *tienes que/has de ir* cargado 'You always *have to carry* a lot'
Non-human	Todo no *ha de/tiene que ser* tampoco estar pensativo y triste todo el tiempo 'One *doesn't have to be* thoughtful and sad all the time, either'
Level of (im)personality	
Active sentences	Creo que *tendrán que/habrán de avanzar* las otras fuerzas 'I think that the other forces *will have to* move forward'
Passive/Impersonal sentences	No hay tanta gana de fiestas, aunque *se tiene que/ha de pasar* 'There's no great desire for celebrating, but it just *has to be got over* with'
Lexical aspect	
Speech verbs	*He de/Tengo que confesar* mi pasión por Granada y Ronda '*I have to confess* my passion for Granada and Ronda'
Stative verbs	La caja *ha de/tiene que ser* sólida 'The box *must be* solid'
Motion verbs	*He de/tengo que ir a verte* aunque no estés en Pamplona '*I have to go see you* even if you are not in Pamplona'
Other (dynamic) verbs	*Tenemos que/hemos de hacer* cuanto podamos '*We must do* all we can'
Type of clause	

Subordinated	Es posible que sea la inteligencia la que *haya de/tenga que entrar* en juego 'It is possible that it is intelligence which *has to come* into play'
Other	Después de todo les *tengo que/he de dar* muchas gracias por el paquete 'After all, I *have to thank you* very much for the package'
Contextual modalisation	
Intensified contexts	Es duro *tener que/haber de explotar* esta maldita esclavitud 'It's hard *having to exploit* this damned slavery'
Other	Un compañero no podía ir al trabajo a la hora que *tenía que/había de ir* 'A fellow could not go to work when *he had to go*'
Syntax of the main verb	
Simple	Hace ocho días que *tengo que/he de dormir* boca abajo 'It's been eight days that *I have had to sleep* face down'
Compound	*Tenía que/había de haber visto* la verdad '*He must have seen* the truth'
Syntax of subject	
Omitted	*Has de/Tienes que saber* que el día 17 Adonis me escribió '*You have to know* that on the 17th Adonis wrote to me'
Explicit	Aquí cada cual *tiene que/ha de* atender a su trabajo 'Here everyone *has to attend* to his work'
Lexical priming	
Same periphrasis	*Tiene que usar* gafas, padece del higado y *tiene que trabajar* mucho 'He *has to wear* glasses, suffers from a liver condition and *has to work* a lot'
Other modal periphrasis	Pastora *tiene que estar* bajo órdenes del yerno y siente disgusto. Quien *ha de saber* bien eso es el marido de Olivia 'Pastora *has to be* under the orders of her son in law and she is unhappy. The one who *must be well aware* of this is Olivia's husband'
None	Le queda poco para terminar, pero para el año que viene *tiene que/ha de ir* al Servicio Militar 'He'll soon be finished, but next year he *has to do* his military service'

For the quantitative analysis, we apply the sociolinguistic comparative method (Poplack & Tagliamonte, 2001), in which two independent multivariate analyses (with identical factor groups) are carried out and then compared. By comparing the data of two historical periods, we can trace not only the fate of the emerging and the receding variants, but also the path along which they enter or leave the system, i.e., the trajectory of their functions, which is of particular interest for the study of grammaticalisation (Poplack 2011:215). These analyses are carried out using Goldvarb X (Sankoff, Tagliamonte, & Smith, 2005), with the periphrasis *haber de* as the application value.

In this study, the variable context is both form and function-based. On the one hand, we consider tokens of two different constructions, but at the same time we limit the multivariate analysis exclusively to the expressions that are clearly modal, such as those exemplified in (1) and (2) above. We do not include purely temporal meanings,

expressing future rather than modal senses, most commonly when the speaker or author uses the periphrases to situate events in a more or less defined future, i.e., competing with other prospective forms such as the morphological or the periphrastic future tense:

(3) Tú, María, no te muevas de ahí mientras no vaya yo; si no, no *he de ir a verte* (=iré a / voy a ir a verte) aunque estés en Pamplona (*Once cartas de mi padre*, 9-2-1936)

'You, Maria, don't you move from there until I go there; otherwise, I shall not go and see you even though you are in Pamplona'

(4) Así es que se me va la vista escribiendo y no tengo aliento para levantarme. Así es que si llego a salir, no *tengo que valer* (=valdré / voy a valer) nada (*Solo habremos muerto si vosotros nos olvidáis*, 9-1-1940*)*

'So here I am losing my eyesight writing and I don't have the strength to get up. So if I do manage to leave, I will surely be of no use for anything'

The prospective uses of *haber de* in (3)—equivalent to the morphological and the periphrastic future—were far more frequent in medieval and early modern Spanish (NGRALE, 2009:2146). On the other hand, although in the Spanish of times gone by these non-modal, temporal meanings were also common in the periphrases with *tener* (Blas Arroyo & González, 2014; Yllera, 1980), five centuries later, their uses as a future variant have practically disappeared, although we still come across the occasional case, as in (4) above. Consequently, in the pages that follow we will focus our attention exclusively on the semantic field of modality, the main locus of variation and change between our periphrases. As a result, the number of tokens finally considered in the analysis is reduced to 1206: 224 from the 19th century and 982 from the 20th century corpus.

RESULTS AND ANALYSIS

Overall results

In the 19th century, the frequency of *haber de* + infinitive versus *tener que* is 46%

(N=104/224) but in the 20th century it is 22% (N=215/767).[6] These figures confirm that,

by the mid-20th century, *haber de* + infinitive had already lost much of the prevalence it

had enjoyed in previous centuries, whereas it still accounted for just over half of all

modal periphrases in the 19th century. Figure 2 shows that the rate of *haber de* remains

very steady throughout that century, where the temporal axis is divided into shorter

periods of 33 years (compare Figure 1). In contrast, the same graph shows a sharp

decline in the use of *haber de* as the 20th century advances; from the 1930s onwards,

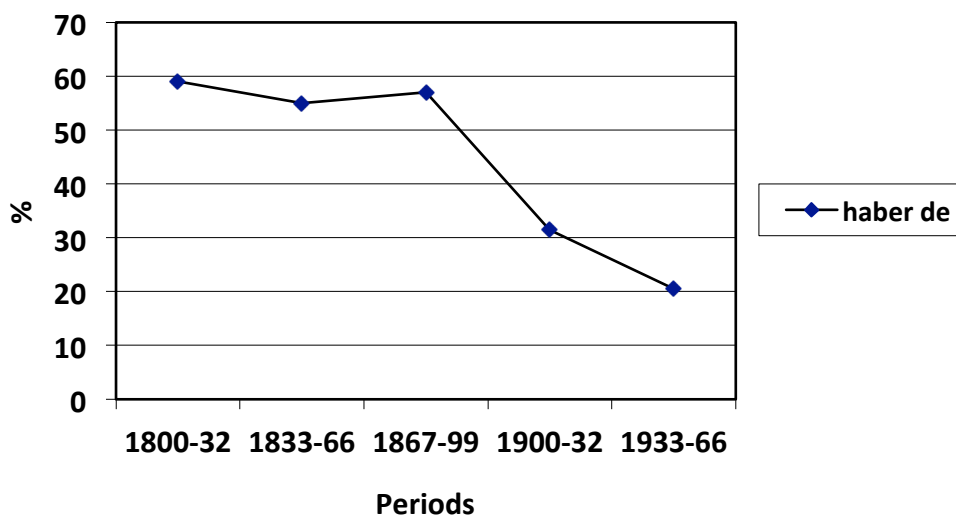the overall frequency of *haber de* barely reaches 20%.

FIGURE 2. Usage frequency (%) of *haber de* + infinitive relative to *tener que* + infinitive over the 19th

and 20th centuries.

The results of the two independent multivariate analyses, from which all tokens with

future non-modal values are excluded, are presented in Table 4. In the following

sections, we will examine the factor groups and constraints that have a statistically

significant effect on the choice of *haber de*.

Table 4. *The effect of linguistic factors on the selection of the verbal periphrasis* haber de + *infinitive in modal (non-future) contexts based on two independent variable rule analyses of the data for the 19th and 20th centuries, respectively*

	19th century N = 224, Input= .45			20th century N = 982, Input= .18		
Factor Groups	**FW**	**%**	**Total N**	**FW**	**%**	**Total N**
Modal senses						
Other (non-deontic)	.83	74	19	.87	64	61
-Epistemic		(60)	(10)		(65)	(52)
-Expressive (surprise…)		(89)	(9)		(56)	(9)
Internal obligation	.55	57	28	.50	24	161
External obligation	.49	42	133	.45	16	537
Necessity/Advisability	.32	39	43	.48	22	209
Range	*52*			*42*		
Sentential modality						
Affirmative sentences	.54	48	189	[.49]	21	882
Negative sentences	.18	17	24	[.53]	24	79
Range	*36*					
Tense						
Present indicative	.64	58	132	.56	25	583
Imperfect indicative	.56	52	25	.50	21	85
Future indicative	.41	40	15	.61	29	72
Other	.24	15	52	.33	15	192
Range	*40*			*28*		
Person/number						
3rd person	.62	60	101	.58	32	382
Other	.38	38	111	.44	16	551
Range	*25*			*14*		

Agentivity (3rd person)						
Non-human	.59	73	30	.61	50	135
Human	.46	55	71	.44	23	247
Range	*13*			*17*		
Level of (im)personality						
Passive/Impersonal sentences	.82	90	19	.92	77	39
Active sentences	.46	42	205	.46	20	944
Range	*36*			*46*		
Lexical aspect						
Stative verbs	.62	57	60	.51	25	392
Other (dynamic) verbs	.59	51	77	.48	20	339
Speech verbs	.40	39	41	.71	31	130
Motion verbs	.30	31	45	.29	7	122
Range	*32*			*41*		
Type of clause						
Subordinate	.55	50	143	.54	24	552
Other	.40	41	80	.44	19	430
Range	*15*			*10*		

Note: The factor groups lexical priming, next phonological context, length of the verbal group, contextual modalisation, syntax of the main verb, and subject expression are not included here as they are not significant.

Modal meanings

We begin our analysis by examining the factor group that has virtually monopolised the debate on this case of grammatical change among Spanish linguists, namely modality. Among the modal meanings associated with the use of both periphrases, in the literature

we find a very strong connection with deontic modality, which encompasses a range of meanings including obligation, permission and necessity (Bybee, Perkins, & Pagliuca, 1994; Fernández de Castro, 1999; García Fernández, 2006, 2013; Gómez Torrego, 1988, 1999; Keniston, 1937; López Izquierdo, 2008; Martínez Díaz, 2008; Olbertz, 1998; Yllera, 1980; , among others). Indeed, a clear majority of all tokens in the corpus (88%) express deontic modality. However, other modal meanings conveyed by these periphrases are the emphatic expression of notions such as surprise, indignation, recrimination (Gómez Torrego, 1999:3356). This is illustrated in the following examples, where the writer verbalises his surprise (5), or emphatically confirms something obvious (6). These expressive values seem to be more closely associated with *haber de* (72%, N=13, combining both centuries due to small raw frequencies):

(5) Su descripción es demasiado lacónica pero a pesar de ello muy favorable por reunir todas las cualidades que prefiero en la mujer. ¿Por qué no *ha de mandarme* su fotografía? (*Madrina de guerra*, 2-9-1938)

'Her description is too laconic but nevertheless very favourable, since it includes all the qualities I prefer in a woman. Why wouldn't she send me a photograph of herself?'

(6) … para mí eres mucho más que un sobrino pero porque te ayudé a criar, te tuve muchas veces en mis brazos y entonces cómo no *tengo que tenerte* cariño (*As cartas do destino*, 7-7-1958)

'… you are far more than a nephew to me, but because I helped bring you up, I held you in my arms many times, so how can I not be fond of you?'

Tokens with an epistemic meaning have also been collapsed with these non-deontic tokens. In these cases, the speaker makes reference to an event or state of affairs that he believes to be probable, presumable or approximate. Those uses appeared for *haber de* in earlier periods (Yllera, 1980), but today they seem more frequent in Latin

America than in European Spanish (NGRALE, 2009:2147). As for *tener que*, according to López Izquierdo (2008:802) "… they began to spread from the late 18th century onwards, and above all during the 19th century" (our translation). (7) and (8) are examples of these usages:

(7) Quien *ha de saber* bien eso es el marido de Olivia (*Una familia y un océano de por medio*)

'The one who must be well aware of this is Olivia's husband'

(8) … el cartero leía las cartas y no faltaban más que tres sin haber aparecido la tuya aún, aunque no perdía la confianza de que *tenía que estar* allí por ser ya hoy jueves (*Once cartas de mi padre*, 9-10-1936)

'… the postman was reading the letters and when only three were left, yours had still not appeared, but I remained confident that it had to be there because today is Thursday'

In the 20th-century corpus, the preference for *haber de* (65% N=34/52) over *tener que* for the expression of epistemic modality remains. This trend of using *haber de + infinitive* in epistemic contexts is also aided by the increasing favoring with non-human subjects (see '*Person/number and degree of animacy of the subjects*' below), which are a long way from the most prototypical sense of obligation, since commands need a human being to be required to obey an order. In fact, cross-tabulation of these factors shows that in the epistemic-non-human contexts, the uses of *haber de* exceed (71%; N=22/31) those of *tener que* by a wide margin. In any case, the proportion of epistemic tokens in the corpus is very low in comparison with *deber* (*de*), the default periphrasis for expressing conjecture in the first half of the 20th century (Blas Arroyo & Vellón, 2014).

In the end, Table 4 shows that non-deontic contexts are by far the best allies of *haber de*, with factor weights very high in both centuries (19th c.: .83; 20th c.: .87).With

regard to deontic modality, which is expressed by the vast majority of tokens analysed here, a number of authors have attempted to pinpoint its different shades of meaning (Bybee, Perkins, & Pagliuca, 1994; Fernández, 1999; García Fernández, 2006, 2013; Gómez Torrego, 1988, 1999; Keniston, 1937; López Izquierdo, 2008; Martínez Díaz, 2008; Olbertz, 1998; Yllera, 1980). However, this is by no means a straight-forward task; among other things, the imposition of the analyst's own subjective interpretation can lead to circularity (Tagliamonte & Smith, 2006:345). In order to prevent this from occurring to the extent possibe, we have divided the deontic axis into two categories that can be measured more objectively:

a) the degree of obligation/necessity or advisability imposed; [7] and

b) the agent that imposes that obligation/necessity.

Combining these two categories results in the following main values:[8]

Subjective or internal necessity or obligation ("internal obligation" in Table 4). This refers to duties that are generated by inner conviction, or by the subject's will or intention based on reasons that may be of a religious, ethical or philosophical nature, or that arise from gratitude, respect or any other internal motivation. It is therefore based on the convictions or the desire of an agent. Thus, the need to fulfil the obligation is felt primarily by the agent, which places these periphrases closer to those of a volitional nature (Roca Pons, 1980:73; Yllera, 1980:114). In case of no coincidence between the speaker and the subject of the clause, it is usually the former who refers to a moral obligation of the latter based on the above mentioned moral values, as in *has de/tienes que* obedecer a Dios 'you must obey God'). (9) and (10) are typical examples of tokens expressing internal obligation.

(9) Creo que la política de ahora no *ha de ser* de engaños ni es cuestión de forjarnos vanas ilusiones que después la realidad de los hechos ha de desvanecer. (*Un catalanófilo de Madrid*, 4-27-1930)

'I believe the politics of today must not be about deceiving, nor is it a question of dreaming up illusions in vain, only for the reality of the facts to make them fade away afterwards.'

(10) He pensado que ya que tú te habías tomado esta molestia, por educación me la *tenía que tomar* yo también, para darte las gracias por lo que has hecho por hacerme agradable la vida, escribiéndome al frente. (*Madrina de Guerra*, 6-9-1939)

'I was thinking that as you had gone to all that trouble, out of courtesy I had to do the same, to thank you for all you have done to make life more pleasant for me by writing to me on the front.'

Agent-oriented or external obligation ("external obligation"). This is obligation in its most literal sense, that is, the unavoidable necessity or the imperative/coercive advice are of an *external* nature for the agent of the action described by the verb. Hence, we are dealing with directive statements, among which we can distinguish different shades of meaning, such as obligations imposed by norm, agreement, social convention, legal code, etc., as in (11), or those forced upon by external circumstances beyond the subject's control, as in (12):[9]

(11) Le mando un periódico que tiene que devolvérmelo porque *he de entregarlo* otra vez. (*Epistolario a Federico García Lorca*, 12-2-1926)

'I have enclosed a newspaper for you that you must return to me because I have to give it back.'

(12) …el local de mi exposición es pequeñito y *he tenido que suprimir* muchos cuadros (*Regoyos: sus cartas inéditas*, 3-30-1908)

'The venue of my exhibition is tiny and I've had to leave out many paintings'

Necessity or advisability felt by the speaker, and therefore with far less coercive power than that expressed in cases such as those above. Thus, unlike example (12), in (13) and (14) the senders of the respective letters say that it is advisable to perform certain actions that they themselves will benefit from:

(13) Allí mismo escribí otro volumen que título "Cinco hombres", impresiones sobre Pablo Iglesias, Jaime Vera, Tomás Meabe, Largo Caballero y Julián Besteiro. Son a la vez crítica de un libro de cada uno de ellos. *He de completarlo* con algo más de lectura. (*Dramas de refugiados*, 9-6-1946)

Right there I wrote another volume entitled "Five Men", impressions about Pablo Iglesias, Jaime Vera, Tomás Meabe, Largo Caballero and Julián Besteiro. At the same time they are a review of a book by each one of them. I must finish it with a little more reading.'

(14) Esto que dices de la ropa, si me hace falta, pues no me hace falta que aquí solo se tiene que tener lo más necesario porque si no siempre *tienes que ir* cargado y con poca ropa hay bastante y no padezcas por mí. (*Cartas del iaio*, 10-12-1938)

'What you say about clothes, whether I need any, well I don't because here all you need are the bare essentials, otherwise you always have to carry a lot and you can get by with few clothes and don't worry about me.'

Table 4 shows, in addition to the overall reduction in the use of *haber de* in the passage from the 19th to the 20th century, that some of the factors within the Modal senses factor group are ordered in a similar way, with non-deontic meaning in the lead, followed at a considerable distance by internal obligation. However, these cases of persistence in the semantic distribution of the receding variant stand in contrast to changes in the two remaining contexts: In the 20th-century corpus, agent-oriented

obligation is the least favourable setting for *haber de*, whilst in the 19th century the least favourable context is necessity.

It should be noted that the considerable magnitudes, indicated by the ranges, of the modality factor group in both periods can be explained mainly by the high association of *haber de* with non-deontic meanings, as previous stated. On the other hand, there is far less difference within the deontic field, especially in the 20th century, which confirms the pervasive extension of the emergent *tener que* in this modal use.

Summing up, it can be observed that by the middle of the 20th century, *haber de* has found a special refuge in non-deontic modal use, such as the expression of probability and conjecture, as well as in several expressive and emphatic usages. In contrast, in the far more common domain of deontic modality, the use of this periphrasis decreases considerably. It seems, thus, that the regression of *haber de + infinitive* as a result of the pressure of its competitor *tener que + infinitive* took a similar path to that followed in the lexical area. *Tener* first discarded *haber* as a possessive verb in the area of prototypical possession, that is to say, in those contexts in which the possessor was [+human, +agentive, +volitive, +controller] and the possessed was [-human, -agentive, -volitive, -controller]. Little by little, *tener* reached the non-prototypical side of the meaning of possession. In a similar way, *tener que + infinitive* beat *haber de + infinitive* in most prototypical modal meanings, that is to say, deontic senses, whilst *haber de + infinitive* maintained meanings connected to conjecture and probability.[10]

However, there are also other structural factors that explain the process of variation and change we are dealing with here and which have gone practically unnoticed in the linguistic literature of Spanish. It is on these that we will focus in the following sections.

4.3 *Sentential modality*

As in the case of modal senses, both convergent and divergent tendencies are observed within this factor group from the diachronic perspective, as illustrated in figure 3:
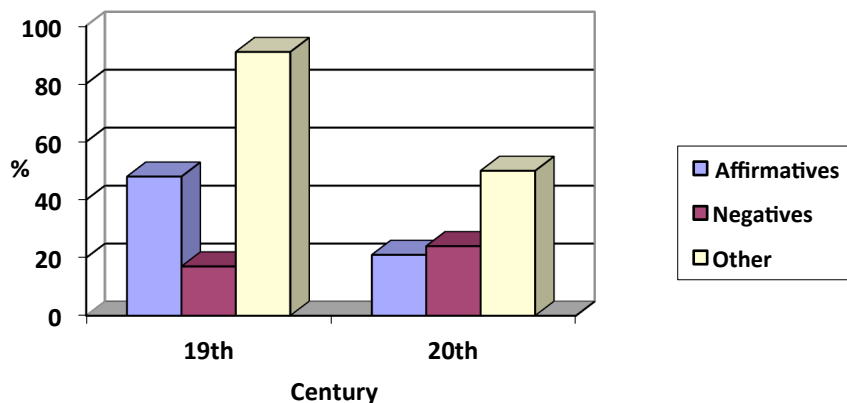
FIGURE 3: Rate of *haber de* relative to *tener que*, broken down by sentential modalities in the 19th and 20th centuries.

Clearly, both in the 19th and the 20th century, non-declarative sentences (mainly interrogatives and exclamatives) are associated with *haber de* (20th c.: 50%; 19th c.: 91%). Notable, however, is the very low number of these syntactic contexts (19th c.: N=11; 20th c.: N=22) compared with declarative sentences (19th c.: N=213; 20th c.: N=961)—not to mention the noteworthy interaction with other factors such as semantic modality. Indeed, the cross-tabulation between these two factor groups shows that a majority of non-declarative sentences contain periphrases with non-deontic modal meanings, i.e., those that are very highly associated with *haber de*, as seen above. In the end, this explains why we have focused the multivariate analysis exclusively upon the two groups of declarative sentences: affirmatives and negatives.

Unlike semantic modality, this factor group is now only selected in the 19th century, where negative contexts strongly disfavoured the selection of *haber de* (.18; 17%) (on the role of negation in grammaticalization, see Givón, 1979; Poplack & Dion,

2009:575; Tagliamonte, Durham, & Smith, 2014; Torres-Cacoullos & Walker, 2009). Nevertheless, differences between affirmative and negative polarity are entirely lost in the 20th century, when the frequency differences of *haber de* in these contexts are completely neutralised.

Tense and mood

The explanatory power of Tense and mood, unlike that of Sentential modality, is maintained in the 20th century. *Haber de* manages to hold its own in the more favouring conjugation paradigms already observed in the previous century, although now with a lower magnitude. This is what happens in the case of both the present (19th c.: .64, 58%; 20th c.: .56; 25%) and, to a lesser extent, the imperfect indicative (19th c.: .56, 52%; 20th c.: .50, 21%), the two most favourable contexts for the periphrasis in the 19th century, though with considerably higher frequencies at that time. Amongst the remaining tense/mood paradigms, some cases of almost categorical avoidance of the use of *haber de* can be observed, for instance with non-finite forms of the auxiliary verb (infinitive, gerund, past participle) (19th c.: 8%, N=1; 20th c.: 2%, N=1), as well as compound tense forms (*haber* + past participle) such as the present perfect indicative, with no tokens in the 19th century and only 1 (out of 37) in the 20th century. It appears likely that the reason for this avoidance of *haber de* in combination with the present perfect could be a stylistic one, such as the speakers' reluctance to repeat the same verb twice in a single auxiliary slot (*he habido de* + infinitive).

. The most important change in this factor group affects the future indicative, whose role in the selection of *haber de* changes significantly from one century to the next. From a moderately unfavourable position in the 19th century (.40), this tense becomes the most favourable context for *haber de* in the 20th-century corpus (.61),

exceeding even the present or imperfect indicative, traditionally the most favourable environments for this periphrasis. One possible hypothesis for this change would be that future uses were connected with the tendency of the periphrasis to be employed in epistemic contexts (see above), given the specialisation of this indicative tense in expressing probability and conjectural meanings.

In any event, what can be seen in the 20th century data is that the future indicative emerges as the strongest factor within this partial re-organisation of the variable context. In sum, while frequency drops, *haber de* becomes favoured in a context that previously disfavoured its use.

Subject Person/number and degree of animacy

Another factor group selected in both centuries is the grammatical person and number of the periphrasis. Table 5 shows the usage frequencies of *haber de*, combining three person and two number paradigms, as well as in the non-finite forms of the main verb (infinitive, gerund and past participle).

TABLE 5: *Rate of* haber de *relative to* tener que*, broken down by grammatical person and number, in the 19th and 20th centuries*

	19th century		20th century	
	%	**Σ**	**%**	**Σ**
1st person singular	31	68	20	327
1st person plural	50	20	5	110
2nd personal singular	48	23	18	91
2nd personal plural	--	--	17	23
3rd person singular	60	81	29	273

3rd person plural	60	20	39	109
Non-personal forms	8	12	2	49

Note: No cases of the 2nd person plural were found in the 19th-century corpus.

The table shows that in the 20th century, the 3rd person verb forms, both in the singular (29%) and even more so in the plural (39%), seem to be most likely to trigger the selection of *haber de*. This tendency is already visible in the 19th-century corpus, although at that time with far higher frequencies. Diametrically opposed to this, non-finite verb forms have a clearly unfavourable effect on use of the traditional periphrasis (2%), a result also observed in the 19th century (8%). The other person/numbers paradigms are situated in between these two poles, with a substantial decline in frequency between the 19th and the 20th century (in particular with the 1st person plural).

The multivariate analysis confirms the relevance of this factor in both centuries: 3rd person verb forms continue to be the most favourable for the selection of *haber de* (.58), when compared to all other contexts (.44). This result relates to the fact that the 3rd person is not the personal paradigm that is most prototypically connected to obligation. Although our interpretation of the deontic modality is a broader one, as emphasized above (see note 7), it should be recognised that some directive speech acts, such as orders and the like, cannot strictly be given to a 3rd person. This probably explains why in all the deontic senses considered, 3rd persons show systematically lower uses of *haber de* than other persons. Conversely, in the non-deontic senses, the proportions are inverted: the uses of *haber de* are double (67%; N=31) those of *tener que* (33%; N=15) in 3rd person, surpassing the figures of the other personal paradigms (*haber*: 57%, N=8 vs. *tener*: 43%, N=6).

The continuity of the interaction between 3rd person and the degree of animacy is another revealing fact. According to some authors, in Old Spanish constructions with the verb *haber* seem to have shared a common deagentivising factor. As stated by Stengaard (2003:1151): "… by means of the periphrasis with *aver*, the subject of the action expressed by the infinitive either loses its possible role as the subject–agent or reinforces its role as subject–recipient or patient involved in the verbal action in question" (our translation). This semantic effect is related to the meaning of the verb *haber*, which originally implied non-agentive or receptive possession in which the subject does not exert any control over the possessed object, unlike the verb it competes with: *tener* (Seifert, 1930). From a cognitive perspective, Garachana (1997) explains this opposition in terms of the prototypicality of the possession, according to which *haber* experienced semantic bleaching of the figurative control over the possessed object, whilst this process did not reach *tener* (see also Garachana & Rosenmeyer, 2011).

In order to verify these hypotheses, we analysed the influence of animacy of the subject on the selection of *haber de*, with a distinction being drawn between human and non-human subjects of these 3rd person verb forms.[11] The data from this analysis confirms that the preference for *haber de* is somewhat greater with non-human subjects, making it one of the few contexts studied here in which this periphrasis reaches frequencies similar to those of *tener que* in the 20th-century corpus (.61; 50%). Conversely, this figure is considerably lower with human 3rd person subjects (.44; 23%). Essentially, these figures are similar to those found in the 19th-century corpus, in which a greater association of non-human subjects with *haber de* (.59, 73%) than among the human subjects (.46; 55%) can be identified.

4.6 *Level of (im)personality of the sentences*

A similar trend can also be observed in the area of semantic (im)personality, a factor group in which active sentences on the one hand contrast with passive and impersonal ones on the other. As illustrated in Table 4, it is among the latter, in which the subject is syntactically and/or semantically camouflaged, and consequently less prototypically connected to obligation, that the use of *haber de* is prevalent (77%), whilst only 20% of active sentences use this periphrasis in the 20th century. By contrast, the differences between these two contexts in the 19th century are considerably lower despite the almost categorical use of *haber de* in passive and impersonal sentences (90%).

It should be noted that the overwhelming majority of tokens in the corpus occur in active constructions (95%; N= 944), and it is precisely these structures that favour *haber de* in both the 19th and 20th centuries. Indeed, even as *haber de* recedes, this constraint increases its strength across the corpus, becoming the strongest constraint affecting the selection of this periphrasis in the 20th century.

Lexical aspect and lexicalisation

The lexical aspect of the main verb is also a significant predictor on variant choice. In the analysis of this factor group, patterns of both continuity and divergence can be identified, similar to those discussed with reference to other factor groups (i.e., semantic and syntactic modality, polarity).

Some factors reflect a pattern of continuity. For instance, among the different semantic verb types, motion verbs are the least favourable contexts for the selection of *haber de* in both periods (19th c.: .30, 31%; 20th c.: .29, 7%). In this regard, it is interesting to note that in the 20th-century corpus some of the most frequent motion verbs—*volver* 'to return', *venir* 'to come', *llevar* 'to take', *pasar* 'to pass', *traer* 'to

bring', *salir* 'to leave'—hardly ever appear with *haber de*. An exception to this rule is *ir* 'to go', as shown in Table 6.

The same hierarchy of constraints is also shared by stative verbs, which clearly favour *haber de* in the 19th century (.62), whilst a century later they have lost some of their favourable influence (.51). Yet some particular stative verbs still show an attachment to this receding periphrasis in the 20th-century corpus, with values that are much higher than the mean for the group. This is mainly the case for *ser* 'to be', which appears with *haber de* in 36 out of 76 tokens (47%).

However, the main novelty within this factor group can be observed with the speech verbs, the group of verbs most clearly associated with *haber de* in the 20th century (.71; 31%). This association seems to derive from some kind of lexicalisation of the receding periphrasis with specific lexical verb types, in a process that has also been described for other grammatical variables (cf. Poplack & Dion, 2009; Poplack & Malvar, 2007; Poplack & Tagliamonte, 2001). With this in mind, it is revealing that just over half of the top 15 most frequent verbs appearing with *haber de* in the 20th-century corpus belong to this group of verbs (see Table 6). In descending order of frequency, these verbs are *decir* 'to say', *saber* 'to know', *confesar* 'to confess', *reconocer* 'to recognise', *expresar* 'to express', *escribir* 'to write', *agradecer* 'to thank' and *juzgar* 'to judge'. Especially relevant is *decir* 'to say', which was already one of the verbs co-occurring most frequently with *haber de* in the 19th century data. However, its relative frequency has increased considerably, rising from 13th place in the 19th century to 2nd place century later. On the other hand, a more detailed analysis of these verbs shows that many of their periphrastic uses appear in what can be called "phatic" contexts. In these periphrases, which are close to fossilised lexical expressions and which are common within the epistolary genre, speakers use this type of verb to "enter into

communication" with the addressee or to divide the pieces of information in small parts (Gómez Torrego, 1999: 3354), as illustred in (15) and (16):

(15) Por lo que a nuestro querido y llorado José María se refiere, *he de decirles* que lo he tenido muy presente en la Santa Misa. (*Cartas de dos hermanos requetés*, 5-4-1937)

'As far as our beloved and lamented José María is concerned, I must tell you that I bore him very much in mind in the Holy Mass.'

(16) … pues *has de saber* que el día 17 Adonis me escribió diciéndome que tenia libre y si quería o podía que le saliera en Vigo… (*Una familia y un océano*, 8-1, 1961)

'…because you must know that on the 17th Adonis wrote to me saying that he had time off and if I was able or wanted to I could go to pick him up in Vigo…'

TABLE 6: *Distribution of the main verbs most frequently co-occurring with* haber de *in the 20th century corpus*

Main verb	N
Ser	36
Decir	13
Saber	9
Ir	8
Dar	7
Estar	7
Confesar	6
Reconocer	5
Encontrar	5
Manifestar	4
Escribir	4
Agradecer	4
Parecer	4
Ver	4

Table 6 shows that *haber de* co-occurs particularly frequently with a small number of main verbs, such as *ser* 'to be', *ir* 'to go', *dar* 'to give' and *estar* 'to be', which indicates that a process of specialisation has taken place. As noted by Elsig (2009:19), a receding variant's lexicalisation in some restricted lexical contexts may be related to the loss of .productivity elsewhere. This is supported by the fact that, in our case at least, lexical specialisation takes place, primarily, in a specific set of grammatical settings. Thus, of the 36 tokens of *haber de ser*, 26 (72%) occur with the 3rd person present indicative, i.e., *ha de ser*. With *decir* 'to say', on the other hand, this periphrasis is found primarily in 1st person singular present indicative contexts (54%; N=7), i.e., *he de decir*.

Clause type

In spite of a lower explanatory power than the factor groups already discussed, the selection of *haber de* also seems to be affected by the type of clause that the verbal periphrases appear in, both in the 19th century (range 15) and 20th century (range 10). *Haber de* was weakly favoured in both centuries in subordinate syntactic settings, as opposed to non-subordinate contexts which exert a slight disfavouring effect. This constraint is thus stable across time, supporting the retentive role of subordination in processes of language change noted elsewhere (Blas Arroyo, 2008; Matsuda, 1993; Tarallo, 1989; but some counterexamples in Tagliamonte, Durham, & Smith, 2014; Torres-Cacoullos & Walker, 2009).

DISCUSSION

By the mid-20th century, use of the periphrasis *haber de* had undergone a striking decline in comparison with previous centuries. This change took place both in the

domain of deontic modality, where this periphrasis had been leading for centuries, as well as in the future temporal domain, in alternation with other verbal variants, such as the synthetic and periphrastic future forms. This decrease is particularly visible when comparing usage to early modern Spanish, but also in comparison to the 19th century, in which *haber de* + infinitive was still very much alive. Moreover, an analysis of the diachronic axis has confirmed the continuing downward progression of the periphrasis throughout the first six decades of the 20th century. Judging by more contemporary studies, this decline seems to have continued where *haber de* appears to have been relegated primarily to some formal registers of the written language (Fernández de Castro, 1999; García Fernández, 2006; Gómez Torrego, 1999), as well as a few dialectal uses in some bilingual areas as a consequence of language contact (Blas Arroyo, 2014; Sinner, 2003). The loss of prominence of what had been the dominant periphrasis since the Middle Ages occurs in virtually all the linguistic contexts analysed resulting in replacement by its competitor, *tener que*, as the prevalent form nearly across the board.

This change does not only affect frequencies but also the relative ranking of the factor groups selected across time. As can be seen in Table 7, while some groups keep a similar rank in each period (i.e., person/number, type of clause), the relevance of others clearly changes from one century to another. Thus, if in the 19th century modal meanings appear as the most significant predictor in variation, this role is occupied by the level of (im)personality in the 20th century data. On the other hand, both agentivity and the lexical aspect of main verbs take on a more relevant role in the 20th century than in the previous one. Another change concerns the sentential modality group, with negative clausal polarity disfavouring *haber de* in the 19th century, but the difference between affirmative and negative contexts vanishing in the 20th century.

Table 7. *Relative ranking of the factor groups selected as significant in the 19th and the 20th centuries*

Ranking	19th century	20th century
1	Modal senses	Level of (im)personality
2	Tense	Modal senses
3	Sentential modality	Lexical aspect
4	Level of (im)personality	Tense
5	Lexical aspect	Agentivity
6	Person/number	Person/number
7	Type of clause	Type of clause
8	Agentivity	

However, despite all these differences in frequencies as well as in the ordering of factor groups, the data of this study also show continuity with the underlying grammar. Thus, it is of particular interest that most of the factor groups selected as predictors in the 19th century remain significant in the 20th century, thereby confirming the principle of persistence which characterises many processes of grammaticalisation (Hopper & Traugott, 2003). This is the case for semantic modality, tense, person, agentivity, level of (im)personality, lexical aspect and clause type. And no less significant is the fact that those factor groups not selected by the statistical model in the 19th century (lexical priming, phonological context, length of the periphrasis, contextual modalisation, syntax of the main verb, syntax of the subject) remain the same in the following period.

Moreover, a similar hierarchy of constraints is observed in both centuries in at least four of the factor groups selected across time. In other words, there is remarkable consistency in the operation of the variable grammar throughout obsolescence. For example, we have seen that 3rd person verb forms, both singular and plural, are the most favourable context for *haber de*, ahead of the remaining finite and particularly the non-finite forms, the latter occurring almost categorically with *tener que* in both the

periods examined here. Similar patterns in both centuries have also been identified for the factor groups clause type (with subordinate clauses more likely to contain *haber de*), agentivity (confirming the preference for *haber de* with non-human subject, as noted in the literature), and the level of (im)personality (maintaining a historical association of passive/impersonal sentences with *haber de* in the first half of 20th century).

By contrast, in the remaining factor groups, both convergent and divergent tendencies can be observed in the hierarchy of constraints; a general trend of continuity in the direction of change is broken by some specific factors whose figures change markedly when compared to the previous century. One such case is the factor group tense/mood. While the present indicative and, to a lesser extent, the imperfect indicative continue to have a favourable effect on the choice of *haber de* in the 20th century, it is the future (indicative) that now shows the clearest association with *haber de*. At the same time, these indicative forms run counter to the continuous development of most of the remaining tense/mood paradigms, which clearly disfavour the receding periphrasis *haber de* in the 19th and 20th century. The lexical aspect of the main verbs also shows a degree of continuity as a factor group, with stative verbs constituting a particularly favourable context for the selection of *haber de*, whilst motion verbs have a negative effect on the selection of this periphrasis in both periods. However, within the same factor group we also find an important change in the relative ranking of one specific factor, namely speech verbs, which in the 20th century are almost twice as likely to appear with *haber de* as in the 19th century. This is mainly the result of lexicalisation of the use of these verbs with *haber de*, a process that can, to some extent, already be observed in the 19th century, but that sharply accelerates in the subsequent period examined here. Last but not least, similar patterns can be identified when analysing semantic modality, the factor group that has almost monopolised the debate about this

syntactic variable in the literature. Indeed, in both centuries modality appears as one of the most relevant factor groups, with the scarce cases of non-deontic modality being the most favourable environment for *haber de*, followed by internal obligations imposed by the subject on him or herself due to inner convictions. Conversely, the influence of obligations resulting from necessity or advisability differs between the two centuries examined. Whilst these types of deontic meaning are clearly the least favourable modal settings for *haber de* in the 19th century, this is no longer the case in the 20th century. Instead, the lowest proportion of periphrases with *haber de* is now found in the domain of external obligations (imposed from the outside by forces or circumstances beyond the subject's control). As external obligation is by far the most frequently expressed type of modal meaning in the corpus (both with *haber de* and *tener que)* and diffusion of the emergent variant *tener que* is particularly strong in this frequent environment, our data confirms the often-observed relationship between frequency and grammaticalization (Bybee, 2003).

We have observed that, in the 20th century, *haber de* takes refuge in some very specific and restricted areas of the grammar and lexicon while its frequency of use drops dramatically in all other contexts. This is the case, for instance, for non-deontic modality, expressing notions such as conjecture, as well as expressive meanings such as surprise, indignation, recrimination, etc. In the 20th century, these types of modality, of which there are very few tokens in the corpus, seem to be acting as barriers to *tener que* taking over entirely in the field of modality. The same pattern can be observed in other comparatively low-frequency areas of the grammatical system, such as non-human 3rd person subjects, or, even more so, in passive/impersonal and non-declarative sentences, where *haber de* remains entrenched. Complementing these cases, we have seen several instances of lexicalisation, i.e. the preferred use of this periphrasis with a small number

of main verbs, such as *ser* and *estar* 'to be', *dar* 'to give' and *ir* 'to go', as well as certain speech verbs (*decir* 'to say', *saber* 'to know', *confesar* 'to confess', *reconocer* 'to recognise', etc.), mainly when used in a phatic sense, and in some specific grammatical contexts.

Summing up, many of the contexts that are still favourable to *haber de* by the mid-20th century belong to areas of the grammar that are comparatively infrequent and/or marked. By contrast, the more frequent/unmarked contexts lead to a significant reduction in the use of the receding periphrasis, in a process that, moreover, accelerates over the course of the century. External obligation, declarative and active sentences, as well as human subjects are among the increasingly favourable environments for the now dominant periphrasis *tener que*, paving the way for its generalisation.

Other cases of grammatical change in Spanish, in which the use of an older variant becomes increasingly restricted to some specific contexts, while a newer alternative variant replaces it in its more traditional functions, have been discussed in the literature (e.g., Aaron, 2006; Company, 2003; Klein-Andreu, 1991; Torres-Cacoullos, 2008). Similar patterns have also been identified for the diachronic development of modal periphrases in Spanish during other historical periods. Thus, in previous studies we have shown that, over the past five centuries, *deber de* + infinitive has also been favoured over the alternative construction without the prepositition (*deber* + infinitive) in more marked (and usually less frequent) contexts (Blas Arroyo & Porcar, in press; Blas Arroyo & Vellón, 2014). This is the case, for instance, for epistemic-conjectural meanings, for which modal periphrases are used far less commonly than for the expression of deontic modality as observed above. The same is true for negative clauses, which consistently favour the use of *deber de* + inf. at all times, whilst affirmative clauses, which are far more frequent, are a more favourable environment for *deber* + inf.

Intensification (which can involve different intensification strategies on the speaker's part) also favours the use of the older variant, *deber de* + infinitive.

From the theoretical perspective of grammaticalisation, it has been noted that linguistic change is not abrupt, but comes about through a series of small adjustments affecting both the emerging and the declining variant. For instance, referring to the evolution of future temporal reference in Portuguese over the past centuries, Poplack (2011:219) notes that: "This change was driven by the gradual expropriation by the incoming P[eriphrastic]F[uture] of the preferred contexts of the older layers, culminating in the contemporary situation in which PF has become the default choice everywhere but in the remaining few bastions of P [the (futurate) present]." Similar patterns can also be observed in this study, where the receding variant for the expression of modality, *haber de*, becomes entrenched in some specific and restricted contexts, at the same time as the emergent variant, *tener que*, is gaining a strong foothold in other, more frequent settings such as affirmative clauses, the expression of agent-oriented obligations, as well as active clauses, especially those with humans subjects.

However, our case study also shows some clear differences to the grammaticalisation processes observed, for instance, in the evolution of the way that the future is expressed in several Romance languages (Poplack & Turpin, 1999; Poplack & Malvar, 2007; Poplack & Dion, 2009). In these studies, constraint hierarchies associated with each variant rarely remain constant across time, and relevant factors gain and lose importance as variants grammaticalise. In our case, whilst some significant changes in the importance and the relative ranking of specific constraints have been identified, it cannot be ignored that there are certain clear similarities between the two periods examined in this study. In fact, many of the factor groups selected as significant—and not significant—in both periods remain the same. This raises the question why a higher

degree of persistence can be observed in the grammaticalisation of the system of modal periphrases in Spanish than in the abovementioned studies on the future.

The answer to this question is not an easy one, and there may not be a single reason, but rather a set of factors. For instance, it cannot be ruled out that the binary nature of the variable (i.e., the fact that we compare only two competing constructions) causes the results to differ from studies examining the choice between three or more different forms, as is frequently the case in studies of the future tenses. At the same time, it should be noted that the perspective of our analysis also differs significantly from those studies, in the sense that we focus our attention on the outgoing variant (*haber de*) and not on the incoming one (*tener que*), as is often the case. It should be taken into account that this change of perspective could have repercussions on analysis and interpretation.

However, despite all these cautions, it has also been observed that old distribution patterns may persist, even into the most advanced stages of grammaticalisation (Poplack, 2011:223). This has been seen, for instance, in the distribution of *must* in the modal system of some dialects of English (Tagliamonte & Smith, 2006:372), except that this form may never have been firmly established in these regional varieties, while *haber de* was the star of the Spanish modal system over centuries. In any event, we must not forget that the period of the 20th century analysed in this study coincides precisely with the beginning of the sharp increase in the use of *tener que*, which replaces its previously dominant competitor in environments in which it was previously absent. As discussed above, this change accelerated as the 20th century progressed, gaining momentum just around the time that the period covered by our corpus ends. Our picture of the substitution process is thus incomplete, and more recent data will be needed to gain insights into how the development continues. Nevertheless, there is still a possibility that cannot be excluded: both periphrastic constructions had already grammaticalised

by the time considered in this study, so what we are seeing here is a mere change in their distributions.[12]

Finally, we would like to highlight the benefits of using a corpus consisting of epistolary and autobiographical texts closely reflecting the spoken language. Indeed, it is possible that certain variation phenomena may vary significantly depending on the genre and text type analysed. In this regard, elsewhere we have already suggested that certain differences in the distribution of the prepositional and non-prepositional variants of *deber (de)* + inf. observed in different studies published recently might have be due to the different types of documents the data was drawn from (cf. Balasch, 2008, 2012; Blas Arroyo & Porcar, in press). Likewise, our data on the variation between *haber de* and *tener que* in the 19th century differs markedly from some previous studies in which the surprising prominence of *haber de* (see Figure 1) may well be related to the text types analysed whose formal nature contrast with those that we have analysed in these pages (López Izquierdo, 2008; Martínez, 2003).

Notes

[1] We would like to express our gratitude to the three anonymous referees of this journal for their helpful comments on earlier drafts. Also, thanks are due to our colleague Kim Schulte for his help with English.

[2] Within parentheses is the title and date of the document from which the respective fragments are taken (see Appendix).

[3] This study is part of the "Variación y cambio lingüístico a través de textos de inmediatez comunicativa: un proyecto de sociolingüística histórica [Linguistic variation and change through texts of linguistic immediacy: an historical sociolinguistic research project]", funded by the University Jaume I (Ref. P1·1B2013-01) and the Spanish Ministry of Science and Technology and carried (Ref. FFI2013-44614-P) by a research team led by the first author (2013-2016).

[4] Unlike the original table provided by López Izquierdo, here we have not included occurrences of the impersonal periphrasis *haber que*, since its impersonality does not allow it to alternate with other modal periphrases in the same contexts.

[5] For the sake of a more accurate comparison between the periphrases with one or the other auxiliary verb, *tener que* and *tener de* are collapsed in this graph.

[6] We have included 15 occurrences of *tener* with the preposition *de* as a complementiser (*tener de* + infinitive). Despite its notable vitality in the past (Blas Arroyo & González, 2014; Yllera, 1980), usage of *tener de* in modern Spanish is now limited to some very specific dialects.

[7] In line with Martínez Diaz (2008:1285), we understand obligative modality in a broad sense, i.e., "as an expression of the subjectivity of the utterance", which implies that the syntactic subject of a clause does not necessarily have to be identical with the speaker. Otherwise, we could consider only sentences with a first person singular subject for some modal categories, such as moral, internally motivated obligation.

[8] In our case, the task has not been simple either. In any event, each of the examples was encoded by both authors independently. At a later stage, all cases in which there was any discrepancy (less than 10%) were reviewed jointly or submitted to a third party for evaluation, with the aim of reaching a decision about those tokens on which the authors still could not agree.

[9] Other common cases of agent-oriented or external obligation are those of commanding or ordering someone to perform actions [*Has de/tienes que* entregar este documento 'You have to deliver this document', as well as those characterised by a sense of inevitability in which the idea expressed by the verb is felt to be so certain that its occurrence is considered necessary and inevitable (Todos *hemos de/tenemos que* morir algún día 'We all have to die some day').

[10] We are grateful to an anonymous reviewer for making this point.

[11] We exclude all tokens with 1st and 2nd person verb forms because their subjects are unavoidably human (or viewed as human-like in the respective context).

[12] I am grateful to an anonymous reviewer for making this point.

REFERENCES

Aaron, Jessi E. (2006). Variation and change in Spanish future temporal expression. Ph.D. dissertation, University of New Mexico.

Blas Arroyo, José Luis. (2008). The variable expression of future tense in Peninsular Spanish: The present (and future) of inflectional forms in the Spanish spoken in a bilingual region. *Language Variation and Change* 20(1):85-126.

Blas Arroyo, José Luis. (2014). The scope of language contact as a constraint factor in language change: the periphrasis haber de + infinitive in a corpus of language immediacy in modern Spanish. *International Journal of Bilingualism* (Published online before print March 27, 2014, doi:10.1177/1367006914524644).

Blas Arroyo, José Luis, & Porcar, Margarita. (in press). Patrones de variación y cambio en la sintaxis del Siglo de Oro. Un estudio variacionista de dos perífrasis modales en textos de inmediatez comunicativa. *RILCE. Revista de Filología Hispánica.*

Blas Arroyo, José Luis, & Vellón, Javier. (2014). En los entresijos de la gramática: la supervivencia de *deber de* + inf. en el español moderno (siglos XVIII-XX). *Revista de Filología Española* 94 (1):9-38.

Blas Arroyo, José Luis, & González, Juan. (2014). 'Qué *tengo que/de hazer*?': variación y cambio lingüístico en el seno de las perífrasis de infinitivo a partir de textos escritos de impronta oral en el español clásico. *Studies in Hispanic and Lusophone Linguistics* 7.2:241-274.

Balasch, Sonia. (2008). *Debe (de) ser*: evolución de la variación. In M. Westmoreland & J. A. Thomas (eds.), *Selected Proceedings of the 4th Workshop on Spanish Sociolinguistics*. Somerville, MA: Cascadilla Proceedings Project:109-119.

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*. Santiago de Compostela, España:607-620.

Bybee, Joan. 2003. Mechanisms of change in grarnmaticization: The role of frequency. In B. D. Joseph & R. D. Janda (eds.), *The handbook ofhistoricallinguistics*. Oxford: Blackwell:602-623.

Bybee, Joan L., Revere D. Perkins, & William Pagliuca. 1994. *The evolution ofgrammar: Tense, aspect, and modality in the languages ofthe world.* Chicago: University of Chicago Press.

Bybee, Joan. 2006.Bybee, Joan & Torres Cacoullos, Rena. (2009). The role of prefabs in grarnmaticization: how the particular and the general interact in language change. In R. L. Corrigan, E.A. Moravcsik, H. Ouali, & K. Wheatley (eds), *Formulaic language; volume* 1, *Distribution and historical change,* 187-217. Amsterdam: John Benjamins.

Company Company, Concepción. (2003). La gramaticalización en la historia del español. *Medievalia* (Número monográfico, Gramaticalizacion y cambio sintáctico en la historia del español) 35:3-61.

Davies, Mark. (2002). *Corpus del español (100 millones de palabras, siglo XIII - siglo XX)*. Available at http://www.corpusdelespanol.org (last search: 10-21-2012)**.**

Elsig, Martin. (2009). *Grammatical variation across space and time. The French interrogative system*. Amsterdam/Philadelphia: John Benjamins.

Elpass, Stephan. (2012). The use of private letters and diaries in sociolinguistic investigation. In J. M. Hernández-Campoy & J. C. Conde Silvestre (eds.), *The Handbook of historical sociolinguistics.* Malden, MA: John Wiley & Sons Inc.: 156-169.

Elspass, Stephan, Nils Langer, Joachim Scharloth, & Wim Vandenbussche (eds.). (2007). *Germanic language histories 'from below' (1700-2000).* Berlin: Walter de Gruyter.

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 Filoloxia* 24:203-235.

Garachana Camarero, Mar, & Rosenmeyer Malte. (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 (dir.). (2006). *Diccionario de perífrasis verbales*. Madrid, Gredos.

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

Gili Gaya, Samuel. (1970). *Curso superior de sintaxis española*. Barcelona: Vox..

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

Gómez Torrego, Leonardo. (1988). *Las perífrasis verbales.* Madrid: Arco Libros.

Gómez Torrego, Leonardo. (1999). Los verbos auxiliares. Las perífrasis verbales de infinitivo. In I. Bosque, & V. Demonte (dirs.), *Gramática Descriptiva de la Lengua Española.* Madrid: Espasa-Calpe:3223-3390.

Hernández Díaz, Axel. (2006). Posesión y existencia. La competencia de haber y tener y haber existencial. In Company Concepción (coord.), *Sintaxis histórica de la lengua española*. Vol 1: Tomo II (Primera parte, la frase verbal):1055-1164.

Hernández García, Carmen. (1998). Una propuesta de clasificación de la interferencia lingüística a partir de dos lenguas en contacto: el catalán y el español. *Hesperia* 1:61-80.

Hopper, Paul J., & C. Traugott, Elisabeth. (2003). *Grammaticalization,* second edition. Cambridge: Cambridge University Press.

Keniston, Hayward. (1937). Verbal Aspect in Spanish. *Hispania* 19:163-176.

Klein-Andreu, Flora. (1991). Losing ground: a discourse-pragrnatic solution to the history of *-ra* in Spanish. In S. Fleischrnan, & L.R. Waugh (eds.), *Discoursepragmatics and the verb: the evidence from Romance.* London: Routledge: 164-178.

Kroch, Anthony. (2003). Syntactic change. In M. Baltin, & C. Collins (eds.), *The handbook of contemporary syntactic theory.* Ma1den, MA, and Oxford: Blackwell.

Labov, William. (1982). Building on empirical foundations. In W. P. Lehmann & Y. Malkiel (eds.), *Directions for Historical Linguistics*. Austin: University of Texas Press: 79-92.

Lapesa, Rafael. (2000). *Estudios de morfosintaxis histórica del español*. R. Cano, & M.T. Echenique (eds.). Madrid: Gredos.

López Izquierdo, Marta. (2008). Las perífrasis modales de necesidad: emergencia y renovación. In C. Company & J. G. Moreno (coords.), *Actas del VII Congreso Internacional de Historia de la Lengua española.* Vol. 1 Yucatán, Mexico. 789-806.

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

Martínez Díaz, Eva. (2008). La expresión de la modalidad de obligación en el Corpus del español conversacional de Barcelona y su área metropolitana (2001). In A. Moreno (coord.), *El valor de la diversidad (meta)lingüística*. *Actas del VIII congreso de Lingüística General*. Referencia: http://www.lllf.uam.es/clg8/actas/index.html.

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.

Okulska, Urszula. (2010). Performing the world of politics through the discourse of institutional correspondence in Late Middle and Early Modern England. In U. Okulska, & P. Cap (eds.), *Perspectives in Politics and Discourse*. Amsterdam: John Benjamins: 173-197.

Olbertz, Hella. (1998). *Verbal periphrases in a Functional Grammar of Spanish*. Berlin: Mouton de Gruyter.

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

Poplack, Shana (2011): Grammaticalization and Linguistic Variation. In B. Heine, Bern & N. Heiko (eds.)*, The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press:209-224.

Poplack, Shana, & Turpin, Danielle. (1999). Does the FUTUR have a future in (Canadian) French? *Probus* 11:133-164.

Poplack, Shana, & Tagliamonte, Sali. (2001). *African American English in the diaspora: tense and aspect.* Oxford: Blackwell.

Poplack, Shana, & Malvar, Elisabete. (2007). Elucidating the transition period in linguistic change: the expression of the future in Brazilian Portuguese. *Probus 19:*121-169.

Poplack, Shana, & Dion, Nathalie (2009). Prescription vs. praxis: the evolution of future temporal reference in French. *Language* 85(3):557-587.

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

Roca Pons, José. (1980). *Introducción a la gramática*. Ed: Anexo XVII. Madrid: Vergara.

Rojo, Guillermo. (1974). Las perífrasis verbales en el gallego actual. *Verba, Anejo 2*.

Sankoff, David, Tagliamonte, Sali A. & Smith, Eric. (2005). *Goldvarb X: A variable rule application for Macintosh and Windows*. Department of Linguistics: University of Toronto.

Schwenter, Scott A., & Torres Cacoullos, Rena. (2008). Defaults and indeterminacy in temporal grammaticalization: the "perfect" road to perfective. *Language Variation and Change* 20:1-39.

Seifert, Eva. (1930). *Haber* y *tener* como expresiones de la posesión en español. *RFE* XVII:233-276; 345-385.

Serrano, María José. (1994). Del pretérito indefinido al pretérito perfecto: un caso de cambio y gramaticalización en el español de Canarias y Madrid. *Linguística Española Actual* 16:37-57.

Sinner, Carsten. (2003). Valores y usos de haber de + infinitivo y tener que + infinitivo en diferentes variedades del castellano. In C. Pusch & A. Wesch (eds.), *Perâifrasis verbals en les llengèues (ibero-)romániques*. Madrid: Iberoamericana-Vervuert:193-206.

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

Tagliamonte, Sali A. (2006). *Analysing sociolinguistic variation.* Cambridge: Cambridge University Press.

Tagliamonte, Sali, & Smith, Jennifer. (2006). Layering, competition and a twist of fate. Deontic modality in dialects of English. *Diachornica* 23(2):341-380.

Tagliamonte, Sali, Durham, Mercedes, & Smith, Jennifer. (2014). Grammaticalization at an early stage: future *be going to* in conservative British dialects. *English Language and Linguistics* 18.1:75-108.

Tarallo, F. (1989). Inside and outside relative clauses. In R. Fasold & D. Schiffrin (eds.), *Language change and variation*. Washington: Georgetown University Press: 255–274.

Torres Cacoullos, Rena. (2000). *Grammaticization, synchronic variation, and language contact: a study of Spanish progressive -ndo constructions.* Amsterdam: John Benjamins.

Torres Cacoullos, Rena. (2001). From lexical to grammatical to social meaning. *Language in Society* 30:443-478.

Torres Cacoullos, Rena. 2008. La evolución de *estar + V-ndo* hacia expresión obligatoria del aspecto progresivo en español. In C. Company Company, & J.G. Moreno de Alba (eds), *Actas VII Congreso Internacional de Historia de la Lengua Española.* Madrid: Arco Libros: 1147-1162.

Torres Cacoullos, Rena. (2009). Variation and gramrnaticisation: the emergence of an aspectual opposition. In S. Tsiplakou, M. Karyolemou, & P. Pavlou (eds.), *Studies in Language Variation: European Perspectives* II. Amsterdam: John Benjamins: 215-224.

Torres Cacoullos, Rena. (2011). Variation and Grammaticalization. In M. Díaz-Campos (ed.), *The Handbook of Hispanic Sociolinguistics*. Malden MA: Blackwell: 148-167.

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

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

APPENDIX: The corpus

Alonso Montero, Xesús. (2009). *Cartas de republicanos condenados a muerte (1936-1948)*. Xerais. (120 letters; period:1936-1948).

Arbeloa, F. J. (2002). Once cartas de mi padre desde el frente (1936-1937), *Historia contemporánea* 48: 103-119 (11 letters; period:1936-1937).

Blasco Martínez, Rosa M., & Rubalcaba Pérez, Carmen. (2003). *Para hablarte a tan larga distancia .... Correspondencia de una familia montañesa a ambos lados del Atlántico (1885-1883)*, Madrid: Ediciones de Librería Estudio (210 letters; period:1855-1883).

Butrón, & Cortés J. (1897). *Carta familiar de D. José Butron y Cortés dirigida a su hijo el Guardia Marina D. Rafael Butron y Pareja*. Madrid: Librería de Fernando Fé (1 carta; period: 1816).

Butrón y Cortés J. (1897). Extracto de las cartas recibidas por D. Emilio José Butrón acusando recibo de la carta familiar de su señor abuelo. In E. J. Butrón & de la Serna (ed.), *Carta familiar de D. José Butron y Cortés dirigida a su hijo el Guardia Marina D. Rafael Butron y Pareja*. Madrid: Librería de Fernando Fé. (55 letters; period: 1890-1897).

Casanova Pujol, Josep. (s.f.). *Cartas del iaio*, La Roca del Vallès: Arxiu de la Memòria Popular (16 letters; period: 1938).

Chávez Palacios, Julián. (1998). *La pérdida de Filipinas narrada por un soldado extremeño (1896-1899). Memorias del sargento Deogracias González Hurtado*. Cáceres: Editorial Regional de Extremadura (Memorias; period: 1896-1900).

Doña, Juana. (2003). *Querido Eugenio. Una carta de amor al otro lado del tiempo*. Madrid: Lumen. (3 letters; period: 1940-1941).

Espuny, Tomás. (2002). *De Gallur a Nueva York: diario de viaje*. Gallur: Ayuntamiento de Gallur (Diario; period:1929).

Fernández, Víctor. (2008). *Cartas de Vicenta Lorca a su hijo Federico*. Barcelona, RBA Libros. (34 letters; period: 1920-1933).

Cemboráin Mainz, Luis. (2003). Cartas de dos hermanos navarros requetés en 1937, *Príncipe de Viana* 66: 477-512. (17 letters; period: 1937).

Gener, Eduardo. (1976). *Diario de viaje de un comerciante gaditano (1829).* Cádiz: Instituto de Estudios Gaditanos. (Diario; period: 1829).

Genovés Amorós, Vicente. (1981). *El epistolario de Navarro Reverter con Cirilo Amorós*. Valencia: Federico Doménech. (86 letters; period: 1879-1886).

Gómez de Avellaneda, Gertrudis. (1989). *Poesías y epistolario de amor y de amistad*. Madrid: Castalia. (74 letters; period: 1839-1867).

Gómez Seibane, Sara. (2011). Textos para la historia del castellano en Galicia, *Moenia* 17, 367-420. (8 letters; period: 1800-1806).

Gonzàlez i Vilalta, Arnau. (2007). *Un catalanófilo de Madrid: epistolario catalán de Ángel Ossorio y Gallardo (1924-1942)*. Barcelona: Universitat Autònoma de Barcelona. (280 letters; period: 1924-142).

Ibarra y Manzoni, Aureliano. (1995) *Diario de mi prisión. Desde el día 8 de Noviembre de 1866 hasta el día 21 de Marzo de 1867*. Elche: Ajuntament. (Diario; period:1866-1867).

Isasi, Carmen, & Ramírez Luengo, José Luis (eds.). (2013). *Una muestra documental del castellano norteño en el siglo XIX*, Lugo, Axac. (168 letters; period: 1800-1894).

Lajo Cosidó, Benjamín. (1998). *Sólo habremos muerto si vosotros nos olvidáis. De la libertad al muro sólo hubo un atardecer (1936-1939)*, Benicull de Xúquer, 7 i mig Editorial (23 letters; period: 1939-1940).

López Alvarez, Juaco. (2000). Cartas desde América: La emigración de asturianos a través de la correspondencia: 1864-1925, *Revista de dialectología y tradiciones populares* LV 1: 81-120. (25 letters; period:1864-1925).

Madrazo, José. (1998). *José de Madrazo: Epistolario*. Madrid: Fundación Marcelino Botín. (241 letters; period: 1833-1853).

Manrique Romero, Alberto. (1996). *Cartas muertas: la vida rural en la posguerra*. Madrid: Ediciones SA. (65 letters; period: 1942-1943).

Márquez Macías, R. (1994). *Historias de América. La emigración española en tinta y papel*. Huelva: Ediciones Ertoil. (151 letters; period: 1820-1824).

Martínez, Mª Luisa. (2007). *Noticias desde Cuba: cartas de emigrantes vallisoletanos en la segunda mitad del siglo XIX*, Valladolid: Universidad de Valladolid. (30 letters; period: 1878-1898).

Martínez Martín, Laura. (2010). *'Asturias que perdimos, no nos pierdas'. Cartas de emigrantes asturianos en América (1863-1936)*. Gijón: Museu del Pueblu d'Asturies. (101 letters; period: 1863-1936).

Núñez Selxas, Manoel, Xosé, & Soutelo, Raúl. (2005). *As cartas do destino. Unha familia galega entre dous mundos, 1911-I971*. Vigo: Editorial Galaxia. (221 letters; period: 1911-1971).

Ollaquindia, Ricardo. (1997). *Cartas de un requeté del tercio del rey: José María Erdozain*, Madrid: Ed. Actas. (28 letters; period: 1932-1937).

Orden, María Liliana da. (2010): *Una familia y un océano de por medio*: *la emigración gallega a la Argentina: una historia a través de la memoria epistolar*. Madrid: Marcial Pons. (48 letters; period: 1910-1968).

Ossó, F. (1997). *Cartas de San Enrique de Ossó y Cervelló al Dr. Félix Sardá y Salvany*. Barcelona: STJ. (129 letters; period: 1865-1895).

Otazu, Fausto. (1995). *Fausto de Otazu a Iñigo Ortés de Velasco: cartas, 1834-1841*. Álava: Diputación Foral de Álava. (50 letters; period: 1834-1841).

Pérez Murillo, Mª Dolores. (1999). *Cartas de emigrantes escritas desde Cuba*, Cádiz: Universidad de Cádiz. (92 letters; period: 1808-1829).

Prado, Mercedes. (1994). *Darío de Regoyos: sus cartas inéditas*. Bilbao: Ed. Mercedes Prado Vadillo (178 letters; period: 1895-1913).

Ramón y Carmen Ortiz, Manuel de. (2003). *Madrina de guerra: cartas desde el frente*. Madrid: La Esfera de los libros. (96 letters; period: 1937-1942).

Rubalcaba Pérez, Carmen. (2006). Libros de cuentas de Pedro Jado. *Entre las calles vivas de las palabras*. Santander: Ediciones Trea:55-74. (Libro de cuentas; period: 1840-1857).

Sánchez Forcada, Manuel. (2003). Diario de campaña de un requeté, *Príncipe de Viana* 64:641-682. (Diario; period: 1936-1939).

Sanza, Marcelino. (2008). *Francia no nos llamó: correspondencia de un campesino aragonés en la tormenta de la guerra*, Vinaròs: Antinea. (70 letters; period: 1939-1940).

Soutelo, R. (2001). *De América para casa: correspondencia familiar de emigrantes galegos no Brasil, Venezuela e Uruguay (1916-1969)*, Santiago de Compostela: Consello da Cultura Galega. (89 letters; period:1916-1966).

Tinell, Roger. (2001). *Epistolario a Federico García Lorca desde Cataluña, la Comunidad Valenciana y Mallorca*, Granada: Ed. Comares. (101 letters; period: 1921-1935).

Tomás, Facundo. (2009*). Epistolarios de Joaquín Sorolla Vol. 1*. Barcelona: Anthropos. (533 letters; period: 1886-1933).

Vargas, Bruno, & Moreno Sáez, Francisco. (2007). *Dramas de refugiados: epistolario de Rodolfo Llopis y otros dirigentes socialistas alicantinos (1939-1947)*. Madrid: Centro Francisco Tomás y Valiente. (78 letters; period: 1939-1947).

Vives i Clavé, Pere, & Bartra, Agustí. (1972). *Cartes des dels camps de concentració*, Barcelona, Edicions 64. (23 letters; period: 1939-1940).