MDPI

Variant Choices of Future Time Reference in Galician: The Grammaticalization of [*haber* (*de*) + INFINITIVE] as a Window to Diachronic Change

Esther L. Brown * and Javier Rivas *

Department of Spanish & Portuguese, University of Colorado Boulder, Boulder, CO 80309, USA
* Correspondence: brownel@colorado.edu (E.L.B.); rivasrod@colorado.edu (J.R.)

Abstract: Compared to neighboring Romance languages, Galician currently maintains a more ubiquitous usage of the construction [*haber* (PRESENT) + (*de*) + INFINITIVE] as a future marker in variation with the periphrastic construction with *ir* 'go' and the morphological future. We examine this understudied construction to gain a better understanding of Galician grammar and also contribute new data with which to consider diachronic change regarding the grammaticalization of the future from obligation markers. We conduct a variationist analysis of 1589 tokens of future forms in recorded conversations (CORILGA) in order to determine the frequency of usage, patterns of variation, linguistic conditioning and degree of grammaticalization of the periphrastic forms with *haber* and *ir* in contrast to the morphological variant. We find evidence to suggest that the periphrastic construction with *haber* is highly grammaticalized as a future marker and we identify factors of the production context that modulate the grammaticalization process.

Keywords: future; Galician linguistics; grammaticalization; obligation; periphrastic constructions; variationist methodology; usage-based linguistics

1. Introduction

Languages frequently have more than one grammaticalized form for the expression of future time (Bybee et al. 1994; Dahl 2000) and often share similar or identical lexical sources for the markers of future. Romance languages such as Spanish, Portuguese, and Galician possess a synthetic future marker that arose from the fusion of a periphrastic construction from Latin [INFINITIVE + *habere* 'have'] (e.g., *cantare habeo* 'I shall sing/I am to sing'). Together with this form, in the history of these languages, periphrastic constructions emerged that competed with the morphological form to express future time. To mark this meaning, the most frequently occurring periphrastic constructions are those with a movement verb such as *ir* 'go' or with a verb of obligation such as *haber/haver* 'have' in combination with an infinitive.

To differing degrees, both periphrastic constructions competed as innovative options to express future meaning alongside the morphological future that they eventually replaced in spoken vernacular. In the case of Spanish, the morphological future was replaced principally by the [*ir* (PRESENT) + *a* + INFINITIVE] construction. The construction [*haber* (PRESENT) + (*de*) + INFINITIVE], although already attested in medieval times as a future marker, was more commonly used to express modal meanings rather than temporal future (Blas Arroyo and Schulte 2017, p. 11) until its decline in the first half of the 20th century (Blas Arroyo 2018, p. 214). The periphrastic construction with *ir* arose in the 13th century (Aaron 2006, p. 268) and has gradually increased in frequency throughout the history of Spanish, being the current-day default option (Aaron 2010). In Portuguese, the periphrastic construction with *haver* was effectively the only alternative to the synthetic form to express futurity until the 18th century (Poplack and Malvar 2007, p. 127). However, in the 19th century, [*ir* (PRESENT) + INFINITIVE] entered into competition for future marking and

eventually prevailed ([Poplack and Malvar 2007](#), p. 157), relegating the *haver* option to formal written contexts (Brazilian Portuguese) or to constructions usually expressing a strong modal meaning (European Portuguese). Consequently, futurity is typically expressed by means of the periphrastic construction with *ir* in present-day Portuguese ([Gradoville 2019](#); [Cunha 2021](#), [2022](#)).

Similar to Spanish and Portuguese, Galician utilizes these periphrastic markers of future with the same lexical sources. Unlike these aforementioned Romance languages, however, in the case of Galician, both *ir* and *haber* periphrases are maintained as possible options to express posteriority. As [Cidrás Escáneo](#) ([2009](#), p. 43) notes, "En galego a situación é algo máis matizada. A pesar do gran incremento de uso de <*ir* + infinitivo>, dáse un notable mantemento da construción <*haber* (*de*) + infinitivo>, que está moi lonxe da marxinalidade acadada nas linguas veciñas como forma de futuro[1]". Although maintained in the language, the future markers including *haber* periphrasis remain relatively unexplored quantitatively.

In this study, using variationist methodology ([Labov 1994](#); [Tagliamonte 2012](#)), we conducted a usage-based analysis of Galician future markers in oral language in order to understand the nature of the variation currently. We identified common forms used to express future, their relative frequency of occurrence in speech, and the linguistic constraints that condition their production. Our goal was to gain a deeper understanding of Galician and to contribute the first large-scale, quantitative analysis of the form–function correlations across future markers in this language.

By analyzing this variation in Galician, we provided new data with which to consider and view grammaticalization processes generally as well as those affecting future markers in Romance specifically. Given that Galician maintains periphrastic constructions with *haber* as a viable marker of future in present-day usage, its study also provides a synchronic window into the diachronic process through which lexical sources expressing obligation acquire future meaning, especially in competition with future markers with lexical sources in movement verbs.

2. Background
2.1. Galician Future Markers

Galician has a synthetic (or morphological) future form that originated from the fusion of the Latin periphrastic construction [INFINITIVE + *habere* 'have'], as shown in (1). Together with the morphological future, different periphrastic constructions have arisen throughout the history of Galician as grammaticalizing markers of future time. The most frequent periphrases are [*ir* 'go' (PRESENT) + INFINITIVE] (excerpt (2)) and [*haber* 'have' (PRESENT) (*de*) + INFINITIVE] (excerpt (3)). Additionally, as (4) illustrates, the simple present also expresses future time in Galician.

(1)   OICO-URB-CBAS-SANTIAGODECOMPOSTELA-02-2014[2]
    e [mañá] xa che dirá | Inma | non-o fixeches ben | |
    'and tomorrow Inma will tell you: you did not do it right'

(2)   OIED-RUR-CHEN-BUEU-02-1967
    mañá vou ver unha orquesta
    'I am going to see a band tomorrow'

(3)   OIED-RUR-CHEN-BUEU-02-1967
    se te meteche aquí | mañán á mañán | hei-te de coller
    'If you got in here, I will catch you tomorrow morning'

(4)   OICO-URB-CBAS-SANTIAGODECOMPOSTELA-04-2014
    hoxe deitámo-nos e mañán dígo-che
    'Today we go to bed and I will tell you tomorrow' (lit. 'tomorrow I say to you')

Galician follows the common tendency found in Romance languages to replace synthetic forms with periphrastic forms cyclically over time ([Hopper and Traugott 2003](#), p. 9). The Latin synthetic form, *cantabo* 'I will sing', which was itself in origin a periphrastic

form, did not make its way into the Romance languages and was instead replaced by a periphrastic future construction [INFINITIVE + *habere* 'have'] (e.g., *cantare habeo* 'I shall sing'). This periphrastic form underwent a process of fusion in Romance languages, whereby *habeo* gradually lost phonological and morphosyntactic autonomy (*habeo* > **aio*), until it became an inflectional suffix indicating future: French *chanter-ai*, Galician/Portuguese *cantar-ei*, Italian *canter-ò*, Provençal *cantar-ai*, Spanish *cantar-é*. In turn, these morphological future forms are currently being replaced with periphrastic constructions in many Romance languages. In present-day Spanish, for example, the [*ir* 'go' (PRESENT) *a* + INFINITIVE] construction has become the default marker of future time, in both monolingual (Sedano 2006; Orozco 2018; Giordano 2022) and contact varieties (Blas Arroyo 2008; Orozco 2018; de Prada Pérez et al. 2021).

2.2. Grammaticalization of Future

Bybee et al. (1994, p. 244) define future in the following terms: "We regard the focal use of future as equivalent to a prediction on the part of the speaker that the situation in the proposition, which refers to an event taking place after the moment of speech, will hold". In order to reach the focal meaning of prediction, grammaticalizing forms generally go through a previous stage in which their core meaning is 'intention', first of the speaker and then of the agent of the construction (not only the speaker, but also the hearer and third persons). It is only when the form may express intention of an agent other than the speaker that the meaning of prediction can be reached, since speakers may be well aware of their own intentions, but they can make predictions only about someone else's future actions. Once the form expresses prediction, its use may extend to inanimate subjects, which by definition, lack intention(s).

Cross-linguistically, languages consistently use the same lexical sources to convey future grammatical meanings (Bybee et al. 1994; Kouteva et al. 2019): general verbs of 'movement towards' (e.g., English *go*, Spanish *ir*), verbs of obligation such as Old English *sceal* 'owe' > present-day English *shall* and Latin *habere* 'have to', and verbs of willingness, including Old English *willan* 'want' > present-day English *will* and Greek *thelo* 'want'. The grammaticalization of these lexical items as future markers occurs in a very specific usage context, as is typical of grammaticalization processes (Hopper and Traugott 2003; Lehmann 2015). For example, the verb *go* is used to express future when it occurs in the present progressive construction (*be going*) and is followed by an infinitival clause of purpose headed by *to*.

(5) Constructed example
I cannot talk right now. I am on my way to Boulder. I am going to teach a class.

The purpose construction has an inference of futurity (Hopper and Traugott 2003, p. 3), which leads the speaker to interpret the event as taking place in the future. This interpretation is possible because one of the original meanings of *go* is 'movement in time'. As it becomes a marker of future time, then, *go* loses only part of its original meaning, 'movement in space'. Intention was already part of the original meaning of the construction, which, once it loses the meaning of 'movement in space', follows the semantic pathway from intention to prediction presented earlier.[3]

Similarly, in the case of Latin *habere* 'have', the verb becomes grammaticalized as a marker of future when it occurs as part of a periphrastic construction with an infinitival form. This construction was attested with both *habere* occurring before and after the infinitive. However, it will be the order [INFINITIVE + *habere*] that will lead to the grammaticalization of *habere* as an inflectional ending for future in Romance languages. Originally, the verb *habere* indicated 'possession', but when it was used with an infinitival form, it expressed a modal meaning of obligation (Fleischman 1982, p. 59) or predestination (Bybee et al. 1994, pp. 261–63). The meaning of obligation or predestination entails intention in the first person, and consequently, prediction, when used in the third person.

The morphological future has come to express, in many Romance languages, including Galician, non-temporal epistemic-related meanings (Álvarez and Xove 2002, pp. 298–99; Freixeiro Mato 2006, pp. 350–54). For example, in (6), *terá* 'she will be (lit.: have)' indicates present time with an additional modal meaning of 'probability'.

(6)   OIED-RUR-CHEN-BUEU-13-1967
      terá dous meses a rapaciña ǀ non sei se os terá ǀǀ
      'The little girl is probably 2 months old. I am guessing she must be.'

This epistemic meaning is a final step in the grammaticalization cline of future markers (Bybee et al. 1994, p. 279).

2.3. Grammaticalization of Galician Future with Haber

Present-day Galician illustrates a similar grammaticalization process to that of Latin [INFINITIVE + *habere*] by means of the periphrastic construction [*haber* (*de*) + INFINITIVE], which is used to indicate future time. The lexical source of this construction is the verb *haber*, which comes from Latin *habere*, whose meaning was 'have, possess'. When used in combination with an infinitive, it often occurs with the preposition *de* 'of', which is also used to indicate possession in Galician (*o libro de Xiana* 'Xiana's book, lit. the book of Xiana'). The original meaning of [*haber* (*de*) + INFINITIVE] in Galician is obligation (Rojo 1974, p. 93). From this meaning of obligation, this construction underwent a semantic change in order to express, first, intention, and then prediction, as already discussed.

In fact, as is noted in Galician grammars (Álvarez and Xove 2002, p. 340; Freixeiro Mato 2006, p. 441), it is very hard to tease apart the meanings of 'obligation' and 'futurity' when studying the usage patterns of [*haber* (*de*) + INFINITIVE] in present-day Galician. In fact, García Represas (2001, p. 109) indicates that the obligation meaning of this construction is linked to a future time. However, the degree of grammaticalization of this construction in Galician is not clear. For Rojo (1974, p. 158), [*haber* (*de*) + INFINITIVE] may express futurity without the additional meaning of obligation. In contrast, Álvarez and Xove (2002, p. 356) argue that the periphrasis cannot express 'neutral' future, i.e., there is always a modal meaning associated with the construction. Thus, the meaning of *haber* (*de*) is highly variable, ranging from a clearer deontic meaning to a more abstract future meaning.

3. Data and Methods

Our analysis is based on the oral data extracted from *Corpus Oral Informatizado da Lingua Galega (CORILGA)* 'Computerized Spoken Corpus of Galician Language,' compiled at the Instituto de Lingua Galega-Universidade de Santiago de Compostela (Fernández Rei and Regueira Fernández 2019). This corpus consists of approximately 1,400,000 words of spoken Galician from conversations recorded between 1964 and 2018 with both men and women of different ages. The corpus contains spoken samples recorded in rural and urban areas of Galicia, ranging from informal casual conversations to interviews, formal speeches, and lectures. Speaker-related information (e.g., gender, age, socioeconomic status) is not yet available to the authors. Therefore, this study will be primarily concerned with the linguistic conditioning of future marker forms.

We based our analysis on approximately 840,000 words of informal oral Galician distributed across 173 different recordings chosen randomly. We chose the informal corpus because, as is noted by Rojo (1974, p. 85), the periphrastic construction with *haber* seems to be more frequent in informal than in formal, literary language. Table 1 shows the distribution of words by decades.

Table 1. Data distribution across different corpora and decades.

Decade	Dates	# Words	% of the Corpus
1960s and 1970s	1964–1975	395,084	47%
1990s	1992–1999	354,228	42%
2000s–	2008–2018	88,089[4]	11%
Total	1964–2018	837,401	100%

From this corpus, we extracted a total number of 2224 examples, including morphological future forms, [*haber* (PRESENT) (*de*) + INFINITIVE] constructions, and [*ir* (PRESENT) + INFINITIVE] constructions[5]. In the case of periphrastic constructions with *ir*, we also included examples in which the speaker uses *a* 'to' between *ir* and the infinitive (27%, N = 298), a construction that is generally attributed to Spanish influence (Freixeiro Mato 2006, p. 343). We excluded from the analysis unclear and truncated examples (N = 22), cases of the *vamos (a) ver* 'Lit: 'we are going to see' as a discourse marker (N = 57), and 60 examples in which the periphrastic construction with *ir* has a habitual meaning or is being used as a historical present [see excerpt (7)]. These exclusions left us with 2085 examples.

(7)   OIED-URB-CDUB-SANTIAGODECOMPOSTELA-07-1995
        van un día á semana | fan-lle compañía | falan | van-lle busca-los medicamentos ‖
        'They go once a week, they keep him company, they talk, they go get his medicines'

The future forms in the Galician corpus are also used to express non-temporal (modal) epistemic meanings. Overwhelmingly, the morphological future form was the one selected by speakers to make conjectures. The periphrastic forms can also convey this meaning, but to a significantly lower degree. Table 2 summarizes the usage patterns of future forms with epistemic meanings.

Table 2. Periphrastic and morphological future forms expressing epistemicity (N = 495).

Future Form	N	%
morphological	468	95
ir	6	1
haber	21	4
Total	495	100

The morphological form more often expresses epistemic (59% of tokens) rather than future meaning (41% of the tokens). This distribution suggests that it is specializing as a modal non-future marker nowadays.

The interpretation of meaning for the periphrastic forms is, at times, inconclusive. Often, the meaning can be interpreted as either future or obligation (in the case of *haber*) or future and movement (in the case of *ir*). Given that the intended meaning of the speaker is not accessible to researchers, we considered, within the envelope of variation, all examples of these periphrastic constructions in which future interpretation is possible, in line with previous studies on tense–aspect–mood markers (e.g., Schwenter and Torres Cacoullos 2008; Aaron 2010; see also Torres Cacoullos 2011, p. 148). This provided 1589 examples of forms expressing future meaning for analysis. We coded each of these tokens for the following factors, which were known to constrain variation in variant forms:

a.   Subject grammatical person: first person, second person, third person. As noted above, according to Bybee et al. (1994, p. 244), the basic meaning of future is prediction. Clear examples of prediction can be achieved only when the form is used with third person subjects (because you can know your intentions and only predict the actions of others). Since the meaning of prediction is reached only when the construction is being used with third person subjects, we will interpret a disfavoring effect on the third person as evidence of a less advanced degree of grammaticalization.

Although not included in the statistical analysis, within third person subjects, we distinguish between animate and inanimate referents. Excerpts (8), (9), and (10), respectively, provide examples of periphrastic *ir*, *haber*, and morphological future with a third person subject:

(8)　OIED-RUR-CHEN-BUEU-07-1967
　　　E di que vai chegar estes días
　　　'And he says he is going to arrive within the next few days'

(9)　OIED-RUR-CHEN-BUEU-13-1967
　　　Á noite ha chega-lo meu home
　　　'My husband is arriving tonight'

(10)　OIED-RUR-CHEN-ORDES-02-1964
　　　Pasará por aquí ás 9.30
　　　'It will pass by at 9.30'

b.　Temporal adverb: yes, no. We coded each target token for whether the speaker uses a temporal adverb (i.e., *mañá* 'tomorrow', *dentro de pouco* 'soon') or does not explicitly frame the verbal event with a temporal marker. The overt temporal adverb is a salient cue situating the verbal event in time. If future time is part of the meaning of the future form, the speaker may rely less upon temporal adverbials to express this meaning. For example, in Spanish, the morphological future correlates with the presence of a temporal adverb when it expresses future (Aaron 2010, p. 27) because it is becoming specialized as an epistemic marker. As such, we interpreted any significant correlations between a future marker and the presence of a time adverb as evidence of reduced grammaticalization of the future marker. Excerpts (11) and (12) provide examples of a periphrastic construction with *ir*. The first one has the temporal adverb *mañá* 'tomorrow', and the second one lacks a temporal adverb:

(11)　OICO-URB-CBAS-SANTIAGODECOMPOSTELA-03-2014
　　　Eu vou mañá vou ver unha orquesta, tío
　　　'I am going tomorrow I am going to see an orchestra, dude'

(12)　OICO-URB-CBAS-SANTIAGODECOMPOSTELA-06-2014
　　　Non volo vou contar
　　　'I am not gonna tell you'

c.　Meaning: Bybee et al. (1994, p. 274) indicated that "apodoses are prime environments for future grams". Following this line of work, we describe 'contingent' those tokens that refer to an event that will take place depending upon the fulfillment of a condition. We contrast these future forms found in the apodosis of a conditional clause with 'assumed' futures in which the speaker is not framing the future event as dependent upon the occurrence of another. This difference is significant in studies on future expression in other languages. In general terms, the apodosis of a conditional clause is a favoring environment for the conservative variant [synthetic future in Portuguese (Poplack and Malvar 2007) and *will* futures in English (Torres Cacoullos and Walker 2009)]. We coded all target tokens for whether they are expressed in the apodosis (contingent) or not (assumed). We interpreted correlations between contingency and future forms as evidence of increased grammaticalization. The meanings of 'contingency' and 'assumed' are, respectively, exemplified in (13) and (14) with periphrastic *haber*:

(13)　OIED-RUR-CORILGA-LALIN-03-2014
　　　Vas come-las verzas queiras non quieras, que senón ha vivir Jesús de Laro e léva-te
　　　'You are gonna eat the kale whether you like it or not, because, if you don't, Jesús de Laro will come and take you'

(14)　OIED-RUR-CORILGA-MAZARICOS-01-2013
　　　E despós a comida, de sempre ha sobrar
　　　'And then the food, there will always be more than enough'

d. Century: The future forms were extracted from a randomly selected set of files recorded across three primary decades (see Table 1). Given the relatively small dataset and the number of linguistic factors we considered, in order to simplify our statistical model, we contrasted recordings from the 20th century (1960s, 1990s) with those from the 21st century.

In addition to the previously described fixed effects, we also included a corpus file (as a proxy for individual speaker) and verb infinitive as random effects in the analysis. The results of our statistical analyses, conducted in R (The R Project for Statistical Computing n.d.), allowed us to determine the probabilistic grammar of each of these future variants. Specifically, the analyses allowed us to explore the synchronic usage of *haber* in this Romance language and determine its degree of grammaticalization as a marker of future time.

4. Results

4.1. Distribution of Future Forms

The results of our analysis reveal patterns of usage for morphological future and periphrastic constructions with *haber* and *ir*. Table 3 provides a percentage of use of these three forms in our data.

Table 3. Distribution of future forms used to express future meaning in Galician.

Future Form	N	%
morphological	326	20
haber	297	19
ir	967	61
Total	1589	100

The form overwhelmingly chosen for use to express future time is the periphrasis with *ir* (61% of the tokens). Both morphological future and the *haber* periphrasis occur approximately in similar proportions.

The periphrastic construction with *ir* is not only the most frequent one in the corpus as a whole, but also the most prevalent one throughout recent decades. Table 4 represents the frequency per million of the three future forms across the time span represented in the corpus. In the most recent decades, the frequency per million for *ir* is nearly twice what it was in the 1960s, whereas for *haber*, the frequency per million is a quarter of what it was in the sixties. We note that these corpora differ in ways other than date. For example, the percentage of urban vs. rural data is much higher in the 2000s than in the 1960s. The speakers interviewed likely differed in important ways with regard to extralinguistic factors not currently available to the authors from the corpus, and variation is likely also constrained by stylistic factors such as degree of formality. The point is that the differences in frequency per million may reflect differences across the corpora independent of time.

Table 4. Frequency per million of *ir*, *haber*, and morphological form across time.

Verb	60s	90s	2000s
haber	595	138	148
ir	866	1341	1691
morphological	473	313	318

4.2. Linguistic Conditioning

Since the *haber* periphrastic form has persisted to present day as a future marker in Galician, we are able to explore these data as a way to understand future expression currently, and additionally, these data could potentially serve as a metric with which to consider the diachronic change evidenced in late Latin and several other Romance languages. As is evident in Table 4, *haber* use in the corpus declines. Resultantly, some

speakers do not use this periphrasis in the speech sample. In order to understand which factors incite a speaker to choose one periphrasis over another, for the statistical modeling we limited the data to all future tokens spoken just by speakers who minimally used the *haber* future once. This reduced the number of target tokens (N = 956). The following tables present the results of three multiple linear regression analyses of the data predicting *haber* usage.

In order to determine the ways in which the innovative periphrastic forms (*haber*, *ir*) differ from the more conservative morphological future, we conducted separate linear regression models using the lme4 function to predict the periphrastic form(s) [see Tagliamonte (2012) for a description of the R analysis]. A summary of these two independent analyses [*haber* vs. morphological, *ir* vs. morphological] is provided in Table 5, with the *haber* results shaded on the left and *ir* on the right.

Table 5. Two multiple linear regression models predicting *haber* (shaded) or *ir* vs. morphological *.

Fixed Effects	N *haber*	Est. Coef.	SE	*p*	N *ir*	Est. Coef.	SE	*p*
Intercept	297	1.243	0.246	***	457	2.094	0.314	***
Century: 21st	13	0.826	0.690	n.s.	25	0.852	0.853	n.s.
Century: 20th	284				452			
Person: third	88	−0.716	0.281	*	139	−0.607	0.304	*
Person: second	56	−0.772	0.297	**	54	−1.469	0.336	***
Person: first	153				264			
Adverb: yes	52	−0.050	0.297	n.s.	78	−0.305	0.316	n.s.
Adverb: no	245				379			
Meaning: contingent	32	−1.017	0.308	***	39	−1.301	0.352	***
Meaning: assumed	265				418			

* *Haber* vs. morphological: N = 499, random effects (speaker: Var. 0.378, Std. Dev. 0.615, infinitive: Var. 0.560, Std. Dev. 0.748). *Ir* vs. morphological: N = 659, random effects (speaker: var. 1.070, Std. Dev. 1.035, infinitive: Var. 1.613, Std. Dev. 1.270). Positive coefficients are associated with increased *haber/ir*: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1., n.s. = not significant.

Importantly, the results of the two multiple linear regression analyses suggest that the *haber* periphrasis and the *ir* periphrasis differ from the morphological form in similar ways. In both models, the morphological future is favored (compared to a periphrastic form) with second and third persons and when it is in the apodosis of a conditional clause expressing contingency. In neither analysis is century[6] or adverb use a significant predictor of future expression. Unsurprisingly, the morphological future is more grammaticalized than the periphrastic form. This is evident from the finding that the periphrastic *haber* and *ir* forms are favored in first person, which is associated with intention and, as has been previously shown, is a former step to the focal meaning of future (prediction). Further, the morphological form is also favored in the apodosis with contingent meaning, which we interpret as evidence of increased grammaticalization (cf. Bybee et al. 1994, p. 274).

The results of the two analyses summarized in Table 5 show marked similarities between the periphrastic forms in relation to the morphological future. Although morphological future presents characteristics of greater degree of grammaticalization, the results do not allow a determination of the degree of grammaticalization of the periphrastic forms in relation to one another. To make this determination, we compared factors constraining *haber* periphrasis compared to *ir*. This informed our research goal of determining a cline of grammaticalization for these future forms. Table 6 summarizes the results of this multiple linear regression analysis.

This analysis, predicting the *haber* compared to the *ir* periphrastic form, demonstrates that with the exception of grammatical person, no significant differences are found between the two forms. As is evident in Table 6, third person subjects do not differ significantly from first person subjects. This implies that both periphrases can be used to express prediction in similar ways. Further evidence in favor of this interpretation is that both occur in combination with inanimate subjects to the same extent in our dataset [*haber* 8% (N = 24), *ir* 9% (N = 42)]. When compared to *ir*, the *haber* periphrasis is more likely in the second person.

With regard to the expression of obligation, Coates (1983, p. 37) notes, "it is generally true that examples with second person subjects are stronger" than those with first and third person subjects. This result may suggest evidence of the retention of the original meaning of obligation for the *haber* in a context (second person subject) that is particularly prone to expressing deontic modality. Given the potential for competition between *ir* and *haber* forms throughout the history of the language, future studies should determine whether *haber* may disappear as a future marker or might specialize as expressing deontic modality meanings.

Table 6. Linguistic factors favoring *haber* periphrasis over *ir* periphrasis (N = 754).

Fixed Effects	N	Est. Coef.	SE	*p*
Intercept	297	−0.6889	0.203	***
Century: 21st	13	−0.331	0.590	n.s.
Century: 20th	284			
Person: third	88	0.163	0.228	n.s.
Person: second	56	0.570	0.267	*
Person: first	153			
Adverb: yes	52	0.106	0.242	n.s.
Adverb: no	245			
Meaning: contingent	32	0.353	0.311	n.s.
Meaning: assumed	265			

Random effects (speaker: Var. 0.741, Std. Dev. 0.861, infinitive: Var. 0.636, Std. Dev. 0.797). Positive coefficients are associated with increased *haber*: Signif. codes: 0 '***' 0.001 '*' 0.05 ' ' 1, n.s. = not significant.

4.3. Haber as a Grammaticalizing Form

As grammaticalizing forms move from lexical forms toward grammatical markers, they reduce in phonetic form (e.g., Phillips 2006; Bybee 2010). Consider, for example, the grammaticalization of Latin *cantare habeo* as Galician *cantarei* in which the auxiliary resulted in an inflectional ending (*-ei*). Phonetic reduction is also evident in periphrastic forms in present-day languages. In Spanish, for example, the construction [*ir* 'go' (PRESENT) *a* + INFINITIVE] undergoes phonological reduction, both in terms of segment lenition and deletion (Lipski 2008, p. 113; Silva-Corvalán and Enrique-Arias 2017, p. 237) and durational shortening (Brown and Rivas 2022). We ask, then, whether there is evidence of phonetic reduction in the *haber* periphrasis in Galician.

Based upon the orthographic transcription of the CORILGA conversations, we note that the Galician *haber* periphrastic form is variably realized both with the preposition (*haber de* + INFINITIVE) and without (*haber* + INFINITIVE), as is shown in (15) and (16) respectively:

(15) OIED-URB-CDUB-SANTIAGODECOMPOSTELA-36-1995
Sea tarde ou sea cedo, eu hei de vir, usté tranquilo
'Sooner or later I will come back, do not worry'

(16) OIED-URB-CDUB-SANTIAGODECOMPOSTELA-36-1995
Despois han-lle vir os da comisión
'After that, the committee members will come'

The forms lacking the preposition are more phonetically reduced (exhibit less phonetic substance) compared to instances of *haber* accompanied by *de*. The subset of the tokens of use of *haber* provide insights, therefore, into the gradual phonetic erosion of a form having gone through grammaticalization. During the process of grammaticalization, do phonetic conditioning factors conspire to propel the reduction further? A positive result might provide impetus to reconsider historical data through this lens.

To our knowledge, an analysis of this type has not been conducted on Galician data. Based on the extent to which comparisons can be drawn to Spanish and Portuguese, these data also provide synchronic evidence with which to consider the processes that took place historically. Such analyses could help identify more specific pathways and

mechanisms of change. We examined the target tokens with phonetic conditioning factors in mind. We coded the accompanying infinitive for whether it has a consonant onset or not, hypothesizing that the reduction in the preposition *de* would be favored by a following consonant due to coarticulatory effects of production. Additionally, abundant evidence suggests that a high token frequency accelerates reductive change (Bybee 1999). We estimated the frequency of the infinitive based upon the future tokens in our corpus. The mean raw frequency for infinitives in our study was 5. Tokens with a count above 5 were coded as 'high', and those 5 and below, as 'low'. For the same reasons as described previously, we suspect corpus differences (across the decades). Table 7 summarizes the rate of preposition use (as opposed to omission) with the *haber* tokens in our dataset. The preposition *de* is omitted more commonly before consonant-initial and high-frequency infinitives and its omission is highest in the data of the most recent recordings.

Table 7. Preposition use with *haber* tokens.

		N	% Preposition
Phonetic Context	Pre-Vocalic	51	96
	Pre-Consonantal	246	48
Word Frequency of Infinitive	Low	64	73
	High	233	51
Decade/Corpus	1960s	235	56
	1990s	49	61
	2000s	13	31

In order to determine which factors condition preposition use (maintenance vs. reduction), we submitted the *haber* tokens to a multiple linear regression model using R, with the speaker and infinitive as random intercepts. We included frequency as a continuous factor (log of raw frequency). Table 8 summarizes the results of this model. If the infinitive following the periphrastic future begins with a vowel, preposition use is highly favored. In contrast, a following pre-consonantal phonetic context favors omission of the *de*. The corpora reflecting speech samples from earlier decades (1960s, 1990s) favor de use compared to the interviews conducted in the twenty-first century. As is evident in Table 7, in just 4 of the 13 cases of *haber* used in the later corpora is a preposition employed between *haber* and the infinitive. Word frequency does not significantly predict preposition omission.

Table 8. Linguistic factors predicting preposition *de* use in *haber* future constructions (N = 297).

Fixed Effects	N	Estimate Coef.	SE	Sig.
(Intercept)	166	–3.308	1.806	.
Word frequency	297	–0.707	0.562	n.s.
Post-phonetic context (vowel)	51	5.253	1.484	***
Post-phonetic context (consonant)	246			
Decade: 2000s	13			
Decade: 1960s	235	4.736	1.939	*
Decade: 1990s	49	4.997	2.065	*

Random effects: Speaker (N = 81) Var. = 3.470, Std. Dev. 1.863, Infinitive (N = 98), Var. = 4.253, Std. Dev. 2.062. Positive coefficients are associated with increased *haber de*: Signif. codes: 0 '***' 0.001 '*' 0.05 0.1 ' ' 1.

The findings summarized above suggest that as forms grammaticalize, in addition to the contributions of the commonly cited effects of frequency of use (token frequency), phonetic factors of the production context are, of course, operative. Effects of context of use, therefore, can accumulate in memory (Bybee 2002) and shape the pattern of change. These results support the notion that the phonetic reduction in the grammaticalizing forms is enhanced in certain production contexts, while phonetic reduction in other words is inhibited.

5. Discussion and Conclusions

Based upon our analyses of the forms expressing future in Galician, we are able to propose the following cline of grammaticalization (*haber* >> *ir* >> morphological). There are several pieces of evidence in support of this cline. Morphological future, as anticipated, demonstrates signs of the most significant degree of grammaticalization. Compared to periphrastic constructions, morphological future is also significantly preferred with second and third persons and in contingency contexts (Table 5). These two results also reveal a more advanced degree of grammaticalization. Additionally, as demonstrated in Table 2, the morphological future form is more frequently used as an epistemic marker compared to other future forms, suggesting a further stage of grammaticalization. Epistemic meanings also appear with the periphrastic forms, but with a much reduced frequency of use. What is more, as previously discussed, the morphological form occurred more frequently in our data to express epistemicity than to express futurity. The morphological future, then, is not the default marker of future time.

Both periphrastic forms seem to be highly grammaticalized, having been bleached of their original semantic content. We find examples, for instance, in which both verbs *haber* and *ir* are used in conjunction with the matching lexical counterpart [ha *haber*, vou *ir*]. They both also occur with inanimate subjects, which indicates auxiliarization, since inanimates are not associated with intentionality. Additionally, the results summarized in Table 5 show that the periphrastic forms behave similarly when compared to the morphological future form. However, when contrasted with one another, *haber* is shown to favor the second person, which suggests a retention of its original meaning of obligation, indicating that *ir* is further grammaticalized. Additional evidence in favor of this cline is found in the analysis of the type and token frequency of each. In our full dataset of future tokens (N = 1589), *haber* occurs with 98 different infinitives, whereas *ir* collocates with 195 different infinitives.[7] A greater type frequency suggests greater grammaticalization (Brinton and Traugott 2005, p. 109), placing *ir* further along the cline than *haber* in this regard. Additionally, *haber* forms are used much less frequently than *ir* forms (Table 2).

As a grammaticalized form, the *haber* periphrasis is demonstrating variable phonetic reduction (as measured using the presence vs. absence of the preposition *de*). Frequency of use alone (as measured within the corpus) does not account for the variable lenition of the form. Conditioning factors present in the production context, such as phonetic context, can help account for the patterns of development of the reduction. As words' usage patterns accumulate in memory, such distributions in discourse can manifest themselves as significantly different rates of change. In line with Brown and Raymond (2012), future studies could estimate lexical distributions in Galician discourse by analyzing historical data.

The variationist analysis of the Galician data presented in this work revealed that the *haber* periphrastic construction has all the makings of a perfect marker of futurity. The *haber* future form is highly grammaticalized (semantically bleached, phonologically reduced), yet it appears to be falling into disuse when we compare it to other markers of future and its frequency of use across decades. However, unlike in other neighboring Romance languages, its loss will not be due to a lack of grammaticalization; as we have demonstrated, it has achieved a high degree of grammaticalization. Given the widespread use of a periphrastic *ir* as a future marker in neighboring languages (Spanish, Portuguese) in lieu of *haber* forms, a language contact explanation supporting this (potential) loss of *haber* as a future marker could be explored in subsequent studies.

Notes

1   'In Galician, the situation is a little more nuanced. In spite of the great increase in the use of <*ir* + infinitive>, the construction <*haber (de)* + infinitivo> is significantly maintained, being far from the marginal status it has acquired as a future form in neighboring languages' (authors' translation).

2   Our analysis is based on data taken from the *CORILGA* corpus (see §3). This corpus provides the following information for each token: text type (OICO: oralidade informal conversa 'informal speech, conversation', OIED: oralidade informal entrevista dirixida 'informal speech, guided interview'), habitat (RUR: rural, URB: urban), source corpus (CBAS: Basanta Corpus, CDUB: Dubert Corpus, CHEN: Henningsen Corpus), town, recording number and year.

3   The Galician periphrastic construction [ir (PRESENT) + INFINITIVE] follows a similar grammaticalization path with the caveat that *ir* occurs in the simple present form (e.g., *vou* 'I go', *vas* 'you go', *vai*—'s/he goes'...) and not in the present progressive (*estou indo* 'I am going', *estás indo* 'you are going', *está indo* 's/he is going'...), like in English.

4   This number reflects all the informal 2000s data available in the corpus.

5   We identified four strategies to express future in this data: the periphrastic forms with *haber* and *ir*, morphological future, and simple present. Using a subset of the same data source (N = 646), we determined that the simple present is a relatively common strategy used to express future meaning in Galician (17%). Future uses of the simple present are frequently accompanied by an adverb indicating future time (N = 36, 41%). In fact, in these data, the percentage of adverb expression is statistically significantly higher (χ^2 = 9.2645, p = 0.0023) with the simple present than with any of the other future markers (N = 71, 19%). In line with Torres Cacoullos and Walker (2009, p. 335), we interpreted this result as evidence that future is not part of the basic meaning of the simple present tense, since it often requires the help of a temporal adverb in the same clause or in the surrounding discourse to express this meaning. This suggests that the future meaning resides outside the verb form itself. Therefore, we dropped the simple present from the analysis. Future research can examine possible additional ways in which present forms express futurity, differing from the object of study in this work (periphrastic future forms).

6   Given the distribution of the use of future markers across decades summarized in Table 4, an external reviewer questioned the decision to analyze the data distinguishing century instead of contrasting the 1960s data with the latter two decades [1960s vs. other (1990s, 2000s)]. We repeated the three statistical models that we reported in this work following this suggestion. The data coded in this way do not change our results, and the interpretations remain the same.

7   Both *haber* and *ir* occur with a multiplicity of semantic types (e.g., dynamic, stative, movement).

References

Aaron, Jessi E. 2006. *Me voy a tener que ir yendo*: A Corpus-based study of the grammaticization of the ir a + INF construction in Spanish. In *Selected Proceedings of the 9th Hispanic Linguistics Symposium*. Edited by Nuria Sagarra and Almeida Jacqueline Toribio. Somerville: Cascadilla Proceedings Project, pp. 263–72.

Aaron, Jessi E. 2010. Pushing the envelope: Looking beyond the variable context. *Language Variation and Change* 22: 1–36. [CrossRef]

Álvarez, Rosario, and Xosé Xove. 2002. *Gramática da Lingua Galega*. Vigo: Galaxia.

Brown, Esther L., and William Raymond. 2012. How discourse context shapes the lexicon: Explaining the distribution of Spanish f-/h/ words. *Diachronica* 99: 139–61. [CrossRef]

Brown, Esther L., and Javier Rivas. 2022. Phonological reduction of Spanish periphrastic future forms. Paper presented at the 10th International Workshop on Spanish Sociolinguistics. Georgia Tech University, Atlanta, GA, USA, April 4–5.

Blas Arroyo, José L. 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: 85–126. [CrossRef]

Blas Arroyo, José L. 2018. Comparative variationism for the study of language change: Five centuries of competition amongst Spanish deontic periphrases. *Journal of Historical Sociolinguistics* 4: 177–219. [CrossRef]

Blas Arroyo, José L., and Kim Schulte. 2017. Competing modal periphrases in Spanish between the 16th and the 18th centuries. A diachronic variationist approach. *Diachronica* 34: 1–39. [CrossRef]

Brinton, Laurel J., and Elizabeth C. Traugott. 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.

Bybee, Joan. 1999. Usage-based phonology. *Functionalism and Formalism in Linguistics* 1: 211–42.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14: 261–90. [CrossRef]

Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, Joan, Revere D. Perkins, and William Pagliuca. 1994. *The Evolution of Grammar. Tense, Aspect, and Modality in the Languages of the World*. Amsterdam and Philadelphia: John Benjamins.

Cidrás Escáneo, Francisco. 2009. Os camiños da gramática: Unha aproximación á variación e o cambio lingüístico desde a sintaxe do galego. *Estudos de Lingüística Galega* 1: 27–46.

Coates, Jennifer. 1983. *The Semantics of the Modal Auxiliaries*. London: Croom Helm.

Cunha, Luís F. 2021. Propriedades temporais do futuro simples em Português Europeu. *Estudos de Lingüística Galega* 13: 29–66. [CrossRef]

Cunha, Luís F. 2022. The expression of futurity in Spanish and Portuguese: Similarities and differences. *Borealis. An International Journal of Hispanic Linguistics* 11: 101–39. [CrossRef]

Dahl, Östen. 2000. The grammar of future time reference in European languages. In *Tense and Aspect in the Languages of Europe*. Edited by Östen Dahl. Berlin and New York: Mouton de Gruyter, pp. 309–28.

de Prada Pérez, Ana, Inmaculada Gómez Soler, and Nick Feroce. 2021. Variable future-time expression in Spanish: A comparison between heritage and second language learners. *Languages* 6: 206. [CrossRef]

Fernández Rei, Elisa, and Xosé L. Regueira Fernández. 2019. CORILGA: Un corpus para o estudo da variación e do cambio lingüístico no galego falado. In *Corpus y Construcciones. Perspectivas Hispánicas*. Edited by Marta Blanco, Hella Olbertz and Victoria Vázquez Rozas. Santiago de Compostela: Universidade de Santiago de Compostela, Servizo de Publicacións, pp. 221–43.

Fleischman, Suzanne. 1982. *The Future in Thought and Language: Diachronic Evidence from Romance*. Cambridge: Cambridge University Press.

Freixeiro Mato, Xosé R. 2006. *Gramática da Lingua Galega II. Morfosintaxe*, 2nd ed. Vigo: A Nosa Terra.

García Represas, Delio. 2001. A necesidade e maila obriga en galego. *Cadernos da Lingua* 23: 89–116. [CrossRef]

Giordano, Roberta. 2022. "*Ir a* + infinitivo" y su potencial comunicativo. *Círculo de Lingüística Aplicada a la Comunicación* 89: 145–54. [CrossRef]

Gradoville, Michael. 2019. The disappearance of the morphological future from educated spoken Carioca Portuguese. In *Lusophone, Galician, and Hispanic Linguistics. Bridging Frames and Traditions*. Edited by Gabriel Rei-Doval and Fernando Tejedo-Herrero. New York: Routledge, pp. 227–45.

Hopper, Paul, and Elizabeth C. Traugott. 2003. *Grammaticalization*, 2nd ed. Cambridge: Cambridge University Press.

Kouteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog, and Seongha Rhee. 2019. *World Lexicon of Grammaticalization*, 2nd ed. Cambridge: Cambridge University Press.

Labov, William. 1994. *Principles of Linguistic Change. Internal Factors*. Oxford: Blackwell, vol. 1.

Lehmann, Christian. 2015. *Thoughts on Grammaticalization*, 3rd ed. Berlin: Language Science Press.

Lipski, John. 2008. *Afro-Bolivian Spanish*. Madrid: Iberoamericana Vervuert.

Orozco, Rafael. 2018. *Spanish in Colombia and New York City. Language Contact Meets Dialectal Convergence*. Amsterdam and Philadelphia: John Benjamins.

Phillips, Betty S. 2006. *Word Frequency and Lexical Diffusion*. New York: Palgrave Macmillan.

Poplack, Shana, and Elisabete Malvar. 2007. Elucidating the transition period in linguistic change: The expression of the future in Brazilian Portuguese. *Probus* 19: 121–69. [CrossRef]

Rojo, Guillermo. 1974. *Perífrasis Verbales en el Gallego Actual*. Santiago de Compostela: Universidade de Santiago de Compostela, Secretariado de Publicaciones.

R: The R Project for Statistical Computing. n.d. About R. Available online: https://www.r-project.org/ (accessed on 8 August 2022).

Schwenter, Scott, and Rena Torres Cacoullos. 2008. Defaults and indeterminacy in temporal grammaticalization: The 'perfect' road to perfective. *Language Variation and Change* 20: 1–39. [CrossRef]

Sedano, Mercedes. 2006. Importancia de los datos cuantitativos en el estudio de las expresiones de futuro. *Revista Signos* 39: 283–96. [CrossRef]

Silva-Corvalán, Carmen, and Andrés Enrique-Arias. 2017. *Sociolingüística y Pragmática del Español*, 2nd ed. Washington, DC: Georgetown University Press.

Tagliamonte, Sali A. 2012. *Variationist Sociolinguistics. Change, Observation, Interpretation*. Malden and Oxford: Wiley-Blackwell.

Torres Cacoullos, Rena. 2011. Variation and grammaticalization. In *The Handbook of Hispanic Sociolinguistics*. Edited by Manuel Díaz-Campos. Malden and Oxford: Blackwell, pp. 148–67.

Torres Cacoullos, Rena, and James Walker. 2009. The present of the English future: Grammatical variation and collocations in discourse. *Language* 85: 321–54. [CrossRef]