

*Me voy a tener que ir yendo: A Corpus-Based Study of the Grammaticization of the *ir a* + INF Construction in Spanish*

Jessi Elana Aaron¹
University of Florida

1 Introduction

1.1 Grammaticization

Much of the recent work in usage-based linguistics has focused on two fundamental issues: Where does grammar come from, and how does it change? The theory of Grammaticization (e.g. Hopper & Traugott 1993; Bybee, Perkins & Pagliuca 1994) offers a refreshing perspective: language changes as frequent routinizations become conventionalized. One example of this is when lexical morphemes evolve into grammatical morphemes, such as English *will*, which originally meant ‘want’, and later became a future marker. This transformation is possible when pragmatic-discursive strategies, such as common inferences in certain frequent contexts, come to be understood as a part of the semantics of the context with which it frequently co-occurs. In the case just mentioned, then, when the speaker would talk of desire, a hearer would often infer intention, and from intention, prediction, or futurity (see Bybee, Perkins & Pagliuca 1994). These inferences would then become conventionalized, and thus become a part of the grammar. Such transformations are often accompanied by an increase in frequency, increased fixedness and phonological reduction.

Given such a usage-based mechanism for grammar formation, it is not surprising to find overwhelming cross-linguistic similarities: forms with certain lexical meanings, if they grammaticize, often have similar grammatical functions. In the case of the future, various common universal paths have been found. Three of the most common are ‘desire’, as in English *will*, and ‘go’ and ‘possession’ (Bybee, Perkins & Pagliuca 1994). It is the latter two that have produced the two morphological Spanish Future forms we have today. The goal of this study is to offer empirical evidence for the diachronic process of the grammaticization of the Spanish Analytic Future within a variationist perspective.

1.2 Analytic Future and Synthetic Future

One ideal site for the examination of grammaticization in progress is the realm of the future, which is notoriously unstable (Cartagena 1995-1996) and for this reason susceptible to change. In present-day Spanish, we find two morphological future forms: the Analytic Future (AF) in (1) and the Synthetic Future (SF) in (2).

(1) Analytic Future (AF): *ir (a) + INF*

...usted **va a poder** tener su propio vagón personalizado (COREC, CECON004B)
‘you will be able to (AF) have your own personalized [train] car’

(2) Synthetic Future (SF): *INF + haber*

Usted **podrá** elegir el vagón que quiere (COREC, CECON004B)
‘you will be able to (SF) choose the [train] car you want’

¹ I would like to thank Clancy Clements, Rena Torres Cacoullós, Catherine Travis, HLS participants, and three anonymous reviewers for their comments and suggestions. I alone, however, am responsible for any errors in fact or analysis.

Both forms have their lexical origins in unsurprising—yet distinct—sources. The AF has its origins in a construction that indicated an *agent on a path toward a goal* (Bybee, Perkins & Pagliuca 1994), as in (3).

- (3) Original meaning of AF: an agent on a path toward a goal
 estudia , mientras **voy** yo , **a** le **dezir** tu pena (Celestina, 104)
 ‘study, while I go to tell (AF) him your sorrow’

The SF, on the other hand, has in origins in a construction indicating *possession* and, later, *obligation* (Bybee, Perkins & Pagliuca 1994), as in (4).

- (4) Original meaning of SF: possession > obligation
 nunca por vicio nin por folgura *dexaredes* de fazer tales cosas, porque, aun desque vos murierdes, siempre viva la fama de los vuestros fechos. (Lucanor)
 ‘never due to vice or sloth should you stop (SF) doing such things, so that, even after you are dead, the fame of your deeds will always live.’

Despite these different origins, these forms, during their grammaticization, have shared uses that are diachronically related: the temporal space of futurity.

If these two forms do, indeed, share certain meanings, a handful of pressing questions arises: What is the difference? Why do speakers choose one form over the other? How has this choice changed over time? In this paper I will show that these questions have yet to be satisfactorily answered. Furthermore, I will argue that the subtle differences between these forms can only be discovered through quantitative analyses. These analyses demonstrate that, while AF frequency has skyrocketed since the mid-13th century, grammatical constraints on AF have been slower to change, and still show evidence of its original lexical meaning.

1.3 The problem: overlapping forms

The majority of the literature on the Spanish and other Romance futures has paid particular attention to the first question raised in the previous section: what is the difference between these forms? Taking a synchronic approach, most scholars have claimed that these forms appear in different contexts or with different meanings. For instance, the AF has been labeled an inchoative, meaning that it refers to an action that has already begun to be carried out, either in intention or "objective reality," (Cartagena 1995-1996:80), or to an action whose preparatory stages have already been accomplished (Vet 1994), as in (5), where the interlocutor has already done something that may lead the picture to break.

- (5) ¡Cuidado! **Vas a romper** el - el cuadro. (COREC, CACON026B)
 ‘Careful! You’re gonna break (AF) the – the picture’

This hypothesis is consistent with the origins of the AF, which denoted an agent on a path toward a goal (Bybee, Perkins & Pagliuca 1994:280; Hopper & Traugott 1993:3).

The AF is also claimed to differ from the SF in that it is used to refer to events that are soon or that are relevant to the present (Cartagena 1978:384; Fleischman 1982:87, 181), as in (6)-(8).

- (6) ¿Quieren que les cuente qué es lo que **vamos a tener hoy** en el programa? (COREC, CBCON032A)
 ‘Do you want me to tell you what we are going to have (AF) today on the program?’
- (7) para sacar dinero porque me he acordado que ya que venía aquí me **voy a ir ahora mismo** a - a esa bodega de ahí de Azcona a comprar unos vinos para un médico (COREC, CBCON014B)
 ‘to take out money because I remembered that as I was on my way over here I am going to go (AF) right now to – to that store there in Azcona to buy some wine for a doctor’

- (8) Y **vamos a ir** el **miércoles** a ensayar con la orquesta. (COREC, CCON018B)
 ‘And we’re going to go (AF) on Wednesday to rehearse with the orchestra’

Others have claimed that the AF is used to refer to distant events that will happen within a specified time frame (Helland 1997:73-75), as in (9).

- (9) Luego que Yamaha va a volver - porque y , las motos - italianas están apretando muy fuerte. - pues ya está a la altura de las Hondas - y **va a empezar** el año próximo también a competir oficialmente - en el mundial de 250.... (COREC, CECON006B)
 ‘So Yamaha is going to return – because and, the Italian motorcycles are really pushing hard. – well it’s already at the level of the Hondas – and next year it’s going to start (AF) to compete officially – in the world 250’

However, no consensus has been reached on whether these forms really carry different meanings or if they are interchangeable (Poplack & Turpin 1999). When faced with patterns found in natural data, however, there are some critical problems with the analyses presented above. Firstly, many of them, such as the idea that a speaker may simply perceive the action as having a preparatory state already completed (Cartagena 1995-1996:80), rely on speaker intentions as a key component. While this is an attractive explanation, speaker intention is most often unavailable to the analyst, and as such cannot serve as the basis of an analysis (DuBois 1987:811-812). Secondly, natural patterns reveal problems with dichotomous separation of the AF and the SF. As we will see in the following section, these forms do overlap in some contexts.

1.4 Language change and variation

The fact that these forms overlap should not be cause for surprise. As Bybee and Hopper (2001:19) aptly point out, "grammar is not fixed and absolute with a little variation sprinkled on top, but it is variable and probabilistic to its very core" (cf. "inherent variation" Labov 1972:204). One of the consequences of this variability, then, is what has been called *layering*, which is found when a form develops new meanings without immediately (or ever) replacing other forms within the same functional domain (Hopper 1991:22).

This layering effect is what we find in the Spanish future today: two forms sharing space within the same semantic domain. So, given what we know of the AF, we might expect to find the SF in complementary distribution, i.e., not in the contexts that, according to the literature, pertain to the AF. A look at natural data, nonetheless, reveals a different picture. Among other contexts, the SF can also be found—though less frequently—as an inchoative, as in (10).

- (10) un proyecto más y se nos van a esa Semana Blanca Canaria en el Pirineo y **de forma ya inminente** **estarán** por allí estos expedicionarios que van a ser bastantes e insisto, también la prensa se dará por allí una vueltita, a ver qué pasa. (COREC, CCON012B)
 ‘one more project and they’re going to leave us for that Semana Blanca Canaria in the Pyrenees and imminently these sight-seers, who are going to be a bunch, will be (SF) around, and I insist, the press will show up around there, too, we’ll see what happens’

The SF can also be found to be used to refer to events that are soon or that are relevant to the present, as in (11)-(13).

- (11) Lo que pasa que **hoy** **estará** muy lleno a lo mejor. (COREC, CACON006D)
 ‘the thing is that today it will be (SF) really full maybe’
- (12) Muy bien, Domingo, nos hemos pasado de tiempo, así es que **volveremos** **dentro de un momento** - con "Lucky Strike" (COREC, CECON006B)
 ‘very well, Domingo, we’ve gone over time, so we will return (SF) in a moment – with Lucky Strike [cigarettes]’

- (13) Pues que - vamos a ver. El martes, yo daré clase a estos niños pues seguramente de seis a siete (COREC, CCON029D)
 ‘well – let’s see. Tuesday, I will give (SF) class to these kids well surely from six to seven’

And, once again, the SF can also refer to distant events that will happen within a specified time frame, as in (14).

- (14) Se dice - que Honda el año próximo no hará motos oficiales en doscientos cincuenta, al cabo de dos días dicen: "No, que habrá - dos". (COREC, CCON006B)
 ‘they say – that Honda next year will not make (SF) official motorcycles in 250, then after two days they say: “No, that there will be – two”’

It appears, then, that the meaning differences between AF and SF—if, in fact, there are any—are not perceptible through qualitative analysis of discourse data alone. Instead, I suggest, we may want to dig deeper. One tangible point of entry for the discovery of possible differences may be found in the distinct lexical origins of these two forms. As forms generalize and develop new meanings, older meanings of the form do not disappear, but rather co-exist with newer meanings (Lichtenberk 1991, Hopper & Traugott 1993:3, Bybee, Perkins & Pagliuca 1994:17, Schwenter 1994:81). This means that we should expect to find *persistence*, or *retention* of older meanings.

Such persistence may take the obvious form of polysemy, where one form has more than one meaning. The consequences of persistence, however, may be more subtle. Instead of (or alongside) polysemy, there may be *distinctive patterns of variation with competing forms*: synchronic syntactic or semantic constraints on the more grammaticized uses of the form (Poplack & Tagliamonte 2001:207-235), i.e., on the AF as a future marker, will reflect its lexical origins and diachronic trajectory.

1.5 Research questions

This study presents a quantitative account of the traces of lexical meaning found in patterns of AF-SF variation. If it is the case for the AF that lexical meaning persists, then we will see evidence of its original “agent on a path toward a goal” meaning in its tendencies to occur in certain linguistic contexts. More specifically, the following hypotheses can be made:

- i. AF will be used more with subjects that are capable of autonomous motion, i.e. animate subjects.
- ii. AF will appear less often in negation and more often to refer to events that will be realized.
- iii. AF will be preferred when referring to a specified time.
- iv. AF will eschew verb classes that are inconsistent with the original goal meaning (e.g. stative, psychological) and the redundant *ir* (e.g. *voy a ir*).

Diachronically, as the AF distances itself from its original lexical meaning, we would expect to find a weakening of the effect of these constraints over time.

2 Data and methodology

2.1 Data

Since this study focuses on patterns of distribution, a corpus of natural language was selected. The corpus was limited to peninsular Spanish in order to minimize the added complexities found in cross-dialectal studies. The texts used range from the mid-13th through the early 21st centuries, and contain both written and, for the 20th century, oral data. Texts were separated by a gap of 100-150 years in order to allow enough time to pass between periods examined for changes to be detectable. A full list of the texts used, in diachronic order, can be found in the Corpus section of the References. These 17 texts (16 written, one oral) made up approximately 935,500 words, and produced a total of 5,579 occurrences of AF and SF, with 1,072 occurrences of the former and 4,507 of the latter.

All instances of AF and SF—whether or not they had future meaning—were extracted from the corpus. However, if our driving question here is the *difference* between these two forms, i.e. the motivation behind the choice of one over the other in language use, then we must only concern ourselves with the instances in which *both forms are possible*. In this case, the site of the relationship between these two forms is futurity; that is, their only link to each other is diachronic, since they both happened to evolve as futures (on diachronic relationships between forms, see Torres Cacoullos 2001). This overlapping space is thus the focus of this study.

If this is the case, then occurrences that fall somewhere outside this overlapping space must be excluded from an examination of AF-SF variation. These exclusions are shown in Table 1, and include: *haber* + PP in SF, which carries a past tense meaning (e.g. ‘she will not have found out’) or epistemic modal past meaning (e.g. ‘she must not have found out’); discourse markers, such as *vamos a ver* ‘let’s see’; non-future uses, including the increasingly common epistemic modal meaning of SF² and habitual uses of the AF; and truncated utterances or utterances whose meaning was unclear. This amounted to 464 tokens, which left the total token count at 5,115.

Table 1. *Exclusions*

<i>Haber</i> + PP (e.g. <i>no se habrá enterado</i>)	85
Discourse markers, nominalizations (e.g. <i>vamos a ver</i>)	103
Non-future uses (e.g. habitual, modal)	243
Truncated utterances, unclear meaning	33
Total	464

The remaining tokens were coded for various contextual factors and subjected to statistical analyses. These methods will be discussed in section 2.2.

2.2 Method: factors considered and Variable Rule Analysis

In the search for possible meaning and/or distributional differences between the Analytic Future and the Synthetic Future in Spanish, we may first glean clues from previous hypotheses on the matter. As noted in the Introduction, there have been many prior attempts to differentiate these forms. Though these have rarely been based on quantitative evidence, and taken as whole they are, at best, inconclusive, it is possible to extrapolate empirically measurable contextual factors from some of these hypotheses. A further guide taken into consideration is the quantitative variationist work by Poplack and Turpin on the Canadian French future (1999).

Given previous findings, then, six linguistic factors were considered to possibly have a statistically significant effect on the choice between AF and SF in Spanish. These were: subject animacy, sentence type (declarative or interrogative), clause type (main or subordinate), polarity (positive or negative), verb class, and the presence of a temporal adverbial. In all but the 17th century (where temporal adverbials were highly infrequent), adverbials were further divided into “specific”, i.e., referring to a specific time frame, such as *mañana* ‘tomorrow’ or *a las seis* ‘at six o’clock’ and “nonspecific”, as in *algún día* ‘someday’ or *siempre* ‘always’.

The statistical effect of each of these factors when considered simultaneously was tested using the computer program GoldVarb (Rand & Sankoff 1990). GoldVarb is a multivariate analysis program that provides a wealth of information about the probabilistic constraints on variation in natural data. First, GoldVarb provides the *Range* for each contextual factor considered, indicating the *magnitude of effect*, i.e. how strong the effect of any given factor group (e.g. verb class) is in predicting the occurrence of a form. Second, it provides the *Probability weight* for each factor, i.e. the relative importance of any given contextual feature (e.g. stative verbs) in the co-occurrence of a form.

² Work not yet published by this author has shown a dramatic increase in modal SF over the past two centuries.

3 Results

Before entering into the question of the tendencies in AF-SF variation over time, however, it is important to step back for a moment to look at the bigger picture. While both forms have, indeed, been possible as futures since the 13th century, in is by no means the case that they have been on an even playing field. In fact, a quick glance at Table 2, which shows absolute and relative frequencies for each form by century, reveals that AF has emerged from a fledgling construction with nearly imperceptible frequency in the 13th century to a thriving, even preferred, construction in 20th century speech.

Table 2. *Absolute and relative frequencies of temporal AF and SF by century, raw and normalized per 10,000 words*

Century (word count)	AF		SF		AF:SF
	N	per 10,000	N	per 10,000	Ratio (%AF)
13th c. (~200,000)	1	0.05	1058	52.9	1:1058 (<1%)
15th c. (~111,000)	12	1.1	770	69.4	1:63 (2%)
17-18th c. (~232,000)	59	2.5	1270	54.7	1:22 (4%)
19th c. (~71,500)	75	10.5	432	60.4	1:7 (12%)
20th c., written (~79,000)	79	10.0	212	26.8	1:3 (27%)
20th c., spoken (~242,000)	768	31.7	379	15.7	1:0.5 (66%)

As we see in Table 2, AF's normalized absolute frequency increased from 0.05/10,000 in 13th-century texts to 31.7/10,000 in 20th-century speech. The diachronic blossoming of the AF was accompanied by a 20th-century decline in the SF, which, after centuries of occurring at a steady rate of 52-60 per 10,000 words, drops to only 26.8 in 20th-century texts, and to an even lower 15.7 in 20th-century speech. In 20th-century speech, in fact, we see a reversal: the AF is now twice as frequent as the SF, and makes up 66% of all 20th-century spoken tokens. We see a similarly astonishing rise in AF relative frequency, from less than 1% in the 13th century, a mere 2% in the 15th, 4% in the 17th, up to 12% in the 19th, and then jumping to 27% and 66% in written and spoken 20th-century data, respectively. These results point to a drastic change in the face of the Spanish future; all evidence seems to indicate that the AF has encroached upon, and indeed is taking over, SF territory.

Given such sweeping changes in frequency, we may suspect to see parallel changes in the internal structure of AF-SF variation. Might we see this change reflected in how speakers choose one form over the other? Is it possible that the AF could retain traces of its lexical origins even in the face of such generalization? The answers are to be found in quantitative analyses of this variation over time. In this study, I have used variable rule analyses to take a synchronic snapshot of the four data sets in which the AF shows more than minimal presence, 17th, 19th and 20th-century literature and 20th-century conversation. Each snapshot reveals the probabilistic constraints on the choice between AF and SF at that moment in time. By comparing these synchronic portraits, we may pinpoint the ways in which the constraints on this choice have changed or remained the same over time.

Table 3 shows the results of the stepwise multiple regression VARBRUL analyses for the four data sets. The first four columns, under "Percent AF," show once again the same trend seen in Table 2: a sharp rise in AF frequency across the board. It is the second set of columns, under "Probability," on which I would now like to focus.

Table 3. VARBRUL results: Factors contributing to the choice of the AF in the 17th, 19th and 20th centuries

17th c. input probability: 0.031 (4%), N = 1329

19th c.: 0.132 (15%), N = 507

20th c. written: 0.250 (27%), N = 291

20th c. spoken: 0.694 (66%), N = 1147

Factor Group	Percent AF				Probability			
	17 th	19 th	20 th w	20 th s	17 th	19 th	20 th w	20 th s
Co-occurring temporal adverbial								
Absent	5	15	31	73	.56	[.52]*	.56	.57
Present	1	19	17	59	.22	[.58]	.38	.43 (specific)
		7	12	37		[.33]	.31	.22 (nonsp)
	<i>Range</i>				34		25	35
Sentence type								
Interrogative	6	14	53	88	[.61]	[.52]	.72	.77
Declarative	4	15	24	64	[.49]	[.50]	.48	.46
	<i>Range</i>						24	31
Polarity								
Affirmative	5	16	29	66	.55	[.52]	[.52]	[.49]
Negative	1	8	16	74	.24	[.36]	[.40]	[.59]
	<i>Range</i>				31			
Clause type								
Main	4	13	28	65	[.47]	.47	[.51]	[.49]
Subordinate	5	22	24	70	[.55]	.63	[.47]	[.53]
	<i>Range</i>					16		
Verb class								
Other, dynamic	6	22	33	74	.62	.65	.59	.58
Motion	9	16	38	69	.73	.55	.60	.56
<i>Ir</i>	--	--	--	68	--	--	--	.49
Stative/percep./psych.	1	6	15	55	.33	.32	.34	.38
	<i>Range</i>				40	33	26	20
Animacy								
Singular, animate	4	17	32	73	[.49]	[.54]	[.56]	.56
Other	4	11	20	59	[.52]	[.45]	[.42]	.43
	<i>Range</i>							13

17th c.: Log likelihood = -221.721; $p < .05$; Chi-square/cell = 0.591219th c.: Log likelihood = -197.382; $p < .05$; Chi-square/cell = 0.962520th c. w: Log likelihood = -155.312; $p < .05$; Chi-square/cell = 0.798420th c. s: Log likelihood = -643.641; $p < .000$; Chi-square/cell = 1.0307

*Square brackets [] indicate that this effect does not achieve statistical significance.

Each number between 0 and 1 in these columns represents the probability that the AF will occur in a certain context. For instance, in the 17th century, the AF was likely to occur a little more than the SF when there was no temporal adverbial, such as *para siempre* 'always', with a probability weight of .56. In contrast, the AF is strongly disfavored in the 17th century when there is a temporal adverbial, with a probability weight of .22. Indeed, we see a similar pattern for adverbials in all four data sets.

A look at the other factors considered reveals a strikingly similar stability in constraints over time. Interrogatives always tend to favor use of the AF, while declaratives appear to neither favor nor disfavor this form. Clause type is significant only in the 19th century, where the AF tends to occur more in subordinate clauses, though the effect is minimal, showing a low range of only 16. Verb class

also shows remarkable stability: active verbs (e.g. *comer* ‘eat’, *buscar* ‘look for’) and motion verbs (e.g. *salir* ‘go out’, *entrar* ‘enter’ consistently favor the AF, while stative (e.g. *estar* ‘be’, *tener* ‘have’), perceptual (e.g. *ver* ‘see’, *oír* ‘hear’) and psychological verbs (e.g. *creer* ‘believe’, *querer* ‘want’) consistently disfavor the AF. These patterns show evidence of the AF’s original lexical meaning, which was inconsistent with the latter three verb classes. Lastly, while the animacy of the subject does not reach statistical significance until the 20th-century spoken data, we do see a general tendency for the AF to occur a bit more often with animate subjects.

The changes we see here are more subtle. The most obvious change is, of course, the entrance of *ir* ‘go’ as a variable context in 20th-century spoken data; before this data set, there were no examples of *voy a ir* ‘I am going to go’ found in the data. Another change we can see is the loss of a polarity constraint: the AF is used increasingly in negated contexts. This could be due to a weakening of the “path toward a goal” meaning of this construction, since a goal may be understood to be meant to be realized. We also see a weakening of the verb class constraint. While this factor has the highest range—and thus the strongest magnitude of effect—in the 15th century, it ranks third of four significant factor groups in 20th-century spoken data, with both temporal adverbials and sentence type showing higher ranges.

This is perhaps the most interesting aspect of these results. In the beginning, the AF was most likely to occur with motion verbs (excluding the fully redundant *ir*), with a probability weight of .73 in the 17th century. In this context, the lexical meaning of *ir* in AF becomes less salient, since the co-occurring motion verb provides the same semantics of motion. This redundancy, perhaps, encourages semantic change, where (non-motion) the inferences of the construction may become more salient to hearers. While the preference for motion verbs, and later other dynamic verbs, continues throughout the form’s trajectory, we also see that this constraint becomes less effective over time, and contexts semantically incompatible with the lexical meanings of *ir* become ever more common. This relaxation of verb class constraints can be seen even more clearly if we examine in more detail the proportion of each verb class within each form by century, shown in Table 4.

Table 4. *Proportion of verb classes by future form, percentages*

Verb class	13 th		15 th		17 th		19 th		20w.		20 s.	
	AF	SF	AF	SF	AF	SF	AF	SF	AF	SF	AF	SF
Other, dynamic	100	53	67	48	64	44	69	42	72	53	58	41
Stative	--	27	--	30	5	34	8	30	8	23	20	34
Motion	--	7	17	6	17	8	14	11	9	5	8	7
<i>Ir</i>	--	3	--	3	--	2	--	3	--	1	6	6
Perception	--	4	17	6	10	7	4	6	2	8	4	6
Psychological	--	6	--	7	3	5	7	8	9	10	4	5
Total	<1	99	2	98	4	96	15	85	27	73	67	33

In Table 4, we see a gradual encroachment of AF into the territories traditionally claimed by SF. While most verb classes show low percentages with both forms, and are as such not as impressive, the pattern shown with stative verbs is unmistakable. While these verbs make up about 30% of SF usage over the past seven centuries, they have had minimal presence with the AF. Instead, we see a slow climb: they do not occur with AF until the 17th century, at a mere 5%. This rises to 8% in the 19th and 20th-century written data sets, and nearly catches up to its rate with the SF in 20th-century spoken data, at 20%. Thus, while the distribution of verb classes remains steady over time for SF, AF gradually spreads to more classes.

4 Summary and general implications

Through a diachronic quantitative analysis of the variation between the Spanish Synthetic Future and Analytic Future, I have shown that the AF has gradually encroached upon the territory of the SF,

leading to a subsequent fall in SF frequency. This encroachment did not entail a general rise in frequency across all linguistic contexts at the same time. Instead, since its inception, the AF has been preferred in certain linguistic contexts, namely with motion and other dynamic verbs without a temporal adverbial—contexts consistent with its original lexical meaning of “agent on a path toward a goal.” Thus, while AF frequency has risen dramatically since the 13th century, from less than 1% to 66%, there are still traces of its lexical origins found in the constraints on its variation with SF. These constraints, while evidencing subtle adjustments as AF generalized to new contexts, have remained relatively consistent over 500 years.

While this study offers a new portrait of AF grammaticization as seen through the evidence found in AF-SF variation patterns, this portrait is by no means complete. The exaggeratedly low rates of AF in the 17th and even 19th centuries make the collection of a statistically relevant number of tokens a challenge, especially when it is important to record AF frequency relative to the SF. Based on the results here, collecting a sample of just a few hundred tokens of 17th-century AF would require the extraction of many thousands of tokens of SF and a corpus of well over 1,000,000 words for the 17th century alone. I have not attempted such a grand-scale study here. A second limitation of this study is that I have not addressed here the use of these forms outside the variable context of futurity. In the case of the AF, such use is minimal, but non-future use of the SF is increasingly frequent. A fuller portrait of the competition within and development of Spanish future expression—beyond the scope of this paper—would examine the semantic changes in these forms that fall outside the realm of future temporal expression.

Despite these limitations, the implications of this study are exciting: even half a millennium after a form begins to grammaticize, quantitative analyses can still show evidence of its original lexical meaning. This shows us that changes in frequency do not necessarily imply similar changes in the internal structure of variation. In fact, the hierarchy of constraints on variation may remain relatively stable even in the face of radical frequency changes. Likewise, we may consider that, during the grammaticization of a form, the subtle changes that emerge in a variety’s grammar may take place at a slower pace than the more noticeable changes in frequency. These results show us, once again, that the very essence of our language is found in its intricately structured variability.

Corpus

- Calila (1250) = Anónimo. *Calila e Dimna*. ed. J. M. Cacho Bleuca and M. J. Lacarra. Madrid: Castalia.
- Zifar (1320) = Anónimo. 1929 [1320]. *El libro del Cavallero Zifar (El libro del caullero de Dios)*, ed. Ch. Ph. Wagner. Ann Arbor: University of Michigan Press.
- Lucanor (1350) = Don Juan Manuel. 1969 [1350]. *El conde Lucanor o Libro de los enxiemplos del conde Lucanor et de Patronio*, ed. J. M. Bleuca. Madrid: Castalia.
- Cárcel (1492) = de San Pedro, D. 1972. *Cárcel de Amor*, ed. Keith Whinnom. Madrid: Castalia.
- Celestina (1499) = de Rojas, F. 1987. *La Celestina*, ed. D. S. Severin. Madrid: Cátedra.
- Quijote (1605-1616) = de Cervantes, Miguel. 1996. *Don Quijote de la Mancha*, ed. F. Sevilla Arroyo & A. Rey Hazas, vols. 2 and 4 of *Obras completas* by Miguel de Cervantes. Madrid: Alianza editorial-Centro de Estudios Cervantinos.
- Dama boba (1562-1635) = Vega, Lope de. 2000. *La dama boba*, ed. Alonso Zamora Vicente. Alicante: Biblioteca Virtual Miguel de Cervantes.
- Príncipe Inocente (1562-1635) = Vega, Lope de. 2003. *Comedia del Príncipe Ynocente*, transcription by Silvia Santos Galiana. Alicante: Biblioteca Virtual Miguel de Cervantes.
- La derrota (1789) = Fernández de Moratín, Leandro. 2002 [1789]. *La derrota de los pedantes*. Alicante : Biblioteca Virtual Miguel de Cervantes.
- El sí (1790-1810) = Fernández de Moratín, Leandro. 1975. *La comedia nueva; El sí de las niñas*, ed. J. Dowling and R. Andioc, Madrid: Castalia, 1975.
- El afán (1831) = Carnerero, José María de (1784-1843). 2000. *El afán de figurar: comedia en cinco actos, en verso*. Alicante: Biblioteca Virtual Miguel de Cervantes.
- La conjuración (1834) = Martínez de la Rosa, Francisco 2003 [1834] *La conjuración de Venecia, año de 1310 / Francisco Martínez de la Rosa ; edición de Marisa Payá Lledó*. Alicante : Biblioteca Virtual Miguel de Cervantes, 2003.
- Billy (1987) = Guerra de Aranguiz, Alicia. 2004 [1987]. *Billy escupe la muerte o Un fin de semana en casa de los Dupont*. Alicante: Biblioteca Virtual Miguel de Cervantes.

- Matar con Mozart (1989) = Cerdán Tato, Enrique. 2004 [1989]. *Matar con Mozart y 29 atrocidades más*. Alicante: Biblioteca Virtual Miguel de Cervantes. Otra ed.: Alicante, Aguaclara.
- Pobres diablos (1999) = Zamora Vicente, Alonso . 2002 [1999]. *¡Estos pobres diablos ... !* Alicante : Biblioteca Virtual Miguel de Cervantes. 1ª ed. en [s.l.], Fundación Antonio Nebrija.
- COREC = *Corpus de Referencia de la Lengua Española Contemporánea: Corpus Oral Peninsular*, director F. Marcos Marín. Available at: www.llf.uam.es/~fmarcos/informes/corpus/corpusix.html (género conversacional)
- Algunos modos (2003) = Sánchez Soler, Mariano (1954-). 2003. *Algunos modos de vivir o de morir*. Alicante : Biblioteca Virtual Miguel de Cervantes.

References

- Bybee, Joan and Paul J. Hopper. 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.
- Bybee, Joan, Revere Perkins and William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.
- Cartagena, Nelson. 1978. Acerca de las categorías de tiempo y aspecto en el sistema verbal español. *Revista Española de Lingüística*, 8.373-408.
- Cartagena, Nelson. 1995-1996. La inestabilidad del paradigma verbal de futuro, ¿hispanoamericanismo, hispanismo, romanismo o universal lingüístico? *Boletín de Filología*, 35.79-100.
- Dubois, John W. 1987. The discourse basis of ergativity. *Language* 63:805-855.
- Fleischman, Suzanne. 1982. *The future in thought and language: Diachronic evidence from Romance*. Cambridge: Cambridge University Press.
- Helland, Hans Peter. 1997. Futur simple et futur périphrastique. *Moderna Sprak*, 91.67-76.
- Hopper, Paul. 1991. On some principles of grammaticization. *Approaches to grammaticalization*, ed. by Elizabeth Closs Traugott and Bernd Heine, 17-35. Amsterdam: Benjamins.
- Hopper, Paul, and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Lichtenberk, Frantisek. 1991. On the gradualness of grammaticalization. *Approaches to grammaticalization*, ed. by Elizabeth Closs Traugott and Bernd Heine, 37-80. Amsterdam: Benjamins.
- Poplack, Shana, and Sali Tagliamonte. 2001. *African American English in the Diaspora*. Oxford: Blackwell.
- Poplack, Shana, and Danielle Turpin. 1999. Does the FUTUR have a future in (Canadian) French? *Probus*, 11.133-64.
- Rand, David, and David Sankoff. 1990. GoldVarb: A variable rule application for Macintosh. Version 2. Montréal: Centre de recherches mathématiques, Université de Montréal.
- Schwenter, Scott. 1994. The grammaticalization of an anterior in progress: Evidence from a Peninsular Spanish dialect. *Studies in Language*, 18.71-111.
- Silva-Corvalán, Carmen, and Tracy D. Terrell. 1989. Notas sobre la expresión de futuridad en el español del Caribe. *Hispanic Linguistics*, 2.190-208.
- Torres Cacoullos, Rena. 2001. From lexical to grammatical to social meaning. *Language in Society*, 30:3.443-478.
- Vet, Co. 1994. Future Tense and Discourse Representation. *Tense and Aspect in Discourse*, ed. by Co Vet and Carl Vetters, 49-76. Berlin Federal Republic of Germany: Mouton de Gruyter.

Selected Proceedings of the 9th Hispanic Linguistics Symposium

edited by Nuria Sagarra
and Almeida Jacqueline Toribio

Cascadilla Proceedings Project Somerville, MA 2006

Copyright information

Selected Proceedings of the 9th Hispanic Linguistics Symposium
© 2006 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 1-57473-413-X library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Aaron, Jessi Elana. 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*, ed. Nuria Sagarra and Almeida Jacqueline Toribio, 263-272. Somerville, MA: Cascadilla Proceedings Project.

or:

Aaron, Jessi Elana. 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*, ed. Nuria Sagarra and Almeida Jacqueline Toribio, 263-272. Somerville, MA: Cascadilla Proceedings Project.
www.lingref.com, document #1384.