Language Variation and Change, **11** (1999), 143–170. Printed in the U.S.A.
© 1999 Cambridge University Press 0954-3945/99 $9.50

Construction frequency and reductive change: Diachronic and register variation in Spanish clitic climbing

RENA TORRES CACOULLOS
University of Florida

ABSTRACT

A comparison of Old Spanish and present-day Spanish data provides evidence that reductive change in grammaticizing forms may be manifested not only as a diachronic process but also as a synchronic difference between formal and informal registers. Clitic climbing frequencies in Spanish auxiliary + gerund sequences have increased diachronically as part of a series of reductive changes and in tandem with construction frequency increases. In calculating construction frequency, both the token frequency of auxiliary + gerund sequences and their frequency relative to lone-standing gerunds turn out to be important. Differences between present-day conversational and written data show that clitic position is stylistically stratified. Register differences in clitic climbing are found to be linked to construction frequency as well, suggesting that frequency effects operate on a more general synchronic level as well as on the level of particular texts. These text-level frequency effects may be related to parallel structure "birds of a feather" effects.

A generally accepted idea is that clitic climbing in Spanish periphrastic constructions correlates with the degree of grammaticization of the construction. Thus, both Lenz (1925:385,§247) and Keniston (1936:163) suggested that the position of object pronoun clitics indicates whether a verb in a construction is more of a lexical item (main verb) or more of a grammatical morpheme (auxiliary). More recently, Myhill (1988a, 1988b) found that preverbal clitic position is most strongly favored when the verb conveys epistemic, future, or progressive meaning: that is, meanings that languages tend to express by means of bound morphemes (Bybee, 1985).

One set of constructions that this idea has been applied to is *-ndo* constructions, commonly referred to as "perífrasis de gerundio" in Spanish or "progressives" in English. These are periphrastic expressions composed of a locative or motion verb (most frequently, *estar* 'be located', *ir* 'go', *andar* 'go around', *venir* 'come', *seguir* 'follow, continue') plus a gerund (*-ndo*) form. The examples in (1)

A shorter version of this article was presented at the 27th meeting of the Linguistic Association of the Southwest, Phoenix, October 1998. I would like to thank Joan Bybee and Alan Hudson (University of New Mexico) as well as three anonymous referees of this journal for helpful comments on earlier drafts. Also, thanks are due to Anthony Higgins (Duke University) for his help with the essay translations.

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

(from Myhill, 1988a) illustrate clitic climbing (CC) and lack of clitic climbing (non-CC) with *estar* + *-ndo*.

(1) a. CC
Me estoy cansando
1s-REFL am-1s get tired-GERUND
'I'm getting tired'
b. Non-CC
. . . toda la cosecha de cebada estaba asoleándo**se** en el solar
was-3s lie-in-sun-GERUND-3s REFL
'. . . the whole barley harvest was lying in the sun in the yard'

Myhill (1988a:239) proposed that CC is disfavored in (1b) because *estar* retains more of its original locative meaning, whereas in (1a) the locative sense is bleached and *estar* contributes only progressive meaning.

In this study I compare frequencies of CC in Old Spanish (c. 1200–1500), present-day conversation, and formal written texts. First, the diachronic evidence shows that CC frequencies have increased in tandem with other indices of reductive change in the form of periphrastic *-ndo* constructions. Second, the synchronic data show that CC frequencies are lower in essays and academic prose than in conversation or novels. These results suggest that formal reduction in grammaticization is manifested not only as a diachronic process but also as a difference between formal and informal registers.[1]

Recent literature has begun to pay attention to the role played by frequency in propelling the changes that occur in grammaticization. Bybee and Thompson (1997) and Bybee (forthcoming) argued that frequency of use operates in language change, in support of the view put forward by Haiman (1994) that grammatical constructions emerge as conventionalized patterns of language use through mechanisms effective more generally in ritualization. The present data show that differences in CC—both diachronic and synchronic—are linked to increases in construction frequency. What is interesting is that frequency effects may operate both on a more general synchronic level and on the level of particular texts, as suggested by register variation in CC. One mechanism through which these text-level frequency effects may operate is a parallel structure "birds of a feather" effect, such as has been found in studies of morphosyntactic variation (Labov, 1994:577–578; Pereira Scherre & Naro, 1991).

DIACHRONIC INCREASES IN CLITIC CLIMBING FREQUENCIES

Given the flexible word order of Old Spanish, the expression "clitic climbing" can actually be misleading. Rivero (1986) provided evidence that Old Spanish clitics have the same distribution as independent words, unlike their modern counterparts, which are more like bound morphemes. For example, non-tonic object pronouns in Old Spanish may be separated from the verb by intervening open class material, as in *quien te algo prometiere* 'the one who would promise something to you', in which *algo* separates dative *te* from the verb *prometiere* (Rivero, 1986:777). Even more relevant is Rivero's (1986:783) observation that, in Old

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

Spanish, not only non-tonic object pronouns but also full-blown NPs can climb in periphrastic expressions. The examples in (2) (from Rivero, 1986) with *querer* 'want' + infinitive illustrate this.

(2) a. quando … Jesuchristo **las sus divinales bodas** quisyiere celebrar (Corb)
'when Jesus **his sacred wedding** would wish to celebrate'
b. antes que **la** quería complir (Corb)
'before he **it** wanted to fulfill'

Nevertheless, I will use the traditional term "clitic climbing" to refer to the positioning of an unstressed object pronoun preposed or postposed to the auxiliary (as opposed to postposed to the gerund) to make my results more readily comparable to those of other studies.

Object clitics appear in one of three positions in Old Spanish *-ndo* constructions. Following familiar formal syntax terminology, possible landing sites for clitics that have climbed from right to left are either preposed to the auxiliary or midway between the auxiliary and the gerund. A third possibility is for the clitic to remain attached to the gerund (i.e., to the verb whose original argument it is). The examples in (3) illustrate the three clitic positions with *estar* + *-ndo*.

(3) Old Spanish *-ndo* construction clitic positions
a. Preposed CC
Desque el negro esto dixo, otro que lo oyó dixo esso mismo, e assí **lo** fueron diziendo … (Luc, XXXII, p. 216)
'once the Negro had said this, another person who heard it said the same thing, and soon everyone [**it**] was saying it …' (cf. Keller & Keating, 1977:132–133).
b. Midway CC
E […] estávan**le** esperando a media legua de aquella su casa (Luc, XVIII, p. 144)
'And those who were to kill him were **[for him]** waiting for him a half a league from his house' (cf. Keller & Keating, 1977:85)
c. No CC
… esta diziendo **le** alla su corac'on … (CO1-32r28, Act IV)
'… his heart is telling **him** there'

I counted CC cases in a corpus of Old Spanish texts spanning three centuries, beginning with the *Poema de Mio Cid* (1140/1207) and ending with the *Celestina* (c. 1500). The texts, editions, and abbreviations used are listed in Appendix 1. I also counted CC cases in a corpus of conversational Mexico City Spanish (UNAM, 1976).

Table 1 shows the positioning of clitics in the Old Spanish and Mexico City corpora. Let us begin with *estar* + *-ndo*. In the Old Spanish corpus, there were 11 cases of CC to a midway position between *estar* and the gerund—39% of all cases where CC was possible.[2] In contrast, midway clitic position was limited to infinitive *estar* + *-ndo* in the Mexico City corpus for a total of 4 tokens (3%), as in (4).

(4) Entonces, se bajan de la cama descalcitos o sucios; lo que sea, tengo que estar**los** mirando. (UNAM, 1976:430)
'Then they get out of bed barefoot or dirty; whatever the case, I have to be **[them]** watching them.'

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

TABLE 1. *Clitic position in Old Spanish and Mexico City* -ndo *constructions*

	No CC		Preposed CC		Midway CC		Overall CC[a]		Total
	N	%	*N*	%	*N*	%	*N*	%	*N*
Old Spanish (579,800 words)									
estar	8	29	9	32	11	39	20	71	28
ir	2	3	45	57	32	40	77	97	79
Mexico City (172,700 words)									
estar	12	10	99	86	4	3[b]	103	89	115
ir	4	7	49	86	4	7[b]	53	93	57

Note: Differences between Old Spanish and Mexico City in the proportion of preposed CC are significant for both *estar* and *ir* (chi-square = 40.86765 and 18.13532, respectively; $p = .0000$). Differences in overall CC frequencies (preposed and midway CC) are significant for *estar* (chi-square = 6.157023, $p = .0131$) but not for *ir* (chi-square = 1.579956, $p = .2088$).
[a]Overall CC combines preposed and midway CC.
[b]Midway CC in Mexico City limited to cases of infinitive *estar* + *-ndo* and *ir* + *-ndo*.

In addition to CC to midway position, there were 9 cases (32%) of CC preposed to *estar* in the Old Spanish data. In the Mexico City data there were 99 such cases (86%). The difference in the proportion of CC preposed to *estar* between the Old Spanish corpus and the Mexico City corpus is highly significant ($p = .0000$). Now, if we combine the cases of preposed and midway position, we have an overall CC total of 71% in the Old Spanish corpus and 89% in the Mexico City corpus. The difference in this overall proportion of CC to non-CC is also significant ($p = .0131$). In short, not only has the overall CC frequency in the *estar* + *-ndo* construction increased over time, but there has also been a change in landing site, from variation between midway and preposed position to categorical preposing for finite forms of *estar*. This result may be taken as an indication of increasing fusion of *estar* with *-ndo* in the construction.

The results for *ir* + *-ndo* are similar in the decline of CC to midway position. The Old Spanish data show a 40% frequency of the midway position, while the Mexico City data show only 7%, limited again to cases of infinitive *ir* + *-ndo*. For preposed CC, the Old Spanish data shows 57%; this frequency increases to 86% in the Mexico City data, a significant increase ($p = .0000$).

It is interesting that preposed CC is the preferred position for *ir* + *-ndo* even in the Old Spanish texts,[3] while clitics in *estar* + *-ndo* are more evenly distributed across the three positions, with a slight preference for midway CC. This may indicate the earlier grammaticization of *ir* + *-ndo*. In other words, the higher proportion of preposed CC in *ir* + *-ndo* compared with *estar* + *-ndo* suggests that this construction had already grammaticized to a (greater) extent by the time the earliest available Old Spanish texts were written. At the same time, the diachronic difference between the Old Spanish and Mexico City corpora in overall CC frequency is significant for *estar* + *-ndo* (from 71% to 89%) but not for *ir* + *-ndo* (from 97% to 93%). The synchronic difference between *ir* and *estar* in overall CC frequencies (93% and 89%, respectively) is also not statistically significant. Thus,

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

even though *ir* + *-ndo* may have gotten a head start on the grammaticization path, *estar* + *-ndo* seems to have caught up, implying a faster pace of change.[4]

CLITIC CLIMBING AS AN INDEX OF FORMAL REDUCTION

In this section we look at the evidence that supports a reformulation of the relationship between CC and grammaticization as proposed by Myhill (1988a, 1988b). The Old Spanish data suggest that mechanical factors related to co-occurring items as opposed to semantic considerations best account for the position of the clitic in *-ndo* constructions. A comparison of the Old Spanish and Mexico City data shows that CC frequencies have increased diachronically in tandem with other indices of formal reduction, such as positional fixing, a decrease of intervening material, and a decline in multiple gerund constructions. Rather than directly reflecting semantic bleaching, CC develops as part of a series of reductive changes in the form of *-ndo* constructions. That is, diachronic increases in CC do not reflect a direct correspondence between the occurrence of CC and the meaning of the periphrastic expression (grammatical vs. lexical) in any given example but rather indicate the conventionalization of auxiliary + gerund sequences as units whose components are increasingly fused.

Factors in clitic climbing

The Old Spanish examples suggest that the three factors disfavoring CC are the presence of a locative expression, intervening open class material, and a parallel gerund in what I call a multiple-gerund construction.[5] These mechanical factors, which are related to items surrounding the construction, may override semantic considerations. In the next set of examples, the main verb is the same, *catando* 'watching, looking at', yet in each case the object clitic has a different position. In (5a) it is preposed to *estar*, in (5b) it is in midway position, and in (5c) it is postposed to the main verb. In both non-preposed cases we have a parallel gerund in the construction: *catando . . . cuidando* in (5b) and *catando & marauillando* in (5c). In the latter we also have the subject pronoun *el* 'he' intervening between *estar* and the first gerund.

(5) Three clitic positions with *estar* + *catando*

a. E ataron le al cuello una muela de brac'o & echaro⟨n⟩ lo de la puente en el rio. & ando gra⟨n⟩d piec'a sobrel agua fablando con lo[^s] q⟨ue⟩ **lo(^s)** estaua⟨n⟩ catando. (EE1 111v15)
'And they tied a stone to his neck and threw him from the bridge into the river. And he walked for a long while on the water speaking to those who **[him]** were (stood there) watching him.'

b. e quando cesar la uio estudo **la** cata⟨n⟩do gra⟨n⟩d piec'a cuydando. (EE1 5v15)
'and when Cesar saw her he remained (stood there) **[at her]** looking at her for a long while, thinking'

c. E estando el catando **los** & marauillando se dellos; dixol aq⟨ue⟩lla uirgen. (EE1 144r79)
'And while he was (stood there) looking **at them** and marveling at them, that virgin said to him'

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

The meaning seems to be just as ambiguous between a more or less bleached *estar* + *-ndo* in all three examples: that is, either the subject is (merely) watching or the subject is (located) watching. It is not at all apparent that the multiple gerunds in (5b) and (5c) are compatible with lexical rather than grammatical uses of *estar* + *-ndo* or with less bleached rather than more bleached *estar* in the construction.

Support for the idea that mechanical factors related to co-occurring items provide a better explanation than meaning differences for the position of the clitic comes from the next set of examples, all from the *Celestina*. It is interesting that this text has the lowest CC frequency for *estar* + *-ndo* since it is the latest Old Spanish text. There is only 1 case of CC out of a total of 5 where CC is possible.[6] The lack of CC is explainable in light of the factors we have been discussing. Two non-CC cases have intervening open class material and a parallel gerund, as in (6).

(6) E como ella estuuiesse **suspensa:** mirando **me** espantada del nueuo me⟨n⟩saje: **escucha⟨n⟩do** fasta ver quien podia ser: el que assi por necessidad d⟨e⟩ su palabra penaua. o quien pudiesse sanar su lengua: en no⟨m⟩brando tu nombre: atajo mis palabras dio se enla frente vna grand palmada (C01-38r26, Act VI)
'And as she stood looking **at me** in amazement, surprised at this message and waiting to discover the name of the sufferer who could be cured if she but spoke—when, sir, I named your name, she cut my discourse sharply off, gave herself a great slap with the palm of her hand on her forehead' (Singleton, 1958:100)

In this example the relationship between *estar* and the gerund *mirando* is not tight at all. *Estar* appears to be more in construction with the intervening descriptive adjective *suspensa* 'amazed, hanging' than with *mirando* 'looking', which in turn appears to be more linked to the second gerund *escuchando* 'listening', as in "looking and listening." Thus, *estar* and *mirando* barely appear to form a unit. However, even though this is a non-CC case, aspectual meaning is not precluded: that is, the looking is viewed as being in progress for a brief period until the offending name is named.

The other two cases in the *Celestina* where CC fails to occur have co-occurring locatives. It is important, however, that locative meaning does not preclude temporal (aspectual) meaning. The example in (7) illustrates how *estar* + *-ndo* may have both a locative and a temporal sense. Here *estar* + *-ndo* clearly has a locative dimension, as indicated by the motion verb *voy* 'I (will) go' in the main clause and the co-occurring locative *allá* 'there': that is, the speaker will go to where the person represented metonymically by "his heart" is located (*está . . . allá*). At the same time, *estar* + *-ndo* has progressive meaning: that is, the action referred to by the gerund is in progress at speech time, as indicated by *se me figura que* 'I imagine that'.

(7) Yo **voy** co⟨n⟩ tu cordon tan alegre: que **se me figura** que esta diziendo **le alla** su corac'on la merced que nos heziste: & que le te⟨n⟩go de hallar aliuiado. (CO1-32r28, Act IV)
'I will go at once with the cord, and go so happily that I can almost imagine I hear his heart is telling **him** there of the great favor you have done us.' (cf. Simpson, 1955:58; Singleton, 1958:87)

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

The one case of CC in the *Celestina* is reproduced in (8).

(8) Pero como soy cierto d⟨e⟩ tu limpieza d⟨e⟩ s(e)[a]ngre & fechos: **me** estoy remirando si soy yo calisto a quie⟨n⟩ tanto bien se le haze. (C01-73r22, Act XII)
'but now that I am certain of your pure intentions **[reflex.me]** I am wondering whether I am really Calisto for whom such glorious things have been done!' (Simpson, 1955:132)

The meaning does not seem to be more progressive in (8) than in (7), where CC does not occur. What is different is the lack of a locative to which *estar* might be connected, or a second gerund to which the first might be connected, or intervening material separating the two. These kinds of co-occurring items pull *estar* and the gerund apart, making for a less tight connection between the two and detracting from the identification of *estar* + *-ndo* as a unit.

It turns out that such co-occurring items have declined over time. CC frequency increases have occurred in tandem with decreases in both multiple-gerund constructions and intervening material (and co-occurring locatives, as we shall see). In the next two sections we look at indices of reductive change in the form of *-ndo* constructions.

Multiple-gerund constructions

In what I am calling multiple *-ndo* constructions, two gerunds are conjoined explicitly by *y* (Old Spanish *e*) 'and' or *o* 'or', as in (9a), or without a conjunction, as in (9b). In either case the two gerunds are symmetrically coordinated or parallel, in that each one could be construed as being in construction with the auxiliary. This is reflected in the use of a *be* + *-ing* construction in the English translation ('were reposing') of the parallel gerund *yaziendo* in (9b).

(9) Old Spanish multiple *estar* + *-ndo* constructions
a. With a conjunction
E estaua **fabla⟨n⟩do** ante tod el pueblo **& falagando** los. (EE1 74v32)
'And he was **speaking** before all the people **and flattering** them.'
A cabo de algunos días, unos omnes estavan **riendo e trebejando** e escribían todos los omnes que ellos conosçían, cada uno de quál manera era, e dizían (Luc, Ex. XX, p. 153)
'In a few days' time men were **laughing and joking** about the matter and decided to write down a list of the names of all the people they knew together with their qualities' (Keller & Keating, 1977:91)
b. Without a conjunction
Com⟨m⟩o ladron venjste de noche a lo escuro
estando nos **dormjendo yaz'jendo** nos sseguro (LBA, 1192)
'You came like a thief in the night, in the dark,
while we **were asleep and were reposing** free from care' (Willis, 1972:324)
Siempre están **fablando, librando** cosas agenas (Corb, Segunda parte, Cap. XII, p. 195)
'They were always **gabbling and poking** their noses into other people's business' (Simpson, 1959:153)

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

TABLE 2. *Multiple-gerund constructions as a proportion of all* -ndo *constructions*

	All *-ndo* Constructions		Multiple Gerunds		Proportion of All *-ndo* Constructions	
	estar	*ir*	*estar*	*ir*	*estar*	*ir*
Old Spanish	167	230	21	28	13%	12%
Mexico City	340	146	2	2	0.6%	1.4%
Essays	67	101	5	10	7%	10%

Notes: Proportion of multiple-gerund constructions is significantly higher in Old Spanish than in Mexico City for both *estar* ($z = 6.0955$, $p = .0000$) and *ir* ($z = 3.7681$, $p = .0001$). Proportion of multiple-gerund constructions is significantly higher in Essays than in Mexico City for both *estar* ($z = 3.9558$, $p = .0000$) and *ir* ($z = 3.0660$, $p = .0011$). Differences in proportion of multiple-gerund constructions between Essays and Old Spanish are not significant.

In counting multiple-gerund constructions I included cases where the gerunds are separated by a location or manner adverbial, as in the first example in (9a), or by an NP subject or object. What is important is that in multiple-gerund constructions there is no one-to-one relationship between the auxiliary and the *-ndo* form, since the same auxiliary appears to be in construction with two or more gerunds, each one vying for a connection with the auxiliary.

In the 13th century *Estoria de España* alone, 13% (5 of 39) of *estar* + *-ndo* tokens are multiple-gerund constructions, either with or without an explicit conjunction. For *ir* + *-ndo* the average is 11% (13 of 114) and for *andar* + *-ndo*, 17% (12 of 64). Overall, in the Old Spanish corpus, there is a total of 21 cases where a second gerund joins *estar* + *-ndo* and 28 such cases for *ir* + *-ndo*. As shown in Table 2, the average occurrence of these multiple-gerund constructions in the Old Spanish data is 13% for *estar* and 12% for *ir*. That is, of all auxiliary + gerund sequences, 13% in the case of *estar* and 12% in the case of *ir* have a second (if not third) gerund conjoined to the gerund immediately following the auxiliary.

In contrast, there are only 2 multiple *estar* + *-ndo* tokens in the Mexico City corpus (UNAM, 1976:16, 126), less than 1% of all tokens of *estar* + *-ndo*. There are 2 examples (about 1.4%) for *ir* + *-ndo* (UNAM, 1976:138, 422). Even (10) is not a very good example of a multiple *-ndo* construction, since *subir y bajar* 'go up and (go) down' may be said to form a routine unit (i.e., 'go up and down').

(10) la presión que se siente en … de que ha de estar **subiendo y bajando** (UNAM, 1976:16)
'the pressure you feel in … that you have to be **moving up and down**' (when you play center position in soccer)

A better present-day example of a multiple-gerund construction is (11).

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

(11) Pues el chofer se jalaba, y aquellos se iban hasta **riendo, comiéndose** los tamales. (UNAM, 1976:422)
'Well the driver would start the bus and they would go off **laughing, eating up** the tamales.'

The decline in the proportion of multiple gerund constructions between the Old Spanish and Mexico City corpora is statistically significant for both *estar* + *-ndo* and *ir* + *-ndo*. What we see instead in the Mexico City data is a tendency to repeat the auxiliary for each *-ndo* form. The redundant use of *estar* and *ir* in examples (12) and (13) may itself be taken as an indication of the grammaticization of the constructions as aspectual markers (cf. Bybee, Perkins, & Pagliuca, 1994:8). Example (12) is a description of shoe sewing; the subject of the series of *ir* + *-ndo* constructions is a needle.

(12) O sea que **va** agarrando y **va** … enudando en cada pisada. **Va** enudando así, ¿entiende? O sea **va** agarrando así, lo **va** dejando … (UNAM, 1976:39)
'It **is** taking hold and it **is** … knotting at every step. It **is** knotting like this, see? You know, it **is** taking hold like this, it **is** leaving it …' (lit. **goes** taking hold, **goes** knotting, **goes** leaving).

(13) Además, en cualquier época, pues **estar** trabajando y **estar** estudiando, no se puede. (UNAM, 1976:253)
'Also, it's always been the case, well, to **be** working and to **be** studying, it can't be done.'

As we shall see, an important difference in present-day varieties is a higher proportion of multiple-gerund constructions in formal texts than in conversation.

Intervening material and positional fixing

In addition to the decline in multiple-gerund constructions, there has been a diachronic reduction in intervening material. As shown in Table 3, there has been a change in the average occurrence and kind of material separating the auxiliary and the gerund. The relative frequency of intervening open class items, such as subject NPs or locative, temporal, and manner expressions, decreases for both *estar* + *-ndo* and *ir* + *-ndo*. Intervening open class items drop from 25% of all tokens of *estar* + *-ndo* in the Old Spanish data to 7% in the Mexico City data. Even more striking is the drop in two-item intervening material. The total number of cases of *estar* + (x) + (y) + *-ndo* in the Old Spanish corpus is 17, or 10% of the total. There are only 3 cases in the Mexico City corpus, equal to 1% of the total. The results for *ir* + *-ndo* are similar. The relative frequency of *ir* + *-ndo* constructions with intervening open class items decreases from 22% in the Old Spanish data to 8% in the Mexico City data. There are no cases at all of two intervening items in the Mexico City corpus, compared with 8% in the Old Spanish corpus.

Viewed from the opposite perspective, the proportion of *estar* + *-ndo* and *ir* + *-ndo* not separated by any kind of intervening material increases from 58% to 86% and 68% to 89%, respectively ($p < .0001$). The lower probability of intervening open class items indicates a higher degree of fusion between an emerging

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

TABLE 3. *Intervening material with* estar + -ndo *and* ir + -ndo

	None		Closed Class[a]		Open Class[b]		Total
	N	%	*N*	%	*N*	%	*N*
Estar							
Old Spanish	96	58	27	16	42	25	165
Mexico City	292	86	25	7	23	7	340
Essays	62	92	2	3	3	4	67
Ir							
Old Spanish	187	68	27	10	61	22	275
Mexico City	130	89	5	3	11	8	146
Essays	96	95	0	0	5	5	101

Note: Differences between Old Spanish and Mexico City in the proportion of zero intervening material and in the proportion of open class items (including cases of "two items") are significant for both *estar* + *-ndo* and *ir* + *-ndo* ($p < .0001$ in all cases). Differences between Essays and Mexico City are not significant.

[a]Closed class items were subject and object pronouns as well as deictic locatives *y* (Old Spanish), *aquí*, *allí*.

[b]Open class items were mostly subject NPs, locatives, temporals, and manner adverbials. Includes cases of two such items. In Old Spanish: 17 (10%) for *estar*, 21 (8%) for *ir*; in Mexico City: 5 (1%) for *estar*, 0 for *ir*.

grammatical morpheme and the verb it attaches to in the construction (Bybee et al., 1994:113).

A third important change in the form of *-ndo* constructions is positional fixing. In the 12th century *Poema de Mio Cid* there is positional variation between *-ndo* forms postposed and preposed to the auxiliary, as the following examples illustrate. In (14a) *catando* is postposed to *estar*, and in (14b) it is preposed. In (15a) *pesando* follows *ir*, and in (15b) it precedes *ir*.

(14) a. Tornaua la cabec'a & **estaua los catando** (0002)
'he turned his head and stood looking at them' (Hamilton & Perry, 1975)
b. **Catando estan** a myo c'id q⟨u⟩a'ntos ha en la cort (3123)
'everybody in the court was gazing at My Cid' (Such & Hodgkinson, 1987)

(15) a. Alos vnos plaze & alos otros **va pesando** (1837)
'Some of them were pleased, and others dismayed' (Such & Hodgkinson, 1987)
b. Hya va el mandado por las t⟨ie⟩rras todas **Pesando va** alos de monc'on & alos de huesca (940)
'News of this raid spread through the country round about, filling the inhabitants of Monzon and Huesca with dismay' (Hamilton & Perry, 1975)

The average number of preposed *-ndo* forms in *Poema de Mio Cid* is 20% for *estar* and 35% for *ir*.[7] It is interesting that preposing seems to be a routine with certain main verbs, as in *passando ua la noch* 'night passes' and *tornando va la cabeza* 'turns his head'. Despite routinized preposing with certain expressions, in most cases there seems to be no pattern, as shown in (14) and (15). The variable

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

position of the gerund with respect to the auxiliary in *Poema de Mio Cid* indicates that these two elements are combining freely as independent lexical items.

After *Poema de Mio Cid*, there are no tokens of preposed *-ndo* in the other medieval texts surveyed, with the exception of the *Corbacho* and *Celestina*. The position of the gerund with respect to the auxiliary seems to have been pretty well fixed by the end of the Old Spanish period. Keniston (1937a:105) reported that preposing is rare in 16th century prose. In the Mexico City corpus, there are no preposed gerunds: that is, postposing is categorical. As free combinations of lexical items become periphrastic expressions, the word which becomes grammatical comes to have a fixed position. The change from a variable to a fixed position for the auxiliary with respect to the gerund may be taken as another index of loss of autonomy (cf. Bybee et al., 1994:110).

In summary, we see a series of diachronic changes in the form of *-ndo* constructions—the fixing of the position of the auxiliary with respect to the gerund, the decrease in intervening open class material, and the decline of multiple-gerund constructions—all of which contribute to the tighter fusion between the auxiliary and the gerund. The claim here is that the diachronic increase in the frequency of CC is part of this series of formal changes that contribute to the emergence of *estar* + *-ndo* as a unit. That is, the increase in CC is not due to semantic bleaching of *estar* alone so that there is a direct correspondence between the occurrence of CC and the meaning (aspectual vs. locative) of *estar* in any given example. Rather, over time the string (locative) + *estar* + (x) + (y) + *-ndo* (+ *-ndo*) has been conventionalized as the *estar* + *-ndo* construction, increasingly without intervening material (Table 3), without a parallel or closeby gerund (Table 2), and without co-occurring locatives (see Table 9), all of which would detract from the identification of *estar* + *-ndo* as a unit. As *estar* + *-ndo* emerges as a unit in and of itself, the clitic is increasingly preposed to *estar*, as is categorically the case with finite verb forms in present-day Spanish. We can represent the emergence of the *estar* + *-ndo* unit as follows:

(locative) + *estar* + (x) + (y) + *-ndo* (+ *-ndo*)

⇓ decline of { co-occurring locatives; intervening material; parallel gerunds }

estar + *-ndo*

→ object clitics preposed to unit

Loss of internal constituent structure is common in grammaticizing constructions (Bybee & Thompson, 1997; Heine, Claudi, & Hünnemeyer, 1991). In generative syntax, CC (or clitic promotion) has been analyzed as the result of a structure-reducing transformation in the complement clause (e.g., Luján, 1980; Rivero, 1970) or as a restructuring rule that reanalyzes a series of verbs into a single verbal complex, thus transforming two underlying clauses into one (e.g., Rizzi, 1982). The change in CC frequencies between Old Spanish and present-day Spanish shows a downgrading of constituent boundaries from two clauses to

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

one, as locative *estar* and directional motion *ir* go from being full-blown verbs with a verbal complement to becoming more auxiliary-like and the *-ndo* form becomes more tightly attached to these auxiliaries. Thus, we may view CC as an index of reductive change.

THE ROLE OF CONSTRUCTION FREQUENCY

So far the comparison of the Old Spanish and Mexico City data shows that CC frequencies have increased as part of a series of formal changes contributing to the emergence of *-ndo* constructions as fused units. Recognizing the important role of repetition in language, Bybee (forthcoming) proposed a new definition of grammaticization, more traditionally defined as "the process whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions" (Hopper & Traugott, 1993:xv). Based on a review of the effects of token and type frequency in semantic, phonological, and morphosyntactic change, she characterized grammaticization as "the process by which a frequently-used sentence of words or morphemes becomes automated as a single processing unit" (Bybee, forthcoming). The increase in CC frequencies points to this kind of process of automatization for *-ndo* constructions.

In this section we look at how reductive changes in form are linked to an increase in the token frequency of auxiliary plus *-ndo* sequences. We also see that token frequency has increased in tandem with an increase in the proportion of gerunds in construction with an auxiliary: that is, there has been an increase in the frequency of gerunds preceded by an auxiliary relative to those gerunds that are lone-standing or in multiple-gerund constructions. This double frequency change—absolute and relative token frequency—has contributed to the emergence of *-ndo* constructions as more fused and more automatized units.

Let us look first at the dramatic increase in the token (or text) frequency of *-ndo* constructions. Table 4 indicates that the token frequency of *-ndo* constructions has increased over time from an average of 10.6 occurrences per 10,000 words in the Old Spanish corpus to 37.5 in the Mexico City corpus. In other words, these constructions occur more than three times more frequently in the latter than in the former. True, some of the frequency difference between the Mexico City and Old Spanish corpora may be attributed to differences between conversational data and written texts in any given period, however faithfully Old Spanish works like the *Corbacho* and the *Celestina* reproduce popular speech of the time. We return to considerations of register in the next section. However, figures from Keniston (1937a, 1937b), who studied comparable corpora of 16th and early 20th century prose, leave no doubt that there is a diachronic side to the increase in text frequency. These figures show that the token frequency of *-ndo* constructions has doubled from an average of 8.3 occurrences per 10,000 words of text in the 16th century to 16.4 occurrences per 10,000 words of text in the early 20th century.[8]

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

TABLE 4. *Token frequencies of* -ndo *constructions per 10,000 words*

Corpus	Word Count[a]	Total *-ndo* Constructions[b]	Average Frequency
Old Spanish	579,800	616	10.6
16th century prose	300,000	248	8.3
20th century prose	600,000	987	16.4
Mexico City	172,700	648	37.5
HablaCulta	1,300,000	4,409	33.9
Essays	376,300	278	7.4

[a]For Old Spanish, the word count is from the Microsoft Word word count of the electronic versions of the texts (PMC, EE1, LBA, and C01), except for the Lucanor and Corbacho, which are from Davies (1998). For 16th and 20th century prose, see Keniston (1937a, 1937b). For Mexico City and HablaCulta, see Clegg and Rodríguez (1993). For the Essays word count, see Appendix 3 and note 12.
[b]All tokens of *estar*, *ir*, *andar*, *venir*, *quedarse*, *seguir* (only 1 token in Old Spanish), and *continuar* (no tokens in Old Spanish), *salir* (only 1 token in Mexico City) plus a gerund were counted.

Concurrent with the diachronic increase in the token frequency of *-ndo* constructions is a shift in the proportion of *-ndo* forms in construction with an auxiliary. Spanish gerunds not in construction with an auxiliary may be used as manner adverbials, as in *Se pasan la mañana* **hablando** por teléfono 'They spend the morning **talking** on the phone'. Adverbial uses of *-ndo* forms, including manner, temporal, conditional, and concessive uses, add up to 84% of all tokens of lone-standing gerunds in a corpus of present-day educated Mexico City Spanish (Luna, 1980:95–119). Spanish gerunds may also be used as nonrestrictive relatives referring to the subject or, with certain predicates, the direct object of the finite verb, as in *¿A dónde hay mayor porcentaje de mujeres* **trabajando?** 'Where is there a bigger percentage of women **working?**' (Luna, 1980:108).[9] I subsume all these uses under the label "lone-standing gerunds," in opposition to auxiliary + *-ndo* constructions.

Table 5 shows that the proportion of gerunds in construction with *estar*, *ir*, *andar*, or another auxiliary has increased from about 21% in the Old Spanish corpus to 70% in the Mexico City corpus. Conversely, the proportion of *-ndo* forms standing alone has decreased from 79% to 30%. The increase in the frequency of *-ndo* constructions relative to lone-standing gerunds is highly significant. Again, part of the difference must be attributed to register differences, but Keniston's (1937a, 1937b) figures for 16th and 20th century prose, which show a 10% increase in the proportion of *-ndo* constructions with respect to lone-standing gerunds, indicate that there is a diachronic side as well.

In his study of emerging English auxiliaries, Krug (1998), using the term "string frequency" to refer to the frequency of the sequence of a verb (e.g., *want* or *got*) followed by *to* + infinitive, found that string frequency is related to indices of grammaticization, specifically to contraction of forms (e.g., *wanna* and *gotta*). However, string frequency is not readily applicable to *-ndo* constructions. First,

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

TABLE 5. *Proportion of gerunds (*-ndo *forms) in construction with an auxiliary*

	-ndo Constructions		Lone-standing *-ndo*		Total *-ndo*
Corpus[a]	*N*	Average	*N*	Average	*N*
Old Spanish	616	21%	2,320	79%	2,936
16th century prose	248	14%	1,579	86%	1,827
20th century prose	987	24%	3,135	76%	4,122
Mexico City	648	70%	277	30%	925
HablaCulta	4,409	65%	2,398	35%	6,807
Essays	133	28%	340	72%	473

Note: The proportion of gerunds in construction with an auxiliary is significantly higher in Mexico City (70%) than in Old Spanish (21%) ($z = 27.7353$, $p = .0000$), in Mexico City than in Essays (28%) ($z = 14.9407$, $p = .0000$), and in Essays than in Old Spanish ($z = 3.4792$, $p = .0003$). The increase between 16th century and 20th century prose (figures from Keniston, 1937a, 1937b) is also significant ($z = 9.0974$, $p = .000$).

[a]Figures for Old Spanish and Mexico City were obtained with a manual count and Essays, with MonoConc (Barlow, 1997). Figures for 16th and 20th century prose were calculated from Keniston (1937a:469, 552–557; 1937b:207–208, 239–242). Figures for HablaCulta are from Clegg and Rodríguez (1993), except that I calculated the figure for lone-standing gerunds by subtracting the figures given for *estar* and other auxiliaries from the given gerund total.

these constructions have an open class verb slot between the auxiliary and the *-ndo* suffix, and so there is not really a single string to count, as in the case of *want to*. Second, the auxiliaries *estar*, *ir*, and *andar* participate in other constructions: for example, the resultative *estar* + past participle construction or the future *ir* + *a* + infinitive construction.

In the case of *-ndo* constructions, it seems appropriate to measure string frequency backwards, from the perspective of *-ndo* rather than the auxiliary, since the only constructions *-ndo* participates in are with *estar*, *ir*, *andar*, and maybe a few other verbal auxiliaries. That is, we can count the tokens of *-ndo* preceded by *estar* or another auxiliary as compared with the number of tokens of lone-standing *-ndo*. We can think of it in terms of transition probabilities in reverse: that is, right-to-left instead of left-to-right transition probabilities. As Table 5 shows, the proportion of lone-standing gerunds declines between the Old Spanish data and the Mexico City data. The decrease in the proportion of lone-standing gerunds means an increase in the probability that a gerund is in construction with *estar*, *ir*, *andar*, or another auxiliary. In addition, the decline in multiple-gerund constructions shown in Table 2 means an increase in the probability that a gerund and an auxiliary are not only associated in a construction, but also bound in a one-to-one relationship, as opposed to multiple gerunds competing for ties to the auxiliary.

I call the increase in the token frequency of the auxiliary + gerund sequences—in combination with the decrease in the proportion of lone-standing gerunds and multiple-gerund constructions—an increase in "construction frequency." It is important that construction frequency is calculated by counting

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

the tokens of the auxiliary + gerund sequence as well as the tokens of gerunds not in such sequences. Increased construction frequency makes for a higher probability that a gerund is indeed tied to an auxiliary. The tighter connection between the auxiliary and *-ndo* means that increasingly the sequence is identified as a fused unit.

The significance of construction frequency, as we have defined it based on the present data, is that loss of autonomy (Bybee et al., 1994:110) may affect both elements of an emerging construction. The auxiliary loses autonomy as does the *-ndo* form, as shown by the decline in lone-standing gerunds. In other words, it is increasingly likely that when a gerund appears it is associated with an auxiliary. Thus, rather than a "morphologization" of the auxiliaries (cf. Hopper & Traugott, 1993:130), the formal change is better described as a fusion of both elements (the auxiliary and the gerund) or as their automatization as a single processing unit (Bybee, forthcoming; cf. Haiman, 1994).

The idea of construction frequency is also the key to our reformulation of the relationship between CC and grammaticization. The Old Spanish evidence does not contradict Myhill's (1988a, 1988b) conclusion that CC frequencies correlate with semantic bleaching of the auxiliary in periphrastic expressions. But in showing that CC is part of a series of reductive changes in the form of *-ndo* constructions, we put forward a stronger statement about the role of form and its relationship to meaning in grammaticization. The difference lies in the role attributed to construction frequency. Specifically, the diachronic evidence shows that higher CC frequencies correspond to higher construction frequencies.

REGISTER DIFFERENCES IN CLITIC CLIMBING

Biber (1986) demonstrated that the distribution of linguistic features in English (e.g., contractions, passives, nominalizations) patterns along certain textual characteristics or "dimensions" rather than along a written/spoken dichotomy. Specifically, he found that fiction is closer to conversation than to academic prose or planned speeches along the dimension identified as "abstract versus situated content," which corresponds to formal versus colloquial style (Biber, 1986:396, 399). It turns out that CC in *-ndo* constructions is stratified similarly along a formal versus informal dimension.

I counted CC cases in a corpus of essays, formal speeches, and scholarly writings by Mexican authors (Appendix 2). This corpus is referred to as Essays. In approximately 1,350 pages of text, I found 95 *-ndo* constructions where CC is possible. CC occurs in 50 tokens, yielding a CC frequency average of 53%. In contrast, CC frequency for *-ndo* constructions averages 88% in the Mexico City corpus (168 cases of CC out of 192 possible cases).[10] A VARBRUL analysis of the Mexico City data and Essays data using the IVARB program (Pintzuk, 1986) indicates that the occurrence of the token in one as opposed to the other corpus is an important factor in determining clitic position. The difference between corpora was the first factor group selected in the IVARB step-up procedure, before co-

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

TABLE 6. *Factors in* -ndo *construction CC (*IVARB *weights)*

Number of tokens $N = 287$, chi-square/cell = .861	
Input probability = .85, significance = .005	
Application value: CC occurs	
Corpus	
Mexico City	.68
Essays	.17
Co-occurring material	
No intervening material or co-occurring locative	.56
Co-occurring locative	.33
Intervening material	.04
Auxiliary	
estar, ir	.59
seguir	.35
andar, venir	.19
Clitic syntax/person	
Impersonal *se* and *me, te, nos*	.70
Le(s), lo(s), and reflexive *se*	.40

Note: Two knockout factors (100% non-CC): both intervening material and co-occurring locative (in factor group co-occurring material); *continuar* (in factor group auxiliary).

occurring material, differences between auxiliaries, or differences between the syntax/person of the clitic.

The results are summarized in Table 6. IVARB values are probabilistic weights for the variable consisting of CC as one variant and non-CC as the other. Values above .50 indicate that the independent constraint favors the occurrence of CC, values below .50 indicate a disfavoring effect, and values around .50 indicate the absence of an effect. From Table 6 we see that CC is most likely in the Mexico City corpus (.68), in the absence of intervening material and a co-occurring locative (.56), when the auxiliary is *estar* or *ir* (.59) and when the clitic is either an impersonal *se* or a 1st or 2nd person object pronoun (.70).

The results are generally in line with what we would expect. First, with respect to co-occurring material, the IVARB weights indicate that intervening material (.04) and, to a smaller degree, co-occurring locatives (.34) highly disfavor CC. This is the pattern we see in the Old Spanish data. Second, CC is more likely with *estar* and *ir* than with *andar* and *venir* (.19) or *seguir* (.35). Myhill (1988a:238) found that *estar* + *-ndo* and *ir* + *-ndo* have higher CC frequencies than other *-ndo* constructions. Third, 1st and 2nd person clitics and impersonal *se* are more likely to climb than are either direct and indirect object 3rd person clitics or reflexive *se* (.40). Myhill (1988a:241ff.) found that CC is favored when the clitic is higher on an animacy hierarchy than the subject.[11]

Most important for my purposes here, the results indicate that the Essays corpus highly disfavors CC, with a weight of .17. The low IVARB weight for the Essays data is all the more striking, given that many of the contexts of occurrence of the variable in this corpus should favor CC. There are relatively fewer cases of

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

TABLE 7. *Comparison of CC in five corpora*

	Estar	*Ir*	*Andar*	*Seguir*
Old Spanish	71% (20)[a]	97% (77)	90% (9)	
Mexico City	89% (103)	93% (53)	75% (21)	80% (4)
MexCult	89% (16)	93% (13)		80% (4)
Novels	89% (75)	86% (44)	71% (5)	36% (8)
Essays	68% (15)	45% (15)		40% (8)

Old Spanish = approximately 570,000 words from seven texts (Appendix 1)
MexCult = approximately 100 pages from UNAM (1971), which contains the transcriptions of interviews with speakers of educated Mexico City Spanish (cf. Lope Blanch, 1986)
Mexico City = approximately 450 pages or 170,000 words from UNAM (1976), the popular Mexico City Spanish counterpart to UNAM (1971)
Essays = a corpus of essays by Mexican authors totaling about 1,300 pages (Appendix 2)
Novels = figures for 20th century novels from Myhill (1988b:238, 240)

Note: Differences between Essays and other present-day varieties are significant for *estar* and *ir* (at $p < .01$) but not for *seguir*.
[a]Figures in parentheses are the absolute number of CC cases.

intervening material and co-occurring locatives in the Essays corpus than in the Mexico City corpus, as shown in Tables 3 and 9 respectively.

Table 7 compares CC frequencies in five different corpora. CC frequencies in Mexico City (UNAM, 1976), MexCult (a corpus of interviews with speakers of educated Mexico City Spanish; UNAM, 1971), and Novels (a corpus of 20th century literature; Myhill, 1988a, 1988b) are higher than in Essays for *estar* + *-ndo* and *ir* + *-ndo* but not for *seguir* + *-ndo*. For *estar* + *-ndo*, there is a significant difference between average CC frequencies at 89% in Mexico City, MexCult, and Novels and those at 68% in Essays. Indeed, the proportion of CC in Essays is not significantly different from the proportion of CC in Old Spanish. For *ir* + *-ndo*, CC frequencies at 86% to 97% are significantly higher in all data sets, including Old Spanish, than in Essays at 45%. The lower CC frequencies in essays and academic prose than in either conversation or novels confirm that register is an important factor in CC.

Given the putative relation between CC frequency and semantic bleaching (Keniston, 1936; Lenz, 1925; Myhill, 1988a, 1988b), the obvious question is, do lower CC frequencies mean that *estar* + *-ndo* and *ir* + *-ndo* are less bleached in formal than in informal registers? To answer this question, we can measure the degree of bleaching or semantic reduction by the proportion of *-ndo* constructions with co-occurring locatives and inanimate subjects. My reasoning is as follows. If *-ndo* constructions have evolved diachronically from location or motion expressions to periphrastic expressions of aspectual meaning, we would expect the average number of *-ndo* constructions with a co-occurring locative and a human subject to decline over time, as features of the original locative or motion meaning are lost. A comparison of Old Spanish and Mexico City corpora confirms this expectation. Table 8 shows the results for co-occurring locatives and

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

TABLE 8. *Co-occurring locatives with* estar + -ndo *and* ir + -ndo

	Estar			*Ir*		
	Locatives	*N*	Average Locatives	Locatives	*N*	Average Locatives
Old Spanish	49	167	30%	70	248	28%
Mexico City	65	340	19%	24	146	16%
Essays	8	67	12%	5	101	5%

Note: For *estar*, the average number of co-occurring locatives is significantly higher in Old Spanish than in Mexico City ($z = 2.6671$, $p = .0038$), but the difference between Essays and Mexico City is not significant. For *ir*, the average number of co-occurring locatives is significantly higher in Old Spanish than in Mexico City ($z = 2.6512$, $p = .0040$) and in Mexico City than in Essays ($z = 2.7574$, $p = .0058$).

TABLE 9. *Subject animacy with* estar + -ndo *and* ir + -ndo

	Estar		*Ir*	
	Inanimate Subjects	Average Inanimate	Inanimate Subjects	Average Inanimate
Old Spanish	0	0%	35[b]	14%
Mexico City	28[a]	8%	44	30%
Essays	16	24%	43	43%

Note: The proportion of inanimate subjects is significantly higher in Essays than in Mexico City for both *estar* ($z = 3.7695$, $p = .0001$) and *ir* ($z = 2.01118$, $p = .0221$). Mexico City has a higher proportion of inanimate subjects than Old Spanish for *ir* ($z = 3.8367$, $p = .0001$).
[a]Includes 10 cases of inanimate subject with a human dative (e.g., *me está pasando algo a mí* 'Something is happening to me') (UNAM, 1976:75). There are no such cases in Essays.
[b]Includes cases of inanimate subjects metonymic for a human and cases with a human dative. These total 13 in Old Spanish and 6 in Mexico City.

Table 9, for subject animacy. For *estar* + *-ndo*, co-occurring locatives decline significantly from an average of 30% in the Old Spanish corpus to 19% in the Mexico City corpus. Human subjects likewise decrease, or conversely inanimate subjects increase, from none in Old Spanish to 8% in Mexico City. For *ir* + *-ndo*, the average number with co-occurring locatives declines from 28% in Old Spanish to 16% in Mexico City. There has also been a doubling of inanimate subjects from 14% in Old Spanish to 30% in Mexico City.

Now, if *-ndo* constructions are somehow less bleached in formal texts than in conversation, we would expect more co-occurring locatives and fewer inanimate subjects in formal texts, in line with the diachronic evidence, but the results indicate just the opposite. Diachronic changes in the average number of co-occurring locatives and inanimate subjects are echoed synchronically in differences between Mexico City and Essays. However, it is not the conversational data that show higher indices of semantic bleaching but rather the essays and academic

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

prose. Let us look first at *estar* + *-ndo*. The difference between Essays (12%) and Mexico City (19%) in the proportion of *estar* + *-ndo* with a co-occurring locative is not significant, and so the two data sets are about the same on that count. However, the proportion of inanimate subjects is significantly higher in Essays (24%) than in Mexico City (8%). For *ir* + *-ndo*, the Essays corpus has a lower locatives average (5%) than the Mexico City corpus (16%). Essays also has a higher inanimate subjects average (43%) compared with Mexico City (30%).

Of course, it is likely that the proportion of inanimate subjects in the essays and academic prose is generally higher than in conversation. The higher proportion of inanimate subjects with *-ndo* constructions in Essays may well reflect the generally higher proportion of inanimate subjects in formal registers. Even so, it would still be the case that *estar* + *-ndo* and *ir* + *-ndo* are sufficiently bleached to be compatible with the kinds of inanimate subjects we may expect to find in these texts, such as culture in (16), disdainful attitudes in (19), homeland in (20), and fear in (21). Similarly, the lower average of locatives may be partly attributed to the less frequent occurrence of deictic expressions in formal registers than in conversation (nearly half the tokens of locatives in Mexico City are the deictics *aquí* 'here' and *allí* 'there'). Nevertheless, it would also be the case that on this count *estar* + *-ndo* and *ir* + *-ndo* are at least as bleached in the Essays data as in the Mexico City data.

The following set of examples illustrates typical uses of *estar* + *-ndo* in the Essays corpus. Present tense *estar* + *-ndo* is used to indicate not only that a situation is ongoing in the present (progressive aspect), but also that it is a new, noteworthy one in contrast with a past state of affairs. In (16) the subject is *la cultura* 'culture', and *está pidiendo* 'is looking for' indicates a situation that is ongoing in the present, with the further nuance that the situation represents a change with respect to the norm up to now. This nuance is reinforced by the adjective *nuevas* 'new'.

(16) La cultura occidental **ha ido** de crisis en crisis **salvándose** unas veces en las ideas, otras en Dios, otras en la razón, hasta nuestros días en que se ha quedado sin ideas, Dios y razón. La cultura **está pidiendo** nuevas bases sobre las cuales apoyarse. (Leopoldo Zea, En torno a una filosofía americana. In Martínez, 1958:294)
'Western culture **has gone** from crisis to crisis **finding salvation** sometimes in ideas, at other times in God, at other times in reason, up to the present, when it has been left without ideas, God, or reason. Culture **is looking** for new foundations on which to stand.'

Another common use of *estar* + *-ndo* is to express a subjective attitude toward the situation, as in (17) where the writer expresses dismay (the *Mañanitas* is the traditional song for birthdays and other celebrations in Mexico).

(17) Mi voto incondicional por las tradiciones se tambalea. **¿Cómo es posible?** Y sí, sí lo es. La congregación **está leyendo** la letra de "Las Mañanitas," en papelitos obsequiados a la entrada. (Monsiváis, 1955:44; both emphases in original)
'My unconditional vote in favor of tradition is wobbling. How can this be possible? But yes, yes it is. The congregation **is reading** the words of "The Mañanitas," from little pieces of paper handed out at the door.'

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

There are also 8 cases of 1st person present *estoy* + *-ndo* in the authors' procedural comments to the reader, mostly with verbs of speech such as *diciendo* 'saying' and *comentando* 'commenting', as in (18). This use accounts for 12% of all the tokens of *estar* + *-ndo* in the Essays corpus.

(18) Al centrar nuestro análisis en la cultura visual, especialmente en las artes plásticas, **estamos queriendo** demostrar la pérdida de autonomía simbólica . . . (García Canclini, 1990:60)
'In focusing our analysis on visual culture, especially on the plastic arts, we **are seeking** to demonstrate the loss of symbolic autonomy . . .'

The most common use of *ir* + *-ndo* in Essays is to indicate not physical motion but a gradually developing situation toward a culminating endpoint. This use is illustrated in (16), where the inanimate subject *la cultura occidental* 'Western culture' does not move in space but undergoes a process. The sense of gradualness is made explicit by *de crisis en crisis* 'from crisis to crisis', while progression toward an endpoint is made explicit by *hasta nuestros días* 'until our days'.

Let us now look at the position of clitics. Examples (19) and (20) present clear cases of aspectual—rather than more locative or motion—uses.

(19) Pero, en cambio, la gente de los balcones y la de los coches y la de los autos y la de los caballos con arreos domingueros, sólo veía a los manifestantes con asomos de incredulidad o con notorias muestras de desprecio. Para ésos—así **estaban proclamándolo** sus actitudes desdeñosas—nada común existía entre ellos y el rudimentario acto cívico que se desarrollaba a su vista (Castellanos, 1966:92)
'But, in contrast, the people on the balconies and in the carriages and automobiles and on horses with their best Sunday harness, only viewed the demonstrators with looks of incredulity or clear expressions of scorn. For them—so their disdainful attitudes **were proclaiming [it]**—they had nothing in common with the basic civic action that was developing in front of them'

(20) . . . tienen ya conciencia—profética—de la patria inminente que **está gestándose** en las entrañas de la Nueva España. (Gabriel Méndez Plancarte, Humanistas del siglo XVIII. In Martínez, 1958:163)
'. . . they are already conscious—prophetically—of the looming homeland that **is gestating** [reflexive ***se***] in the entrails of New Spain.'

In (19), the subject 'disdainful attitudes' is inanimate or at best metonymic for human agents, and there is little, if any, vestige of locative meaning. In (20), the subject 'homeland' is again inanimate, though it could be considered a case of personification. We have the locative expression 'in the entrails of New Spain', but *está gestándose* 'is gestating' clearly expresses moment-of-speech ongoingness. Yet both are non-CC cases, pace Myhill (1988a, 1988b).

In summary, neither the *estar* + *-ndo* nor the *ir* + *-ndo* construction appears to be less bleached in the Essay corpus than in the Mexico City corpus, suggesting that semantic reduction has kept the same pace in formal and informal registers. But there seems to be a split between form and meaning, in that CC frequencies in Essays have not kept pace with Mexico City. Register variation in Spanish clitic position is interesting with respect to the parallel reduction hypothesis (Bybee et al., 1994:106ff.): grammaticization of form (formal reduction) may lag

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

behind grammaticization of function (semantic reduction) in more formal registers. If confirmed by studies of other variables, this result could be useful for the study of grammaticization processes in the absence of diachronic evidence.

The status of clitic position as a stylistically stratified variable is reflected in statements by various Hispanists. Lenz (1925:§247) observed that "es muy posible que la inclinación del lenguaje literario a añadirlo al final del infinitivo o gerundio se deba a reflexión gramatical" 'it is very possible that the inclination in literary language to postpose [the object clitic] to the infinitive or gerund is due to grammatical reflection'. Spaulding (1926:274) observed that postposed position is found "not [. . .] in [19th and early 20th century] writers like Galdós, Alarcón or Valera, but in rhetorical stylists like Tamayo, or Valle-Inclán." Anecdotal reports of school teachers' remarks about correctness (e.g., Silva-Corvalán, 1994:128, note 18) are also in line with an awareness of clitic position as a mark of formality.

Support for the importance of style or register rather than social factors of class and education in CC variation is provided by the figures for MexCult and Novels in Table 7. Some of the same authors in the Essays corpus also write fiction, yet CC frequencies are higher in Novels than in Essays. All of the Essays authors are also speakers and would presumably speak the variety represented by MexCult (UNAM, 1971; cf. Lope Blanch, 1986). Yet CC frequencies are higher in MexCult than in Essays. Indeed figures for MexCult are identical to those for Mexico City. CC thus appears to be subject to situational rather than social variation.

CONSTRUCTION FREQUENCY AND PARALLEL STRUCTURE EFFECTS

An important fact about *-ndo* constructions in the Essays corpus is their low token frequency. As shown in Table 4, the average occurrence of these constructions in Essays is 7.6 per 10,000 words.[12] Compare this figure with the 37.6 tokens on average per 10,000 words in Mexico City and 33.9 in HablaCulta, a corpus of educated Spanish spoken in major cities (Clegg & Rodríguez, 1993). That is, *-ndo* constructions are nearly five times more likely to appear in conversation than in formal texts.[13]

This result is not really surprising, given that the distribution of verb forms is highly "genre-dependent" (Givón, 1990:963). For example, in about 110 pages of academic prose (García Canclini, 1993:86–196), there are 13 *-ndo* constructions in the words of the authors (*ir*, 7; *seguir*, 3; *estar*, 2; *venir*, 1) but 15 in quotes from oral interviews (*estar*, 7; *ir*, 3; *seguir*, 3; *andar*, 1; *venir*, 1), even though these made up only a small portion of the text. The high proportion of stative verbs and the kinds of things talked about in formal registers explains, at least in part, the low token frequency of *-ndo* constructions, since these constructions generally express progressive or continuous (durative) aspect. That is, there may be good functional motivations (in the sense reviewed in Labov, 1994:547ff.) for the low token frequency of *-ndo* constructions in these texts.

But functional reasons do not explain the low frequency of CC. Even if there are fewer *-ndo* constructions because of what is talked about, why should the pro-

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

portion of CC cases be lower? It is not because the constructions are somehow less bleached in Essays than in Mexico City, as we have seen. Here we find support for the role of construction frequency in formal change. Recall that higher CC frequencies are diachronically linked to higher construction frequencies. On this count the Essays corpus is closer to the Old Spanish corpus than to the Mexico City corpus. Not only is the token frequency of *-ndo* constructions lower in Essays than in Mexico City (Table 4), but the proportion of gerunds in a one-to-one relationship in construction with an auxiliary is also lower (Table 5). Looking at it from the converse perspective, the proportion of lone-standing gerunds in Essays is 72% compared with 30% in Mexico City. In addition, multiple-gerund constructions average 7% and 10% in Essays for *estar* + *-ndo* and *ir* + *-ndo*, respectively, which does not differ significantly from the proportion of multiple-gerund constructions in Old Spanish. These figures contrast with about 1% in Mexico City (Table 2).

By what mechanism might construction frequency have an effect on CC? A well-known effect in psycholinguistic studies is the priming of lexical items by semantically or phonetically similar items. Sociolinguistic studies have shown that priming and parallel structure effects also apply to syntax. Thus, Weiner and Labov (1983) found that agentless passives in English are most strongly favored following another agentless passive, and that coreferential subjects tend to appear in the same syntactic position. Pereira Scherre and Naro (1991) found that agreement markers in Brazilian Portuguese are subject to a parallel processing effect, in that explicit marking is followed by explicit marking and lack of marking is followed by further lack of marking. Spanish CC may be showing the same kind of effect. Since clitics are categorically postposed to lone-standing gerunds, the relatively high proportion of such forms in formal texts may encourage a parallel structure in the much smaller proportion of gerunds that are in construction with an auxiliary.

Initial support for a parallel processing effect in the positioning of clitics comes from the lack of CC in multiple-gerund constructions. There is only 1 CC occurrence in 6 multiple-gerund *ir* + *-ndo* tokens in which CC is possible in the Essays corpus.[14] For example, in the multiple-gerund construction in (21), each gerund has its clitic postposed, *apoderándo**se*** 'taking hold of him' and *empujándo**lo*** 'impelling him'. The second gerund *empujándolo* may be interpreted in two ways. It could be a manner adverbial modifying *ir apoderándose* as a unit, and thus the multiple-gerund construction means something like 'fear gradually takes hold of him, impelling him forward'. Or *empujándolo* could be (indirectly) in construction with *ir*, with continuative meaning, as in *va empujándolo* 'keeps on impelling'; thus the multiple-gerund construction as a whole means something like 'fear gradually takes hold of him and keeps impelling him forward'.

(21) Un miedo informe va apoderándo**se** gradualmente de él, empujándo**lo** hacia adelante siempre. (Rodolfo Usigli, El gran teatro del mundo. In Martínez, 1958:174)
   a. 'A vague fear gradually takes hold of him, impelling him ever forward.'
   b. 'A vague fear gradually takes hold of him and impels him ever forward.'

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

There seems to be no great meaning difference, whether the second gerund is a modifier of the *ir* + first gerund sequence (21a) or in indirect construction with *ir* (21b). But the presence of a second gerund, vying as it were for a relationship with the auxiliary *ir*, makes for a less tight connection between *ir* and the first gerund, and CC fails to occur. In addition, when the second gerund has its own clitic, parallel structure favors postposing of the first gerund's clitic.

Parallel structure effects in clitic position do not preclude the development of clitic position as a sociolinguistic variable subject to register variation. What may have begun as the persistence of older patterns in formal varieties of the language (i.e., the high proportion of lone-standing gerunds encourages postposed clitic position for the relatively small number of gerunds in construction with an auxiliary) may very well end up being a feature that speakers or authors manipulate to convey formality. This idea is not new. The English passive seems to be an example of the same sort of thing. Passives can be viewed as adaptations to the functional demands of information flow in English. At the same time, the circumstances in which they are used are more formal than those in which the active is used, and thus they may acquire an association of formality (A. Hudson, personal communication). Similarly, in Spanish, the kinds of things talked about in formal situations require fewer *-ndo* constructions and a proportionally higher number of lone-standing gerunds, which in turn encourage postposition of clitics because of parallel structure effects. Postposed clitic position then becomes a mark of formality in its own right.

We may summarize some of our conclusions as follows. First, CC is part of a series of reductive changes in the form of periphrastic expressions as they emerge as more fused units. We found that CC frequencies in Spanish *-ndo* constructions increase diachronically in tandem with positional fixing, a decrease in intervening material, and a decline of multiple-gerund constructions. Second, we also found lower CC frequencies in formal texts than in conversational data, though not lower indices of reductive change in meaning or bleaching (co-occurring locatives, inanimate subjects). Thus, formal reduction in grammaticizing constructions may lag behind semantic reduction in more formal registers. Third, in evaluating frequency effects, it may be important to count both the tokens of an item appearing in a construction as well as the tokens of the same item not appearing in the construction. We found that it is not only the token frequency of *-ndo* constructions that varies significantly between corpora, but also the proportion of lone-standing gerunds. We described the combination of increased token frequency and decreased proportion of lone-standing gerunds as increased construction frequency. Construction frequency encourages reductive change: we found that higher CC frequencies in the Mexico City corpus compared with the Old Spanish corpus occur in tandem with higher construction frequencies. Fourth, we found that lower CC frequencies in the Essays Corpus compared with the Mexico City corpus are linked to lower average construction frequencies. Thus, frequency effects may operate more generally in the language at any given time as well as on the level of particular texts. Fifth, lower construction frequency, as we defined it, means a greater proportion of lone-standing gerunds. Invariably postposed clitics in lone-standing gerunds may

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

encourage a parallel structure for gerunds in construction with an auxiliary. Our findings suggest a link between construction frequency and parallel structure effects such as have been found in other studies of syntactic variables. Specifically, one mechanism by which frequency of use may have a reductive effect in formal change is through a parallel structure effect.

Where does all this leave us with respect to a correlation between CC and the degree of grammaticization of periphrastic expressions in Spanish? The diachronic evidence supports the general idea. However, we saw that considerations of register may override this general tendency. Most importantly, we found that CC is linked to construction frequency.

NOTES

1. Halliday (1988:162) defined register in terms of co-occurrence patterns as "a cluster of associated features having a greater-than-random […] tendency to co-occur." For sociolinguistic approaches to register, see Biber and Finegan (1994). For a discussion of formality, see Irvine (1979).
2. Midway position is preferred in the earliest Old Spanish text, *Poema de Mio Cid*: 7 of 10 clitic tokens (70%) with *estar* + *-ndo* and *se*(*e*)*r* + *-ndo* in midway position (see verses 2, 154, 1058, 1746, 2305 for *estar* and verses 122, 1840, 2239, 2532, 3553 for *seer*).
3. An exception is the EE1, where in 18 of 26 *ir* + *-ndo* cases (69%) the clitic is in midway position. This text, a prose chronicle, could be taken to be the most formal in the Old Spanish corpus.
4. In parallel, *estar* + *-ndo* shows a faster pace of change than *ir* + *-ndo* with respect to semantic evolution. The meaning of *ir* + *-ndo* of "gradually developing situation," as in *poco a poco va creciendo* 'it is [goes] growing little by little', is closer to the original directional movement meaning of the construction than is the meaning of *estar* + *-ndo* of (transitory-experiential) habitual situation, as in *ahora estoy trabajando en Futurama* 'now I am working at Futurama', to the original locative meaning (Torres Cacoullos, 1999).
5. Keniston (1937a:104–106; see also Spaulding, 1926:267) included the occurrence of a stressed word before the construction as a factor in clitic position. Preliminary results from the small number of Old Spanish tokens suggest that co-occurring locatives and intervening material disfavor CC. First, of 28 *estar* + *-ndo* cases in which CC is possible, 8 have a co-occurring locative. None of these have preposed CC, 3 (37.5%) have CC to midway position, and 5 (62.5%) have no CC. Second, of 3 cases with intervening open class material, 2 have no CC (both in C01, 38r26 and 61v4), and 1 has an ambiguous reflexive pronoun.

We find many examples in Old Spanish where the reflexive pronoun's host may be either the *-ndo* form or the auxiliary. An example of an ambiguous reflexive pronoun is *Míranse las manos con tantas sortijas e van***se** *los beços mordiendo por los tornar bermejos* (Corb, Parte II, Cap. VIII, p. 182) 'She looks at her hands all covered with rings, and chews her lips to make them red' (Simpson, 1959:140), where it is not clear whether *se* is an argument of *ir* or *morder*. It is likely that this ambiguity of reflexive pronoun clitics promoted the change in favor of CC by adding to the frequency of apparent CC constructions (cf. Labov, 1994:588ff., on probability matching by language learners).
6. There are only 31 *estar* + *-ndo* tokens in the 1499 edition of the *Celestina*. Only 5 are possible CC cases.
7. The 20% figure for preposed gerunds with *estar* in *Poema de Mio Cid* includes *seer* + *-ndo*, which disappears from Old Spanish prose by the 14th century.
8. Keniston's 16th century data are from a corpus of 40 prose texts, totaling 300,000 words (1937a:xiv). Over half the texts "contain a more or less literary version of actual conversation" (Keniston, 1937a:xxi). The 20th century data are from 60 prose texts, totaling 600,000 words, dating from 1900 to 1933 (1937b:7), and again "over half the materials studied represents a literary version of conversational style" (1937b:6). Based on Keniston's statements and the breakdown of text types, I assume that register has been as much as possible controlled for, and so differences can be attributed to diachronic change. I computed the frequency counts in Table 4 from Keniston's (1937a:468–469; 1937b:207–208) absolute figures.
9. In Old Spanish (and some present-day varieties) there is also the *en* 'in' + *-ndo* construction, more or less equivalent to present-day Spanish *al* 'at' + infinitive or English *(up)on* + *-ing*, as in ***endurmiendo,*** *apareciol el nuestro sennor ih*⟨*es*⟩*u xp*⟨*rist*⟩*o* (EE1-112r) 'upon falling asleep our lord Jesus Christ appeared before him'.

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

10. Note the low number of tokens for this variable (cf. Silva-Corvalán, 1989:98). From a total of roughly 1,800 pages of text, I found fewer than 300 tokens, disproportionately distributed: one-third in the Essays corpus (approximately 1,350 pages) and two-thirds in the Mexico City corpus (450 pages).
11. The results of an earlier run separating factors in the clitic syntax/person group gave the following results: impersonal *se*, .77; *me, te, nos*, .69; *lo(s)*, .47; *le(s)*, .38; reflexive *se*, .36. For an example of impersonal *se*, see *se siente* 'one feels' in (10) or *no se puede* 'one cannot' in (13); an example of reflexive *se* is in (1). The low IVARB weight for reflexive *se* is very likely an epiphenomenon of the very skewed distribution of these tokens, since 48 of 81 tokens (59%) are from the Essays corpus. Of these, only 23 (47%) have CC, while of the 33 tokens in the Mexico City corpus, 26 (79%) have CC. On the other hand, the weight for impersonal *se* is probably a truer reflection of its effect. Of 10 tokens, 9 are in Essays, and of these 7 (78%) have CC.

It is striking that direct object pronouns (e.g., *lo*) seem to favor CC more than indirect object pronouns *le(s)*, which are overwhelmingly definite and human. However, in Mexican Spanish the form *le* has advanced on a grammaticization path from object pronoun to agreement marker to intensifier. For example, in *A mí me gusta la política y siempre estoy buscándo**le*** 'I like politics and I'm always looking into things' (UNAM, 1976:302), the clitic *le* does not refer to an argument (either animate dative or inanimate accusative) of *buscar* 'to look for' but is part of the expression *buscarle*, where the erstwhile transitive action of "looking for" has no specific patient. This kind of intensive *le* seems to disfavor CC.
12. I estimated the word count for Essays at 364,000 as follows. I counted the number of words on pages 87–88 of Castellanos (1966), which was the book with the smallest page size and relatively large letter case. The counts were 286 and 282, respectively. Then I multiplied the page total (as shown in Appendix 2) by 280. The result most likely underestimates the word count; the figures given in Table 4 for Essays should be even lower.
13. The Essays texts were selected to include well-known Mexican writers, though clearly others could have been chosen. García Canclini (1993) and Lope Blanch (1983) were chosen to sample research-based academic writing; the remaining texts more strictly belong to the genre known as *ensayo* 'essay'. The individual variation in the text frequency of *-ndo* constructions shown in Appendix 3 is noteworthy. The texts span nearly three-quarters of the 20th century, though no clear pattern emerges with respect to date of publication. The two academic texts show a lower than average frequency. One factor that might be important is the writer's sex (cf. Arnaud, 1998). The second highest frequency is in Castellanos (1966), the one female writer I included.
14. The one case actually has the clitic preposed to auxiliary *haber*:

Se ha ido enriqueciendo y complicando la sintaxis castellana (Lope Blanch, 1983:138)
'Spanish syntax has been getting richer and more complex'

REFERENCES

Arnaud, René. (1998). The development of the progressive in 19th century English: A quantitative survey. *Language Variation and Change* 10:123–52.
Barlow, Michael. (1997). MonoConc for Windows v1.2. Houston, TX: Athelstan.
Biber, Douglas. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language* 62:384–414.
Biber, Douglas, & Finegan, Edward (Eds.). (1994). *Perspectives on register: Situating register variation within sociolinguistics*. Oxford: Oxford University Press.
Bybee, Joan L. (1985). *Morphology: A study on the relation between meaning and form*. Amsterdam: Benjamins.
_____ (forthcoming). Mechanisms of change in grammaticization: The role of frequency. In R. Janda & B. Joseph (Eds.), *Handbook of historical linguistics*. Oxford: Blackwell.
Bybee, Joan L., Perkins, Revere, & Pagliuca, William. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
Bybee, Joan L., & Thompson, Sandra. (1997). Three frequency effects in syntax. *Berkeley Linguistics Society* 23.
Clegg, J. Halvor, & Rodríguez, J. David. (1993). *Progressive constructions in the speech of U.S. Hispanics*. Paper presented at the XIV Spanish in the United States, San Antonio, TX.
Davies, Mark. (1998). Corpus of historical Spanish prose. http://138.87.135.33/~mdavies/espanol.htm.
Givón, Talmy. (1990). *Syntax. A functional-typological introduction* (Vol. II). Amsterdam: John Benjamins.

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

Haiman, John. (1994). Ritualization and the development of language. In W. Pagliuca (Ed.), *Perspectives on grammaticalization*. Amsterdam: Benjamins. 3–28.
Halliday, Michael A. K. (1988). On the language of physical science. In M. Ghadessy (Ed.), *Registers of written English*. London: Pinter. 162–178.
Heine, Bernd, Claudi, Ulrike, & Hünnemeyer, Friederike. (1991). From cognition to grammar: Evidence from African languages. In Elizabeth Traugott & Bernd Heine (Eds.), *Approaches to grammaticalization* (Vol. I). Amsterdam: Benjamins. 149–187.
Hopper, Paul J., & Traugott, Elizabeth Closs. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.
Irvine, Judith T. (1979). Formality and informality in communicative events. *American Anthropologist* 81:773–790.
Keniston, Hayward. (1936). Verbal aspect in Spanish. *Hispania* 19:163–176.
_____ (1937a). *The syntax of Castilian prose: The sixteenth century*. Chicago: University of Chicago Press.
_____ (1937b). *Spanish syntax list*. New York: Holt.
Krug, Manfred. (1998). String frequency: A cognitive motivating force in coalescence, language processing and linguistic change. *Journal of English Linguistics* 26:286–320.
Labov, William. (1994). *Principles of linguistic change: Internal factors*. Cambridge: Blackwell.
Lenz, Rodolfo. (1925). *La oración y sus partes* (2nd ed.). Madrid: Centro de Estudios Históricos.
Lope Blanch, Juan M. (1986). *El estudio del español hablado culto: Historia de un proyecto*. México: UNAM.
Luján, Marta. (1980). Clitic promotion and mood in Spanish verbal complements. *Linguistics* 18:381–484.
Luna Trail, Elizabeth. (1980). *Sintaxis de los verboides en el habla culta de la ciudad de México*. México: UNAM.
Myhill, John. (1988a). Variation in Spanish clitic climbing. In Thomas J. Walsh (Ed.), *Synchronic and diachronic approaches to linguistic variation and change* (GURT '88). Washington, DC: Georgetown University Press. 227–250.
_____ (1988b). The grammaticalization of auxiliaries: Spanish clitic climbing. *Berkeley Linguistics Society* 14:352–363.
Pereira Scherre, Maria Marta, & Naro, Anthony. (1991). Marking in discourse: "Birds of a feather." *Language Variation and Change* 3:23–32.
Pintzuk, Susan. (1986). VARBRUL: Version 2S. Philadelphia: Department of Linguistics, University of Pennsylvania.
Rivero, María-Luisa. (1970). A surface structure constraint in negation. *Language* 46:640–666.
_____ (1986). Parameters in the typology of clitics in Romance and Old Spanish. *Language* 62:774–807.
Rizzi, Luigi. (1982). *Issues in Italian syntax*. Dordrecht: Foris.
Silva-Corvalán, Carmen. (1989). *Sociolingüística: Teoría y análisis*. Madrid: Alhambra.
_____ (1994). *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon.
Spaulding, Robert K. (1926). History and syntax of the progressive constructions in Spanish. *University of California Publications in Modern Philology* 13:229–284.
Torres Cacoullos, Rena. (1999). Variation and grammaticization in progressives: Spanish *-ndo* constructions. *Studies in Language* 23:25–59.
Universidad Nacional Autónoma de México (UNAM). (1971). *El habla de la ciudad de México: Materiales para su estudio*. México: Publicaciones del Centro de Lingüística Hispánica.
_____ (1976). *El habla popular de la Ciudad de México: Materiales para su estudio*. México: Publicaciones del Centro de Lingüística Hispánica.
Weiner, E. Judith, & Labov, William. (1983). Constraints on the agentless passive. *Journal of Linguistics* 19:29–58.

APPENDIX 1

OLD SPANISH CORPUS: ABBREVIATION, TEXT, EDITION, AND ENGLISH TRANSLATIONS

PMC = *Poema* (or *Cantar*) *de Mio Cid*, 1140/1207 (verse, epic).
Electronic version of Menéndez Pidal paleographic edition.
Hamilton, Rita, & Perry, Janet. (1975). *The Poem of the Cid*. With an introduction and notes by Ian Michael. Manchester: Manchester University Press.

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

Such, Peter, & Hodgkinson, John. (1987). *The Poem of my Cid*. Warminster, England: Aris and Phillips Ltd.

Apol = *Libro de Apolonio*, c. 1250 ("cuaderna vía" verse, epic).

Alvar, Manuel. (1976). *Libro de Apolonio. Estudios, ediciones, concordancias* (3 vols.). Madrid: Castalia.

EE1 = *Estoria de España*, Alfonso X, 1270–1289 (prose, chronicle).

Kasten, Lloyd, Nitti, John, & Jonxis-Henkemans, Wilhelmina. (1997). *The electronic texts and concordances of the prose works of Alfonso X, El Sabio*. Madison, WI: HSMS. (CD-ROM).

LBA = *Libro de buen amor*, Juan Ruiz, Arcipreste de Hita, 1330/1343 (verse, sermon, autobiography, exempla).

Electronic version of Salamanca ms.

Willis, Raymond S. (1972). *Juan Ruiz, Libro de Buen Amor*. Princeton, NJ: Princeton University Press.

Luc = *Libro de los enxiemplos del Conde Lucanor e de Patronio*, Don Juan Manuel, 1335 (prose, didactic exempla).

Sotelo, Alfonso I. (1976). *Don Juan Manuel "Libro de los enxiemplos del Conde Lucanor e de Patronio*." Madrid: Cátedra.

Keller, John E., & Keating, L. Clark. (1977). *The book of Count Lucanor and Patronio. A translation of Don Juan Manuel's El Conde Lucanor*. Lexington: The University Press of Kentucky.

Corb = *Arcipreste de Talavera o Corbacho*, Alfonso Martínez de Toledo, 1438 (prose, sermon, exempla).

Gerli, Michael. (1987). *Alfonso Martínez de Toledo "Arcipreste de Talavera o Corbacho*." Madrid: Cátedra.

Simpson, Lesley Byrd. (1959). *Little sermons on sin*. Berkeley: University of California Press.

C01 = *(Tragi)comedia de Calisto y Melibea* or *Celestina*, Fernando de Rojas, 1499 (prose, novel in dialogue).

Corfis, Ivy A., O'Neill, John, & Beardsley, Theodore S., Jr., (1997). *Early Celestina electronic texts and concordances*. Madison, WI: HSMS. (CD-ROM)

Simpson, Lesley Byrd. (1955). *The Celestina* (translated from the Spanish). Berkeley: University of California Press.

Singleton, Mack Hendricks. (1958). *Celestina* (translated from the Spanish). Madison: University of Wisconsin Press.

In the citation of examples, the original text in the appendix is reproduced exactly. Omissions are indicated by ellipsis (. . . .). Numbers in parentheses following examples refer to verse number for the PMC, Apol, and LBA, to the Cátedra edition page number for Luc and Corb, and to the folio location for EE1 and C01.

APPENDIX 2
ESSAYS CORPUS

Castellanos, Rosario. (1966). *Juicios sumarios. Ensayos*. Xalapa, México: Cuadernos de Facultad de Filosofia, Letras y Ciencias, Universidad Veracruzana. 13–113.

García Canclini, Néstor. (1990). *Culturas híbridas. Estrategias para entra y salir de la Modernidad*. México, D.F., México: Grijalbo. 31–93.

García Canclini, Néstor (Ed.). (1993). *El consumo cultural en México*. México, D.F., México: Consejo Nacional para la Cultura y las Artes. 86–196.

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press

Lope Blanch, Juan M. (1983). *Análisis gramatical del discurso*. México, D.F., México: UNAM. 11–175.
Martínez, José Luis (Ed.). (1958). *El ensayo mexicano moderno*. México, D.F., México: Fondo de Cultura Económica. 9–409.
Monsiváis, Carlos. (1995). *Los rituales del caos*. México, D.F., México: Era. 15–113.
Montemayor, Carlos. (1985). *El oficio literario*. Xalapa, México: Universidad Veracruzana. 9–111.
Paz, Octavio. (1959). *El laberinto de la soledad*. México, D.F., México: Fondo de Cultura Económica. 9–105.
Reyes, Alfonso. (1927). *Cuestiones gongorinas*. Madrid: Espasa-Calpe. 5–89, 133–268.

APPENDIX 3
-ndo CONSTRUCTION FREQUENCIES IN ESSAYS

	Estar		*Ir*		*Seguir*		Total[a]	
	N		*N*		*N*		*N*	Average[b]
Reyes (212 pp.)	0		10		8		20	9.4
Martínez (400 pp.)	30		36		34		111	27.7
Paz (96 pp.)	0		0		4		4	4.2
Castellanos (100 pp.)	10		14		4		38	38
Montemayor (102 pp.)	3		3		7		13	12.7
García Canclini (62 pp.)	10		13		9		35	56.4
Monsivías (98 pp.)	8		8		5		22	22.4
Lope Blanch (164 pp.)	4		10		7		22	13.4
García Canclini, ed. (110 pp.)	2		7		3		13	11.8
	N	Average	*N*	Average	*N*	Average	*N*	Average
Total Essays (1,344 pp.)	67	5	101	7.5	81	6	278	20.7
Total Mexico City (450 pp.)	340	75.5	146	32.4	39	8.7	637	141.5

Note: Text listed in order of date of publication.
[a]Total includes cases of *continuar*, *venir*, and *andar* + *-ndo*. I did not include footnotes or quoted material in the count of *-ndo* constructions.
[b]Average per 100 printed pages.

https://doi.org/10.1017/S095439459911202X Published online by Cambridge University Press