

## The evolution of Spanish clitic climbing: A corpus-based approach

Mark Davies

To cite this article: Mark Davies (1997) The evolution of Spanish clitic climbing: A corpus-based approach, *Studia Neophilologica*, 69:2, 251-263, DOI: [10.1080/00393279708588211](https://doi.org/10.1080/00393279708588211)

To link to this article: <https://doi.org/10.1080/00393279708588211>



Published online: 21 Jul 2008.



Submit your article to this journal [↗](#)



Article views: 119



View related articles [↗](#)



Citing articles: 3 View citing articles [↗](#)

# The Evolution of Spanish Clitic Climbing: A Corpus-Based Approach

MARK DAVIES

## 1. Introduction

1.1. Clitic climbing in Spanish, which is the movement of unstressed object pronouns from a postverbal (1a) to a preverbal position (1b), presents researchers with a number of interesting questions.

(1a) *si la criada no quiere [hacerlo]* (México-C:19:259)

(1b) *un colega mío no lo quiere [hacer\_\_]* (San José:22:1)

The most general question, discussed by researchers from a number of different syntactic frameworks, is why there should be any clitic climbing at all in Spanish, since in many other languages it is ungrammatical (cf. Eng. *\*I it want [to do\_\_]*, Fr. *\*je le veux [faire\_\_]*) (see Contreras 1979, Pizzini 1982, Aissen and Perlmutter 1983, Zagana 1986, Myhill 1988, Kayne 1989, Rosen 1989). In addition to this one general issue, however, various researchers have attempted to identify the various factors that influence clitic climbing in Spanish. Several have sought to explain why clitic climbing occurs more often and is more acceptable with certain main verbs (e.g. *ir+a*, *acabar de*, *querer*) than with others (e.g. *renunciar+a*, *tratar+de*, *anhelar*) (cf. Contreras 1979, Luján 1980, Suñer 1980, Myhill 1988, Rosen 1989). Other researchers have considered the role of the clitic itself, such as reflexive vs. nonreflexive and single vs. multiple clitics (Spaulding 1927:346, Aissen and Perlmutter 1983:365, Rosen 1989). It is only within the past decade or so, however, that researchers have begun to use large corpora of written and spoken Spanish to study these questions (Myhill 1988, Davies 1995).

1.2. Just as with the Modern Spanish construction, there are a number of unresolved issues and interesting questions regarding the historical development of Spanish clitic climbing. For example in Section 3 we will see that although clitic climbing was the norm in OSp (2a), by the 1600-1700s it was much less common (2b). It is still unclear, however, precisely when it was that the main decrease in clitic climbing took place.

(2a) & que **lo** non **deuie fazer** por ninguna manera (EstEsp 87v [1270])

(2b) Sin duda, pero yo no **debía hacerlo** (María 650:1 [1867])

A second question concerns the motivation for the shift from (2a) to (2b). The few researchers who have discussed possible motivations have argued for fairly abstract changes in the clausal structure of Spanish (Wanner 1982) or the nature of the clitic itself (Rivero 1991). In Section 4, however, we will suggest that changes in clitic climbing may have been motivated by simple, surface-oriented changes in a related construction. A third question is whether the same factors that influence clitic climbing in Modern Spanish might have motivated the historical shifts, which is an issue that we will consider in Section 5. The fourth and final question is perhaps the most problematic one. In Section 5 we will see that although there was a 500–600 year shift away from clitic climbing and that by the 1700–1800s clitic climbing had almost disappeared, within the past 200–300 years there has been a shift back towards clitic climbing. How can we motivate a syntactic change that moves slowly from X to Y over a period of several centuries, becomes almost complete, and then moves back in the opposite direction?

## 2. Data base and methodology

2.1. Although there have been a number of studies on diachronic clitic climbing in

Spanish, none of the questions that we have just presented have been completely answered. This is due in large part to the fact that the data on which these studies were based were inadequate. Let us review briefly the previous studies and show what data are already available and what data are still lacking. Spaulding (1927) presents clitic climbing data that are based on extracts from just two or three authors in each century from the 1200s to the 1700s, and Colburn (1928) is based on ten different novels from the 1800s. In Keniston (1937), clitic climbing is one of hundreds of constructions under discussion, and the data are taken from just the 1500s. Wanner (1982) is based on just 5000 words of text from nine texts from *El Cid* to *El Corbacho* (mid-1100s to mid-1400s), some of them being in verse, rather than strictly prose. Granberg (1989) considers clitic climbing only peripherally in his examination of the period 1250–1400 and deals with a very small number of total cases of clitic climbing. Finally, Rivero (1991) is a theory-oriented study that is based on previous researcher's studies and contributes little in the way of new data.

2.2. This study is based on a much larger and complete corpus than those listed above, and is an attempt to use these data to answer more completely the questions presented in the introduction. The data that we will use are based on more than 12,500 examples of the clitic climbing construction taken from the largest available corpus of historical Spanish, which contains nearly 5,400,000 words in 118 different texts from the 1200s to the 1800s. Table 1 shows the number of texts and the number of words from each century, and a detailed listing of all of the 118 texts is found in Note 1. In creating the corpus, care was taken to ensure that there were adequate data from each of the seven centuries under examination, which are each composed of approximately 700,000 words of text. In addition, care was taken to ensure that there are at least ten texts from each century, and that no one text accounts for more than 20% of the data for that century. Notice also that the texts for the 1800s are divided into two sections – those from Spain and those from Latin America.

To extract the more than 12,500 tokens from the corpus, the following procedure was used. We first used the *OmniPage* Optical Character Recognition (OCR) software to scan into the computer several of the 46 pre-1500 texts and all of the 72 post-1500 texts. As Note 1 indicates, many of the pre-1500 texts were also taken from Volume 0 of the ADMYTE corpus of historical Spanish texts. After editing and formatting the texts, we used the *WordCruncher* software package to create an 'every word' index of the corpus, and were then able to use the index generated by *WordCruncher* to perform complex Boolean and proximity searches on the data. Proximity searching means that we were able to extract, for example, all cases of a clitic (*me*, *te*, *lo*, etc.) occurring within five or six

Table 1  
Composition of the corpus and number of tokens by verb

	# texts	# words	<i>deber</i>	<i>querer</i>	<i>poder</i>	–PREP	+PREP	Total
1200	14	776,700	644	444	681	34	62	1865
1300	10	744,200	382	377	637	28	121	1545
1400	15	765,200	454	291	728	60	76	1609
1500	19	745,300	228	402	774	146	191	1741
1600	16	701,100	94	332	832	197	290	1745
1700	17	669,800	281	236	834	141	213	1705
1800-ES	13	425,500	89	179	455	109	220	1052
1800-LA	14	550,600	191	159	517	110	279	1256
Total	118	5,300,000	2363	2420	5458	825	1452	12518

–PREP = *saber, pensar, esperar, desear, procurar*

+PREP = *comenzar a, empezar a, dejar de, cesar de, venir a, llegar a, salir a, volver a*

characters of one of the seventeen governing verbs (*poder, querer, desear, començar a, dejar de*, etc.), which was in turn within a certain number of characters of an infinitive (giving, for example, *lo quiero fazer*). Boolean searching means that we were able to combine the results of this search with other searches, for example to include all cases of one of the seventeen governing verbs followed either by a clitic plus infinitive (*quiero lo fazer / quierolo fazer*) or by an infinitive plus clitic (*quiero fazerlo/ quiero fazer lo*). Using this process, we were able to extract more than 12,500 examples of the clitic climbing construction, which are summarized in Table 1.

### 3. Dating the shifts

The first question we need to answer is precisely when it was that the major shifts in clitic climbing took place. To guide us in our discussion, let us consider the three possible placements of the clitic:

- |  |         |
|--|---------|
| (3a) <i>si lo quiere fazer</i> por su medida (Estoria de Espana: 72v [1270]) | initial |
| (3b) <i>yo quierovos dezir</i> el enxenplo (Historia Troyana: 51 [1253])     | medial  |
| (3c) <i>Más quiero avergonzarlos</i> (Guzmán: 459 [1597])                    | final   |

It is generally accepted that in Old Spanish, initial and medial placement were the norm and that final placement was quite rare, and that the major decrease in clitic climbing (i.e. initial and medial placement) started to take place in perhaps the 1500s (see Chenery 1905, Hanssen 1913, Leavitt 1954, Menéndez Pidal 1954). There are only four studies, however, that have provided much original data (Spaulding 1927, Colburn 1928, Keniston 1937, and Wanner 1982), and unfortunately there is a fair degree of disagreement between the studies. This is mainly due to the different corpora on which the studies were based and the different ways in which the data were organized and presented. For example, Wanner shows that final placement fell from 6% in the 1200s to 3% in the 1300s and then increased to 10% in the 1400s. Spaulding, on the other hand, shows an increase in final placement during this period from 0% to 1% to 3%. Keniston suggests that final placement in the 1500s occurred in about 13% of all cases, whereas Spaulding places it at about just 8%. The major gap in these studies is for the period 1600–1900, which is precisely when the major shifts took place. Spaulding (1927) is the only researcher to look at the crucial period from the 1600–1700s, and his data are quite limited, since they are based on short selections from just three authors in each century and are not categorized by verb. Colburn (1928) gives the most complete view of the 1800s, but unfortunately he fails to categorize his examples by main verb and also mixes together cases of climbing with lower clause objects (*lo quiero vender*) and subjects (*lo dejé salir*).

It might seem that the dating of the shifts is an insignificant matter, but if we want to correlate clitic climbing shifts with other shifts in the grammar, it is necessary to be quite certain of just when the changes took place. Thus there is a need to go back and correct inaccuracies where they exist, and to verify the data for those periods in which there is only one previous study. The information in Table 2 and Figure 1, then, represents an attempt to provide data on clitic climbing shifts for the entire period from the 1200s to the end of the 1800s. The table shows the number of cases of initial, medial, and final placement (no clitic climbing) and the corresponding percentages of final placement with both the three most common verbs (*poder, querer, deber*) and the thirteen less common verbs. The corresponding Figure 1 shows in graphical form the increase in final placement and the decrease in clitic climbing with the three main and thirteen other verbs, and also shows the approximate date in which final placement became the most common option.

The data show that with the three main verbs, final placement doubled in each of the periods from the 1200s to the 1700s. Figure 1 shows that with the thirteen less common verbs, final placement became the most common placement in about 1580, and reached the same level with *poder, querer, and deber* in about 1650. Finally, these figures suggest that

Table 2  
Initial, medial, and final placement with all verbs in all periods

	3 Main Verbs						13 Other Verbs					
	1	2	3	% 1	% 2	% 3	1	2	3	% 1	% 2	% 3
1200	984	503	21	0.67	0.30	0.02	31	43	12	0.38	0.48	0.14
1300	880	194	50	0.76	0.18	0.04	37	88	14	0.32	0.56	0.11
1400	811	133	77	0.77	0.14	0.07	68	43	18	0.52	0.33	0.14
1500	850	66	150	0.78	0.07	0.14	166	40	100	0.53	0.14	0.32
1600	613	34	302	0.63	0.05	0.31	160	41	249	0.35	0.10	0.54
1700	351	13	761	0.28	0.01	0.70	68	7	255	0.21	0.03	0.76
1800-ES	127	3	526	0.17	0.01	0.82	34	2	278	0.09	0.00	0.90
1800-LA	63	0	762	0.08	0.00	0.92	27	1	355	0.07	0.00	0.93

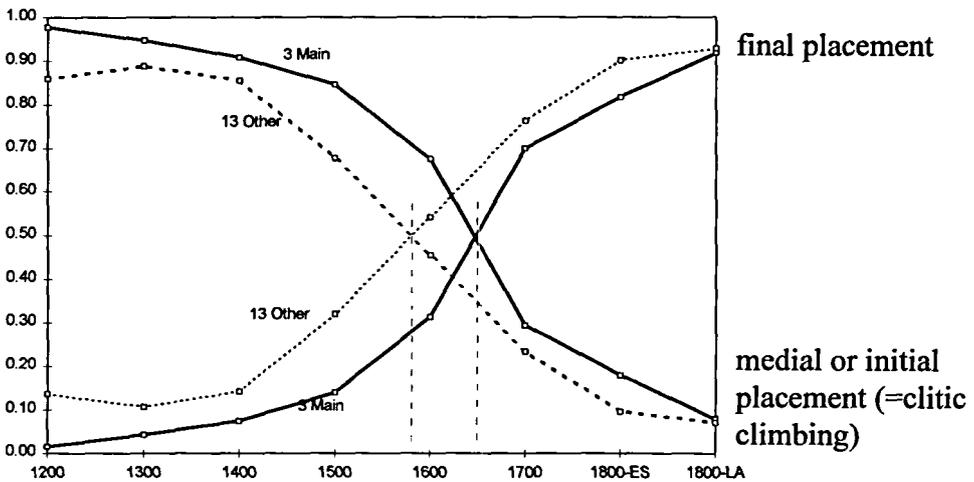


Fig. 1.

final placement was somewhat more common in all periods than had been shown by other researchers, although this may be due in part to the way in which our data were organized (see Note 2).

#### 4. Motivating the decrease in clitic climbing

Now that we have shown *when* the major shift occurred, we can turn to the question of *what* the motivation for the shift might have been. Previous studies have argued that the changes were due to rather abstract shifts in the grammar. Wanner (1982), for example, suggests that there was a general shift from 'loose' clauses in Latin (which permitted movement of the clitic out of its clause) to 'tighter' clauses in the Romance languages (in which the clitic stayed within the embedded clause). Rivero (1991) argues that there was a fundamental change in the nature of the clitics themselves. She suggests that in Old Spanish they were words and that they thus had a high degree of freedom of movement. Since Old Spanish, however, they have evolved into clitics, which means that they are more tightly bound to the governing infinitive.

Rather than using abstract changes in clausal structure or the nature of clitics to motivate the changes in clitic climbing, we propose that the changes result from concrete, surface

Table 3

Percentage of final placement with PREP+INF, VERB+PREP+INF, AND VERB+INF

	PREP+INF de hacerLO	VERB+PREP+INF dejan de hacerLO	VERB+INF quieren hacerLO
1200	0.67	0.13	0.09
1300	0.29	0.08	0.08
1400	0.42	0.13	0.09
1500	0.58	0.29	0.18
1600	0.99	0.50	0.33
1700		0.75	0.57
1800-ES		0.89	0.65
1800-LA		0.95	0.68

level shifts in a related construction. This other construction is PREP+INF (“preposition + infinitive”), in which a preposition is followed by an infinitive, which in turn is either preceded or followed by a clitic:

(4a) él promete ser contigo *para hacerlo* (Pecadores: 291 [1556])

(4b) el tiene grand manera *para lo fazer* (Tristan: 87r [1410])

As Hanssen (1913), Menéndez-Pidal (1954) and others have noted, there were a series of interesting shifts with this construction from the 1200s to the 1500s. Up through the late 1200s, final placement was the norm (4a), followed by a roughly two hundred year period of medial placement (4b), after which placement reverted to the original final placement (4a). Table 3 and the corresponding Figure 2 show that the data from our corpus support these claims. They indicate that final placement (*de hacerlo*) was the rule in 67% of the cases in the 1200s, which decreased to just 29% in the 1300s, and then began a steady rise through the next two hundred years (1400s 42%, 1500s 58%), to become the rule by the 1600s (99%).

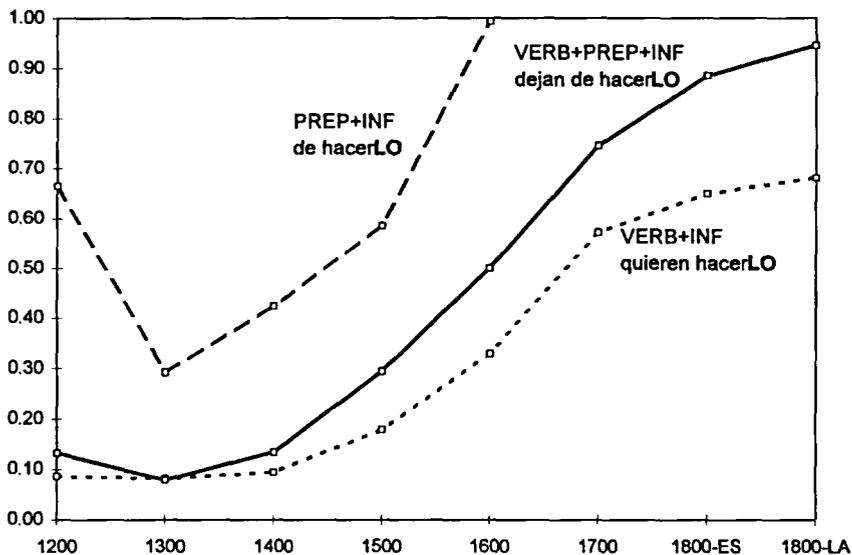


Fig. 2.

For our purposes, however, what is important is the effect that changes with the PREP+INF construction might have had on clitic climbing. In Figure 2 we have divided the main verbs in the clitic climbing construction into two classes—those that take a preposition (*comenzar a, dejar de, volver a*, etc.) and those that do not (*querer, pensar, procurar*, etc.). Since the first class of clitic climbing verbs contains a PREP+INF sequence, it is reasonable to assume that there would be a relationship between these constructions (5b) and the strictly PREP+INF construction (5a), which might in turn extend to the related clitic climbing construction in which there is no preposition (5c). Thus the “locus” of the change is found in cases like (5a), spreads to cases like (5b), and then finally affects cases like (5c):

- (5a) **de hacerlo**                      PREP + INF    1 =>  
 (5b) **dexan de hacerlo**    VERB + PREP + INF    2 =>  
 (5c) **quieren hacerlo**    VERB + INF            3

Examples:

- (6a) Mas por amor **de darles** algo (Castilla 102v [1200-1284])  
 (6b) non **dexan de darles** de comer (Caza 36v [1385-88])  
 (6c) y **quisiese darles** una merienda (Guzmán 222:2 [1597])

Since the data for Table 3 and Figure 2 are based on more than 1700 cases of the PREP+INF construction and 1400 cases of the VERB(+PREP)+INF construction, we are able to map out quite precisely the three shifts, and investigate how they might be related to each other. Let us consider, then, what evidence there is that the locus of the shift towards final placement started with PREP+INF, spread to VERB+PREP+INF, and finally to VERB+INF.

First, concerning the link between PREP+INF and VERB+PREP+INF, it is quite interesting to note that as final placement with PREP+INF decreased from the 1200s to the 1300s, there was a “parallel” (albeit attenuated) shift with the VERB+PREP+INF construction. Then as there was a shift back towards final placement with PREP+INF after the 1300s, there was a corresponding increase in final placement with VERB+PREP+INF. As (5) suggests, the link between PREP+INF and VERB+INF is somewhat less direct. Thus it is not surprising that when final placement with PREP+INF decreases 1200-1300, there is no statistically significant decrease with VERB+INF. Likewise, although the VERB+PREP+INF construction follows PREP+INF and moves towards final placement starting in the 1300s, there is a hundred year lag time before there is any appreciable increase with VERB+INF.

Let us consider for a moment an alternate hypothesis, which is that there was just one single, unitary shift towards final placement with all three constructions, which is the type of “parametric shift” that is proposed in Lightfoot (1991) and elsewhere. Such a scenario would predict three parallel shifts for PREP+INF, VERB+PREP+INF, and VERB+INF. There would of course be a somewhat lower degree of final placement with VERB(+PREP)+INF than with PREP+INF. This is because there are more alternatives to final placement in the VERB(+PREP)+INF construction, since initial placement was common (*lo dexa de fazer*), but was obviously not possible with PREP+INF (*\*lo de fazer*).

There are at least three arguments, however, against this hypothesis of one unified shift. First, final placement with PREP+INF was the norm in two-thirds of all cases in the 1200s, and yet it was much less common with VERB(+PREP)+INF, at only about ten percent. It would be difficult to account for this large difference, even taking into account the lower degree of final placement with VERB(+PREP)+INF due to the possibility of having initial placement with this construction. It thus appears that in the early to mid-1200s, the clitic placement rules for PREP+INF and VERB(+PREP)+INF were separate ones. Second, the dramatic decrease in final placement with PREP+INF in the 1300s was only slightly mirrored in the VERB+PREP+INF construction, and not at all with VERB+INF. Finally,

final placement increases sharply with PREP+INF after the 1300s, is somewhat less marked from the 1300s to the 1400s with VERB+PREP+INF, and takes place only after the 1400s with VERB+INF. This clearly seems to be a case of a syntactic shift that spread from one construction (PREP+INF) to a closely related one (VERB+PREP+INF), and then finally from the second construction to the third (VERB+INF).

## 5. Reversing the shift: why clitic climbing has not disappeared

5.1. As we can see in Figure 2, final placement became the rule with PREP+INF by about the late 1500s. We might therefore logically expect that VERB(+PREP)+INF would follow suit, and that (allowing for some lag time), final placement with these constructions would have become the rule by the 1700s or 1800s. In other words, clitic climbing should have died out by Modern Spanish. Yet as Figure 2 indicates, there is a curious “tapering off” of the shift towards final placement with these two forms starting in about the 1700s.

Furthermore, Modern Spanish data clearly show that initial placement (i.e. clitic climbing) is still very common, as evidenced by the numerous studies that have attempted to account for clitic climbing in Spanish. The best evidence for robust clitic climbing in Modern Spanish comes from Davies (1995), which is based on a corpus of 2,500,000 words of spoken Spanish and 1,000,000 words of written Spanish. His data show that the average degree of final placement in Modern Spanish is 90% in written Spanish, which is a small increase from the 79% percent figure for the 1800s. The data that are much more difficult to explain is the lower figure of 60% final placement in spoken Spanish. Assuming that the spoken register reflects better the popular tendencies of the language, we see that in recent times there has actually been an increase in clitic climbing. In other words, some factor seems to have either slowed or reversed the shift towards uniform final placement, which judging from the 1400s-1700s data, should have become the rule by Modern Spanish. How can we explain such facts?

5.2. One option is to look at Modern Spanish to see what factors are related to clitic climbing, and determine whether any of these factors might have also been active from about the 1600s on in order to reverse the shift towards final placement. A common analysis of clitic climbing in Modern Spanish deals with the nature of the main verb. Napoli (1981) and Myhill (1988) both argue that semantically “weak” or “auxiliary-like” verbs allow clitic climbing more than semantically “strong” or “non-auxiliary” verbs. Napoli defines auxiliary verbs as those that “add conceptually basic or simple information”, “introducing no independent action or state”. In these analyses, a type of “semantic bonding” occurs between semantically simple or auxiliary-like verbs and the infinitive that governs the clitic. Since the two verbs behave as though they were one, both semantically and syntactically, the clitic can easily be placed at the beginning of the verbal complex, as it would with any conjugated verb (7a). In cases where the main verb is more semantically complex, on the other hand, it is both semantically and syntactically more independent from the infinitive, which prevents the clitic from attaching to it (7b).

(7a) *lo* [quiero hacer\_\_]

(7b) ?? *lo* [espero] [hacer\_\_]

The Modern Spanish data clearly support this analysis. These data show that in the spoken *Habla Culta* corpus, the auxiliary-like verbs take clitic climbing in anywhere from 47% (*querer*) to 86% (*ir+a*) of all cases, while some non-auxiliary verbs rarely allow clitic climbing, such as *preferir* (15%), *intentar* (11%), and *esperar* (0%).

If the  $\pm$ auxiliary distinction is important in governing Modern Spanish clitic climbing, then we would want to know when this distinction developed, and what role it might have had in reversing the decrease in clitic climbing. In one of the most complete discussions of this to date, Wanner (1982) argues that the  $\pm$ auxiliary factor began to exert its influence from about the 1500s. He suggest that while clitic climbing was essentially free in Old Spanish, there was a reanalysis in the 1500s, in which for the first time the semantics of the

verb became important. Only those verbs that were truly auxiliary-like in nature continued to allow clitic climbing, and it began to die out with the less auxiliary-like verbs. The only problem is finding the data to support such a model, which would need to show the amount of clitic climbing with different classes of verbs (both +auxiliary and -auxiliary) over a number of centuries. Unfortunately, these data have not previously been available. Researchers have either grouped all verbs together (Spaulding 1927, Ramsden 1963), or have given very imprecise judgments on acceptability with different verbs (Wanner 1982), or are limited to one particular century (Keniston 1937). Our study, on the other hand, is the first to look at a range of verbs over a long period of time.

What we want to find is a situation in which in Old Spanish clitic climbing was equally as common with both +auxiliary and -auxiliary verbs. We then want to find a split developing in the 1500–1600s, in which clitic climbing increases with the +auxiliary verbs. This would provide evidence that the  $\pm$ auxiliary factor was involved in reversing the 500–600 year old shift towards final placement. In Figure 3, we show the percentage of final placement with the three most simple and auxiliary-like verbs (*poder*, *querer*, *deber*) and the corresponding figures for five other semantically more complex verbs (*pensar*, *esperar*, *desear*, *procurar*, *saber*).

As is quite obvious, the data from the more than 11,000 examples with these two classes of verbs fail to support the scenario that we have just proposed. The data show that the  $\pm$ auxiliary factor was already in play in Old Spanish. From the 1200s through the 1600s, final placement (i.e. no clitic climbing) was always at least twice as common with -auxiliary verbs as it was with the +auxiliary verbs. Worse yet, it is precisely in the post-1600 period, when the split was supposed to have developed, that there is actually a convergence between the two classes of verbs. For example, in the 1700s the percentage of final placement with the two classes of verbs narrows to 70% vs. 80%, which converges even more to 87% vs. 92% in the 1800s. Thus it appears that an explanation based on a sudden split, involving the semantics of the main verb, cannot be supported by the data.

As Figure 3 indicates, however, we must still account for the data that clearly show a  $\pm$ auxiliary distinction in the spoken register of Modern Spanish (1900-S). These data show 46% clitic climbing (i.e. 54% final placement) with the +auxiliary verbs, but only

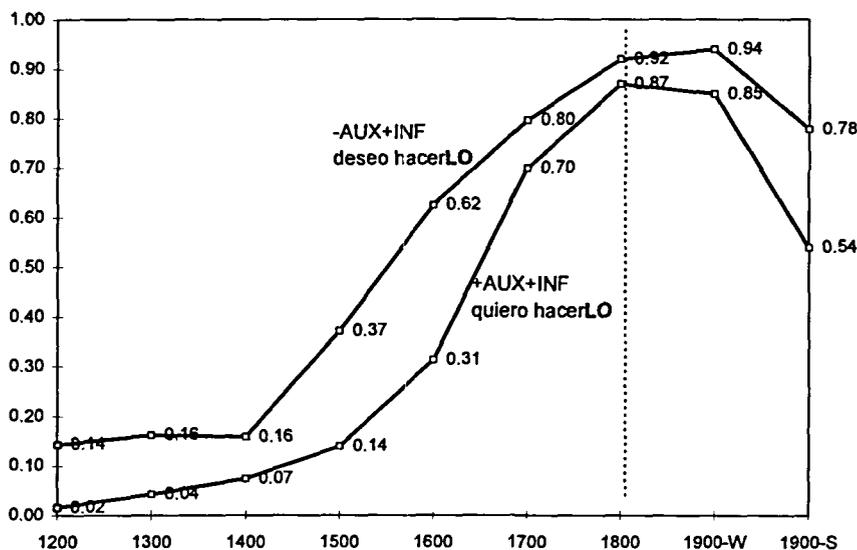


Fig. 3. Percentage of final placement with  $\pm$ auxiliary verbs

22% clitic climbing with the -auxiliary verbs. We are thus faced with data that indicate that the degree of clitic climbing was dependent on the semantics of the main verb at every stage since Old Spanish and is clearly functional in Modern Spanish, but a factor that appears to be the least important precisely when we need it to be a check against the general shift away from clitic climbing.

5.2. In spite of the difficulties, there still may be a solution in sight. Figure 1 shows that there was an approximately 100 year lag-time between shifts in PREP-INF and the corresponding changes in VERB-INF. In other words, it took a considerable period of time before the effect of the PREP-INF construction overcame the "inertia" of preexisting tendencies with the VERB+INF construction. Let us assume likewise, then, that although final placement became generalized with the PREP-INF construction in the late 1500s-early 1600s, as the data suggest, the shift towards final placement with VERB+INF continued on for one to two hundred years after the change was completed with PREP+INF, although there was a gradual leveling off after this time. During this period of 'leveling off', in which the previous effect of PREP+INF became gradually less important, there was a "reanalysis" in the grammar, based on the  $\pm$ auxiliary nature of the verb. Whereas there had always been a higher degree of clitic climbing with the +auxiliary verbs due to semantic/pragmatic constraints, it was only after the competing motivation based on the PREP+INF construction died out that the main factor influencing clitic climbing was the nature of the main verb.

What happened syntactically can be compared quite nicely with the type of phonemic reanalysis that often occurs in language. While a phonetic distinction might exist for a long period of time, it is only when another phonemic change occurs in the language that the existing phonetic distinction becomes important phonemically. For example, the Late Latin intervocalic /d/ (*cadere, fide, pede*) was weaker than the intervocalic /b/ (*bibere, habere, nube*), but speakers were quite likely unaware of the distinction for much of Late Latin. It was only when intervocalic /p,t,k/ became voiced as /b,d,g/ (*lupus~lobo, latus~lado, lacus~lago*), thus "crowding out" the pre-existing /b,d,g/, that we see that these intervocalic voiced stops behave differently in phonemic terms, with /d/ being more easily lost than /b/ (*cadere~caer, fide~fe, pede~pie* vs. *bibere~beber, habere~haber, nube~nube*) (cf. Lathrop 1984, Lloyd 1987).

In the same way that intervocalic /d/ was weaker than /b/ in Late Latin, yet at a merely phonetic level, there was a distinction between +auxiliary and -auxiliary verbs since Old Spanish, and it is only with several thousand tokens taken from large corpora that we are able to see this previously unnoticed "-etic" distinction. Later, in the same way that the 'strength' of the Late Latin stops became phonemically important only after a related phonemic shift had taken place, in the same way, the  $\pm$ auxiliary distinction only became "-emic" once the shift based on PREP-INF had died out. This view is of course different from that of Wanner (1982) and others, who have assumed that the "-etic" distinction between +auxiliary and -auxiliary verbs was new to the Spanish of the 1500-1600s.

If we accept for the moment the correctness of the scenario that we have just presented, we still need to account for the "narrowing" of the  $\pm$ auxiliary distinction from the 1700s to the 1800s (and written Spanish from the 1900s), which is precisely the period in which the newly "-emic"  $\pm$ auxiliary distinction should have caused the two types of verbs to diverge. Returning to the analogy of the Late Latin shift, it is reasonable to assume that once the original /p,t,k/ had become the new /b,d,g/, speakers would then be more conscious of the pronunciation of the words that had the original but now weakened /b,d,g/ (such as *nube, fide, legere*). There might be a period of "overcorrection", in which speakers would "restore" the lost consonant to words such as *fe* and *leer*. In the same way, once the PREP+INF motivation was lost and the only motivation for or against clitic climbing was the preexistent rule of  $\pm$ auxiliary, speakers might have overcorrected and shied away from the older (but still possible) tendency towards clitic climbing, which would lead to the narrowing of the  $\pm$ auxiliary distinction seen in Figure 3. It is not

unreasonable to assume that this would have been the most common in the written language (which is always more conservative). This would explain quite nicely why the  $\pm$ auxiliary contrast is so much less pronounced in written Modern Spanish than in spoken Modern Spanish. Furthermore, although there are obviously no data from spoken Spanish in earlier periods, we might assume that in the spoken register of the 1600–1800s the  $\pm$ auxiliary was also somewhat more pronounced than what is attested in the written texts of the same period.

Using the model presented here, we can explain a number of otherwise problematic data. First and foremost, we can explain why there has been a shift back towards clitic climbing, after a 500 year long shift towards final placement. In previous studies there had been no discussion of this reversal, because there were still no data-based studies such as Myhill (1988) and Davies (1995) to show the high degree of clitic climbing in Modern Spanish, especially in spoken Spanish. Second, the  $\pm$ auxiliary contrast, which is the mechanism for the reversal, is not ad-hoc. This is the same factor that, according to many researchers, is the principal factor governing clitic climbing in Modern Spanish. Finally, by focusing on factors such as overcorrection, the model explains why the  $\pm$ auxiliary contrast did not immediately reverse the shift towards final placement once the shift was completed with PREP-INF.

## 6. Conclusion

The goals of this study have been threefold. First, at a descriptive level, we wanted to map out with more precision than any previous study the shifts in Spanish clitic climbing from the 1200s to the 1800s. The data from the corpus indicate that there was a steady decrease in clitic climbing from the 1400s to the 1700s (after which it tapers off somewhat). Figure 1 shows, however, that there was no one period of dramatic change during this time. This is important in the sense that competing models of syntactic change make different predictions about the rate of syntactic change. For example, the ‘Principles and Parameters’ model of syntax (cf. Lightfoot 1991) predicts rather abrupt shifts in the grammar, whereas the ‘Typological-Functional’ model (cf. Croft 1990) predicts rather gradual shifts in the grammar. The data from this study, at least, tend to support the latter view.

A second goal of the study was to consider how the shifts in clitic climbing might have been influenced by changes in related constructions. We have shown that there was not a general shift with all INF-CL constructions, but rather that the shifts moved from PREP+INF to VERB+PREP+INF to VERB+INF. Thus there is an ‘extension’ of the syntactic shift through related constructions, which are all related at a surface level. This approach contrasts with the Principles and Parameters model, which hypothesizes changes in the settings for abstract ‘parameters’, which usually motivate a number of sudden, simultaneous changes in the language (cf. Lightfoot 1991).

A third goal of the study was to show how the same factors that influence clitic climbing in Modern Spanish may have also affected diachronic shifts in the construction. We have suggested that clitic climbing is very dependent on the  $\pm$ auxiliary nature of the main verb. This factor was at play in all stages of the language, but played a crucial role in reversing the 500 year long decrease in clitic climbing. In summary, we hope to have shown that large computer-based corpora, such as the one used in this study, can provide us with the necessary data to illustrate some of the basic mechanisms of syntactic change.

## NOTES

## 1 Composition of the corpus

The following is a list of the texts contained in the 118 text / 5,300,000+ word corpus of historical Spanish prose. The entries indicate the [date], *title*, {number of words}, and (publication information).

**1200 [14 texts / 776,700 words]**

[1200–1300] *Libro de los halcones* {1,200} (From the ADMYTE CD-ROM of Spanish texts (Vol. O). Other texts from this CD-ROM indicated with “(A)”; [1200–1300] *Tratado de la cetrería* {5,500} (A); [[1200–1284] *Fueros de Castilla* {77,200} (A); [1247] *Fueros de Aragón* {49,300} (A); [1250] *Libro de los animales de caza* {20,100} (A); [c1250] *Poridat de poridades* {15,600} (A); [1253] *Libro de los engaños* {18,100} (U NC P, 1959); [1270] *Estoria de España* {154,700} <Alfonso X> (Hisp. Sem. Med. Stud., U Wisconsin); [1270] *Historia troyana en prosa y verso* {26,200} (S. Aguirre, 1934); [1272–1280] *General estoria* {128,800} <Alfonso X> (Hisp. Sem. Med. Stud., U Wisconsin); [1280–1300] *Libro de los cient capítulos* {23,600} (Indiana UP, 1960); [1292–3] *Castigos y documentos para bien vivir* <Sancho IV> {76,500} (A); [1293] *Libro de consejo y de los consejeros* / Sem Tob {18,300}; [1295] *Gran conquista de Ultramar* {161,000} (A)

**1300 [10 texts / 744,200 words]**

[1300–1400] *Crónica de Sancho IV* / Fernán Sánchez de Valladolid {22,100} (A); [1300–1400] *Crónica de Alfonso X* / Fernán Sánchez de Valladolid {75,400} (A); [1300–1350] *Sumas de la historia troyana* / Leomarte {102,700} (A); [1310] *Leyes del estilo* {39,500} (A); [c1320?] *Crónica de veinte reyes* {140,800} (A); [1330] *El Conde Lucanor* / Don Juan Manuel {84,500}; [1342–1355] *Libro de la montería* <Alfonso XI> {112,200} (A); [c1350] *Historia troyana* / Guido de Columna {119,600} (A); [1355–1360] *Proverbios morales* / Sem Tob {8,900} (A); [1385–1388] *Libro de la caza de las aves* / Pedro López de Ayala {40,100} (A)

**1400 [15 texts / 765,200 words]**

[c1410] *Cuento de Tristán de Leonís* {76,000} (A); [c1412] *Suma de las crónicas de España* / Pablo de Santa María {32,400} (A); [c1419] *Menor daño de medicina* / Alfonso Chirino {50,500} (A); [1420–1433] *Retórica* / Tr. Alonso de Cartagena {65,100} (A); [1422] *De los oficios* / Tr. Alonso de Cartagena {53,100} (A); [1423] *Arte cisoría* / Enrique de Aragón {24,300} (A); [1438] *El Corbacho* {74,100} / Alfonso Martínez de Toledo (Ciceri, 1990); [1448] *Libro de las doñas* / Francesc Eiximenis {90,200} (A); [1453–1467] *Inventionario* / Alfonso de Toledo {78,400} (A); [c1454] *Espejo de medicina* / Alfonso Chirino {33,000} (A); [c1482] *Fábulas de Esopo* {73,800} (A); [c1484] *Breve confesionario* {9,700} (A); [1486] *Claros varones de Castilla* / Hernando de Pulgar {24,500} (Taurus, 1985); [c1490] *Imitación de Cristo* {60,800} (A); [1491] *Arnalte y Lucenda* / Diego de San Pedro {19,400} (A)

**1500 [19 texts / 745,300 words]**

[1492] *Cárcel de amor* / Diego de San Pedro {27,600} (Cátedra, 1974); [c1493] *Libro llamado Infancia Salvatoris* {43,700} (A); [1494] *De las mujeres ilustres en romance* {87,100} (A); [1495] *Grimalte y Gradissa* / Juan de Flores {27,500} (In *The Novels of Juan de Flores...*, Inst. French Studies, 1931); [1495] *Historia de Grisel y Mirabella* / Juan de Flores {18,400} (In *The Novels of Juan de Flores...*); [1498] *Glosa sobre Lux bella* / Domingo Marcos Durán {27,900} (A); [1499] *La Celestina* / Fernando de Rojas {74,100} (Espasa Calpe, 1972); [1508] *Amadís de Gaula* / García Rodríguez de Montalvo {65,100} (Cátedra, 1988); [1526] *Sumario de la natural historia de las Indias* / Gonzalo Fernández de Oviedo {47,700} (Historia, 1986); [1527] *Diálogo de las cosas acaecidas en Roma* / Alfonso de Valdés {26,400} (Cátedra, 1992); [1529] *Diálogo de doctrina cristiana* / Juan de Valdés {38,700} (Nacional, 1979); [1536] *Diálogo de la lengua* / Juan de Valdés {30,700} (Cátedra, 1987); [1542] *Relación, o Naufragios* / Alvar Núñez Cabeza de vaca {30,100} (Scripta Humanistica, 1986); [1549] *Relación de las comunidades de Castilla* / Pero Mejía {37,200} (Múnoz Moya y Montraveta, 1985); [c1550] *Los corsarios Barbarroja* / Francisco López de Gómara {30,000} (Polifemo, 1989); [1550–1560] *Brevísimas relación de la destrucción de Africa* / Bartolomé de las Casas {32,000} (Gobierno de Canarias, 1989); [1552] *Los amores de Clareo y Florisea* / Alonso Núñez de Reinoso {42,500} (Unex, 1991); [1553] *El crotalón* / Cristóbal de Villalón {38,400} (Cátedra, 1982); [1554] *Lazarillo de Tormes* {20,400} (Tarraco, 1976)

**1600 [16 texts / 701,100 words]**

[1556] *Guía de pecadores* / Fray Luís de Granada {37,200} (Planeta, 1986); [1557] *El patrañuelo* / Juan Timoneda {39,000} (Albatrós, 1987); [1571] *Las moradas del castillo interior* / Santa Teresa de Jesús {39,400} (Fraile, 1981); [1583] *De los nombres de Cristo* / Fray Luís de León {42,000} (Espasa-Calpe, 1991); [1583] *La perfecta casada* / Fray Luís de León {42,400} (Taurus, 1987); [1599] *Guzmán de Alfarache* / Mateo Alemán {60,100} (In *La vida de Lazarillo de Tormes ...*, Mundo Actual, 1982); [1604] *La vida del Buscón* / Francisco de Quevedo {42,400} (In *La vida de Lazarillo de Tormes ...*, Mundo Actual, 1982); [1605/1615] *Don Quijote de la Mancha* / Miguel de Cervantes {53,300} (Clásicos y Ensayos, 1977); [1609] *Comentarios reales de los incas* / Inca Garcilaso de la Vega {38,200} (Ayacucho, 1976, Vol. 1); [1612] *Dos novelas ejemplares (Rinconete y Cortadillo, La ilustre fregona)* / Miguel de Cervantes {34,700} (Harrap, 1971); [1617] *Los trabajos de Persiles y Sigismundo* / Miguel de Cervantes {31,100} (Castalia, 1969); [1618] *Vida de Marcos de Obregón* / Vicente Espinel {41,000} (Espasa-Calpe, 1972); [1637/1647] *Novelas amorosas, Desengaños amorosos* / María de Zayas {50,900} (Castalia, 1989); [1640] *Empresas políticas* / Diego de Saavedra Fajardo {44,800} (Planeta, 1988); [1646] *Vida y hechos de Estebanillo González* {49,700} (Doncel, 1972); [1651] *El criticón* / Baltasar Gracián {54,600} (In *Obras de B. Gracián*, Taurus, 1983)

**1700 [17 texts / 669,800 words]**

[1660] *El día de fiesta por la mañana y por la tarde* / Juan de Zabaleta {41,500} (Castalia, 1983); [1663] *Navidades de Madrid y noches entretenidas* / Mariana de Carvajal {39,900} (Franco Angeli, 1988); [1667] *El no importa de España* / Francisco Santos {56,800} / (Tamesis, 1973); [1686] *Historia de la conquista de México* / Antonio de Solís {23,500} (Porruá, 1978); [1726–1760] *Teatro crítico universal* and *Cartas eruditas* / Benito Jerónimo Feijoo {63,900} (In *Obras (selección)*, Taurus, 1985); [1737] *La poética* / Ignacio de Luzán {25,400} (Labor, 1977); [1743] *Vida* / Diego de Torres Villarroel {42,600} (Cátedra, 1980); [1760] *Historia del famoso predicador fray Gerundio de Campazas* / José Francisco de Isla {40,400} (Planeta, 1991); [1762–63] *El Pensador* / José Clavijo y Fajardo {23,100} (Islas Canarias, 1989); [1763–64] <Escritos económicos> / Gregorio Mayans y Siscar {63,500} (In anthology *Epistolario. Vol. 5: Escritos económicos*, Ayuntamiento de Oliva, 1976); [1774] *Cartas marruecas* / José Cadalso {41,200} (Aubi, 1978); [1780–1800] / Gaspar Melchor de Jovellanos {40,500} (In *Obras en prosa*, Castalia, 1970); [1781–87] *El Censor* {41,000} (Labor, 1972); [1789] *La derrota de los pendientes* / Leandro Fernández de Moratín {14,700} (Labor, 1973); [1793] *El rodrigo* / Pedro Montegón {38,200} (Inst Cult Juan Gil-Albert, 1990); [1794–1800] *Diario de Madrid* / Manuel Pardo de Andrade {35,300} (Fundación Pedro Barrie, 1989); [c1798] <Cartas y discursos> / Juan Meléndez Valdés {38,200} (In the anthology *Poesía y prosa*, Planeta, 1990)

**1800 España [13 texts / 425,500 words]**

[1833] *Las palabras* / Mariano José de Larra {32,500} (Austral, 1982); [1834] *Sancho Saldaña* / José de Espronceda {34,700} (Taurus, 1983); [1842] *Escenas matritenses* / Ramón Mesonero Romanos {34,600} (Planeta, 1987); [1844] *El señor de Bembibre* / Enrique Gil y Carrasco {37,800} (Magisterio, 1974); [1847] *Escenas andaluzas* / Serafín Estebañez Calderón {30,100} (Cátedra, 1985); [1852] <Selected prose> / Fernán Caballero {35,900} (In the anthology *El Alcázar de Sevilla, Simón verde, y otras relaciones*, Biblioteca de la cultura andaluza, 1985); [1874] *Pepita Jiménez* / Juan Valera {32,000} (Planeta, 1987); [1882] *La pródiga* / Pedro Antonio de Alarcón {30,900} (Nacional, 1975); [1884] *La regenta* / Leopoldo Alas “Clarín” {31,700} (32,000); [1886] *Los pazos de Ulloa* / Emilio Pardo Bazán {38,900} (Espasa Calpe, 1987); [1887] *Fortunata y Jacinta* / Benito Pérez Galdós {34,700} (Alianza, 1983); [1890] *La espuma* / Armando Palacio Valdez {30,200} (Castalia, 1990); [1895] *La puchera* / José María Pereda {31,600} (Castalia, 1980)

**1800 LatAm [14 texts / 550,600 words]**

[1832] *Don Catrín de la Fachenda* / José Joaquín Fernández de Lizardi {27,200} (In *Novelas selectas de Hispano América, Siglo XIX*, Labor Mexicana, 1959); [1839] *Cecilia Valdés, o la loma del ángel* / Cirilo Villaverde {29,000} (Cátedra, 1992); [1845] *Facundo* / Domingo Faustino Sarmiento {33,200} (Hispanica, 1961); [1846] *Guatimozín, último emperador de México* / Gertrudis Gómez de Avellaneda {61,100} (In *Novelas selectas*); [1863] *El ideal de un calavera* / Alberto Blest Gana {57,600} (In *Novelas selectas*); [1867] *María* / Jorge Isaacs {55,100} (In *Novelas selectas*); [1869] *Clemencia* / Ignacio Manuel Altamirano {50,300} (In *Novelas selectas*); [1879] *Enriquillo* / Manuel de Jesús Galván {61,200} (In *Novelas selectas*); [1882] *Carmen* / Pedro Casera {57,000} (In *Novelas selectas*); [1884] *Juvenilia* / Miguel Cané {24,800} (In *Novelas selectas*); [1885] *Amistad funesta* / José Martí {35,700} (In *Novelas selectas*); [1890] *La noche buena* / José Tomás de Cuéllar {11,800} (In *Novelas selectas*); [??1893] *De sobremesa* / José Asunción Silva {32,200} (In *Obras completas*, Banco de la República, 1965); [1899] *El donador de almas* / Amado Nervo {14,800} (In *Novelas selectas*)

- 2 A close comparison of Table 1 with Tables 2 and 3 show that the total number of tokens was calculated somewhat differently in Table 1 than in the subsequent tables. In Table 1, for example, there are 1865 tokens from the 1200s, while Table 2 shows just 1590. The rationale for this concerns sentences containing an initial *se*. In some of these sentences the *se* is reflexive, and we can assume that the clitic has climbed: *Juan se quiere acostar*. In the majority of the cases, however, it is the impersonal *se* (*se puede nadar, se debe considerar que...*), in which *se* is not the object of the embedded infinitive, and in which there has been no clitic climbing. If we were to count these cases of impersonal *se* as true clitic climbing, as most previous studies have done, we would have a higher percentage of initial placement. Rather than sort through the thousands of examples with *se* and categorize them by type, we simply assigned to initial *se* a value of 3/10 that of a true object clitic (*me pueden ver*). In other words, ten initial *se* only count as much as three initial non-*se* clitics. Since Table 1 does not deal with the differences between initial, medial, and final placement, there was no need to discount the tokens with impersonal *se*, and so the total number of tokens is higher than in Tables 2 and 3, where the adjustment is made.

## REFERENCES

- Aissen, Judith and David Perlmutter. 1983. “Clause reduction in Spanish.” *Studies in Relational Grammar*, ed. by David Perlmutter, 360–403. Chicago: Univ. of Chicago Press.
- Colburn, G.B. 1928. “The complementary infinitive and its pronoun object.” *Hispania*, 11, 424–29.
- Chenery, Winthrop H. 1905. “Object pronouns in dependent clauses: a study in Old Spanish word order.” *PMLA*, 20, 1–151.
- Contreras, Heles. 1979. “Clause reduction, the saturation constraint, and clitic promotion in Spanish.” *Linguistic Analysis*, 5, 161–82.
- Croft, William, et al. 1990. *Studies in Typology and Diachrony: Papers Presented to Joseph H. Greenberg on His 75th Birthday*. Amsterdam: Benjamins.

- Davies, Mark. 1995. "Analyzing Syntactic Variation with Computer-Based Corpora: The Case of Modern Spanish Clitic Climbing." *Hispania*, 78, 370–80.
- Granberg, Robert A. 1989. "Object pronoun position in Medieval and Early Modern Spanish." Unpublished dissertation, UCLA.
- Hanssen, Friedrich. 1913. *Gramática histórica de la lengua española*. Halle a.S.: M. Niemeyer.
- Keyne, Richard. 1989. "Null subjects and clitic climbing." *The null subject parameter*, ed. by O. Jaeggli and K. Safir, 239–261. Dordrecht: Kluwer.
- Keniston, Hayward. 1937. *The Syntax of Castillian Prose: The Sixteenth Century*. Chicago, Univ. Chicago Press.
- Lathrop, Thomas. 1984. *Curso de gramática histórica española*. Barcelona: Ariel.
- Leavitt, Walter D. 1954. The position of the object pronouns in Spanish. Unpublished dissertation. Yale.
- Lightfoot, David. 1991. *How to set parameters: arguments from language change*. Cambridge, MA: MIT Press.
- Lloyd, Paul M. 1987. *From Latin to Spanish. Vol 1: Historical Phonology and Morphology of the Spanish Language*. Philadelphia: American Philosophical Society.
- Lujan, Marta. 1980. "Clitic promotion and mood in Spanish verbal complements." *Linguistics*, 18, 381–484.
- Menéndez Pidal, Ramón. 1954. *Cantar de Mío Cid: texto, gramática, y vocabulario*. Vol 1. 3rd ed. Madrid: Espasa Calpe.
- Myhill, John. 1988. "The grammaticalization of auxiliaries: Spanish clitic climbing." *BLS* 14, 352–63. Berkeley: Berkeley Linguistics Society.
- Napoli, Donna Jo. 1982. "Semantic interpretation vs. lexical governance." *Language*, 57, 47–69.
- Pizzini, Quentin A. 1982. "The Positioning of Clitic Pronouns in Spanish." *Lingua*, 57, 47–69.
- Ramsden, Herbert. 1963. *Weak pronoun position in the early Romance languages*. Manchester: Manchester Univ Press.
- Rivero, María Lusia. 1991. "Clitics and NP climbing in Old Spanish." *Current Studies in Spanish Linguistics*, ed. by Fernando Martínez Gil and Hector Campos, 241–82. Washington, D.C.: Georgetown Univ Press.
- Rosen, Sara Thomas. 1989. *Argument Structure and Complex Predicates*. Unpublished dissertation. Brandeis University.
- Spaulding, Robert K. 1927. "Puedo hacerlo versus lo puedo hacer and similar cases." *Hispania*, 10, 342–48.
- Suñer, Margarita. 1980. "Clitic promotion in Spanish revisited." *Contemporary studies in Romance linguistics: proceedings of the Eighth Annual Symposium on Romance Languages*, ed. by Frank Nuessel, 300–330. Bloomington, IN.: IULC.
- Wanner, Dieter. 1982. "A history of Spanish clitic movement." *BLS* 8, 135–47. Berkeley: Berkeley Linguistics Society.
- Zagona, Karen. 1986. "Evidencia para complementos SV en español." *Aproximaciones a la sintaxis del español: Estudios sintácticos del español y el progreso de la teoría lingüística*, ed. by Magdalena García Pinto and Mario A. Rojas, 149–165. Barcelona: Puvill.