Analyzing Syntactic Variation with Computer-Based Corpora:
The Case of Modern Spanish Clitic Climbing

Mark Davies
Illinois State University

Abstract: A large computer-based corpus provides the data for the first comprehensive investigation of clitic climbing (*lo debe comprar vs. debe comprarlo*) in both written and spoken modern Spanish. The results are based on nearly 15,000 tokens with 32 different main verbs (*ir+a, tener+que, desear,* etc), extracted (using *WordCruncher*) from a computer corpus of 3.5 million words of spoken and written Spanish from ten countries. The large amount of data from both registers permits consideration of previously problematic questions. For example, the data suggests that a semantics-based model accounts well for the continuum-like distribution of clitic climbing with different main verbs, and that the nature of the clitic (reflexivity/animacy) is also important. Although the frequency of clitic climbing varies little from country to country, it is much more common in spoken than in written Spanish, which in turn poses major new questions concerning diachronic Spanish clitic climbing.

Key Words: Spanish language, clitics, clitic climbing, syntax, syntactic variation, corpus-based approach

1. Introduction

One of the most persistent topics of research in Spanish syntax in the past two decades has been that of "clitic climbing", which is the apparently optional movement of unstressed object pronouns from a postverbal (a) to a preverbal position (b), in the context of infinitives subordinated to modal-like main verbs:

(1a) *ella quería* [*comprarlos*]
(1b) *ella los quería* [*comprar___*]

There are two issues dealing with clitic climbing that have been the focus of most previous research. First, investigators such as Contreras (1979), Luján (1980), Napoli (1982), Suñer (1980), Pizzini (1982), and Rosen (1989) have sought to explain why clitic climbing is not possible with all verbs, as it is in (b) above. For example, as the data from the study will show, with many verbs clitic climbing is only partially (a) or completely unacceptable (b):

(2a) ?? *ella los deseaba* [*comprar___*]
(2b) * *ella los anhelaba* [*comprar___*]

Beyond the question of why clitic climbing occurs only with certain verbs, researchers have addressed the more general question of why any clitic climbing at all is possible. A common premise of recent syntactic theory is that there are tight constraints on the movement of elements across clausal boundaries, and the seemingly problematic movement of clitics across such a boundary needs to be explained:

(3) *ella los quería* [*comprar___*]
         | | (clausal boundary)

Some generative grammarians such as Rizzi (1982) theorize a type of clausal reconstruction, in which the clausal boundary is removed and movement can occur. Other generativists such as Luján (1980) and Kayne (1989) suggest that the boundary remains intact, but that the absence or presence of other structural elements either permit or disallow clitic climbing. In the Relational Grammar model (Aissen and Perlmutter 1983), clitic climbing data provides important evidence for a supposedly

universal rule of Clause Union, in which the Argument Structure of both clauses is unified, and all lower clause elements (clitics, in this case) become main clause elements. Regardless of the particular model, explaining clitic climbing has been an important goal of several contemporary syntactic theories.

1.1 Given the potential that clitic climbing has for syntactic theory, one would hope that previous researchers would have collected extensive data for Modern Spanish on which to base their analyses, but unfortunately this is not the case. Nearly all of the previous studies have been theory-oriented, and have provided valuable insight into why even limited clitic climbing occurs in different languages. Nevertheless, because the research questions under consideration often relied on the introspective judgments of one or two researchers, these studies have typically dealt very little with the actual frequency of clitic climbing in the community at large. Yet even among the theory-oriented studies, there are a few that have looked at data beyond the level of the individual researcher. Suñer (1980), for example, bases her conclusions on intuitions from ten native speakers, and Farrán (1990–91) bases his survey to sixteen native speakers from Puerto Rico.

Most of the large scale data-based studies, however, were conducted more than fifty years ago. For example, Colburn (1928) looked at clitic climbing in short selections from ten novels from the 1800s, and Spaulding (1927) expanded the scope somewhat to also include some works from the early 1900s. The most ambitious study was Keniston (1937a), which briefly considered clitic climbing in a more general investigation of syntactic variation in selections from sixty prose works covering 1900–1930. The only extensive data-based study during the past fifty years has been Myhill (1988), which is based on fairly lengthy passages from seven novels by five recent authors. In summary, although there has been some data-based research on clitic climbing in written prose, most of the research is based

on data from eighty to one hundred years ago. More serious is the fact that to date there has been no study made of clitic climbing based on any corpus of spoken Spanish. It is not difficult to see why researchers have found it hard to explain Spanish clitic climbing, when in fact there has been little understanding about how widespread it is and what type of variation exists.

2. Data base for Spanish clitic climbing

The purpose of the present study is to provide the first comprehensive data base of clitic climbing in Modern Spanish, based on corpora from both the spoken and the written registers. A secondary focus will be to consider how this new data helps to clarify several issues that have previously been quite problematic, such as measuring the acceptability of clitic climbing based on the nature of the governing verb or the nature of the clitic itself. Regarding the composition of the data base, the spoken corpus on which the study is based on the complete *"Habla culta"* corpus, which comprises more than 2,500,000 words from 402 conversations with speakers from eleven different cities in ten countries in Latin America and Spain. The size of the eleven corpora ranges from 68,000 to 389,000 words each, with an average size of over 200,000 words. The goal was to also create a large written corpus, achieved by compiling a corpus of more than 1,000,000 words of prose text from 356 short stories by twentieth-century authors from the same ten countries as the *"Habla culta"* corpus. Each of these ten corpora was comprised of approximately 100,000 words. (Note 1 gives a complete listing of the *"Habla culta"* cities and the countries represented in the *"Cuento"* corpus). The advantage of basing our corpus on the same countries for both spoken and written Spanish is that this will permit more accurate comparisons of clitic climbing in the two registers. To extract the clitic climbing data from the 3.5 million word corpus, the texts were scanned into a computer, and then indexed with the *WordCruncher* text

retrieval program. Once an every-word index of the corpus was constructed, the *WordCruncher* program could then carry out rather complicated Boolean and proximity searches on the data. I extracted all cases of any one of thirty two verbs followed by an infinitive, which in turn was either preceded (a) or followed (b) by a clitic pronoun.

(4a) *"El cumplimiento ante todo, señores",*
nos *solía recalcar a diario.*
(Chile2:Bombal:103)
(4b) *Solía comentar***me** *las cartas que recibía.* (Cuba1:Leante:389)

The list of the thirty two verbs under investigation was based on those verbs that were most often included in previous research, as well the most frequently appearing verbs in the corpus. The examples were then further categorized (by main verb) and submitted to distributional analyses, producing the results presented. In total, nearly 15,000 tokens of potential clitic climbing constructions were extracted (as seen in 4a–4b), including more than 10,000 from the *"Habla culta"* corpus and more than 4,000 from the *"Cuento"* corpus.

2.1 The large data base makes it possible to address several aspects of variation in Spanish clitic climbing that have not been adequately treated until this time. First, Section 3 will consider whether clitic climbing is more common in either the spoken or the written register. Some researchers have suggested that it is more common in spoken Spanish, but without a corpus of spoken Spanish there has previously been no way to test this hypothesis. Related to this sociolinguistic variation is the question of whether the frequency of clitic climbing with all verbs varies from one geographical dialect to another (Section 4).

Section 5 treats the question that most previous research has centered on, which is the variation resulting from different main verbs (such as *ir+a, tener+que, desear,* etc.). Particular attention focuses on for the "degrees" or "shades" of acceptability of clitic climbing found with given main verbs. Finally, Section 6 considers other syntactic factors (besides the main verb) that are at play in conditioning clitic climbing, but which have rarely been looked at in previous research. These include the nature of the clitic itself (single vs. multiple clitics, reflexive vs. nonreflexive, and animate vs. inanimate referent) as well as the syntactic environment in which the construction occurs.

3. Variation by register

Most previous research has been based solely or largely on the intuitions of a single speaker or a very limited number of speakers, and there has to date been no large-scale corpus-based investigation for spoken Spanish. As a result, it has been impossible to study whether the phenomenon is more common in one register than in another, although Suñer (1980:328) and others have hypothesized that it is more common in the spoken register. The data from the corpus confirms that this is the case. As the figures from Table 2 below indicate, the average for [+CC] in the spoken register is nearly three times more common than in the written register (eg. *acabar* 29% vs. 85%, *volver+a* 27% vs. 81%). The only exception is the verb *pasar+a,* in which 1/6 tokens in the written register is [+CC], but 0/5 in the spoken register. Nevertheless, there are only 11 cases with *pasar+a,* as opposed to nearly 15,000 with all verbs combined. Note also that *ir+a* is the only verb to take [+CC] in the majority of the cases in both registers and the only verb with which the written language approaches spoken Spanish. In summary, one can clearly say that [+CC] is more common in spoken than in written Spanish. Assuming the uncontroversial notion that the spoken register of a language represents the more popular tendencies of a language than the conservative written register, [+CC] represents a popular (rather than conservative) tendency of Spanish.

4. Variation by geographical dialect

According to (Sankoff 1988), dialectal differences in syntax are generally found at

the register level (vertical dialects) rather than in different geographical divisions (horizontal dialects). Therefore, one should not expect to find nearly as much variation in [+CC] from one city or country to another as between the spoken and written registers. The data seen in Table 1 confirms the hypothesis that the difference in [+CC] between the eleven cities is quite slight. These figures are based on the average [+CC] for the eight most common verbs, which account for 83% of all tokens in the corpus. Note that the highest degree of [+CC] in the spoken register is found in *México*, which is the only city for which we have a *Popular* in addition to a *Culto* corpus. Note also that there is a difference of only four percent separating the seven cities at the lower end of the table.

Table 1. %[+CC] by city, country

	Average	Spoken	Written
México (C/P)	0.46	0.66*	0.26
España (M/S)	0.45	0.61*	0.28
Puerto Rico	0.41	0.56	0.26
Cuba	0.37	0.47	0.27
Colombia	0.36	0.45	0.28
Argentina	0.36	0.59	0.14
Venezuela	0.34	0.53	0.15
Chile	0.34	0.53	0.15
Perú	0.34	0.41	0.26
Bolivia	0.33	0.43	0.22
AVERAGE	0.40	0.56	0.23

* *México (C/P)* refers to the average of *México Culto* and *México Popular*, while *España (M/S)* refers to the average of *Madrid* and *Sevilla*.

5. Variation by verb

As previously noted, very little research has dealt with gradations or shades of acceptability of [+CC] with given verbs. Typically, statements regarding acceptability have been given in binary terms; in other words, a verb either allows or disallows clitic climbing. This study is the first to consider in-depth the "degree of acceptability" of [+CC] with given verbs in both the spoken (and written) register. The results are presented in the following table, which gives for each register the %[+CC] with each verb and the corresponding number of tokens for that verb.

Reagrding the influence of the main verb, Table 2 indicates that verbs do not divide neatly into strictly [+CC] or [-CC] classes. Rather, there is a continuum ranging in spoken Spanish from *ir+a* (86% [+CC]) to *poder* (60%) to *tener+que* (38%) to *intentar* (11%) to *esperar* (0%). The problem is how to account for this continuum-like nature of main verbs, which is an issue has already been addressed in three important studies of Romance clitic climbing (Napoli 1982, Myhill 1988, and Rosen 1989). Their suggestion is that we adopt a meaning or semantics-based model, in which clitic climbing is not binary in nature ([+CC] vs. [+CC]), but one in which the frequency of the construction with different verbs is almost predicted to range over the entire continuum of [+CC] to [-CC].

In the Napoli model, [+CC] is a function of how "auxiliary-like" the main verb is, with highly auxiliary verbs allowing clitic climbing more than weak auxiliaries or non-auxiliary verbs. She defines auxiliary verbs as verbs that "offer supplementary information about the action or state [of the embedded verb]" and that "add conceptually basic or simple information". Myhill casts the issue in terms of "grammaticalization" of the main verb. Those main verbs that mainly carry aspectual information or possess modal-like meaning are the verbs that most allow [+CC] (353–57). Rosen (1989), working from within the generative paradigm, makes reference to "light" vs. "heavy" verbs, in which the distinction arises from the meaning of the verb, and is mapped directly onto syntax.

As for how these semantics-based accounts relate to the data from the corpus, Table 2 shows that of the fifteen verbs that most easily allow [+CC], nearly all of them fall into the auxiliary/aspectual/modal set of verbs. For example *ir+a, acabar+de, volver+a, soler, venir+a, terminar+de, empezar+a, llegar+a, dejar+de,* and *comenzar+a* all serve as aspectual markers (5–6). The modals include *haber+de* and *tener+que* (which express obligation), *poder* and *saber* (ability) and *querer* (desire). The one noticeable exception to this rule is *hay*

Table 2. %[+CC] by verb and register

VERB	AVERAGE/#		SPOKEN/#		WRITTEN/#	
ir+a	0.76	(3421)	0.86	(2838)	0.66	(583)
acabar+de	0.57	(143)	0.85	(92)	0.29	(51)
haber+de	0.56	(134)	0.80	(65)	0.33	(69)
volver+a	0.54	(379)	0.81	(159)	0.27	(220)
soler	0.46	(60)	0.86	(16)	0.05	(44)
poder	0.41	(4764)	0.60	(3633)	0.22	(1131)
venir+a	0.36	(237)	0.55	(129)	0.18	(108)
saber	0.33	(164)	0.52	(119)	0.14	(45)
terminar+de	0.31	(23)	0.55	(13)	0.08	(10)
querer	0.31	(1265)	0.47	(798)	0.15	(467)
empezar+a	0.29	(379)	0.45	(221)	0.13	(158)
llegar+a	0.27	(177)	0.42	(131)	0.11	(46)
dejar+de	0.25	(66)	0.41	(21)	0.10	(45)
tener+que	0.25	(1230)	0.38	(984)	0.13	(246)
comenzar+a	0.25	(219)	0.45	(75)	0.05	(144)
aprender+a	0.22	(24)	0.45	(19)	0.00	(5)
lograr	0.21	(121)	0.41	(45)	0.02	(76)
necesitar	0.20	(63)	0.33	(36)	0.07	(27)
deber	0.20	(714)	0.32	(473)	0.08	(241)
salir+a	0.17	(17)	0.35	(9)	0.00	(8)
pensar	0.15	(98)	0.24	(54)	0.07	(44)
desear	0.12	(39)	0.20	(14)	0.04	(25)
tratar+de	0.10	(216)	0.20	(110)	0.01	(106)
pasar+a	0.08	(11)	0.00	(5)	0.17	(6)
preferir	0.07	(51)	0.15	(30)	0.00	(21)
procurar	0.07	(37)	0.14	(18)	0.00	(19)
intentar	0.06	(52)	0.11	(11)	0.01	(41)
resolver	0.03	(13)	0.05	(7)	0.00	(6)
esperar	0.00	(18)	0.00	(14)	0.00	(4)
insistir+en	0.00	(6)	0.00	(1)	0.00	(5)
soñar+con	0.00	(1)	0.00	(1)	0.00	(4)
haber+que	0.00	(485)	0.00	(385)	0.00	(100)

que, which expresses simple deontic modality, and yet never allows clitic climbing (7). Although further research is needed, the unexpected nature of *hay+que* may be due its origin as a verb of possession plus a locative adverbial element (*HABER* + *Y*), which sets it apart from the more lexically simple verbs.

(5a) *como te decía antes, y lo vuelvo a repitir* ... (Sevilla:16:197)
(5b) *El hace muchos años me vino a ver* (Buenos Aires:33:510)

(6a) *Lo pavimentaron, lo terminaron de pavimentar hace como tres meses atrás* (Santiago:41:223)
(6b) *Y los frecuenta todavía, o ya los dejó de ver?* (México-Popular:6:78)
(7) *Hay que darles buenas oportunidades* (México-Culto:16:213)

Compare now the auxiliary/aspectual/modal nature of these preceding verbs with that of the verbs that least commonly allow clitic climbing, which are verbs that are semantically complex and less modal-like:

tratar+de, preferir, intentar+de, insistir+en, etc.

(8a) *Taquito se acercó al grupo y trató de mirar**la** de la misma manera en que todos la habían mirado.* (Perú:Echenique:361)
(8b) *Y ella no la disimulaba ni intentaba disminuir**la*** (Venezuela1:Mariño Palacio:183)

To further compare the two groups of verbs, consider the series of verbs that expresses desire: *querer, desear, preferir, soñar+con.* As the meaning becomes "conceptually less basic" (cf. Napoli), the percentage of [+CC] decreases accordingly (47%–20%–14%–0%).

(9a) *Estoy trabajando en varias comunidades, y hasta **me** han querido hacer comunera.* (Lima:21:278)
(9b) *Como todo proceso que **se** desea apagar, el tratamiento que me hice yo fue colocarme hielo* (Santiago:22:373)
(9c) *Ahora prefieren ir**se** a bailar, que a go go y la cuestión esa* (México-Culto:32:443)
(9d) *Y no sólo eso, sino que eran codiciadísimas y los mendigos ni soñaban con poner**les** manos encima.* (México1:Puga:272)

In summary, the data from the corpus suggests that a semantics-based account like that already proposed by Napoli (1982), Myhill (1988), and Rosen (1989) provides clues about which type of verbs are most likely to allow clitic climbing, and that this account also allows quite nicely for the continuum-like nature of the phenomena. Although these models, being meaning-based, cannot give precise quantitative statements regarding the probability of [+CC] with any given verb, they much more accurately account for the continuum-like nature of the phenomena than models that simply mark a given verb as [+CC] or [-CC]. Finally, while this is not the first study to suggest this type of semantics-based account for clitic climbing, it does provide strong support in favor of such a model.

6. Variation by other syntactic factors

While most past research has focused on the verb as the principal syntactic or semantic factor influencing clitic climbing, there have been occasional references to other syntactic factors. This section considers four of these: single vs. multiple clitics, reflexive vs. nonreflexive clitics, animate vs. inanimate clitics, and the nature of the syntactic material immediately preceding the clitic climbing construction. The data from the corpus will confirm that each of these factors impacts in some way on the probability that the clitic will climb to the main verb.

6.1 Multiple clitics

Spaulding (1927:346) claims that when multiple clitics are involved (both of them objects of the embedded infinitive) there is a greater probability of [+CC], a claim disputed by Colburn (1928:428). Keniston (1937a:71–72) is the only study to quantify the effect, and shows that [+CC] occurs in 34/45 (75%) of the cases in which there are multiple clitics. In more recent studies this factor has received less attention and the common assumption (cf. Aissen and Perlmutter 1983:365, Luján 1980:383) is that two clitics can as easily stay attached to the embedded infinitive ([-CC]) as can one single clitic:

(10a) *quiero mostrár**telos*** (Aissen and Perlmutter 1983:365)
(10b) *quisiera poder dár**telo*** (Luján 1980:383)

The data from the corpora, however, show that there is a higher incidence of [+CC] with multiple clitics (11a) than with single clitics. In other words the type of construction seen in (11b) is more marked than if there was one clitic.

(11a) *El estudiante sabe que no va a pasar el rato porque ella sencillamente no **se lo** va a permitir* (San Juan:15:296)
(11b) *Los dioses le habían regalado una hermosa vida y quería retornár**sela** antes de matar**la** en su hora.* (Cuba2:Pastora Fernández:54)

Table 3 below summarizes the data for single vs. multiple clitics, as well as the effect of reflexivity, animacy, and the syntactic environment. The table is based on the eight most frequent main verbs, which account for about 83% of all tokens. Regarding single and multiple clitics, the first section of Table 3 shows a difference of 87% (multiple clitics) vs. 68% (single clitic) in spoken Spanish and 62% vs. 31% in written Spanish. Both of these are statistically significant at the .001 level (chi-square values of x=64.87 and x=56.24, respectively).

6.2 Reflexivity

Rosen (1989:181–90) devotes considerable attention to the question of whether [+CC] is equally as common with reflexive and nonreflexive clitics, and concludes that reflexive clitics tend to remain attached to the infinitive, since they are part of the argument structure (i.e. lexicon) of the verb:

(12a) OK *los chicos empezaron a caerse (unos sobre otros)*
(12b) ?? *los chicos se empezaron a caer (unos sobre otros)*

To date there have been no corpus-based studies that have investigated this phenomenon. This study confirms that there is a lower degree of [+CC] with reflexive clitics (13a) than with nonreflexive clitics. In other words, the lexically dependent clitic tends to mirror that dependency in the syntax by staying with the governing infinitive (13b).

The data is summarized in the second section of Table 3, in which the difference between reflexive and nonreflexive clitics is significant at the .001 level.

(13a) *"Mamá se tiene que ir mañana a la ciudad a donde llegamos primero.* (Puerto Rico1:García Ramis:115)
(13b) *¿No le da miedo que vaya a convertirse en un jipi?* (México-Popular:21:286)

6.3 Animacy

Myhill is the only researcher to have looked at how the animacy of the clitic affects clitic climbing, and he finds that animate clitics climb more easily than nonanimate clitics (1988:357–61). For example, *te/os* have 73% [+CC], *me/nos* = 47%, third person animate = 33%, and third person nonanimate = 18%. Myhill attributes this to the fact that there is a relationship between animacy and topicality, and highly topical material ("what were talking about"; often "you" (*te/os*), "us" (*nos*), or "me" (*me*)), tends to be placed towards the beginning of the sentence. This study has followed Myhill's four-way categorization of *te/os*, *me/nos*, third person animate, and third person nonanimate. Although it would have been possible to do so, I have not broken the categories down farther into "person" distinctions like *me* vs. *nos*, both because we want the data to be correlated to Myhill's, and because the four-way categorization already provides evidence for the general effect of "animacy."

Table 3. %[CC] by nature of the clitic and syntactic environment

VARIABLE	SPOKEN			WRITTEN		
	%	[+CC]	[-CC]	%	[+CC]	[-CC]
MULTIPLE						
2 clitics	.87	369	57	.62	89	55
1 clitic	.68	5647	2643	.31	796	1727
REFLEXIVE						
-reflexive	.73	2671	967	.35	431	796
+reflexive	.66	836	440	.27	111	296
ANIMATE						
+animate	.76	2601	1044	.39	466	740
-animate	.62	1222	734	.27	184	493
PRECEDING ELEMENT						
que	.79	507	131	.32	45	96
y	.73	131	71	.30	96	55

The data in the third section of Table 3 indicate that Myhill's hypothesis is confirmed in both the spoken and written portions of the corpus. Clitics with an animate referent (14a) tend to climb more easily than clitics with a nonanimate referent (14b), with the difference again being significant at the .001 level.

(14a) *pero como sólo estaba apoyado, con todo su peso, sobre un codo, los brazos **le** empezaron a temblar.* (México1:Elizondo:203)
(14b) *Como son tan pocos, pues no **lo** pueden hacer, en realidad* (Madrid:23:433)

6.4 Preceding material

One final syntactic factor of Modern Spanish clitic climbing that has not been studied by any previous researcher, but which nevertheless impacts on clitic climbing, is the nature of the syntactic material that immediately precedes the clitic climbing construction. Several researchers have looked at this issue for older stages of Spanish (Ramsden 1963:55–103 and Keniston 1937b:89–96), and they report that in Old and Middle Spanish, clitics with finite verbs nearly always took preverbal placement when the material preceding the verb was a subordinating conjunction (15a). When it was a coordinating element, on the other hand, placement was usually postverbal (15b).

(15a) *que se pudiessen dantes esparzer por la tierra*
(15b) *et dixieron le que uieran una animalia*

Wanner (1991) and Alemán (1985), among others, attribute this to the fact that in older stages of the languages the clitics tended to "rest" on the preceding element (enclisis), and that conjunctions provided a better "landing site" than coordinating elements. Although there are obviously other syntactic elements (such as subjects, adverbials, etc.) that can occupy this preceding slot, these researchers note that the greatest contrast is between subordinating and coordinating conjunctions, as seen in (15a–b).

In applying these Old and Middle Spanish facts to Modern Spanish clitic climbing, this study finds that the difference based on the preceding element is still statistically significant, albeit at only the .05 level. The fourth section of Table 3 shows the degree of [+CC] when the preceding element is *QUE* (the most common subordinating conjunction, =16a) and when it is *Y* (the most common coordinating element, =16b).

(16a) *Y si Ud. va allá pensando **que** se va a curar eso naturalmente lo... lo cura* (Habana:15:489).
(16b) *y **me** tengo que retirar un poco y aplastar aún más la cara contra el soleado cristal* (España1:Sueiro:158).

As mentioned, the investigation in this case treats only one subordinating conjunction (*que*) and one coordinating conjunction (*y*), where, based on Old and Middle Spanish data, we would expect the effect to be the greatest. Future researcher might consider the effect of other preceding elements, where the effect would most likely be more subtle.

7. Conclusion

The size and completeness of the study makes it possible to gain some insights into a number of questions that could not be fully answered in previous studies. The data suggests that a semantics-based model such as Napoli (1982), Myhill (1988), and Rosen (1989) allows quite well for the continuum-like nature of clitic climbing, as well as suggesting what classes of verbs would be most like to allow it. The study also confirms hypotheses of some previous researchers about the influence of the clitic itself and the syntactic environment in which the construction occurs.

7.1 Perhaps the most challenging question relating to clitic climbing, however, is accounting for the difference in acceptability between the spoken and the written registers, and suggesting how this fits in with the historical trajectory of clitic climbing in
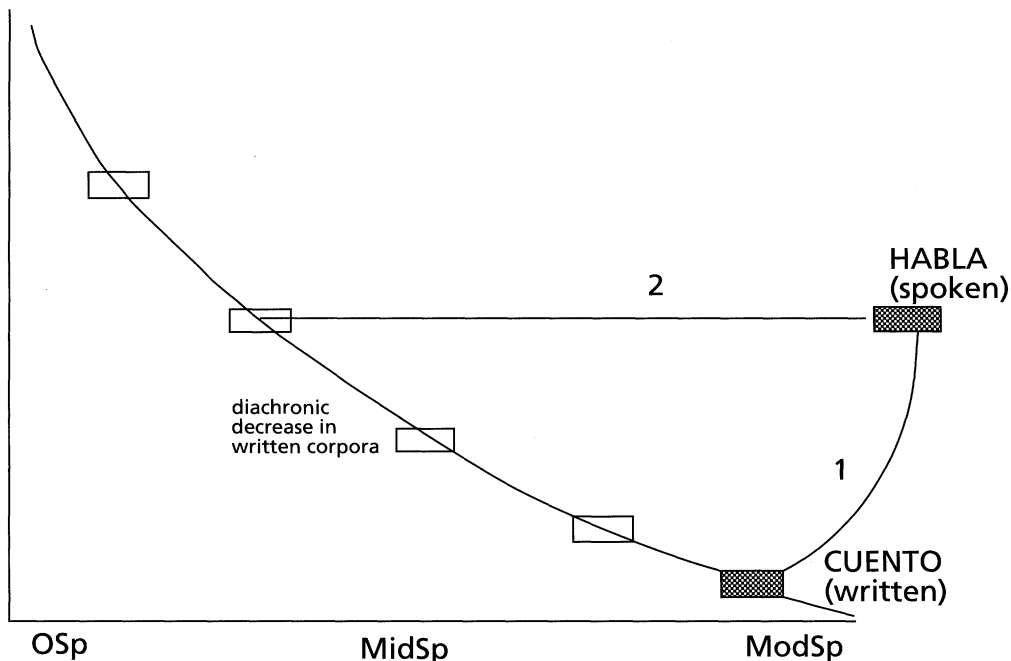
Figure 1

Spanish. This question, of course, could not have been asked until the present time, in which there exists for the first time reliable data on the extent of [+CC] in the two registers. Concerning the historical shifts, Wanner (1982) shows conclusively that in written Spanish there has been a gradual but continual decrease in [+CC] from Old Spanish to the present time. Therefore even in Modern Spanish there should be evidence of decreasing [+CC]. The handful of previous corpus-based studies do confirm the very gradual decrease in clitic climbing in written Spanish even during the past 150 years (cf. Spaulding 1927, Colburn 1928, Keniston 1937, Myhill 1988). This decrease in [+CC] since Old and Middle Spanish is indicated in Figure 1.

Assuming the non-controversial position that spoken Spanish more accurately reflects the popular tendencies of the language (vs. the conservative tendencies of the written language), this would suggest that clitic climbing should be less common in spoken Modern Spanish than in written Modern Spanish. Surprisingly, however, the corpus shows just the opposite to be true; clitic climbing is more common in the spoken than in the written register. To account for this, one must assume one of two things. The first possibility (#1 in Figure 1) is that there has been a very recent and dramatic increase in clitic climbing, reversing the historical trend, and this is evidenced as of yet only in the spoken register. In this scenario written Spanish still has a low degree of [+CC], but soon it will be at the level of spoken Spanish. Assuming the correctness of this model, one would want to explain why there has been such a dramatic reversal in [+CC], and explain why no other studies have found evidence for this.

A more plausible account, perhaps, is that there has been a syntactic bifurcation (probably since Middle Spanish), in which clitic climbing has become progressively less common in the written register, but has remained popular in spoken Spanish, at about the Old and Middle Spanish levels (#2 in Figure 1). This would allow for the difference between the two registers in Modern Spanish and yet suggests that the long-term diachronic shifts towards decreased [+CC] in the written register have continued on, with no sudden reversals in direction. The logical question, obviously, is why there

was a syntactic split between the two registers at an older stage of Spanish, a question that invites further research.

■ NOTES

[1]The *"Habla culta"* texts utilized in the study are: **Bogotá**: *El habla de la ciudad de Bogotá. Materiales para su estudio.* Eds. H. Otálora de Fernández y A. González G. Bogotá: Instituto Caro y Cuervo, 1986; **Buenos Aires**: *El habla culta de la ciudad de Buenos Aires. Materiales para su estudio.* 2 vols. Ed. A. M. Barrenechea. Buenos Aires: Universidad Nacional de Buenos Aires, 1987; **Caracas**: *El habla culta de Caracas. Materiales para su estudio.* Eds. A. Rosenblat y P. Bentivoglio. Caracas: Universidad Central de Venezuela, 1979; **Habana**: From thirty interviews taken from an unpublished manuscript in possession of Prof. Joseph Matluck, Univ. Texas-Austin. **La Paz**: *El habla de la ciudad de La Paz. Materiales para su estudio.* Ed. N. G. Marrone. La Paz: Ediciones Signo, 1992; **Lima**: *El español de Lima. Materiales para el estudio del habla culta.* Ed. R. Caravedo. Lima: Pontificia Universidad Católica del Perú-Fondo Editorial, 1989; **Madrid**: *El habla de la ciudad de Madrid. Materiales para su estudio.* Eds. M. Esgueva y M. Cantarero. Madrid: C.S.I.C., 1981; **México Culto**: *El habla de la ciudad de México. Materiales para su estudio.* Ed. J. M. Lope Blanch. México, D.F.: UNAM, 1971; **México Popular**: *El habla popular de la ciudad de México.* Ed. J. M. Lope Blanch. México, D.F.: UNAM, 1976; **San Juan**: *El habla culta de San Juan. Materiales para su estudio.* Eds. A. Morales y M. Vaquero. Rio Piedras: Editorial de la Universidad de Puerto Rico, 1990; **Santiago**: *El habla culta de Santiago de Chile. Materiales para su estudio.* Eds. A. Rabanales y L. Contreras. Vol 1. Santiago: Universidad de Chile, 1979; Vol 2: Bogotá: Instituto Caro y Cuervo, 1990; **Sevilla**: *Sociolingüística andaluza 2. Material de encuestas para el estudio del habla urbana culta de Sevilla.* Eds. V. Lamiquiz y M. A. de Pineda. Sevilla: Universidad de Sevilla, 1983. I would like to thank George DeMello for his considerable assistance in acquiring these *"Habla culta"* materials. [Citations refer to Ciudad:Muestra:Página]

The short stories are all by authors born after 1900, whose works are found in the following: **Argentina**: *Antología de cuentistas argentinos.* Buenos Aires: Fondo Editorial Bonarense, 1979; **Bolivia 1**: *Los mejores cuentos bolivianos del siglo XX.* Ed. Ricardo Pastor Poppe. La Paz: Los Amigos del Libro, 1989; **2**: *Antología de cuentos: II Concurso nacional.* La Paz: Los Amigos del Libro, 1968; **Chile 1**: *Antología del nuevo cuento chileno.* Ed. Enrique Lafourcade. Santiago: Zig-Zag, 1954; **2**: *"Cuento"s de cabecera.* Ed. Hernán Poblete Varas. Santiago: Zig-Zag, 1967; **Colombia 1**: *El cuento colombiano.* Ed. Eduardo Pachón Padilla. Vol 1. Bogotá: Plaza y Janes, 1980; **2**: *Premios que cuentan.* Bogotá: Universidad Central, 1988; **Cuba 1**: *"Cuento"s cubanos del siglo XX.* Ed. Salvador Bueno. Habana:

Arte y Literatura, 1975; **2**: *Narradores cubanos de hoy.* Ed. Julio E. Hernández-Miyares. Miami: Universal, 1975; **España 1**: *22 narradores españoles de hoy.* Ed. Félix Grande. Caracas: Aviles, 1990; **2**: *"Cuento" español de posguerra.* Ed. Medardo Fraile. Madrid: Cátedra, 1986; **México 1**: *Los mejores cuentos mexicanos.* Ed. Gustavo Sainz. Barcelona: Océano, 1982; **2**: *Antología de cuentos mexicanos.* Ed. María del Carmen Millán. Vol 1. México, D.F.: Nueva Imagen, 1977; **Perú**: *El cuento peruano, 1968-1974.* Ed. Ricardo González Vigil. Lima: Copé, 1984; **Puerto Rico 1**: *Reunión de espejos.* Ed. José Luis Vega. Rio Piedras, P.R.: Cultural, 1983; **2**: *"Cuento"s puertorriqueños de hoy.* Ed. René Marqués. Rio Piedras, P.R.: Cultural, 1968; **3**: *Apalabramiento: cuentos puertorriqueños de hoy.* Ed. Efraín Barradas. Hanover, N.H.: Ediciones del Norte, 1983; **Venezuela 1**: *Narrativa venezolana contemporánea.* Ed. Rafael di Prisco. Madrid: Alianza, 1971; **2**: *Dos siglos de prosa venezolana.* Ed. Mariano Picón-Salas. Madrid: Edime, 1965. [Citations refer to Collection:Author:Page]

Both electronic corpora are privately developed and owned are not available for distribution.

■ WORKS CITED

Aissen, Judith and David Perlmutter. (1983) "Clause Reduction in Spanish." *Studies in Relational Grammar.* Ed. David Perlmutter. Chicago: U Chicago P. 360–403.

Alemán, Iris. (1985) "The Position of Clitics in Old Spanish." *Cornell Working Papers in Linguistics* 7: 1–16.

Colburn, G.B. (1928) "The Complementary Infinitive and its Pronoun Object." *Hispania* 11: 424–29.

Contreras, Heles. (1979) "Clause Reduction, the Saturation Constraint, and Clitic Promotion in Spanish." *Linguistic Analysis* 5: 161–82.

Farrán, Neus Bonet. (1990–91) "Movimiento clítico en el español de Puerto Rico." *Revista de Estudios Hispánicos* 17–18: 401–18.

Kayne, Richard. (1989) "Null Subjects and Clitic Climbing." *The Null Subject Parameter*, ed. O. Jaeggli and K. Safir. Dordrecht: Kluwer. 239–261.

Keniston, Hayward. (1937a) *Spanish Syntax List.* New York: Henry Holt and Company.

_____. (1937b) *The Syntax of Castillian Prose: The Sixteenth Century.* Chicago, U Chicago P.

Lujan, Marta. (1980) "Clitic Promotion and Mood in Spanish Verbal Complements." *Linguistics* 18: 381–484.

Myhill, John. (1988) "The Grammaticalization of Auxiliaries: Spanish Clitic Climbing." *Proceedings of the Berkeley Linguistics Society 14.* Berkeley: BLS. 352–63.

Napoli, Donna Jo. (1982) "Semantic Interpretation vs. Lexical Governance." *Language* 57: 47–69.

Pizzini, Quentin. (1982) "The Positioning of Clitic Pronouns in Spanish." *Lingua* 56: 47–69.

Ramsden, Herbert. (1963) *Weak Pronoun Position in the Early Romance Languages.* Manchester:

Manchester UP.

Rizzi, Luigi. (1982) "A Restructuring Rule." *Issues in Italian Syntax.* Foris: Dordrecht, 1982. 1–48.

Rosen, Sara Thomas. (1989) *Argument Structure and Complex Predicates.* Unpublished dissertation. Brandeis Univ.

Sankoff, David. (1988) "Sociolinguistics and Syntactic Variation". *Linguistics: The Cambridge Survey.* Ed. Frederick Newmeyer. Vol 4. Cambridge: Cambridge UP. 140–61.

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. Frank Nuessel. Bloomington, IN: IULC.

Wanner, Dieter. (1982) "A History of Spanish Clitic Movement." *Proceedings of the Berkeley Linguistics Society 8.* Berkeley: BLS. 135–47.

_____. (1991) "The Tobler-Mussafia Law in Old Spanish." *Current Studies in Spanish Linguistics.* Ed. Fernando Martínez-Gil and Hector Campos. Washington, Georgetown U P. 313–78.

WordCruncher V. 4.5. Computer Software. Johnston & Company Electronic Publishers, 1993.