

HISPANIC JOURNAL

A CORPUS-BASED APPROACH TO DIACHRONIC CLITIC CLIMBING IN PORTUGUESE

Author(s): Mark Davies

Source: *Hispanic Journal*, Vol. 17, No. 1 (Spring, 1996), pp. 93-111

Published by: Indiana University of Pennsylvania

Stable URL: <https://www.jstor.org/stable/44284487>

Accessed: 05-06-2020 18:04 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Indiana University of Pennsylvania is collaborating with JSTOR to digitize, preserve and extend access to *Hispanic Journal*

A CORPUS-BASED APPROACH TO DIACHRONIC CLITIC CLIMBING IN PORTUGUESE

1. Introduction

1.1 In traditional grammars of Portuguese, one of the most debated issues of syntax has been the placement of clitics (unstressed object pronouns), and no less than three entire books by Silveira, Figueiredo, and Sampaio Doria are dedicated to the problem. The issue is not limited just to clitic position with finite verbs, but extends to nonfinite verbs also. The phenomenon of variable clitic placement with nonfinite verbs, in which the clitic moves from a position either following (1a) or preceding the infinitive (1b) to a position preceding the governing finite verb (1c), is known in the literature as “clitic climbing”, and has been the subject of a number of studies dealing with Portuguese within the past two decades, including those by Quicoli, Oliveira, Pizzini, and Comrie.

- (1a) eles querem [ver-*me*]
- (1b) ele querem [*me* ver]
- (1c) eles *me* querem [ver ___]
“they want to see me”

Researchers have focused on clitic climbing for at least three reasons. First, as we will see, there are those who have argued that in certain sociolinguistic and geographical dialects of Portuguese, clitic climbing is unacceptable in all cases. Others have responded, just as strongly, that these same dialects freely permit it. Thus, clitic climbing becomes an issue of dialectal variation in syntax. For those dialects in which clitic climbing is accepted, there is the second question of why it is more acceptable with certain verbs than with others. For example, most researchers claim that it is most acceptable with modal verbs such as *poder*, *dever*, and *querer* (2a), less likely with verbs such as *desejar* and *começar a* (2b-c), and quite unacceptable with verbs such as *aprender* and *anelar* (2d-e).

- (2a) eles *me* podem [ver ___]
- (2b) ? eles *me* desejam [ver ___]
- (2c) ?? eles *me* anelavam [ver ___]

The third reason for interest in clitic climbing, not only for Portuguese but for other Romance and non-Romance languages as well, is the insight that it gives

into current syntactic theory. 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) *eles me podem* [ver ___]

Grammarians working in a number of syntactic frameworks have suggested motivations for the problematic movement of the clitic past clausal boundaries, including syntactic motivations such as “clausal reconstruction” (Rizzi), “intervening” but “abstract” syntactic material (Luján and Kayne), “semantic governance” (Napoli), and “clause union” (Aissen and Perlmutter). Regardless of the particular model, accounting for and explaining clitic climbing has been an important goal of several contemporary syntactic theories.

1.2 Considering the attention that has been paid to clitic climbing in past research, one would hope that previous researchers would have collected extensive data on its frequency in different dialects of Portuguese, and in different historical periods. Unfortunately, this is not the case. Nearly all previous studies of Portuguese clitic climbing have been prescriptive in nature, with the emphasis placed on what should occur in different registers, rather than any data-based study of what actually does occur. Yet even the prescriptive articles give us some insight into clitic placement with infinitives subordinated to main verbs. To summarize this research, let us take as an example the following sentence:

(4) *eu (lhe) queria (lhe) ver (lhe)*
 1 2 3

The general prescriptive rule is that in a sentence like (4), the final <3> position is in contrast with the initial <1> and medial <2> position, and that the contrast between initial and medial position is based on the preceding element (Dunn 271-73, Pereira 262-64, Vázquez-Cuesta and Luz 449-50, Willis 141-42, and Mateus 501-3). A number of these prescriptive grammars acknowledge that although this is the rule for Luso-Portuguese (Portugal) and written Brazilian Portuguese, spoken Brazilian Portuguese favors almost exclusively the medial placement (Bechara 401-2, Brandão 361-65, Teyssier 122-26, and Cunha and Cintra 315-17).

As mentioned, beyond the prescriptive articles there is very little hard data on the frequency of clitic climbing at different historical periods and in different registers. As Comrie points out, even the ‘theory-oriented’ research has typically

been based on the judgments and intuitions of individual speakers, and there have been sharp disagreements regarding these judgments. For example, Quicoli argues that clitic climbing is freely permissible in his dialect of Brazilian Portuguese, while Oliveira just as strongly responds that it is not a feature of any dialect of Modern Portuguese. In the only data-based study to date, Comrie looks at one work by a single Brazilian author, and tentatively determines that both authors are wrong; that clitic climbing is common only in formal/written Brazilian Portuguese, but cautions that his findings are quite limited and tentative. In addition to the basic question of any clitic climbing in certain dialects, the related question of the frequency with certain verbs (*poder, dever, começar a*, etc.) has never been investigated. In summary, although there has been a fair amount of data-based research on clitic climbing in the other Romance languages (cf. articles on Spanish by Davies, Myhill, and Farrán), even the most basic questions about the actual frequency of clitic climbing in Modern Portuguese remain unanswered.

Turning to clitic climbing in older stages of Portuguese, we note that although a few philologists have informally addressed the issue, here also there remain many unanswered questions. Dias (317) and Sampaio Doria suggest that clitic placement in Old and Middle Portuguese was somewhat different from the Modern Portuguese given above in (4). As with the Modern Portuguese construction, they claim that in older stages of the language, certain preverbal elements (or “exordiums”, such as negation, subordinating conjunctions, or certain adverbs) ‘attract’ clitics (5a), while others do not, including the absence of a preverbal element (5b). The difference from Modern Portuguese was that at these older stages of Portuguese, the postinfinitival position (5c) was very uncommon.

(5a) *não me* queriam [ver _]

(5b) (e) queriam *me* [ver _]

(5c) (e) queria [ver *me*]

In terms of actual data to support the claims, Sampaio Doria is the only researcher to date to base his conclusions on an organized corpus of pre-Modern Portuguese texts. Although Doria provides the most valuable data for this period, his study nevertheless suffers from a number of shortcomings. First, the corpus is limited to just five authors for the pre-modern period: João de Barros (1496-1570), Camões (1524-1580), Luís de Sousa (1555-1632), Antonio Vieira (1608-1697), and Manuel Bernardes (1644-1710). Thus it is based on texts from just a two hundred year period of pre-Modern Portuguese (about 1520-1700). In addition, the study is concerned primarily with clitic placement with finite verbs, and the

clitic climbing construction is addressed only peripherally. Finally, there is no discussion of the relative frequency of clitic climbing with the different governing verbs (*poder*, *começar a*, etc). Thus as with Modern Portuguese many of the most basic questions regarding clitic climbing in Old and Middle Portuguese remain unanswered. This again is in contrast to the other Romance languages, in which there has been much more research on diachronic clitic climbing (cf. studies on Spanish by Spaulding, Colburn, Keniston, Ramsden, Wanner “Spanish”, Wanner *Latin*, Granberg, and Rivero). In summary, we see that for both the historical and the modern periods of the language, previous data-based research on clitic climbing in Portuguese has been very weak, compared with the other Romance languages. This lack of data-based research on Portuguese in turn makes it difficult to paint an accurate and complete picture of clitic syntax for the Romance languages as a whole.

2. The database

2.1 The present study is an attempt to provide a comprehensive, data-based investigation of clitic climbing in Old and Middle Portuguese, as well as several registers of Modern Portuguese. It is based on selections from 236 different authors / speakers of Portuguese from these three periods, and comprises a total of more than 3,800,000 words of text. The texts for Old and Middle Portuguese contain texts from the 1300s (7 texts, 460,000 words), the 1400s (12 texts, 465,000 words), the 1500s (25 texts, 446,000 words), and the 1600s (30 texts, 265,000 words). The Modern Portuguese corpora includes 24 transcribed conversations of native speakers of Brazilian Portuguese (the *Linguagem Falada da Cidade de São Paulo* project; 148,000 words), 26 short stories from Brazil (75,000 words), selections from eleven novels from Portugal (239,000 words), and selections from the Ramsey-Borba corpus of written Brazilian Portuguese, containing material from novels, plays, magazines and newspapers, essays, and technical articles (1,670,000 words). Needless to say, the corpora on which the present study is based are many times the size of all those used in the few previous data-based studies of Portuguese clitic climbing. (For a complete listing of texts, please see Endnote 1.)

To extract the clitic climbing data, the following procedure was used. I personally scanned into the computer the majority of the 3,800,000 words of texts, using the *OmniPage* Optical Character Recognition (OCR) software. Afterwards I edited the text and inserted certain formatting codes (page and paragraph divisions, etc.) and then used the *WordCruncher* software package to create an “every word” index of the corpus. Using this index, the *WordCruncher* program was then able to perform complex Boolean and proximity searches on the data. Proximity searching means that I was able to

extract, for example, all cases of a clitic (*me, te, lhe, etc.*) occurring within perhaps five or six characters of one of the eight governing verbs (*poder, querer, dever, desejar, esperar, acabar de, começar a, deixar de*), which was in turn within a certain number of characters of an infinitive (giving, for example, *o quero fazer*). Boolean searching means that I was able to combine the results of this search with other searches, for example to include all cases of one of the eight governing verbs followed either by a clitic plus infinitive (*quero o fazer*) or by an infinitive plus clitic (*quero comê-lo*). Using this process, I was able to extract nearly 5000 examples of the type of construction just described. Endnote 2 provides a complete summary of the tokens extracted from the corpora.

2.2 As a result of the large number of example sentences extracted from the corpora, we are prepared to answer a number of questions about Portuguese clitic climbing that have not been fully explored until this time. First, has there been a decrease in clitic climbing (i.e. a decrease in the initial placement) since Old Portuguese? If so, when did this occur, and what might be the possible motivations for such a shift? A related question is whether clitic climbing is more common in written / formal / Luso Portuguese than in informal, spoken Brazilian Portuguese. Second, some have suggested that the medial position is the norm for spoken Brazilian Portuguese, whereas the post-infinitival final position is the norm for other, more formal dialects of Modern Portuguese. Is this the case, and if it is, what are possible motivations for the difference? Finally, what effect does the particular governing verb have on clitic climbing, both historically and in Modern Portuguese? For example, are the more frequently occurring verbs (*poder, querer, dever*) more likely to allow clitic climbing than the less frequent verbs such as *começar a, deixar de, and desejar*, and if so, why? In summary, the large corpus on which this study is based will allow us to gain insight, for the first time, into a much discussed aspect of Portuguese syntax. This will in turn allow us to examine historical, dialectal and register differences in Portuguese, and discover how Portuguese (at least in terms of this one phenomenon) fits into the larger picture of the Romance languages as a whole.

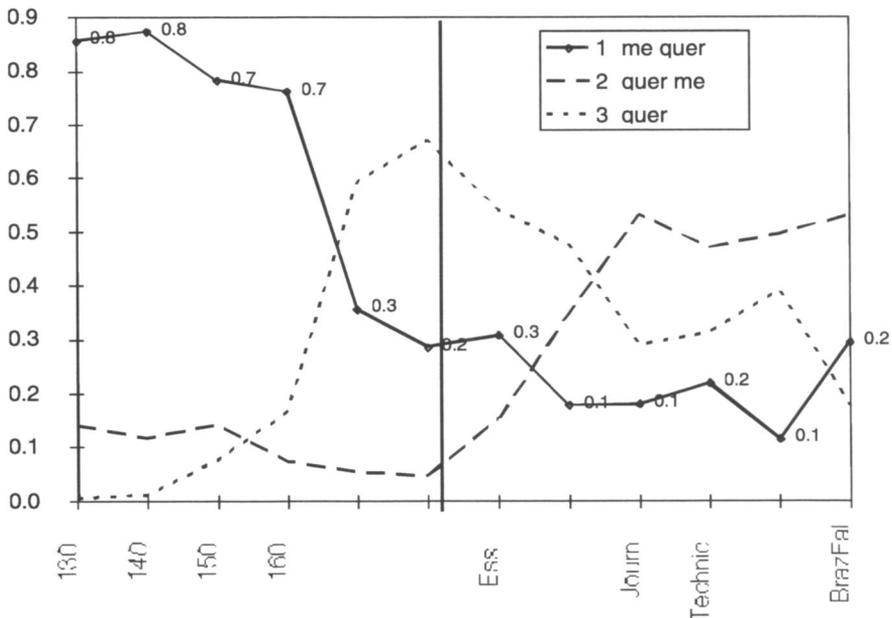
3. General decrease in clitic since Old Portuguese

3.1 A number of researcher working on other Romance languages (Wanner “Spanish”, *Latin*, and Pearce) have noted that there has been a general decrease in clitic climbing since the older stages of these languages. In other words, in Modern Portuguese there might be less of a tendency for the clitic to “climb” to the preverbal position (6a) than in Old or Middle Portuguese (6b):

- (6a) o tratador dos cães *lhe* queria desvendar os milagres do officio (Cegos 124:3)
 (6b) mas elles n^o *lho* quiseram dar (Cron1344 2:102)

The data from our corpus confirm that this is the case for Portuguese also. In Figure 1 we see the percentage of initial <1>, medial <2>, and final <3> placement in the other stages of the language (1300s, 1400s, 1500s, 1600s), as well as for several registers of Modern Portuguese. (Note that in both Figures 1 and 2, the left side of the chart is chronological, and extends from the 1300s to the 1900s. In both figures, these portions of the corpora give the clearest picture of the diachronic shifts in clitic climbing. To the right of the dividing line, the figures indicate the percentage of clitic climbing in several registers of Modern Brazilian Portuguese: a collection of short stories (*BrazSS*), the five different genre of the *Brazil* corpus (*Essay, Novel, Journal, Technical, and Drama*), as well as *BrazFal* (the corpus of Spoken Brazilian Portuguese).

Figure 1. Clitic placement in all periods and registers



This figure shows that the percentage of clitic climbing (initial placement) with the three major verbs (*poder, dever, and querer*) ranges between 86% and 87% in the 1300s and 1400s (657/735 and 935/1014 total cases, respectively; see Endnote 3), to 78% and 76% in the 1500s and 1600s (665/789 and 310/402

cases, respectively). The largest decrease does not come until ModPt, however, where the figures are only 36% in European Portuguese (Luso: 123/301 cases), and 28% in a collection of Brazilian short stories (BrazSS: 23/78 cases). In the other corpora of Modern Brazilian Portuguese, the figures also range between 11% in a collection of plays (*Drama*:37/179 cases) to a high of 31% in essays (*Essay*: 60/151). Restating the facts somewhat, the figures from Modern Portuguese show Luso (European) Portuguese and formal Brazilian Portuguese (*Essay*) to have the highest degree of clitic climbing, and the most colloquial of the written Brazilian corpora (*Drama*) is where there is the lowest degree of clitic climbing.

We should briefly explain the relatively high 29% figure for the corpus of spoken Brazilian Portuguese (*BrazFal*), which is where we might otherwise expect to find the lowest degree of clitic climbing, since it is the most colloquial register. First, the results are skewed somewhat, since there are only 48 total cases of the clitic climbing construction (any of the three placements) in that corpus, which is quite small for a corpus of 148,000 words. The collection of Brazilian short stories (*BrazSS*) has nearly twice that amount in a corpus nearly half the size. The relatively small amount of clitics in any of the three positions is undoubtedly related to the more generalized phenomenon of “object pronoun dropping”, which is one of the distinguishing features of spoken Brazilian Portuguese (see the studies by Kliffer, Wheeler, Raposo, and Kato and Tarallo). In other words, we find few cases of the clitic climbing construction (with the clitic in any of the three positions), simply because object clitics are avoided in the more colloquial registers of Brazilian Portuguese. A second explanation for the relatively high percentage of clitic climbing in the spoken Brazilian corpus is that all of the cases of “clitic climbing” in fact involve the pronoun *se*, with its ‘impersonal’ meaning of “one does, one says”, etc:

(7a) eu acredito que não *se* pode pensar claramente (BrazFal 1:70)

(7b) eu acho que hoje o que *se* pode fazer com a imagem não se deve fazer com palavra (BrazFal 2:257)

There are no cases true clitic climbing, where an object pronoun, governed by the nonfinite verb, climbs to the initial position, as in the following examples from written Brazilian Portuguese:

(8a) o doutor estava de costas, não *me* podia ver (BrazSS 104:6)

(8b) vamos a necrotério, que eu *lhes* quero mostrar uma coisa (BrazSS 38:13)

Thus, the data shows that spoken Brazilian Portuguese, which should have the lowest degree of clitic climbing because it is the most colloquial register, does

in fact follow our expectation. In summary, we have demonstrated to this point that there has been a decrease in clitic climbing since Old Portuguese, with the most pronounced decrease in the period 1700-1900. Related to this is the fact that in Modern Portuguese, the most conservative registers have the highest degree of clitic climbing, while the lowest degree is in spoken Brazilian, and the corpora (e.g. “plays”) which are based on the spoken register.

3.2 Let us now briefly consider a couple of possible motivations for the general decrease in clitic climbing. The first motivation is discussed at length in Wanner (*Latin*), and deals with gradual, long-range shifts in the clausal structure of Latin and the Romance languages. In general terms, he argues that Latin had fairly “loose” clausal structure; in other words elements were not tightly bound to the clause in which they originated. At this stage, clitics originating as objects of the infinitive could freely move out of the clause governed by that infinitive and attach to the finite verb. Over the process of time, however, the clausal structure of the Romance languages has become more rigid and the clitic is now more tightly bound to the clause in which it originates.

There is an additional motivation, one which is based more on syntactic changes particular to Portuguese itself. Kayne argues that there is a relationship between possible deletion or suppression of subjects in a language (“Null Subject Properties”), and the overall clausal structure of that language. Without going into all of the details, he proposes that languages which allow null subjects have “loose” clauses, out of which syntactic elements can move. Languages which do not allow null subjects, on the other hand, have more “rigid” clause structure, and elements stay within the clause.

Regarding Portuguese, Lira and Tarallo document the fact that at the older stages of the language, null subjects were much more common (9a), and that there has been a gradual increase in obligatory subjects since Old Portuguese, and especially since the 1600s. In addition, the colloquial spoken Portuguese of Brazil is the Modern Portuguese dialect / register that has the highest degree of obligatory subjects (9b).

(9a) aquela molher ... e __ era i entam e __ era catecumina e __ nunca viinha à igreja
(Pelagia 211:3)

(9b) *ela* estava ligada a todo o desenvolvimento do indivíduo, *ela* ia estar junto, *ela* é todo o desempenho do indivíduo, *ela* não é simplesmente o uso de intelecto, mas *ela* é toda ... (BrazFal 1:30)

These null subjects facts suggest that Portuguese has developed increasingly “rigid” clauses over time, and has thus experienced a decrease in clitic climbing.

We also find the smallest degree of null subjects and the lowest percentage of clitic climbing in spoken Brazilian Portuguese. Thus, at least at the surface level, there appears to be a link between null subjects and clitic climbing. In addition, this explanation would also explain why French and Brazilian Portuguese, the two Romance languages / dialects that now have no (or few) null subjects, are the ones which have the least amount of clitic climbing (none, in the case of French). On the other hand Spanish and Italian, two Romance languages that still freely allow null subjects, have a much higher degree of clitic climbing.

4. Medial vs. final placement

4.1 With the breakdown in clitic climbing, which we have demonstrated above, clitics in most dialects and registers have become limited primarily to the medial (10a) or the final (10b) position:

- (10a) você devia *me* conhecer melhor, Glicério (Braz-Plays: EL 8211)
 (10b) não deve sacá-*lo* da cintura (Braz-Novels: TG 26259)

According to several investigators (including Bechara 401-2, Brandão 361-65, Teyssier 122-26, and Cunha and Cintra 315-17), there is dialectal / register variation between these two placements, with more colloquial, spoken Brazilian Portuguese preferring medial placement and the European and formal Brazilian dialects and registers preferring final placement. As of yet, however, there has been no hard data on which to quantitatively verify these native speaker intuitions.

The data from our corpus does show variation between the medial and final placement, and substantiates quite well the claims about the nature of the variation between different dialects and registers of Modern Portuguese. As Figure 1 above shows, as initial placement has decreased historically, there has been a corresponding increase in final placement. In other words, from Old and Middle Portuguese until the 1900s, there has been a shift from (11a) to (11b), with (11b) picking up the slack caused by the reduction in clitic climbing. But there has been very little increase in the other alternative, the medial (11b). The following chart provides the exact figures for the increase in final placement:

- (11a) eles *me* querem ver <1>
 (11b) eles querem ver-*me* <3>
 (11c) eles querem *me* ver <2>

Table 1. Final <3> placement (*querem ver-me*)

Text	%	# examples	Text	% <3>	# examples
1300	0%	735	Essay	54%	151
1400	1%	1014	Novel	48%	327
1500	8%	789	Journal	29%	348
1600	17%	402	Technical	31%	329
Luso	59%	301	Drama	39%	179
BrazSS	67%	78	BrazFal	18%	48

What is quite obvious, of course, is that although there had been a historical shift over the past 500-600 years towards final placement, this has been reversed somewhat in the more colloquial registers of Brazilian Portuguese. As Figure 1 indicates, in these dialects and registers the shift has been towards medial placement. In addition, Figure 1 and in the following chart indicate that medial placement remained rather stable from the oldest periods of the language to recent times, but that in several registers of Brazilian Portuguese it is now clearly the norm. The only corpora in which medial placement is fairly uncommon is “Essays”, which tend to be linguistically conservative.

Table 2. Medial <2> placement (*querem me ver*)

Text	% <2>	# examples	Text	% <2>	# examples
1300	14%	735	Essay	15%	151
1400	12%	1014	Novel	35%	327
1500	14%	789	Journal	53%	348
1600	7%	402	Technical	47%	329
Luso	5%	301	Drama	50%	179
BrazSS	5%	78	BrazFal	53%	48

4.2 Let us consider for a moment a possible motivation for the variation we have just discussed. Ideally, we want to find a motivation for medial placement in Spoken Brazilian Portuguese (SBP), as opposed to final placement in other more formal dialects of Modern Portuguese. First, we note that in SBP there is a general avoidance of VERB + CLITIC, and not just with clitic climbing. In SBP, phonetic reduction has led towards a loss of verbal agreement, so that (for example) 1pl, 3sg, and 3sg all have the same morphological form (12a). This word final phonetic instability may be related to the general SBP tendency to avoid placing more material — a clitic — at the end of a finite verb, and the general rejection of enclisis in SBP (12b) (see Azevedo, and Lemle and Naro):

(12a) ele fala, eles fala, a gente / nos fala
 “he / they / we speak”

(12b) ?? ele fala-o, eles fala-o, a gente / nos fala-o

Consider also the fact that in SBP the inflected infinitive is used less than in other Modern Portuguese dialects (cf. studies by King and Koike). One might assume a type of analogy between the verbal ending of the inflected infinitive and the postposed clitic in the clitic climbing construction:

(13a) sem eles falarem

(13b) ele quer falar-me

Finally, consider the link to clitic position in the PREP + INF (preposition + infinitive) construction. We find that SBP prefers medial position in these constructions (14a), whereas the other dialects and registers prefer final position (15a). This of course correlates directly with clitic climbing construction (14b, 15b).

(14a) PREP-CL-INF: com essa expectativa de *me* encontrar (BrazFal 2:108)

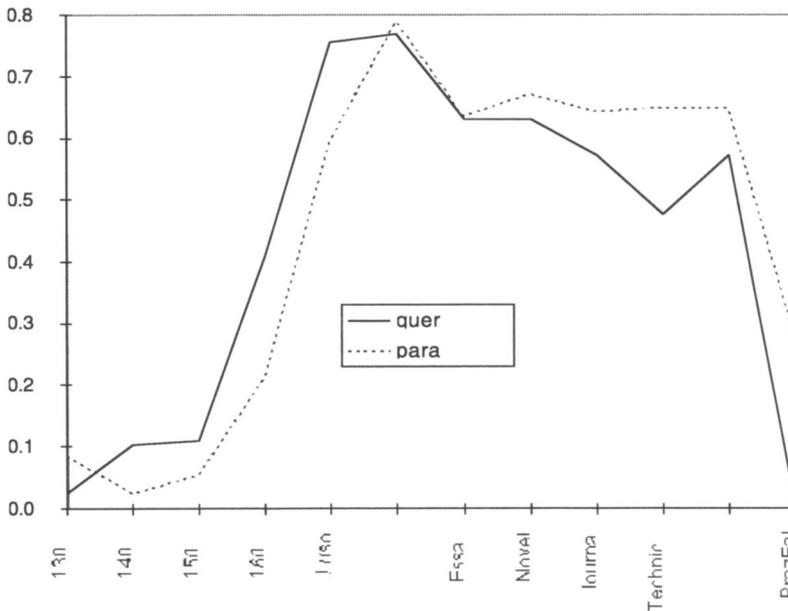
(14b) VERB-CL-INF: tu pode *me* dizer as horas (BrazFal 3:96)

(15a) PREP-INF-CL: E depois de obrigá-*lo* a nova jura (Cegos 105:6)

(15b) VERB-INF-CL: posso matá-*lo*, si quiserem (Alegria 172:2)

The relationship between the PREP + INF and the clitic climbing construction is seen more graphically in the following figure, which shows the percentage of final placement with both type of constructions since Old Portuguese.

Figure 2. Percentage of final placement with preposition (*para fazê-lo*) and verb (*quer fazê-lo*) in all periods and registers



One cannot help but notice how closely the two phenomena are related. The same factor that motivates PREP-CL-INF in the preposition construction is most likely the same one that motivates V-CL-INF in the clitic climbing construction. This may in turn be related to the general avoidance of “verb-final” material with finite verbs and with the inflected infinitive, seen above. In all four cases, SBP prefers the preverbal / medial position, while the other dialects permit or favor verb final material. In summary, this suggests a more general explanation for medial position in SBP, beyond just the phenomena of clitic climbing.

5. Effect of the main verb

5.1 The final question that we want to address is how the main verb (*poder*, *começar a*, etc.) affects clitic climbing. To date there has been virtually no discussion of this issue regarding Portuguese, but it has received quite a bit of attention by researchers working on other Romance languages. There are two main schools of thought regarding the influence of the main verb. One suggests that different verbs trigger different syntactic processes, which either do or do not act as barriers to clitic climbing (cf. Luján, Kayne, and Rivero). The other approach, exemplified by Napoli, Myhill, and others, suggests that it is more semantic in nature, and argues that the verbs that most readily allow clitic climbing are those that are semantically the most basic and most auxiliary-like in nature. They argue that verbs like *ir*, *poder*, and *haver de* have little “semantic weight”, and have a meaning that simply adds aspectual or modal-like

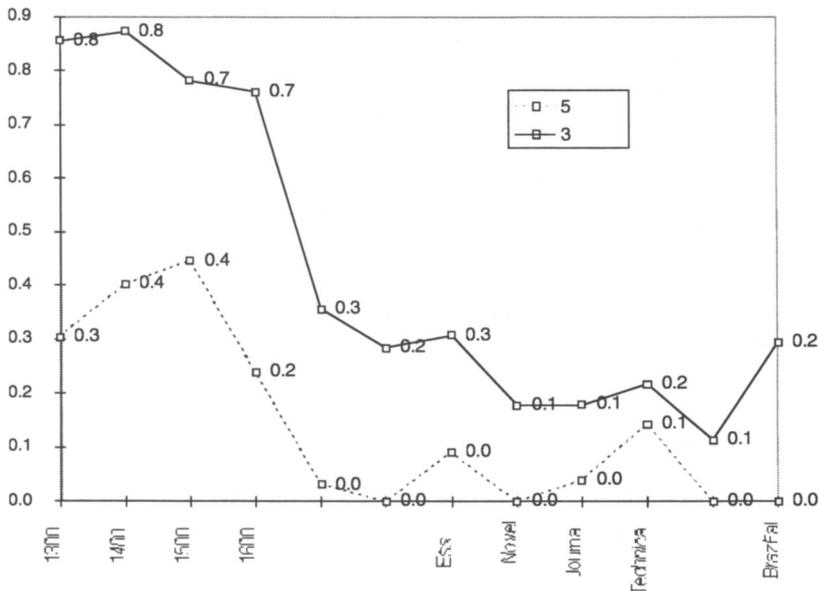
information to the following verb. Since they are not very independent from the governing infinitive in terms of meaning, they are not syntactically very independent either. Thus the two verbs (e.g. “*posso ver*”) are more of a single unit than “*anelava ver*”. Since clitics can be placed before simple finite verbs (*ele o viu*), even clitics that originate as objects of the infinitive can be preposed, since the entire verb unit acts semantically as if it were one single verb (16a). When the main verb is semantically more heavy or independent from the infinitive, however, the two verbs do not act as one unit, and the clitic cannot climb (16b):

(16a) *ele o queria :: ver*

(16b) **ele o anelava || ver*

5.2 In terms of actual clitic climbing data, these researchers have shown that the Romance language cognates of the Portuguese *poder*, *dever*, and *querer* are among those that most often allow clitic climbing. In the present study we have divided these three common verbs from five verbs that occur less frequently (*começar a*, *deixar de*, *desejar*, *acabar de*, and *esperar*). The following figure shows the percentage of clitic climbing (initial <1> placement) with the two sets of verbs in all periods of the language:

Figure 3. Percentage of clitic climbing with most / less common verbs (*me querem ver / me anelam ver*)



The figure indicates that at all periods of the language, and in all modern registers, the degree of clitic climbing is higher with the three most common verbs. This is therefore in agreement with the finding for the other Romance languages. In more exact terms, we find that in Old and Middle Portuguese these three most common verbs allowed clitic climbing two to three times as often as the five less common verbs, and this has continued into nearly all registers of Modern Portuguese.

When we move from initial placement (clitic climbing) to medial placement, we find something else of interest. The following chart indicates that for the five less common verbs, this medial position was the preferred position in Old Portuguese and into Middle Portuguese:

Table 3. Clitic placement with most / less common main verbs (*me querem* / *desejam ver*)

Text	Three main verbs		Five less common verbs			
1300	0.86	0.14	0.00	0.30	0.67	0.02
1400	0.87	0.12	0.01	0.40	0.49	0.10
1500	0.78	0.14	0.08	0.45	0.45	0.11
1600	0.76	0.07	0.17	0.24	0.35	0.41
Luso	0.36	0.05	0.59	0.03	0.21	0.76
BrazSS	0.28	0.05	0.67	0.00	0.23	0.77
Essay	0.31	0.15	0.54	0.09	0.28	0.63
Novel	0.18	0.35	0.48	0.00	0.37	0.63
Journal	0.18	0.53	0.29	0.04	0.39	0.57
Technical	0.22	0.47	0.31	0.14	0.38	0.48
Drama	0.11	0.50	0.39	0.00	0.43	0.57
BrazFal	0.29	0.53	0.18	0.00	1.00	0.00

The importance of initial vs. medial placement with the two groups of verbs is the following. According to “Wackernagel’s Law” (cf. Slawomirski), the variation between initial and medial position is a function of the element that precedes the main verb, with certain elements “pulling” the clitic into initial

position, and others not (cf. Dunn 271-73, Pereira 262-64, Vázquez-Cuesta and Luz 449-50, Willis 141-42, and Mateus 501-3 for ModPt; and Dias 317 and Sampaio Doria for Old and MidPort). For example, subordinating conjunctions create initial placement (17a), whereas coordinating conjunctions or no element at all will lead to medial placement (17b):

(17a) *que o deveades de prouar e[m] maneira que os cristãos nõ ajam que responder contra vos (Arimatea 298:19)*

(17b) *e devo-vos conselhar a meo grado o que seja vossa honra (Demanda 2:338)*

However, according to this model, the particular finite verb (*poder, deixar de, etc.*) should not affect the initial / medial alternation. During the historical periods in which clitic climbing is possible, if a main verb allows the clitic to “climb” then its initial or medial position becomes a function of the element preceding the main verb, but not the verb itself. What our data suggests is that there was a three-way distinction in Old and Middle Portuguese: final placement was rare with all verbs, the less common verbs preferred medial placement, and only the most common verbs took initial placement in the majority of the cases.

6. Summary

Because of the nearly 4,000,000 word corpus on which this study is based, we are able to answer a number of questions about historical and dialectal clitic climbing in Portuguese that have not been fully considered by previous researchers. First, the nearly 5000 examples in the corpus demonstrate a clearly decrease in clitic climbing since the oldest periods of the language. This decrease is most pronounced in the most innovative contemporary dialects of Spoken Brazilian Portuguese. In addition, this decrease may be due to general typological shifts in the grammar, but may also be due to a more formal motivation which involves the decrease in null subjects. Second, with the decrease in initial placement, Spoken Brazilian Portuguese has clearly moved toward medial placement, in contrast with the shift towards final placement in other dialects. This avoidance of final position in Spoken Brazilian Portuguese may be due to a general avoidance of verb final syntactic material in that dialect. Finally, the most common main verbs have favored clitic climbing the most, in all periods of the language. This raises certain questions about Wackernagel’s Law in Old Portuguese, which supposedly motivates the initial / medial alternation in clitic placement. In summary, we see that an extensive, computer data-based study such as this is able to create a rather comprehensive view of certain aspects of historical and dialectal syntax.

Endnotes

¹ The electronic corpora is based on the following texts. With the exception of one block of data, I personally scanned in, edited, and processed for use with *WordCruncher* all of the texts listed below. The one exception is the Ramsey-Borba corpus of Brazilian Portuguese that is listed in the final section, which I received in Folio Views format from Professor Miriam Ramsey of Brigham Young University. Information is given for: [the date of the text], *the title*, {the number of words included in the corpus}, (and the publication information). In most cases, the citations given in the paper refer to the specific text and page within the text.

1300 [7 texts / 479,300 words]: [c1344] *Crónica general de Espanha* {163,800} (Acad Port Hist, 1951-4); [1300-1400] *Joseph de Arimatea* {60,500} (UNC P, 1967); [1300-1400] *Barlaam e Josephat* {22,800} (U Penn P, 1967); [1300-1400] *Historia do muynobre Vespesiano* {18,000} (U Exeter, 1983); [1300-1400] *Vida de Túndalo* {6,800} (In Prosa medieval portuguesa, Ed. Comunicação, 1985); [1300-1400] *Vida de Santa Pelágia* {4,500} (In Prosa medieval portuguesa); [c1400] *Demanda do santo graal* {202,900} (Inst Nac Livro, 1944).

1400 [12 texts / 463,400 words]: [1400-1450] *Orto de esposo* {47,500} (Inst Nac Livro, 1956); [1400-1450] *Boosco deleitoso* {40,800} (Inst Nac Livro, 1950); [c1431] *Cron. do condestável Nuno Alvarez* / ?? Fernão Lopes {37,500} (Acad Port Hist, 1972); [1430-1440] *Livro de montaria* {45,100} (In Obras dos principes de Avis, Lello e Irmão, 1981); [1430-1440] *Leal conselheiro* {39,400} (In Obras dos principes...); [1430-1440] *Livro de bem cavalgar* {39,700} (In Obras dos principes...); [1430-1440] *Virtuosa bemeitoria* {36,700} (In Obras dos principes...); [1430-1440] *Livro dos officios* {36,400} (In Obras dos principes...); [1431-43] *Crónica de D. Fernando* / Fernão Lopes {44,600} (Impr Nac, 1975); [1446 (translated)] *Livro de vita Christi* / tr. Fr. Bernardo de Alcobaça {42,700} {42,700}; [1400-1500] *Vida e feitos de Júlio Cesar* {42,300} {Fund Gulbenkian, 1970}; [1400-1500] *Livro de silolôquio de Sancto Agostinho* {24,300} ({Valle Cintra}, 1957).

1500 [14 texts + 1 anthology (11 authors) / 451,700 words]: [c1500] *Cron. de D. João II* / Rui de Pina {21,500} (Atlântida, 1950); [c1500] *Virgeu de consolação* {24,400} (Livr Golobo, 1959); [early 1500s] *Cron. de D. Afonso Henriques* / Duarte Galvão {24,700} (Portugalia Livr, 1950); [1513] *Flos sanctorum* {50,900} (Inst Nac Invest Cient, 1988); [1520] *Cron. do imperador Clarimundo* / João de Barros {38,900} ({Lisboa}, 1953); [1528] *Estrangeiros* / Sá de Miranda {31,200} (In *Obras completas...*, (Coimbra), 1930); [1533] *Vilhalpandos* / Sá de Miranda {21,000} (In *Obras Completas*); [1529] *Itinerário...* / Antonio Tenreiro {19,100} (Ed Estampa, 1980); [1532] *Rópica pnfema* / João de Barros {28,000} (Inst Nac Invest Cient, 1983); [c1555] *Menina e moça* / Bernadim Ribeiro {46,500} ({Lisboa}, 1960); [1567] *Cron de D. João / Damião de Góis* {23,400} (U Nova de Lisboa, 1977); [1570-78] *Peregrinação* / Fernão Mendes Pinto {18,400} (Sá de Costa, 1961); [c1588] *Soldado prático* / Diogo de Couto {17,300} (Livr Sá de Costa, 1961); [1589] *Diálogos* / Amador Arrais {43,400} (Lello, 1974); *ANTHOLOGY* {64,000} *Novelistas e cronistas portugueses do século XVI*, ed. João Palma Ferreira (Imprensa Nacional, 1982).

1600 [3 texts + 2 anthologies (16 + 10 authors) / 261,300 words]: [c1650] *Sermões* / P. Antonio Vieira {33,100} (Cultrix, 1975); [c1650] *Cartas familiares* / Francisco Manuel de Melo {39,100} (Sá de Costa, 1942); [1652] *Arte de furta* {40,700} (Melhoramentos, 1951); *ANTHOLOGY A* = *Contos portugueses* (Lisboa, 1964); *ANTHOLOGY B* = *Novelistas e cronistas portugueses do século XVII e XVIII*, ed. João Palma Ferreira (Imprensa Nacional, 1981).

LUSO [11 texts / 239,000 words]: [1906] *Os Pobres* / Raul Brandão {21,800} (Comunicação, 1984); [1930] *O homem que matou o diabo* / Aquilino Ribeiro {21,800} (Bertrand, 1972); [1934] *Terra fria* / José Maria Ferreira de Castro {22,400} (Guimarães, 1966); [1945] *Vindima* / Adolfo Correia da Rocha {22,500} (Coimbra, 1966); [1947] *Porta de Minerva* / Braquinho da Fonseca {21,500} (Atica, 1947); [1956] *A corça prisioneira* / Joaquim Paço de Arcos {21,100} (Gumarães, 1988); [1958] *O homem disfraçado* / Fernando Namora {21,800} (Bertrand, 1975); [1961] *Barranco de cegos* / Alves Redol {20,400} (Avante, 1982); [1965] *Alegria breve* /

Vergílio Ferreira {23,000} (Bertrand, 1981); [1982] *Memorial do convento* / José Saramago {21,100} (Caminho, 1982); [1983] *Balada da Praia dos Cães* {21,500} (Projournal, 1983).
 BrazSS [1 anthology (26 authors) / 75,100 words]: *O conto fantástico*, ed. Jerónimo Monteiro. (Brasileira, 1959)
 BrazFal [1 collection (24 speakers / 148,300 words): A linguagem falada da cidade de São Paulo, ed. Atiliba Teixeira de Catilo and Dino Preti. 3 vols. (T.A. Queiroz, 1986).
 Brazil (Essay, Novel, Journal, Technical, Drama) [102 selections / 1,670,300 words]: Selections from the Ramsey-Borba corpus of Brazilian Portuguese. Citations refer to Genre and Corpus Block (their numbering).² Composition of corpus. Number of tokens, by verb, in all periods and registers.

Text	poder	dever	querer	other	Total
1300	337	65	333	84	819
1400	495	301	218	59	1073
1500	440	85	264	97	886
1600	248	61	93	37	439
Luso	165	49	87	53	354
BrazSS	36	9	33	13	91
BrazFal	37	6	5	4	52
Brazil	668	201	205	155	1229
Total	2426	777	1238	502	4943

³ Note that the percentage of initial <1> placement does not correspond exactly to the total cases of initial placement divided by the total number of examples. This is because much of the initial placement occurs with the pronoun *se*, used in its impersonal sense (*se pode fumar*). Since these are not true cases of clitic climbing, their weight in the calculations is discounted. Since only part of the *se* in initial position refer to objects of the embedded infinitive (eles *se acostaram* as sete), each case of initial placement with the pronoun *se* counts for only 30% of the normal weight of object clitics that have climbed to the initial position (o quero fazer). The number given in parenthesis, indicating the total number of cases of initial placement and the total number of all placements, refers to the total number of both *se* and non-*se* cases; in other words, before discounting the *se* cases to 30%.

WORKS CITED

- Aissen, Judith and David Perlmutter. "Clause Reduction in Spanish." *Studies in Relational Grammar 1*. Ed. David Perlmutter. Chicago: Univ. of Chicago Press, 1978. 360-403. (Earlier version in *BLS* 2:1-30.)
- Azevedo, Milton. "Vernacular Features in Educated Speech in Brazilian Portuguese." *Hispania* 72 (1989): 862-72.
- Bechara, Evanildo. *Lições de português pela análise sintática*. Rio de Janeiro: Editora Fundo de Cultura, 1960.
- Brandão, Cláudio. *Sintaxe clássica portuguesa*. Belo Horizonte: Univ. de Minas Gerais, 1963.
- Colburn, G.B. "The Complementary Infinitive and its Pronoun Object." *Hispania* 11 (1928): 424-29.
- Comrie, Bernard. "Remarks on Clitic Climbing in Brazilian Portuguese." *Lingua* 58 (1982): 243-65.
- Cunha, Celso and Luís F. Lindley Cintra. *Nova gramática do português contemporâneo*. 2nd ed. Lisboa: Edições João Sá da Costa, 1984.
- Davies, Mark. "Analyzing Syntactic Variation with Computer-based Corpora: The Case of Modern Spanish Clitic Climbing." *Hispania* 78 (1995): 370-80.
- Dias, Augusto Epiphânio da Silva. *Sintaxe histórica portuguesa*. 2nd ed. Lisboa: Livraria Clássica Editora, 1933.
- Doria, Antônio de Sampaio. *Sintaxe de pronomes*. São Paulo: Companhia Editora Nacional, 1959.
- Dunn, Joseph. *A Grammar of the Portuguese Language*. Washington, D.C.: National Capital Press, 1928.
- Farrán, Neus Bonet. "Movimiento clítico en el español de Puerto Rico." *Revista de Estudios Hispánicos* 17-18 (1990-91): 401-18.
- Figueiredo, Candido. *O problema da colocação de pronomes*. Lisboa, n.p.
- Granberg, Robert A. "Object Pronoun Position in Medieval and Early Modern Spanish." Diss. UCLA, 1989.
- Kato, Mary A. and Fernando Tarallo. "Anything YOU Can Do in Brazilian Portuguese." *Studies in Romance Linguistics*. Eds. Osvaldo Jaeggli and Carmen Silva-Corvalán. Dordrecht: Foris, 1986. 343-58.
- Kayne, Richard. "Null Subjects and Clitic Climbing." *The Null Subject Parameter*. Ed. Osvaldo Jaeggli and Ken Safir. Dordrecht: Kluwer, 1989. 239-61.
- Keniston, Hayward. *The Syntax of Castilian Prose: The Sixteenth Century*. Chicago: Univ. of Chicago Press, 1937.
- King, Larry D. "On Predicting the Portuguese Infinitive." *Proceedings of the Tenth Anniversary Linguistic Symposium on Romance Languages*. Ed. Heles Contreras and Jürgen Klausenburger. Seattle: U Washington, 1981. 133-45. (*Papers in Romance*. Vol 3. Supp 2.)
- Kliffner, Michael D. "The Infinitive in Portuguese: A Generative Approach." *Orbis* 27 (1978): 779-2.
- Kliffner, Michael D. "The Case of the Missing Reflexive." *Hispania* 65 424-27.
- Koike, Dale. "Register, Social Variables, and Variation of the Infinitive in Brazilian Portuguese." *Hispania* 68 (1986): 134-42.
- Lemle, Miriam, and Anthony J. Naro. *Competências básicas do português*. Rio de Janeiro: Fundação Movimento Brasileiro de Alfabetização, 1977.
- Lira, Solange de Azambuja. *Nominal, Pronominal, and Zero Subject in Brazilian Portuguese*.

- Diss. U Pennsylvania, 1982.
- Lujan, Marta. "Clitic Promotion and Mood in Spanish Verbal Complements." *Linguistics* 18 (1980): 381-484.
- Mateus, Maria Helena Mira, et al. *Gramática da língua portuguesa*. Coimbra: Livraria Almedina, 1983.
- Myhill, John. "The Grammaticalization of Auxiliaries: Spanish Clitic Climbing." *BLS* 14. Berkeley: Berkeley Linguistics Society, 1988. 352-63.
- Napoli, Donna Jo. "Semantic Interpretation vs. Lexical Governance." *Language* 57 (1982): 47-69.
- Oliveira, Mario Antônio de. "Clitic Placement in Portuguese." *Ensaio de Linguística 1*. Eds. Yara Goulart Liberato and Mario A. Perini. Belo Horizonte: Univ. de Minas Gerais, 1978. 44-79.
- Pearce, Elizabeth. *Parameters in Old French syntax: Infinitival Complements*. Dordrecht: Kluwer, 1990. (Diss. U Illinois, 1985)
- Pereira, Eduardo Carlos. *Gramática expositiva: curso superior*. 5th ed. São Paulo: Editôra Nacional, 1940.
- Pizzini, Quentin. "The Positioning of Clitic Pronouns in Spanish." *Lingua* 57 (1982): 47-69.
- Quicoli, Carlos. "Conditions on Clitic Movement in Portuguese." *Linguistic Analysis* 2 (1976): 199-223.
- Ramsden, Herbert. *Weak Pronoun Position in the Early Romance Languages*. Manchester: Manchester UP, 1963.
- Raposo, Eduardo. "On the Null Object in European Portuguese." *Studies in Romance Linguistics*. Eds. Osvaldo Jaeggli and Carmen Silva-Corvalán. Dordrecht: Foris, 1986. 373-390.
- Rivero, María Luisa. "Clitics and NP Climbing in Old Spanish." *Current Studies in Spanish Linguistics*. Eds. Fernando Martinez Gil and Hector Campos. Washington, D.C.: Georgetown UP, 1991. 241-82.
- Rizzi, Luigi. "A Restructuring Rule." *Issues in Italian Syntax*. Foris: Dordrecht, 1982. 1-48.
- Silveira, Agenor. *Colocação dos pronomes (regras e notas explicativas)*. São Paulo: Rosa, 1920.
- Slawomirski, Jerzy. (1986) "Wackernagel's law in Old Spanish." *Revista de la Sociedad Española de Linguística* 16:91-100
- Spaulding, Robert K. "Puedo hacerlo versus lo puedo hacer and Similar Cases." *Hispania* 19 (1927): 343-48.
- Tarallo, Fernando. *Tempos linguísticos: itinerário histórico da língua portuguesa*. São Paulo: Africa, 1990.
- Teyssier, Paul. *Manuel de langue portugaise (Portugal-Brésil)*. Paris: Klincksieck, 1977.
- Vázquez Cuesta, Pilar and Maria Albertina Mendes da Luz. *Gramática portuguesa*. 2nd ed. Madrid: Gredos, 1961.
- Wanner, Dieter. "A History of Spanish Clitic Movement." *BLS* 8. Berkeley: Berkeley Linguistics Society, 1982. 135-47.
- Wanner, Dieter. *The Development of Romance Clitic Pronouns from Latin to Old Romance*. New York: Mouton de Gruyter, 1987.
- Wheeler, Dana. "Object Deletion in Portuguese." *Current Researches in Romance Languages*. Eds. James P. Lantolf and Gregory B. Stone. Bloomington, IN: IULC, 1981. 207-20.
- Willis, R. Clive. *An Essential Course in Modern Portuguese*. London: Harrap, 1971.