

The Application of Clitic Climbing in European Portuguese and the Role of Register

Aroldo Leal de Andrade
Universidade Estadual de Campinas

1. Introduction

Clitic climbing (CC), the manifestation of a clitic pronoun semantically related to a non-finite verb together with a governing verb, is a much-debated issue in corpus studies.¹ Many researchers have explored the topic with the intent of identifying the logic of the variation between CC and non-CC. In this paper we want to address a specific problem pointed out in the literature about CC in Spanish: its incidence is far less frequent in written texts. This can be interpreted as a problem in face of the diachronic evolution of the phenomenon, which showed higher frequency in medieval times. Therefore, the tendency for grammatical innovations (non-CC) to be more common in speech seems to be contradicted. I want to shed light on this problem by assigning the difference between registers to the dependence of the phenomenon to Information Structure. In order to do so, I demonstrate not only that register is a relevant variation factor in European Portuguese (EP), but also that it is connected to internal factors which are relevant to understand variation.

The paper presents the results from a EP corpus made up of about 1,000 data from different registers and sources: we have collected sentences from interviews given to linguistic databases and newspapers through the *Corpus do Português* (Davies & Ferreira 2006), and from contemporary Portuguese novels, all listed in the Appendix. The sentences include a finite verb followed by an infinitive, once CC is obligatory in EP in verb sequences involving a gerund or a participle.

Before moving along, it is necessary to differentiate between two constructions that show CC: “restructuring” and “clause union”.² While restructuring is usually linked to governing verbs of the modal, aspectual and movement classes which usually display semantic bleaching – see (1) – in “clause union” the governing verb is causative or perceptive. A verb of the latter type selects a complement with a non-coreferent subject. This embedded subject occurs in accusative or dative form, and climbs obligatorily, alongside other clitics – (2)b. Once there is no variation with respect to CC in clause union, we will only deal with restructuring data, similarly to previous studies about variation in CC.³

* This research was financed by Fapesp (DR 2006/50256-8). I would like to thank Daniel Silva for help with the translation of examples and Charlotte Galves, the organizers and audience of the *Hispanic Linguistics Symposium 2008* and two anonymous reviewers for their valuable comments. Any remaining errors are my responsibility.

¹ I use the term “governing verb” following Lo Cascio’s (1970) classical work on Italian. It implies that the verb is higher in a bottom-up tree, according to the Government and Binding Theory framework.

² The term “restructuring” refers to the transformation proposed by Rizzi (1982) that optionally reanalyzes the terminal chain $V_x(\mathbf{P})\mathbf{V}$ as a complex predicate (where V_x is a governing verb). Cf. González López (2008) and de Andrade (2010) for a review of proposals about restructuring in Generative Grammar. On the other hand, the term “clause union” appeared in Relational Grammar (Aissen & Perlmutter 1983).

³ Some authors contend that Spanish split clitics are possible in the “clause union” construction - *Le hice leerlo* - instead of moving both clitics to the higher domain - *Se lo hice leer* - both meaning “I made him read it”, where the dative clitic suffers allomorphy in the second case (Luján 1980:385). In EP clitic groups are usually realized as contracted forms, so there is no possibility of split clitics: *Fiz-lho reparar* (‘I made him/her fix it’).

- (1) Restructuring⁴
- a. *Pois quem o podia fazer, foi-se embora.* (Gullander)
So who it=could do, went=INHT away
'So whoever could do it went away'
 - b. *É extremamente importante começarmo-nos a abrir a isso,*
is extremely important begin.1PL=REFL to open to this
porque nem tudo é betão, nem matéria, nem interesses económicos. (JN)
because not everything is concrete nor matter nor interests economical
'It is extremely important for us to begin to get opened to this, because not everything is concrete, or matter, or economical interests'
 - c. *Saberei que me enganei quando me vieres dizer*
will.know.1SG that self=deceived.1SG when to.me=come.2SG say
que não se vê um único vulto de soldado cristão (Saramago)
that not IMPL=see.3SG an only shadow of soldier christian
'I will know I was wrong when you come to tell me that no one sees even a shadow of a christian soldier'
- (2) Clause union (Brito, Duarte & Matos 2003:857 (48)a; 860 (55)a)
- a. *O patrão mandou lavar o chão aos empregados antes de saírem.*
the boss made wash the floor to.the employees before of leave.INF.3PL
 - b. *O patrão mandou-hes lavar o chão antes de saírem.*
the boss made=them wash the floor before of leave.INF.3PL
'The boss made the employees wash the floor before leaving.'

The text develops in four sections. In section 2 I tackle the data on register differences and compare them to the results in Spanish. Section 3 presents quantitative results on internal factors. Section 4 includes a discussion where we try to give a unified analysis to the problem of variation with CC. Finally section 5 displays the concluding remarks.

2. Register and CC

The relevance of register has been pointed out in the literature about variation in Spanish CC. Davies (1995:373f) has noticed a systematic difference between registers in a corpus composed of texts from ten Spanish-speaking countries. His results on verb sequences with an infinitive show that the distance between registers with respect to CC indices can be as high as 30%. A similar methodology was applied by Cacoullos (1999) to study data with verb sequences with gerunds in Mexican Spanish. For instance, with *estar* +gerund, she reports 89% of CC in sociolinguistic interviews, against 68% in essays, thus confirming the relevance of register in CC.

First of all, it is necessary to verify whether the EP data replicate this expectation. In order to do so, I have grouped data according to genre and register. Therefore, sociolinguistic interviews are placed in the **informal** register. On the other hand, newspaper interviews and novels are considered as pieces of **formal** register. An inescapable observation that stems from this pattern is that the types of texts available in the corpus favor a connection between register and mode: the informal register is mainly found in the spoken mode.⁵ Having this in mind, let us observe the results taking registers and genres as a criterion in Table 1. There is a clearer application of CC in sentences expressing the informal register: 56.4%, much higher than the percentages found with formal interviews and novels: 36.2% and 32.1%. Such a distinction is much higher when we compare weights: .80 against .51 and .43, respectively.

⁴ In the glosses, clitics appear with the symbol '=' to show connection with their hosts. Verbal affixes shown are limited to agreement in person/number (e.g. 1SG, 2PL). This is the list of abbreviations used: INF: inflected infinitive; INHT: inherent clitic; IMPL: impersonal clitic; PASS: passive clitic; PL: plural; REFL: reflexive clitic; SG: singular.

⁵ This fact about corpora composition has driven some authors to use the term *register* to refer to the *mode* continuum, such as Davies (1995).

Table 1. The application of CC according to register and genre

	<i>N</i>	+CC %	<i>weight</i>	<i>Total</i> <i>N</i>
Informal: interviews	57	56.4	.80	101
Formal: interviews	138	36.2	.51	381
Formal: novels	171	32.1	.43	532

$$\chi^2 = 21.721; p < 0.001$$

As a first approximation, the meaning of register differences is related to markedness, understood in terms of context dependency. Word order is more flexible and pragmatic in the informal register, and more rigid and grammatical in the formal register (Givón 2001). If non-CC represents the most rigid option - once the clitic is positioned together with the verb that selects it - it should be the non-marked option for formal registers (and most of the texts in the written mode). The opposite is expected in the informal register, with the assignment of CC as a non-marked option.

In the diachronic axis, the Spanish data on register were interpreted by Davies (1995) with surprise. Once CC was (almost) obligatory in medieval Romance, the innovative character of non-CC should rather be more frequent in the informal register, but the opposite result is found. However, if the problem is framed in terms of the processing dimension, Davies' expectation can be reinterpreted. In order to show this, I will look into internal factor groups judged relevant for the application of CC. To this aim I will explore only the data linked to the formal register. This choice is due to the possibility of bias stemming from the inclusion of data from sociolinguistic interviews, together with the fact that they represent the smaller part of the sample.

3. Internal factors to CC

A plausible working hypothesis is that the same general principles govern CC in EP and in Spanish. The following factor groups were chosen for analysis, after the specialized literature:

- The clitic grammatical function (Myhill 1989);
- The presence and type of intervening elements (Cacoullos 1999);
- The frequency of the governing verb (Davies 1997; Cacoullos 1999).

Apart from these, we have also looked at the syntactic contexts for clitic placement. This is an important difference between European Portuguese and Spanish: the first one shows a "conservative" pattern of clitic placement, according to Uriagereka (1995). Proclisis (i.e. preverbal placement) is not determined by verb finiteness, but by the presence of proclisis operators. On the other hand, enclisis (i.e. postverbal placement) is found in all contexts lacking the relevant operators, including sentences with the verb in first position, or preceded by the subject, or by sentence adverbs. We will look more carefully into the meaning of proclisis operators below.

At this point, it is useful to make a digression for the sake of clarity. The EP clitic placement pattern makes us distinguish clitic **position** and **placement** as distinct operations: the first one is responsible for the syntactic category where the clitic appears, while the second is responsible for the ordering of the clitic with respect to this category (Galves, Ribeiro & Torres Morais 2005). This distinction, although not so evident in contemporary Spanish, is important to understand the case of European Portuguese, where both preverbal and postverbal CC are possible, apart from (postverbal) non-CC:

- (3) a. [...] *declaração que os pajens contestaram argumentando que sempre as deveremos matar (Saramago)*
kill
'such a declaration was contested by the boy servants who argued that we shall always kill them'

- b. “*Devem-nas destruir*” – disse a mulher da renda e da criança. (Jorge)
 must.3PL=them destroy – said the woman of.the lacework and of.the child
 ‘You should destroy them’ said the woman holding a lacework and a child’
- c. “*Devias divertir-te e fazer as senhoras felizes*” (Gullander)
 should.2SG have.fun=REFL and make the ladies happy
 ‘You should have fun and make the ladies happy’

Besides, when a preposition *de* is selected by the finite verb, preverbal non-CC is also possible. Therefore, CC can occur with proclisis or enclisis, as much as non-CC:

- (4) [...] *estamos sempre a tempo de o fazer e havemos de o fazer* (Público)
 are.1PL always in time of it=do and will.1PL of it=do
 ‘We always have time to do it and we will do it’

In the table below we group the results of the mentioned factor groups, ranked according to their relevance. Relevance was assessed by means of the range of factor weights.

Table 2. Multivariate analyses of the contribution of internal factors for the probability of CC

<i>Contemporary European Portuguese</i>			
<i>Corrected mean = .26</i>			
<i>Log likelihood = -418.242</i>			
<i>Total N</i>	<i>913</i>		
	<i>N</i>	<i>%</i>	<i>weight</i>
Clitic function			
Clitic group	8/ 9	88.9	.96
Ethic	16/ 31	51.6	.73
Dative	113/ 223	50.7	.72
Accusative	117/ 359	32.6	.48
Reflexive/ inherent	55/ 291	18.9	.30
<i>Range</i>			<u>.66</u>
Syntactic context			
Proclisis context	267/ 499	53.5	.76
Enclisis context	42/ 414	10.1	.20
<i>Range</i>			<u>.56</u>
Intervening elements			
V-V Adjacency	251/ 636	39.5	.55
Particle <i>a</i>	44/ 179	24.6	.46
Lexical Phrase	4/ 11	36.4	.40
Particle <i>de</i>	9/ 70	32.1	.25
Adverbial Phrase	1/ 17	5.9	.09
<i>Range</i>			<u>.45</u>
Verb frequency			
Most frequent verbs	273/ 721	37.9	.54
Less frequent verbs	36/ 192	18.8	.34
<i>Range</i>			<u>.20</u>

In the ensuing subsections I explain these results and assign the factor groups to one of two variation axes: clitic function and syntactic context are related to the marking of **topicality**; and intervening elements and verb frequency are connected to **predicate integration**.

3.1. Factors pointing to the relevance of topicality

In this section I intend to show that the results on clitic function and syntactic context are related to the notion of topicality. Topicality is assessed by means of two concepts, according to Givón (2001):

referential accessibility and thematic importance. These can be applied to referential expressions by means of hierarchies. For instance, Farkas & Kazazis (1980) have proposed two hierarchies to account for the ordering of the elements inside a clitic group in Romanian:

- (5) Topicality hierarchies (Farkas & Kazazis 1980: 77)
- a. Ethic > Goal > Theme
 - b. 1 > 2 > 3

I contend that the same preference orders are relevant for the application of CC in European Portuguese, in terms of a comparison between the topicality of the clitic to that of the subject. In fact, the results on Table 2 conform to the first hierarchy, which involves semantic roles, applied to clitic function: CC is much higher with ethics – 51.6% –, followed by datives (which receive the goal role) – with 50.7% – and then by accusatives (corresponding to the theme role), with 32.6% of CC.⁶

Apart from these, there are two other factors classified together with clitic function, which show the biggest and the lowest incidence of CC in this factor group: **clitic groups** and **reflexive/inherent clitics**. The high level of CC with the former (88.9%) can be also understood from the notion of topicality: the expression of two arguments by the clitic, one of those being necessarily human (the dative) raises its topicality.⁷ On the other hand, once the reflexive clitic is coreferent with the subject, it tends to show a lower topicality; the same can be argued for inherent clitics, which are a mere requirement of the verb and do not have a semantic value on their own. We will turn back to this issue in section 3.

- (6) a. *A razão que conheço não lha posso dar.* (Jorge)
the reason that know.1SG not it+to.you=can.1SG give
'The reason I know, I cannot give it to you'
- b. [...] *afastai de mim este cálice, disse o outro, e não Lhe adiantaria nada*
remove.2PL from me this cup, said the other, and not to.him=help nothing
que outra vez lho tornariam a impor. (Saramago)
because another time it+to.him=come.back.3PL to impose
'Remove this cup from me, said the other, and it would not be helpful to him, because they would impose it on him another time.'

Turning now to the syntactic contexts for clitic placement, I have found a bigger preference for CC when the sentence shows a proclisis operator: 53.5%, while clauses representing enclisis contexts show only 10.1% of CC. This is especially meaningful if we consider that in independent clauses enclisis is the non-marked option, while proclisis is obligatory in sentences showing any of the following elements in the beginning of the sentence (to be more precise, in a position “governing” the clitic; for detailed lists of these elements, see Salvi 1990; Duarte & Matos 2000, among others):

- complementizers (*que, se...* ‘that, if’);

⁶ Ethics include **ethic datives** - which impose a reading according to which the person referred to is regarded with interest - and **possessive datives**, which correspond semantically to possessive pronouns:

- (i) *Eu até lhe posso juntar um bocadinho de água, tá um bocadinho grossa.* (CRPC)
I even to.you=can put a little.bit of water, is a little.bit thick
'I can even put a little bit of water in it, it is a little bit thick'
- (ii) [...] *agora poderia pôr-lhe um braço sobre os ombros* (Saramago)
now I could put=to.him an arm over the shoulders
'Now I could put an arm on his shoulders'

⁷ Clitic person (and animacy) are the primary factors pointed out by Silverstein (1976) in order to assess topicality, as shown in the scale in (i). The quantification of topicality in terms of person was not statistically significant in my data, because of the high quantity of *se* clitics.

- (i) 1st person pronouns > 2nd person pronouns > 3rd person pronouns > proper names > kinship terms > human names > animate names > inanimate names

- some adverbs (*nunca, já, bem...* ‘never, already, indeed’);
- some negated or quantified phrases (*nenhum/algum/todo X...* ‘no/some/all X’);
- interrogative and focalized phrases (*quem, só/até X...* ‘who(ever), only/even X’);
- the sentential negation marker *não* (‘not’).

Such elements have in common the activation of an affective operator, in the sense of Klima (1964). These impose a relation with a variable, represented by the clitic, which then tends to assume a presupposition of existence acquired from this semantic relation. Since topics are elements necessarily presupposed, we can again establish a connexion between CC and high topicality of the clitic.

These results demonstrate that clitic position and placement, although being separate concepts, are strongly connected. As far as I know, no other study noticed this fact before. The null hypothesis would be that restructuring could occur independently of the clitic position with respect to the governing verb. This connection gets clearer when we look at the results combining clitic position and placement:

Table 3. Clitic position and placement in European Portuguese restructuring predicates

	<i>Preverbal CC</i>	<i>Postverbal CC</i>	<i>Preverbal Non-CC</i>	<i>Postverbal Non-CC</i>
Proclisis context	294, 54%	0, -	17, 3%	225, 42%
Enclisis context	0, -	64, 14%	23, 5%	355, 80%
Total	294, 30%	64, 6%	40, 4%	580, 59%

As a consequence of the connection between proclisis contexts and CC, the number of cases of postverbal CC is much lower than those of preverbal CC (preverbal non-CC has low indices across the board because there are few verbs creating the context for this placement option, i.e., that select the preposition *de*).

3.2. Factors pointing to the relevance of predicate integration

Two other relevant factor groups were presented in Table 2: the presence and type of intervening elements between the verbs composing a complex predicate; and governing verb frequency. These factors point to the relevance of predicate integration, which will be higher the closer the relevant verbs are, and the less meaning the governing verb displays (in other words, the more auxiliary it is).

The loss of unity between governing verb and infinitival verb can be intuitively understood from the fact that there are intervening elements that break their adjacency. While Spanish allows adverbs, focalizing elements and floating quantifiers between the two verbs, European Portuguese is more liberal, once it also allows subject NPs and PPs.

- (7) Intervening elements in Spanish complex predicates (Moore 1996:53-54)
- a. *Lo suelen **siempre** terminar antes de la hora.*
*it=tend.3PL **always** finish before of the hour*
 ‘They tend to always finish before time is up’
 - b. *Lo quisiera **por supuesto** comprar barato.*
*it=wanted.3SG **of course** buy cheap*
 ‘S/he would like of course to buy it cheaply’
 - c. *Te lo hacen **hasta** aborrecer.*
*to.you=it=make.3PL **even** hate*
 ‘They make you even hate it’
 - d. *Los niños lo quieren **todos** comer.*
*the children it=want.3PL **all** eat*
 ‘The children all want to eat it’
- (8) Additional intervening elements in European Portuguese complex predicates
- a. *porque lhe queria **Deus** mostrar a grandeza da sua misericórdia (Saramago)*
*because to.him=wanted **God** show the greatness of.the his mercy*
 ‘Because God wanted to show him the greatness of his mercy’

- b. *Bem me queria a mim parecer que a história não é a vida real*
 well to.me=wanted.3SG **to me** seem that the history is not the life real
literatura, sim, e nada mais. (Saramago)
 literature yes and nothing more
 ‘I preferred to believe that history is not real life; it is literature, yes, and nothing more’

According to Hawkins (2004), processing complexity increases the bigger the domain where a dependency relation occurs. Since the climbed clitic and the infinitival verb enter in a dependency relation (for instance, because of semantic role assignment), if an XP apart from the governing verb occurs between the clitic and the infinitive, there will be a delay in the acknowledgment of such a relation. Apart from this general consideration, I observe that the type of intervening element is also relevant for the possibility of disruption of V-V adjacency: an argument XP is more acceptable than an adjunct XP. This can explain the higher acceptance of CC with lexical phrases if compared to adverbial phrases, as shown in Table 2: 36.4% versus 5.9%. Inside the group of lexical phrases, which includes NPs and PPs, the same trend whereby arguments allow CC is found.

Turning now to the role of the governing verb in variation with CC, the literature has approached the level of grammaticalization of auxiliary verbs by means of its frequency. One approach, which considers an indirect relation between grammaticalization and frequency, is represented by Myhill (1989). He proposes that CC “should be more likely when the finite verb represents a meaning which is commonly represented grammatically in the languages of the world” (p. 354). The proposed semantic hierarchy is the following:

- (9) Grammaticalization hierarchy of restructuring verbs (Myhill 1989, with adaptations)
 Progressive aspect (*estar, ir, andar, venir* + gerund ‘be/ end up+gerund’) > Future tense (*ir* ‘will’) > Epistemic modality (irrealis *ir, haber de, poder* ‘can, shall’) > Movement (*ir, venir* ‘go, come’) > Deontic/root modality (*tener, querer, poder, deber* ‘must, want, may’) > Inceptive aspect (*empezar, comenzar* ‘begin, start’)

The second approach to grammaticalization considers a direct relation with frequency, and was applied to Portuguese by Davies (1997) and to Spanish by Cacoulios (1999). This perspective derives from Bybee’s (2003) theory according to which the repetition of a lexical item has a relevant role in the implementation of grammaticalization. She has illustrated the relevance of repetition for the generalization of meaning observed with some English auxiliaries. During this process, not only has verb frequency increased, but also the syntactic contexts where the verb occurs.

The second approach was chosen, relating governing verb frequency and CC. In order to establish a threshold for verb frequency, I have divided the amount of data (913) by the number of lexical entries (26). Therefore, verbs presenting more than 35 tokens in the corpus were classified amongst the most frequent.⁸ This is the exhaustive list of predicates, classified according to this methodology:

- (10) a. Most frequent verbs: *querer* (‘want’), *poder* (‘can’), *dever* (‘should’/‘must’), *ter de* (‘must’), *ir* (‘will’), *vir a* (‘end up+gerund’), *começar a* (‘begin’), *estar a* (‘be+gerund’);
 b. Less frequent verbs: *tentar* (‘try’), *desejar* (‘desire’), *esperar* (‘hope’), *pretender* (‘intend’), *procurar* (‘seek’), *conseguir* (‘get’), *saber* (‘know’), *haver de* (‘will’/‘let’), *tornar a/voltar a* (‘do X again’), *costumar a* (‘be used to’), *chegar a* (‘end up’), *acabar de* (‘just+participle’), *deixar de* (‘stop’), *ficar a* (‘stay+gerund’), *continuar a* (‘keep+gerund’), *andar a* (‘be+gerund’), *precisar* (‘need’).

As expected, the results on verb frequency in Table 2 show a higher preference for climbing with the most frequent verbs, which show 37.4% of CC against 18.8% of CC with the less frequent verbs. The

⁸ I have checked for correspondence between verb frequency in the corpus and their overall frequency in language use by looking for similar data in EP texts from the 19th and 20th centuries in the *Corpus do Português*. The same distribution is repeated at the level of 1500 tokens as a threshold.

relevance of this classification - based on language use and not on formal features - could be confirmed because the quantification using the raising/control divide was not significant ($\chi^2 = 1.531$; $p = 0.215$).

Therefore, predicate integration seems to be a relevant aspect for the application of CC, as noticed from the results on intervening elements and verb frequency. Now the following questions remain to be answered:

- Is there any relation between topicality and predicate integration as regards CC?
- Is there a general logic that can explain the variation results in terms of internal factors and the higher incidence of CC in the informal register?

4. In search of a unified analysis

In this section I intend to develop two connections. First I want to show that topicality and predicate integration are related variation axes. Second, this unified analysis of internal factors to the incidence of CC should give a clearer meaning to the register differences mentioned in the beginning of this paper. Before tackling these problems, the comprehension of semantic and pragmatic values related to CC can pave the way for the necessary unified analysis of CC.

The results in section 3.1 have suggested that the climbed clitic displays high topicality. Uriagereka (1988) has presented a test related to *de re/de dicto* interpretations that concurs to this fact.

- (11) CC and coercion of *de re* interpretation (Uriagereka 1988:375 (104))
 ? *El senador le queria dar unas hostias a Samuel Clemens por suas arengas politicas*
 the senador him=wanted give some punches to Samuel Clemens for his harangues political
 # *pero se rajó especificamente porque uno le dijo que Clemens era Mark Twain.*
 but self=chickened.out specifically because one him=said that Clemens was Mark Twain
 ‘The senator wanted to punch Samuel Clemens for his political harangues, but he chickened out specifically because someone told him Clemens was Mark Twain’

This sentence becomes marginal because of the inclusion of the coordinated clause initiated by *pero*, which imposes a *de dicto* interpretation, thus contradicting the *de re* interpretation coerced by the climbed clitic, whereby ‘Samuel Clemens’ receives a presupposition of existence. Therefore, the scope of *Samuel Clemens* is higher in the clause, because it is coreferent with the dative clitic, higher than the accusative argument *unas hostias*. It is important to notice that in a sentence showing non-CC, the inclusion of the coordinated sentence is uneventful to its acceptability.

This fact stems from the concept of intensionality assigned to the complement of restructuring verbs, with consequences for the interpretation of referential expressions in its domain. For instance, Chierchia & McConnell-Ginet (2000: 307) explain that in the sentence *Bond believes that the author of this letter is a spy*, the referential expression *the author of this letter* can be understood as a specific person in the world, e.g. *Bill*, or can be only a part of the Bond’s belief. In the latter case, Bond believes that the actual world is in a set of worlds where someone wrote this letter, whoever this person is. One more time, the scope of the referential expression determines its interpretation.

This semantic interpretation implies that CC imposes the high scope interpretation of the clitic referent with respect to intensional operators. This has a consequence in the pragmatic level: the climbed clitic has a direct connection with another topical expression of the clause, which usually happens to be the subject. In the expression given by Chierchia & McConnell-Ginet, *Bond* and *Bill* must have been both previously mentioned in the discourse, so that the anaphoric meaning of the referential expression turns out to be possible. We want to identify the informational status of the clitic in this case as a **secondary topic**.

The possibility of multiple topics in a clause follows from the concept of topic as an element pragmatically presupposed for the utterance to be construed about it. According to Erteschik-Shir (1997), every clause has a primary (or sentence) topic, which is the element to which the truth value of a sentence is evaluated. A secondary topic, on the other hand, is an entity such that the utterance is construed to be about the relationship between it and the primary topic (Nikolaeva 2001:26). We contend that this is the case when we find a sentence with CC: the utterance involves at least two

potential topics, usually the subject and the clitic, both salient elements and therefore already available in the discourse.

The informational role of CC is a relatively unexplored subject. A significant study in this line is Gill (2000), who proposes that while non-CC represents the default position of the clitic, CC “will function to keep the discourse topic salient, where there is a new sentence topic or to ground new information by linking it to the previous referent” (p.86). Observe the following excerpt:

- (12) Clitic saliency in the discourse (Gill 2000: 170; example from *Pepita Jiménez*, 75)
Antes me canso yo que él_i, y no queda vericueto, ni lugar agreste ni cima de cerro escarpado
 first INHT=tire I than he and not stays faraway place nor place wild nor peak of hill steep
*en estas cercanías, adonde no lleguemos. [El señor Vicario]_j, **me_i=va reconciliando mucho***
 in this vicinity where not arrive.1PL. The Mr. Vicario me=goes reconciling a.lot
con el clero español, a quien algunas veces he tildado yo_i, hablando con usted, de poco ilustrado.
 with the clergy spanish to who sometimes have criticized I, talking with you of little illustrated
Cuanto más vale, me digo_i, a menudo, [este hombre]_j, lleno de candor y de buen deseo
 How.much more worth me=say.1SG at tiny this man full of candor and of good desire
 ‘First I tire before him, and there is no rough, pathless place, nor wild place, nor a steep peak
 of a hill in this vicinity, where we don’t go. Father Vicario takes me reconciling a lot with the
 Spanish clergy, those whom I sometimes have criticized, speaking with you, of being little
 illustrated. How much more valuable, I say frequently, this man, filled with candor and good
 desire’

The bold clause in (12) presents two participants who had been already introduced and referred to in the previous discourse. Therefore, the clitic has a high saliency and tends to be interpreted as a secondary topic, a situation marked in word order; in other words, through CC.

I assume that the same principle holds for CC crosslinguistically. A corollary of this principle is the high incidence of CC with passive *se* (not included in the EP corpus because the clitic is said to be generated in the high position, depending on the approach), and the low frequency of CC with reflexive clitics.

- (13) *Os milagres se podem repetir sem que padeça minguar a potência miraculosa. (Saramago)*
 The miracles PASS=can repeat without that suffer shortage the power miraculous
 ‘Miracles can be repeated without the miraculous power suffering any shortage’
- (14) *Os colonos procuravam envolver-se pouco na guerra. (CRPC)*
 The colonists tried involve=REFL little in.the war
 ‘The colonists tried to avoid getting involved in the war’

If passive *se* is interpreted according to Raposo & Uriagereka (1996), it has an argument function as the agent of the sentence. Therefore, it tends to be interpreted as a secondary topic, once another element taking the role of subject (*os milagres* ‘miracles’) is identified as the primary topic. On the other hand, since reflexive clitics are coreferent with the subject, they cannot be identified as secondary topics. Remember that, in order to get this interpretation, a clitic must have a distinct referent from the primary (or sentence) topic.

In the proposed framework the results on variation that point to the relevance of topicality to the application of CC are expected. Now we can turn to the questions posed in the end of the last section. First of all, how does predicate integration fits into this picture? If the complex predicate has signs of disruption, either by the lexical status of the governing verb or by the insertion of intervening elements between the verbs forming it, the clitic cannot be interpreted as a topical element but as a necessary part of the complement, without a presupposed existence. Therefore, this factor group is not completely independent of topicality, but can help or hinder the assignment of a topical interpretation to the clitic. This is what the examples below suggest.⁹

⁹ I thank an anonymous reviewer for pointing these examples to me.

- (15) *Voy a recoger a los niños.*
 go.1SG to pick.up to the kids
 ‘I am going to pick up my kids’
- (16) a. *Voy al colegio a recogerlos*
 go.1SG to.the school to pick.up=them
 b. **Los voy al colegio a recoger*
them=go.1SG to.the school to pick.up
 ‘I am going to pick them up at school’

In the examples above, the verb *ir* (‘go’) varies between its lexical meaning of movement and its grammaticalized meaning pointing to future tense. While in (16)a the verb varies among both readings, in (16)b the only interpretation available is the future tense, because of CC. However, this requirement is in conflict with the presence of a locative PP which imposes the lexical meaning of *ir*. Consequently, in this example the clitic referent *los niños* does not necessarily acquire the presupposed interpretation of existence; instead, it tends to be interpreted as part of the adjunct clause introduced by *a recoger* (‘to pick up’), an event in a possible world.

Second, how do register differences derive from the Informational Structural status of CC? I propose that the strategies for presenting background information in different registers alongside with the connexion between backgrounding and proclisis contexts ensure a better understanding of the role of register in CC. Barry (1987) proposes, with respect to clitic placement in Old Spanish, that proclisis is linked to the contexts in which one uses discursive backgrounding, while enclisis is linked to foregrounding contexts. While the latter develop a narrative, the former bring details and additional information. This can be done by the insertion of dialogues, remarks to the reader and adverbs making reference to mentioned facts. I believe that such correlation is valid for clitic position in EP, also indirectly to clitic placement, since proclisis operators are grammaticalized in this language. Once a strong correlation between syntactic contexts for clitic placement and CC was found, I examine the distribution of +CC data according to syntactic contexts in Table 4.

Table 4. Distribution of +CC data according to register and syntactic context

<i>Register/genre</i>	<i>Context</i>	<i>N</i>	<i>%</i>
Informal: interviews	Proclisis context	32	56.1%
	Enclisis context	25	
Formal: interviews	Proclisis context	108	79.4%
	Enclisis context	28	
Formal: novels	Proclisis context	154	93.3%
	Enclisis context	11	

The number of sentences in proclisis contexts is proportional to the level of formality of the text: 56.1% for informal texts, 79.4% in formal interviews and 93.3% in novels. Such results reflect the fact that formal texts are less dependent on context; therefore, the recourse to backgrounding must be more frequent. On the other hand, this means that the proportion of +CC in the informal register reflects the topicality of the arguments, independently of the existence of an affective operator. This is also expected under Givón’s (2001) observation about informal registers, where strategies of topicalization are expected to be much more frequent.

A consequence of this asymmetry is that the informal register shows a higher number of data of postverbal CC with respect to the formal register. We believe at least part of these results can be linked to the concept of backgrounding if in the relevant excerpts the speaker has his/her attention turned to the interlocutor, not to the topic of conversation. This is illustrated in (17). In (18), on the other hand, the clitic shows high topicality because it refers to a referent just reactivated in the discourse: *os teus [filhos]* ‘your children’.

- (17) *Vens? Vou-te dar de lanchar, anda! (CRPC)*
 come.2SG? will.1SG=to.you give of snack come
 ‘Will you come? I will give you something for a snack, hurry up!’
- (18) — *Eu é que não tenho vida, não é, não tenho criada, não tenho vida para...*
 I is that not have life, not is, not have maid, not have life for
E os teus? Vens-os buscar? (CRPC)
 and the.PL yours? come.2SG=them pick.up
 ‘It’s me who does not have a life, you know, I don’t have a maid, I don’t have a life for... and
 your (children)? Are you coming to take them?’

Therefore, the distinction between backgrounding and foregrounding seems to be more directly related to clitic position, and partly related to clitic placement in EP. The meaning of register differences to the application of CC can be revealed from the occurrence of syntactic contexts associated with backgrounding, which seem to be higher proportionally to the formality of the text. Besides, the function of pure topicality marking by the climbed clitic is higher in the informal register.

5. Concluding remarks

This paper has presented a description and analysis of variation in the application of CC in contemporary European Portuguese, with focus on register differences and its meaning with respect to internal factors. The results confirm the observation made by other authors according to which CC is much more frequent in the informal register. Besides, they give support to an analysis that assigns an informational role to CC.

I have characterized the climbed clitic as a secondary topic. As a consequence, the results pointing to the relevance of topicality, such as clitic function and the syntactic contexts for clitic placement, receive a straightforward account. On the other hand, the factors related to predicate integration are also relevant, once a clitic which is too far away from the governing verb will be probably interpreted as part of the intension of the verb complement instead of having a presupposed existence.

If CC is observed through these lenses its diachronic evolution is not strange: medieval Romance languages showed frequent recourse to topicalizations, a signal of a V2 grammar for some researchers. A prediction of the mentioned analysis is that the higher a language (or register) makes use of informationally marked constructions, the higher the incidence of CC will be. Since the discussion of diachronic data would lead us too far afield, we leave this matter unexplored here.¹⁰

Similarly to previous research, our main conclusion is that variation in CC is not free. In other words, we contend that it is not dependent on stylistic preferences. Register differences found in CC show that this phenomenon is primarily dependent on Information Structure.

Appendix. European Portuguese Corpus

Novels

Saramago	Saramago, José. (1989) <i>História do Cerco de Lisboa</i> . Lisboa: Editorial Caminho.
Jorge	Jorge, Lídia. (1995) <i>A Costa dos Murmúrios</i> . Lisboa: Dom Quixote.
Reis	Reis, Patrícia. (2007) <i>Morder-te o coração</i> . Rio de Janeiro: Língua Geral.
Gullander	Gullander, Miguel. (2007) <i>Perdido de volta</i> . Rio de Janeiro: Língua Geral.

Interviews

CRPC	Corpus de Referência do Português Contemporâneo [sociolinguistic interviews]
JN	<i>Jornal de Notícias</i> [newspaper]
Público	<i>Público</i> [newspaper]
Web	Websites [interviews]

¹⁰ However, I refer the interested reader to de Andrade (2010) for a development of this idea in terms of the change from Classical Portuguese to Modern European Portuguese.

References

- Aissen, Judith & David M. Perlmutter (1983) Clause reduction in Spanish. In: David Perlmutter (ed.) *Studies in Relational Grammar 1*. Chicago: The University of Chicago Press, 360-403.
- de Andrade, Aroldo. (2010) *A subida de clíticos em português: um estudo sobre a variedade europeia dos séculos XVI a XX*. PhD dissertation. Campinas, Brazil: Universidade Estadual de Campinas.
- Barry, A. K. (1987) Clitic Pronoun Position in Thirteenth-Century Spanish. *Hispanic Review* 55:2, 213-220.
- Brito, Ana Maria; Inês Duarte & Gabriela Matos. (2003) Tipologia e distribuição das expressões nominais. In: Maria Helena Mira Mateus et al. *Gramática da Língua Portuguesa*. 5th. ed. Lisboa: Caminho, 795-867.
- Bybee, Joan. (2003) Mechanisms of change in grammaticization: The role of frequency. In Brian D. Joseph & Richard D. Janda (eds.) *The Handbook of Historical Linguistics*. Oxford: Blackwell, 602-623.
- Cacoullos, Rena Torres. (1999) Construction frequency and reductive change: diachronic and register variation in Spanish clitic climbing. *Language variation and change* 11, 143-170.
- Chierchia, Gennaro & Sally McConnell-Ginet. (2000) *Meaning and grammar: an introduction to semantics*. 2nd. ed. Cambridge, MA: MIT Press.
- Davies, Mark. (1995) Analyzing Syntactic Variation with Computer-Based Corpora: The Case of Modern Spanish Clitic Climbing. *Hispania* 78: 2, 370-380.
- Davies, Mark. (1997) A corpus-based approach to diachronic clitic climbing in Portuguese. *Hispanic Journal* 17, 93-111.
- Davies, Mark & Michael Ferreira. (2006) *Corpus do Português* (45 million words, 1300s-1900s). [<http://www.corpusdoportugues.org>]
- Duarte, Inês & Gabriela Matos. (2000) Romance Clitic and the Minimalist Program. In: João Costa (ed.) *Portuguese Syntax: New Comparative Studies*. Oxford: Oxford University Press, 116-142.
- Erteschik-Shir, Nomi. (1997) *The Dynamics of Focus Structure*. Cambridge: Cambridge University Press.
- Farkas, Donka & Kostas Kazazis. (1980) Clitic pronouns and topicality in Rumanian. In: Jody Kreiman & Almerindo E. Ojeda (eds.) *Papers from the sixteenth regional meeting of the Chicago Linguistic Society*. Chicago, 75-82.
- Galves, Charlotte; Ilza Ribeiro & Maria Aparecida Torres Morais. (2005) Syntax and morphology in the placement of clitics in European and Brazilian Portuguese. *Journal of Portuguese Linguistics* 4:2, 143-177.
- Gill, Deborah Jean. (2000) *An anaphoric approach to clitic position in Spanish*. PhD dissertation. Los Angeles, CA: University of Southern California.
- Givón, Talmy. (2001) *Syntax: An Introduction*. Amsterdam/Philadelphia: John Benjamins.
- González López, Verónica. (2008) *Spanish Clitic Climbing*. PhD dissertation. University Park, PA: Pennsylvania State University.
- Hawkins, John A. (2004) *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Klima, Edward S. (1964) Negation in English. In: Jerry A. Fodor & Jerrold J. Katz. (eds.) *The Structure of Language: Readings in the Philosophy of Language*. Englewood Cliffs, NJ: Prentice Hall, 246-323.
- Lo Cascio, Vincenzo. (1970) *Struttura pronominali e verbali italiane*. Bologna: Zanichelli.
- Luján, Marta. (1980) Clitic Promotion and Mood in Spanish Verbal Complements. *Linguistics* 18, 381-484.
- Moore, John C. (1996) *Reduced constructions in Spanish*. New York & London: Garland.
- Myhill, John. (1989) The grammaticalization of auxiliaries: Spanish clitic climbing. In: Shelley Axmaker; Annie Jaissner & Helen Singmaster (eds.) *Proceedings of the 14th. Annual meeting of the Berkeley Linguistics Society: General session and Parasession on Grammaticalization*. Berkeley, CA, 352-363.
- Nikolaeva, Irina. (2001) Secondary topic as a relation in information structure. *Linguistics* 39:1, 1-49.
- Raposo, Eduardo & Juan Uriagereka. (1996) Indefinite SE. *Natural Language and Linguistic Theory* 14:4, 749-810.
- Rizzi, Luigi. (1982) A restructuring rule. In: *Issues in Italian Syntax*. Dordrecht: Foris, 1-48.
- Salvi, Giampaolo. (1990) La sopravvivenza della legge di Wackernagel nei dialetti occidentali della Penisola Iberica. *Medievo Romanzo* XV: 2, 177-210.
- Silverstein, Michael. (1976) Hierarchy of features and ergativity. In: R. M. W. Dixon (ed.) *Grammatical categories in Australian languages*. Canberra: Australian Institute of Aboriginal Studies, 112-171.
- Uriagereka, Juan. (1988) *On Government*. PhD dissertation. Storrs, CT: University of Connecticut.
- Uriagereka, Juan. (1995) An F Position in Western Romance. In: Katalin É. Kiss. *Discourse Configurational Languages*. Oxford: Oxford University Press, 153-175.

Selected Proceedings of the 12th Hispanic Linguistics Symposium

edited by Claudia Borgonovo,
Manuel Español-Echevarría,
and Philippe Prévost

Cascadilla Proceedings Project Somerville, MA 2010

Copyright information

Selected Proceedings of the 12th Hispanic Linguistics Symposium
© 2010 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-440-9 library binding

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

Ordering information

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

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

Web access and citation information

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

This paper can be cited as:

de Andrade, Aroldo Leal. 2010. The Application of Clitic Climbing in European Portuguese and the Role of Register. In *Selected Proceedings of the 12th Hispanic Linguistics Symposium*, ed. Claudia Borgonovo, Manuel Español-Echevarría, and Philippe Prévost, 97-108. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2409.