

Clitic climbing in Caracas Spanish: A sociolinguistic study of *ir* and *querer**

Aarnes Gudmestad
Indiana University

Abstract

The present study investigates clitic climbing in spoken Modern Spanish of Caracas, Venezuela. The data come from a corpus of oral interviews. All examples of clitics appearing with the finite verbs *ir* 'to go' and *querer* 'to want' followed by a nonfinite verb are coded for linguistic and social factors. Linguistic variables include verb tense and type and clitic animacy, person, and number. The social variables are socioeconomic class, gender, and age. Results show that clitic placement was significantly more frequent in the pre-finite position than in the post-nonfinite position. The two verbs also behave differently. Linguistic features influence clitic placement for *ir*, whereas social factors influence clitic placement for *querer*.

Introduction

Clitic climbing (CC) is a phenomenon in which an unstressed object pronoun is moved from a post-nonfinite verb position to a pre-finite verb position, as in *va a hacerlo* to *lo va a hacer* (Napoli 1981). Previous research on CC in Spanish has studied its historical development and has characterized its use according to syntactic and semantic features. Although some researchers have also examined this phenomenon across genres (Davies, 1995; Torres Cacoullos, 1999) and geographical region (Davies, 1995), to the researcher's knowledge, other social factors had not been addressed when the present study began. Thus, the current investigation not only adds to the body of research on linguistic properties influencing clitic placement but also extends the analysis to previously neglected social variables. Specifically, the use of CC with *ir* 'to go' and *querer* 'to want' in spoken Modern Spanish of Caracas, Venezuela is studied.

Background

Limited research on CC has been conducted on naturally occurring data, though Myhill (1988) and Davies (1995, 1997) have addressed this need. Myhill (1988) examined the grammaticalization of this phenomenon. A quantitative analysis of Modern Spanish prose revealed that semantic properties of the finite verb and pragmatic factors of the subject and the clitic interacted to affect CC. With respect to the meaning of the finite verb, Myhill argued that when the verb had an auxiliary or non-basic meaning, CC was favored, whereas a non-auxiliary or basic meaning disfavored CC. CC with verbs expressing an auxiliary meaning signified grammaticalization. Even though detailed criteria for distinguishing basic from non-basic meaning were not provided, analyses of various verbs showed that only the latter experienced a higher rate of CC. Specifically, when *ir* conveyed progressive meaning in *ir + gerund* constructions or futurity in *ir + a + infinitive* constructions, CC was more likely than when *ir* expressed the basic meaning of motion. Similarly, the auxiliary meaning of antagonism for *querer* favored CC while the non-auxiliary meaning of desire favored the post-nonfinite verb position. Moreover, this study showed that the animacy hierarchy influenced CC. According to this hierarchy in Spanish, Myhill stated that animacy is ranked as follows for CC: second person > first person > third person-singular-human > third person-plural-and/or nonhuman. In other

* A previous version of this paper was presented at the Second Annual Graduate Student Conference on Luso-Brazilian and Hispanic Literature, Linguistics and Culture, Indiana University – Bloomington, February 2005.

words, second person is more animate than first person; first person is more animate than third person-singular-human; and third person-singular-human is more animate than third person-plural-and/or nonhuman. The data showed that CC was favored when the clitic was more animate than the subject of the sentence and disfavored when the clitic was lower on the hierarchy than the subject.

Davies (1995) continued the study of CC with written prose, extended the analysis to spoken conversational corpora from the 1970s and compared the rate of CC use between the two registers. Ten Spanish speaking countries and 11 cities were represented in the data. Results showed that CC was more frequent in the spoken register but that the difference among regions was minor. The study also examined linguistic factors. Finite verbs did not categorically permit or prohibit CC, but rather there appeared to be a continuum of acceptability. For example, *ir* and *querer* exhibited high rates of CC whereas *intentar* 'to try' and *esperar* 'to hope/wait' did not. Higher rates of CC occurred when more than one clitic was present, when the clitic was not reflexive, when the clitic was animate, and when the preceding element was *que* 'that' or *y* 'and'.

In an attempt to account for the difference between the registers, Davies (1995) assumed that spoken data more accurately reflected language use and postulated that the most likely explanation was that the two registers have not evolved in a similar fashion. While CC has decreased in the written register since Middle Spanish, it has remained relatively constant in the oral register. He left the question of why the two registers show different diachronic development to future research.

Finally, Davies (1997) investigated the historical development of CC using an extensive corpus of written texts ranging from the 1200s to the 1800s. The analysis showed that CC gradually decreased from one century to the next and then began to increase in the 1800s. The decrease in CC developed first in *preposition + infinitive* constructions, then in *verb + preposition + infinitive* constructions, and finally in *verb + infinitive* constructions. The data also revealed that the auxiliary versus non-auxiliary meaning distinction of the finite verb influenced clitic placement at all stages and played an important role in reversing the decrease in CC. These studies illustrated that oral data have not been extensively studied and that many social factors have yet to be investigated.

The Present Study

Research Questions and Hypotheses

Previous research on CC has guided the research questions for the present investigation. The current study continues the analysis of various linguistic variables and initiates the examination of three social variables. Hypotheses follow each research question.

- (1) Is clitic placement more frequent in the post-nonfinite verb position or the pre-finite verb position for *ir* and *querer*?

It is hypothesized that clitic placement will be more frequent in the pre-finite verb position. Results from an analysis of oral data presented in Davies (1995) showed that 53% of clitics were produced in the pre-finite verb position in Caracas speech. Clitics in the pre-finite verb position occurred at a rate of 56% when data from the 10 Spanish speaking countries were averaged together.

(2) Do linguistic factors predict CC for *ir* and *querer*?

Hypotheses are given for each of the five linguistic variables (verb type, clitic animacy, clitic person, clitic number, and verb tense) and will be discussed in turn. First, for the variable of verb type, it is expected that CC will occur more frequently with *ir* constructions than with *querer* constructions. According to the averages for the spoken data in Davies (1995), CC occurred in 86% of the *ir* tokens and 47% of the *querer* tokens.

Second, it is hypothesized that clitic animacy will influence CC. Myhill (1988) and Davies (1995) both concluded that animate clitics favored CC. Myhill's data showed a strong effect for the animacy hierarchy. CC was more common when the clitic was more animate than the subject on the animacy hierarchy. Davies (1995) did not examine the relationship between the subject and the clitic but the oral data he analyzed revealed that animate clitics were more likely to be preposed. The current study followed Davies (1995): It limits the analysis to the clitic and does not investigate the relationship between the clitic and the subject.

Third, it is expected that clitic person will also influence CC. The hypothesis for this linguistic factor is related to the animacy hierarchy; it can be recalled that Myhill (1988) provided the following hierarchy for CC in Spanish: second person > first person > third person-singular-human > third person-plural-and/or nonhuman. Although Myhill (1988) found this hierarchy to be significant when both the subject and the clitic were considered, the present study does not investigate the role of subject animacy on CC. Instead, it is hypothesized that second-person clitics will be more likely to be placed in the pre-finite verb position than first-person clitics and first-person clitics will be more likely to occur in the pre-finite verb position than third-person clitics (all third-person clitics will be examined together in the present investigation).

Fourth, no previous research on CC has investigated the effect of the number of the clitic, so no hypothesis is formulated at this time. The present study aims to initiate the analysis of potential similarities or differences between singular and plural clitics and clitic placement.

Fifth, previous research has not investigated the role of verb tense in CC. However, the discussion of auxiliary (non-basic) and non-auxiliary (basic) meaning of the finite verb in Myhill (1988) provided the motivation for this research question. Myhill coded for basic and non-basic meaning but it was unclear how such an analysis was carried out. According to his study, the non-basic meaning of *ir* depended on the syntactic structure of the sentence. The auxiliary meaning of *ir* in the *ir + a + infinitive* periphrastic construction expressed futurity and in the *ir + gerund* construction constituted a progressive meaning. He classified 'to want' as the non-auxiliary meaning for *querer* and a meaning of antagonism as an auxiliary meaning. For both verbs the results showed that the auxiliary meaning favored CC and that the non-auxiliary meaning disfavored CC.

With a lack of specific criteria for determining finite verb meaning, the assumptions put forth in Myhill (1988) are questioned in the current study. First, it is unclear how a future reading could be avoided in the periphrastic future construction when *ir* is in the present tense. For example, the sentences *voy a buscarlo* or *lo voy a buscar*, 'I am going to look for it/him', a future reading is clearly conveyed regardless of

any motion meaning that may be derived from the infinitive.¹ The one example Myhill provided with *ir* in the present tense conveyed futurity and occurred with a clitic placed in the pre-finite verb position. Myhill did cite one example in which *ir* expressed motion: “yo fui a decirLE”, ‘I went to tell him’ (p. 237, original capitalization). It seems that in this construction *ir* can express futurity or intention in the present and imperfect and directionality in the preterit. Similarly, after the researcher of the current study consulted native speakers of Spanish, it appears that *querer + infinitive* constructions convey a meaning of ‘want’ unless they are used in the preterit. Consequently, the present investigation has coded for verb tense instead of meaning. If this linguistic factor is significant and is related to Myhill’s (1988) auxiliary versus non-auxiliary meaning distinctions, present and imperfect *ir* and preterit *querer* may favor CC.

(3) Do social factors predict CC for *ir* and *querer*?

To the researcher’s knowledge, when the current investigation began, previous studies had not examined CC with the social variables of socioeconomic level, gender and age. However, Davies (1995) demonstrated that CC was more common in speech than in written data, a more conservative register. If placing clitics after the nonfinite verb is indeed a more conservative construction, CC may be less common in the upper class (Labov 1972b). Women may also use CC less frequently than men if the construction is stigmatized (Silva-Corvalán, 2001). Finally, younger and older speakers may be more likely to use the innovative forms, while middle-aged speakers may tend to use more conservative forms (Labov, 1972b).²

Although sociolinguistic variation is a possibility, the current study adopts the null hypothesis for each of the social variables. Research has shown that syntactic variables typically do not exhibit social variation (see for example, Silva-Corvalán, 2001). Thus, clitic placement is believed to be an indicator (Labov 1972a), or a linguistic variable that has little or no social value.

Method

Corpus

The data for this study come from the *Estudio sociolingüístico de Caracas* project, a corpus of 30-minute interviews from 160 speakers from Caracas, Venezuela carried out in 1987 and 1988 (Bentivoglio & Sedano, 1993). The semi-guided interview included topics such as festivals in Caracas, religious beliefs, education and past or present political situations. Socioeconomic level was determined by participants’ occupation, level of education and living conditions, their total and average family income and their parents’ occupation. 48 speakers; evenly distributed among three socioeconomic levels (lower, middle and upper), gender (female and male), and age (14-29 and 61 and older); were selected for the current investigation.

Data Coding

All tokens were coded for linguistic and social variables. A token was defined as an utterance in which a single clitic was produced with a finite *ir* or *querer* verb followed by a nonfinite verb. Examples appearing with two clitics were not included in the data. When two clitics were present in a sentence, they were never separated (i.e. one in the pre-finite verb position and one in the post-nonfinite verb position). Thus, coding the

token twice for each clitic would have skewed the results, and choosing to code for one of the clitics and not the other would have been arbitrary. The binary dependent variable was clitic placement: pre-finite verb position or post-nonfinite verb position.

Independent linguistic variables. There were five independent linguistic variables. Three of these variables characterized clitics: animacy (animate or inanimate), person (first, second or third), and number (singular or plural). Animacy of the clitic was determined by contextual information in the interviews. Clitics that refer to people, animals, and true reflexives were classified as animate. All other nouns and non-anaphoric clitics were coded as inanimate. Criteria established in Clements (2006) were employed to distinguish between anaphoric (true reflexive) and non-anaphoric clitics. Specifically, agentivity and intentionality distinguish true-reflexive from non-anaphoric clitics.

The other two linguistic variables deal with the finite verb. The categorization for the verb type variable followed Myhill (1988) and Davies (1995). *Ir* was classified as a motion verb and *querer* as a modal. *Ir* and *querer* were also coded for tense/aspect. In this corpus the present, preterit, imperfect, present subjunctive, and imperfect subjunctive were used in the linguistic context under investigation.

Independent social variables. The three social variables were socioeconomic level, gender, and age. Speakers had already been classified according to these variables by Bentivoglio and Sedano (1993). The categories for socioeconomic level were upper, middle, and lower class. Gender included female and male speakers. Age was also divided into two categories: 14-29 and 61 and over.

A sample of the coding procedure is provided using a sentence from the corpus: *Ella quiere dominarte todo el tiempo*, 'She wants to dominate you all the time'. The variant for the dependent variable is the post-nonfinite verb position since the clitic *te* follows the non-finite verb *dominar*. The clitic is animate, second person, and singular. The verb is the modal verb *querer* and is in the present tense. The social variables for this speaker are lower socioeconomic level, male, and the 14-29 year age group.

Data Analysis

All tokens were coded for the linguistic and social variables and entered into the Statistical Package for the Social Sciences 11.0 (SPSS), and a multivariate analysis was carried out. The first statistical tests included distributional cross-tabulations and chi-square correlations. These tests revealed which individual independent variables were significantly related to clitic placement, although not every significant factor necessarily predicted the dependent variable when analyzed together with other linguistic and social factors in the regression model. Next, a binary forward stepwise regression analysis was performed for each verb type. This logistic regression analysis examined all independent variables in relation to each other and then selected the variables that best predicted the dependent variable. Variables that did not improve the prediction of clitic placement were not included in the model. The alpha level was set at $p \leq 0.05$.

Results

The data included a total of 480 tokens. Table 1 reveals that CC overwhelmingly prevailed in the corpus: 89.2% of the clitics (428/480 tokens) occurred in the pre-finite verb position, whereas only 10.8% (52/480 tokens) were produced in the post-nonfinite

verb position. Table 1 also shows that tokens with *ir* (n=397) were much more frequent in the corpus than *querer* tokens (n=83) but that both verb types favored CC. The speakers produced 92.7% of the clitics (368/397 tokens) with *ir* and 72.3% (60/83 tokens) of the clitics with *querer* in the pre-finite verb position. Moreover, the chi-square test showed that the two verb types were significantly different from each other in terms of clitic placement ($X^2=29.592$, $df=1$, Cramer's $V=0.248$, Significance ≤ 0.000). In other words, CC occurred with *ir* significantly more often than with *querer*. Because of this statistical difference, the analyses of the remaining independent variables were carried out separately for *ir* and for *querer*.

Table 1
Distribution of Clitic Placement according to Verb Type

Verb type	Pre-finite		Post-nonfinite		Total tokens
	Tokens	%	Tokens	%	
<i>Ir</i>	368	92.7	29	7.3	397
<i>Querer</i>	60	72.3	23	27.7	83
Total	428	89.2	52	10.8	480

$X^2=29.592$, $df=1$, Cramer's $V=0.248$, Significance ≤ 0.000

Linguistic Factors

Ir. Three linguistic variables were significantly related to clitic placement for the motion verb *ir*. According to Table 2, the chi-square tests for clitic animacy ($X^2 = 6.343$, $df=1$, Cramer's $V=0.126$, Significance <0.05), clitic person ($X^2 = 6.226$, $df=2$, Cramer's $V=0.125$, Significance <0.05), and clitic number ($X^2 = 25.037$, $df=1$, Cramer's $V=0.251$, Significance <0.001) were significant. However, verb tense was not statistically significant.

Table 2
Chi-square Tests for Linguistic Factors for 'Ir'

Factor	x^2	df	Cramer's V
Animacy	6.343*	1	0.126*
Person	6.226*	2	0.125*
Number	25.037***	1	0.251***
Tense	3.999	4	0.1

* $p<0.05$. *** $p<0.001$.

Cross-tabulation distribution results provided more details about the three significant variables. This distribution revealed that both animate and inanimate clitics favored the pre-finite verb position but that the rates were significantly different. CC occurred more frequently with inanimate clitics (96.2% or 178/185 tokens) than with animate clitics (89.6% or 190/212 tokens).

Table 3

Distribution of Clitic Placement with 'Ir' according to Clitic Animacy

Animacy	Pre-finite		Post-nonfinite		Total	
	Tokens	%	Tokens	%	Tokens	%
Animate	190	89.6	22	10.4	212	100.0
Inanimate	178	96.2	7	3.8	185	100.0

In addition, first, second, and third person clitics favored the pre-finite verb position, but the frequency of CC varied significantly (Table 4). CC occurred most frequently with second person clitics (98.6% or 71/72 tokens), followed by first person (93.9% or 108/115 tokens), and finally third person (90.0% or 189/210 tokens).

Table 4

Distribution of Clitic Placement with 'Ir' according to Clitic Person

Person	Pre-finite		Post-nonfinite		Total	
	Tokens	%	Tokens	%	Tokens	%
First	108	93.9	7	6.1	115	100.0
Second	71	98.6	1	1.4	72	100.0
Third	189	90.0	21	10.0	210	100.0

Lastly, once again, CC occurred more frequently with both categories for clitic number, and the difference between singular and plural clitics was statistically significant. Clitic placement in the pre-finite verb position was highest with singular clitics (95.7% or 313/327 tokens). 78.6% (or 55/70 tokens) of plural clitics were also produced in the pre-finite verb position. Nevertheless, when interpreting the statistics, it is important to note that there were considerably more singular than plural clitics produced in the corpus (327 singular clitics and 70 plural clitics).

Table 5

Distribution of Clitic Placement with 'Ir' according to Clitic Number

Number	Pre-finite		Post-nonfinite		Total	
	Tokens	%	Tokens	%	Tokens	%
Singular	313	95.7	14	4.3	327	100.0
Plural	55	78.6	15	21.4	70	100.0

Querer. The results for the modal verb *querer* were very different from the *ir* data. No linguistic variables were significantly related to clitic placement. Table 6 illustrates that the chi-square tests for clitic animacy, person, and number and verb tense were not statistically significant.

Table 6
Chi-square Tests for Linguistic Factors for 'Querer'

Factor	x^2	<i>df</i>	Cramer's V
Animacy	1.047	1	0.112
Person	1.22	2	0.121
Number	2.196	1	0.163
Tense	9.14	4	0.332

Social Factors

Ir. Although three linguistic factors were significant for *ir*, no social variables were significant. Table 7 shows that the chi-square tests for socioeconomic level, gender, and age were not statistically significant.

Table 7
Chi-square Tests for Social Factors for 'Ir'

Factor	x^2	<i>df</i>	Cramer's V
Socioeconomic level	0.635	2	0.04
Gender	2.681	1	0.082
Age	0.365	1	0.03

Querer. In contrast to the results for *ir*, two social variables were significantly related to clitic placement for *querer* (Table 8). The chi-square tests for socioeconomic level ($X^2 = 9.346$, $df=2$, Cramer's $V=0.336$, Significance <0.05) and gender ($X^2 = 5.486$, $df=1$, Cramer's $V=0.019$, Significance <0.01) were statistically significant. The chi-square test for age was not significant.

Table 8
Chi-square Tests for Social Factors for 'Querer'

Factor	x^2	<i>df</i>	Cramer's V
Socioeconomic level	9.346*	2	0.336*
Gender	5.486**	1	0.019**
Age	2.617	1	0.178

* $p < 0.05$. ** $p < 0.01$.

Cross-tabulation distribution results illustrate the frequency of CC for the significant variables. First, each socioeconomic level favored CC, but the rate varied significantly across levels (Table 9). CC occurred most frequently for the speakers of the lower socioeconomic level (87.5% or 35/40 tokens), followed by the middle level (63.2% or 12/19 tokens), and lastly the upper socioeconomic group (54.2% or 13/24 tokens).

Table 9

Distribution of Clitic Placement with 'Querer' according to Socioeconomic Level

Socioeconomic level	Pre-finite		Post-nonfinite		Total	
	Tokens	%	Tokens	%	Tokens	%
Upper	13	54.2	11	45.8	24	100.0
Middle	12	63.2	7	36.8	19	100.0
Lower	35	87.5	5	12.5	40	100.0

Both males and females also produced more clitics in the pre-finite verb position, although there was a significant difference between the two genders. Table 10 reveals that CC was more frequent among females (82.6% or 38/46 tokens) than among males (59.5% or 22/37 tokens).

Table 10

Distribution of Clitic Placement with 'Querer' according to Gender

Gender	Pre-finite		Post-nonfinite		Total	
	Tokens	%	Tokens	%	Tokens	%
Female	38	82.6	8	17.4	46	100.0
Male	22	59.5	15	40.5	37	100.0

Logistic Regression

In order to determine if the linguistic and social factors worked together to predict CC, a binary logistic regression for each verb type was also performed. The results for *ir* show clitic animacy, person, and number worked together to predict CC (Table 11). In other words, when these three factors were included in the model, they better predicted CC than when these variables were by themselves. However, adding verb tense or any of the social variables did not improve the predictive power of the model. Table 12 also illustrates that this model predicted 92.7% of clitic placement for the motion verb *ir*.

As in the case of *ir*, the same factors that were significant in the chi-square tests for *querer* were also included in the regression model for *querer*. Table 11 reveals that socioeconomic level and gender worked together to better predict CC than when these features were by themselves. Neither the linguistic variables nor age improved the predictive power of the model. Table 12 shows that the regression model for *querer* had less predictive power than the model for *ir*. It explained 77.1% of CC for the modal verb.

Table 11
*Significant Predictors of Clitic
 Placement according to Verb Type*

Factor	<i>Ir</i>	<i>Quer</i>
Animacy	X***	
Person	X*	
Number	X***	
Tense		
Socioeconomic level		X*
Gender		X**
Age		

Note. X = variable included in the model.
 *p<0.05. **p<0.01. ***p<0.001.

Table 12
*Details of the Statistical Tests for the Logistic
 Regression Models according to Verb Type*

Statistical tests	<i>Ir</i>	<i>Quer</i>
χ^2	38.256	16.800
df	4	3
Significance	<0.001	<0.01
2 Log likelihood	169.337	81.173
Nagelkerke R ²	0.226	0.264
Percent predicted	92.7	77.1

Discussion

The results presented in the previous section will now be discussed in relation to the research questions and hypotheses.

- (1) Is clitic placement more frequent in the post-nonfinite verb position or in the pre-finite verb position?

Clitic placement was clearly more frequent in the pre-finite verb position. CC occurred at a rate of 89.2% in the corpus. This result was much higher than the results in Davies (1995) in which 56% of clitics in the spoken data of 10 countries from the 1970s were produced in the pre-finite verb position. It should be noted, however, that the results for the current study are limited to two verbs whereas the data in Davies' research included a larger number of verbs. Furthermore, it can be recalled that *ir* and *querer* were among the verbs in the Davies corpus that exhibited high rates of CC in comparison to other verbs investigated. Nevertheless, while the differences in percentages of CC between Davies (1995) and the present study may suggest linguistic change, discussion later on in this section will not support this hypothesis.

(2) Do linguistic factors predict CC for *ir* and *querer*?

The results for each linguistic variable are discussed individually.

Verb Type

The high use of CC seen in the entire corpus was maintained when the verbs were analyzed separately. Although clitics in the pre-finite verb position occurred significantly more often with *ir* (92.7%) than with *querer* (72.3%), both verbs favored CC. These rates of CC were higher than the rates found in the oral data from the 1970s in Davies (1995). Davies found that CC occurred in 86% of the clitics with *ir + a + infinitive* and in only 47% of the clitics with *querer*. Once again, although the increase in CC with each verb could be evidence for linguistic change, results for other variables contradict this notion.

Clitic Animacy

In light of previous research, the results of the current study for animacy were particularly interesting. First, this linguistic factor was one of the significant linguistic variables included in the regression model that predicted CC for *ir* but was not significantly related to clitic placement for *querer*. While Myhill (1988) and Davies (1995) analyzed all verbs together, they also concluded that animacy favored CC.

Second, in the present study, inanimate clitics were actually produced in the pre-finite verb position more frequently than animate clitics. Following Myhill (1988) and Davies (1995), this result would not be predicted. However, neither study mentioned coding a distinction between anaphoric and non-anaphoric *se* (Clements, 2006). Thus, it is possible that the tokens coded as non-anaphoric in the current investigation were coded as animate instead of inanimate in previous studies. This coding difference could have influenced the results. Future research could test this hypothesis.

Clitic Person

It was hypothesized that the results for person of the clitic would correspond to the animacy hierarchy. The results for *ir*, but not for *querer*, confirmed this prediction and also revealed that clitic person is a significant predictor of CC in the regression model. Second-person clitics exhibited the highest rate of CC followed by first-person clitics, and finally third-person clitics. However, if these results are in fact related to the animacy hierarchy, they are interesting because, although animacy was significant for *ir*, inanimate clitics were more often produced in the pre-finite verb position than animate clitics. Future research needs to be carried out on animacy and person of the clitic to determine if the animacy hierarchy truly influences clitic placement.

Clitic Number

No hypothesis was formulated for this variable. The results showed that clitic number was a significant predictor of CC for *ir* but not for *querer*. While both singular and plural clitics occurred more frequently in the pre-finite verb position, CC was significantly higher with singular clitics. It was also noted that the speakers used many more singular than plural clitics. This observation leads to a possible explanation for the significance of clitic number: clitics that are more frequent in the language are more likely to be produced in the pre-finite verb position. General frequency data appear to

only partially support this hypothesis. Although Alameda and Cuetos (1995) reported that the singular indirect object pronoun (*le*) was more frequent than the plural (*les*), the data for most direct object pronouns and definite articles were tabulated together (feminine singular and plural and masculine plural direct object pronouns and definite articles: *la*, *las*, and *los*, respectively). Thus, it is not possible to know whether direct object pronouns exhibit a similar frequency of use pattern as indirect object pronouns. More research needs to be conducted to confirm the possible relationship between the frequency of singular and plural clitics and CC.

Verb Tense

Verb tense was not a significant predictor of CC for either verb. These results also suggest that verb tense does not correspond to the meaning distinctions identified by Myhill (1988). Further research needs to be carried out to better understand Myhill's auxiliary and non-auxiliary meaning distinctions and their relation to CC.

(3) Do social variables predict CC for *ir* and *querer*?

The results for each social variable are discussed individually. It can be recalled that no significant results were expected for the social variables since syntactic dependent variables have not typically exhibited social variation in previous research (Silva-Corvalán, 2001).

Socioeconomic Level

Socioeconomic level was a significant predictor of CC for *querer* but not for *ir*. Although the three levels all favored CC, their use of CC varied significantly. As the socioeconomic level decreased, the frequency of CC increased. While the highest socioeconomic level only slightly favored clitics in the pre-finite verb position, the lowest group's use of CC was considerably higher than that of clitics in the post-nonfinite verb position. The middle socioeconomic group's use of CC fell between these two groups.

A possible explanation for this difference among socioeconomic levels could be linguistic change. Linguistic change often begins with lower socioeconomic groups and then spreads to higher levels (Labov, 1972b). However, considering the results for Caracas speech in Davies (1995), this explanation seems unlikely. An average for all verbs showed that clitics occurred in the pre-finite verb position in 53% of the data. These data were collected in the 1970s from speakers in higher socioeconomic groups. Similarly, the highest socioeconomic group in the current study produced clitics in the pre-finite verb position in 54.2% of the tokens with *querer*. Even though it is not known if the two groups are entirely comparable, the data from Davies (1995) and the current investigation seem to suggest that CC has not increased substantially over the decade for speakers in the higher socioeconomic group. Consequently, the difference among socioeconomic levels may be a stable difference. Nonetheless, these results must be interpreted with caution due to the low number of tokens obtained for *querer*.

Gender

Gender was also a significant predictor of CC in the regression model for *querer* but was not significant for *ir*. Both females and males favored CC, but females produced clitics in the pre-finite verb position significantly more often than men. Higher use of an

unstigmatized linguistic form by females can be a sign of change in progress (Silva-Corvalán 2001). However, as will be discussed shortly, linguistic change still appears to be an unlikely explanation. As in the case of socioeconomic level, a possible explanation for the results for gender may be that it is a stable social difference. These results also remain tentative since only a small number of *querer* tokens were produced by the speakers in the current study.

Age

This variable is not a significant predictor of CC for either verb. As mentioned earlier in the discussion, the data for the overall rate of CC, for the frequency of CC with *ir* and *querer*, for socioeconomic level with *querer* and for gender with *querer* could suggest that clitic placement is undergoing linguistic change. However, this hypothesis seems improbable since chi-square tests and logistic regression analyses indicated that age was not a significant predictor of CC. Furthermore, linguistic change often occurs in younger speakers first, but the data in the current study did not support this prediction (Labov, 1972b). These results lend support to Davies's (1995) hypothesis that CC has remained high over the years. They also supported conclusions that the differences among socioeconomic levels and between genders are stable social differences.

Conclusion

The current study clearly revealed that clitics in constructions with a finite *ir* or *querer* verb followed by a nonfinite verb occur predominately in the pre-finite verb position in spoken Modern Spanish of Caracas. Although there was a higher rate of pre-finite verb clitics with both verbs, CC was more common with *ir*. The sociolinguistic variables predicting CC also varied between the two verbs. Linguistic factors of clitic animacy, person, and number predicted CC with *ir*, whereas the social factors of socioeconomic level and gender predicted CC with *querer*. Thus, this research contributed to the field of Hispanic linguistics by extending the study of linguistic factors influencing CC to the analysis of a specific variety of Spanish and by initiating the investigation of previously neglected social factors.

Furthermore, the most significant contribution to the general field of linguistics concerns the results for the social variables. Although syntactic structures do not typically vary according to social factors (Silva-Corvalán, 2001), the present study has suggested that CC is more common among women and the lower socioeconomic level. Although these results are based on a small data sample of verbal constructions with *querer*, it can be argued that research on syntactic variation should not ignore social variables.

Several areas of future research emerge from the current investigation. Foremost, because the number of tokens for *querer* is small, more extensive data collection and analyses on this verb are needed to confirm the results of the present study. In fact, coding the entire Caracas corpus would not only strengthen or refute the conclusions for *querer* but also for *ir*. Extending this research to other motion and modal verbs is also valuable. Since linguistic factors were significant for *ir* and social factors were significant for *querer*, it would be interesting to learn if these results are characteristic of the distinction between motion and modal verbs.

In addition, to the researcher's knowledge, detailed comparisons of CC use across varieties of Spanish have not been carried out. Although Davies (1995) suggested CC

does not exhibit regional differences, stronger claims could be made if more comprehensive analyses of different language varieties were conducted. The results from this study could be compared to an investigation analyzing CC in another variety if that study were to examine comparable oral data and similar linguistic and social factors. Finally, although the results of this study do not seem to provide evidence for linguistic change, conducting research on more recent Caracas speech could confirm or refute this conclusion. Over a decade has past since the corpus examined in this study was collected, thus carrying out sociolinguistic interviews during the current decade that parallel those in 1987 would permit a reliable comparison.

Notes

1. Neither the present investigation nor Myhill (1988) addresses the role the nonfinite verb may play in clitic climbing. This issue is left to future research.
2. Middle-aged speakers were not included in the present investigation. An examination of CC for these speakers is left to future studies.

References

- Alamada, J.R. & Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. 2 Vol. Oviedo: Universidad de Oviedo Servicio de Publicaciones.
- Bentivoglio, P. & Sedano, M. (1993). Investigación sociolingüística: sus métodos aplicados a una experiencia venezolana. *Boletín de Lingüística*, 8: 3-35.
- Clements, J.C. (2006). Spanish non-anaphoric *se*. In J. Clancy Clements & Jiyoung Yoon (eds.) *Functional approaches to Spanish syntax: Lexical semantics, discourse, and transitivity* (pp. 236-264). London: Palgrave.
- Davies, M. (1995). Analyzing syntactic variation with computer-based corpora: The case of modern Spanish clitic climbing. *Hispania*, 78.2: 370-380.
- Davies, M. (1997). The evolution of Spanish clitic climbing: A corpus-based approach. *Studia Neophilologica*, 69.2: 251-263.
- Labov, W. (1972a). *Language in the inner city*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1972b). On the mechanism of linguistic change. In *Sociolinguistic patterns*, 160-182. Philadelphia: University of Pennsylvania Press.
- Myhill, J. (1988). Variation in Spanish clitic climbing. *Georgetown University Round Table on Language and Linguistics*, 227-250.
- Napoli, D. J. (1981). Semantic interpretation vs. lexical governance: Clitic climbing in Italian. *Language*, 57.841-87.
- Silva-Corvalán, C. (2001). *Sociolingüística y pragmática del español*. Washington, D.C.: Georgetown University Press.
- Torres Cacoullos, R. (1999). Construction frequency and reductive change: Diachronic and register variation in Spanish clitic climbing. *Language Variation and Change*, 11.2: 143-170.