

Extending the sociolinguistic variable to the syntactic level: The case of *para* + infinitive/subjunctive in Venezuelan Spanish*

Scott Lamanna
Indiana University

Abstract

Previous studies have analyzed variable use of the infinitive versus the subjunctive in clauses following the preposition *para* ('for, in order to') in Spanish in terms of discourse, syntactic, semantic, and pragmatic factors. However, no study has analyzed this phenomenon according to extralinguistic factors, nor determined which linguistic factors are the most significant predictors of this variation when considered together. The current study presents a variationist analysis of the choice of infinitive versus subjunctive after *para*. Social class, gender, and age are examined to address the role of extralinguistic variables. Six linguistic variables related to the *para* and/or matrix clause are also included (type and person of the *para* clause subject, coreferentiality between *para* and matrix clause subjects, transitivity of the *para* clause verb, and discourse theme and informational status of the *para* clause). Subjects are speakers of Venezuelan Spanish comprising the 1987 corpus *Estudio Sociolingüístico de Caracas*. Results show that none of the extralinguistic variables are significant. Coreferentiality, person and type of subject, and discourse theme, however, are significant. The relationship between these findings and prescriptive Spanish grammar is discussed, along with the study's relevance to the debate concerning the extension of the sociolinguistic variable to the syntactic level.

0. Introduction

In Spanish, both the infinitive and the subjunctive are possible in clauses following the preposition *para* ('for, in order to'). An example of each of these two verb forms is given in (1) and (2) below, respectively:

(1) *Tengo_i que regresar_i a la casa para estudiar_i.*
have-1SG-PRES that return-INF to the house for study-INF
'I have to return home (in order) **to study.**'

(2) *Tengo_i que regresar_i a la casa para que mi esposa_j estudie_j.*
have-1SG-PRES that return-INF to the house for that my wife
study-3SG-PRES-SUBJ
'I have to return home so that my wife **can study.**'

According to the rules of prescriptive Spanish grammar (Bello 1921, Butt & Benjamin 2000, Dominicis & Reynolds 2007), the infinitive is used after *para* when the

* A previous version of this paper was presented at the Fifth Annual Graduate Student Conference on Luso-Brazilian and Hispanic Literature, Linguistics and Culture, Indiana University – Bloomington, March 2008.

subjects of the matrix clause and the subordinate clause are coreferential. This is illustrated in (1) above, where the subject of both the matrix verb *tengo* ‘I have’ and the infinitive *estudiar* ‘to study’ is *yo* ‘I’. When the matrix and subordinate clause do not have the same subject, the subjunctive is the prescribed usage. This can be seen in (2), where the subject of the subordinate clause verb *estudie* ‘he/she studies – subj.’ is *mi esposa* ‘my wife’, which is not coreferential with the matrix clause subject *yo*.

In several dialects of colloquial Spanish, including Caribbean Spanish (Suñer 1986) and Venezuelan Spanish (Bentvoglio 1987a), there is variable use of the infinitive and the subjunctive in clauses following the preposition *para* in the same syntactic context. Examples (3) and (4) below both exhibit contexts in which the subjects of the matrix and subordinate clauses are not coreferential, and can be translated identically into English. In (3) the infinitive is used, and in (4) the subjunctive is employed:

(3) *Para yo_i estudiar_i en ese colegio, mi papá_j tiene_j que conseguir_j otro trabajo.*
 for I study-INF in that school my dad have-3SG-PRES that get-INF another job

(4) *Para que yo_i estudie_i en ese colegio, mi papá_j tiene_j que conseguir_j otro trabajo.*
 for that I study-1SG-PRES-SUBJ in that school my dad have-3SG-PRES that get-INF another job
 ‘(In order) for me **to study** in that school, my dad has to get another job.’

The type of variation seen in these examples has been analyzed in terms of discourse factors (Bentvoglio 1987a, Serrano 2004) as well as syntactic, semantic, and pragmatic factors (DeMello 1995, Kirschner 1992, Morales 1989, Murillo Medrano 2003). The role of these linguistic factors is discussed in detail in section 1.

While the influence of linguistic factors on morphosyntactic variation is not controversial, the role of extralinguistic (social) factors in such variation has been the subject of some debate in the sociolinguistics literature. The debate has centered on the possibility of extending the sociolinguistic variable from the phonological level, where both linguistic and extralinguistic factors have been shown to play a role in variation, to the syntactic level (Labov 1978, Lavandera 1978, Romaine 1984).

The current study contributes to this ongoing debate by presenting a variationist analysis of the choice of the infinitive versus the subjunctive after *para*. Section 1 includes a discussion of previous literature on the use of one and/or both of these forms after *para* and ends with the research questions that motivate the current investigation. Section 2 discusses the method employed for the current study, including the corpus used and how the data were coded and analyzed. The results are presented and discussed in section 3. Finally, section 4 includes conclusions and suggestions for future research.

1. Previous research

A number of studies have addressed the use of the infinitive and/or the subjunctive after *para* from several different angles. For example, an attempt has been

made to explain the variation between the two structures in terms of discourse factors. Bentivoglio (1987a) examined all clauses introduced by *para* in a corpus of Venezuelan Spanish, whether the verb was in the infinitive or the subjunctive, and whether a lexical subject was present or absent. She coded for referentiality by using the following three categories: (1) same reference (SR): the subjects of the matrix and subordinate *para* clause are identical; (2) different reference 1 (DR1): the subject of the *para* clause is coreferential with some other part of the matrix clause (often a clitic); and (3) different reference 2 (DR2): the subject of the *para* clause is not coreferential with any constituent of the matrix clause. She also coded for whether the subject of the *para* clause was personal or impersonal. She found that in cases of SR, the overwhelmingly preferred structure (used 92% of the time) was *para* \emptyset inf (*para* + infinitive without an overt lexical subject). Among the remaining SR cases (n=9), four were instances of *para* + lexical subject + infinitive (PLSI) and five contained *para que* + subjunctive. The fact that both the infinitive and the subjunctive were used in cases of SR led the researcher to conclude that some factor in addition to referentiality plays a role in the variation between these two structures.

Bentivoglio (1987a) found that *para* \emptyset inf was also used in 32% of cases of DR1 and 66% of cases of DR2, where prescriptive grammar indicates that only the subjunctive should be used. She explained this fact by appealing to the difference between personal and impersonal subjects. A combined analysis of DR1 and DR2 cases revealed that 69% of impersonal subjects were used with *para* \emptyset inf and 73% of personal subjects appeared with either PLSI or *para que* + (lexical subject) + subjunctive. Bentivoglio (1987a) claimed that *para* \emptyset inf could be used with impersonal subjects because in these cases, the identity of the subject is irrelevant. Bentivoglio (1987a) next addresses cases in which PLSI or *para que* + (lexical subject) + subjunctive were found instead of the expected *para* \emptyset inf (8% of SR subjects and 31% of DR impersonal subjects). She found that in these cases, “the use of the marked forms... is motivated by the presence of at least one contextual element threatening the right identification of the *para* clause subject” (Bentivoglio 1987a:9). Thus, both PLSI and *para que* + (lexical subject) + subjunctive were found to have a disambiguating function. Due to the limited number of tokens of PLSI found, Bentivoglio (1987a) does not attempt to account for how speakers choose between PLSI and *para que* + (lexical subject) + subjunctive when the need to disambiguate arises.

Serrano (2004) examined infinitive/subjunctive variation after *para* in a corpus of spoken Peninsular Spanish in terms of three discourse factors: management of theme, management of information, and management of persons in the discourse. She examined four types of themes: argumentative, narrative, expository and descriptive. Under management of information, she took into account the distinction between old and new information. Finally, she divided persons in the discourse into three categories: appellative (1st person), elocutive (2nd person referents) and impersonal (3rd person or impersonal referents).

For management of theme, Serrano (2004) found that the infinitive was used more frequently with expository and descriptive themes, whereas the subjunctive was used more often with argumentative and narrative themes. When examining old versus new information, she found that the infinitive was used more often with old information and the subjunctive with new information. The most noteworthy finding in the area of persons

in the discourse was that impersonal referents were used 64% of the time with the infinitive, but only 5% of the time with the subjunctive.

Morales (1989) examined the subjunctive/infinitive distinction after *para* by applying grammaticality judgments to invented sentences. She argued, as did Bentivoglio (1987a), that presence or absence of coreferentiality is not sufficient to account for this variation and that semantic and pragmatic factors must also be taken into account. For example, she argued that a speaker may choose to use the infinitive to emphasize the semantic traits of the verb as a lexical item. Her conclusion is that three factors need to be included when analyzing constructions with *para*: (1) the type of syntactic structure formed (e.g. whether the *para* clause modifies the entire sentence or just the VP), (2) referentiality between the arguments of the subordinate and matrix clause, and (3) the type of temporality indicated by the subordinate verb (e.g. whether it indicates future time or is more temporally neutral).

Kirschner (1992) administered a questionnaire to a group of Spanish-English bilingual university students in the United States, and found that they preferred PLSI to *para que* + (lexical subject) + subjunctive after impersonal expressions such as *es importante* 'it is important'. Although the analysis is carried out from a Government and Binding perspective, Kirschner (1992) appears to be subscribing to functionalist syntax when he claims that the shift from subjunctive to infinitive is "due primarily to semantic or contextual judgments which modify the bilingual's rules for mood choice" (Kirschner 1992:91). Two specific examples of these judgments are "evaluation of redundancy and the identification of an explicit subject (of the infinitive)" (Kirschner 1992:104).

Murillo Medrano (2003) analyzed tokens from a corpus of cultivated Costa Rican speech to determine which syntactic, semantic, and pragmatic factors influence the expression of an overt subject in *para* + infinitive constructions. He did not find many cases of such marking, and concluded that this structure was not very common in cultivated speech. However, the fact that any such structures were used at all was posited to be due to the influence of colloquial speech and he suggested future research on such speech in Costa Rica to determine the frequency of PLSI structures.

One final study relevant to the current investigation is DeMello (1995). He examined PLSI in cultivated speech from twelve Spanish-speaking cities, and discovered that its use was significant only in Caracas and San Juan. He found that 82% of subjects in this construction were pronominal and that coreferentiality of matrix and subordinate clause subjects was less common than the occurrence of non-coreferential subjects. DeMello (1995) also compared cultivated and popular Mexican speech and found evidence to support the claim that PLSI is more common in popular speech.

All of the above studies have investigated the role of internal linguistic factors in the use of the infinitive and/or the subjunctive after *para*. To date, however, no study has examined the variation between these two verb forms in terms of the extralinguistic (social) variables of gender, age, and social class. In addition, no attempt has been made to determine which linguistic variables are the most significant predictors of this variation when considered together in a single study design. The current study addresses this gap in the literature by utilizing the relatively large 1987 corpus *Estudio Sociolingüístico de Caracas* 'Sociolinguistic Study of Caracas' (Bentivoglio & Sedano 1993). This corpus allows us to carry out a variationist analysis of the choice of infinitive versus subjunctive in *para* clauses.

1.1 Research Questions

The present study attempts to answer the following research questions:

- (1) Does the use of the infinitive versus the subjunctive following *para* vary significantly according to gender, age, and social class?
- (2) Which linguistic variables are the most significant predictors of this variation when considered together in a single study design?

2. Method

2.1 Corpus

Subjects for the current study are 40 speakers of Venezuelan Spanish from the 1987 corpus *Estudio Sociolingüístico de Caracas* ‘Sociolinguistic Study of Caracas’, which consists of 30 minute semi-guided interviews. The subjects are equally distributed between males and females, four age groups, and five social classes (see description of factor groups on pages 9-10 for details). The Caracas corpus includes four speakers for each of the forty possible combinations of these variables (for example, male, 14-29 years old, upper class). Each speaker is identified within his/her own group by the letters a, b, c, and d. For the current study, one speaker out of four was randomly selected from each group by drawing one of the four letters from an envelope.

2.2 Data Coding and Analysis

All clauses introduced by *para* that were produced by the 40 speakers were included in the analysis. This is the same approach that was followed by Bentivoglio (1987a). In order to answer the research questions for the study, a statistical analysis of the data was performed using the program *Goldvarb 2.0*. Eleven factor groups were included. The first group represents the dependent variable. Group two coded for the individual speakers included in the study. Groups three through five represent the extralinguistic (social) variables included in the investigation and groups six through eleven the linguistic variables. The following is a brief description of each of these factor groups, with examples included from the corpus where applicable.

Factor group 1: This factor group takes into account the type of verb that appears in the *para* clause. Factors: i) infinitive, ii) subjunctive.

- (5) infinitive

Tuve tamaño para jugar basketball
have-1SG-PAST size for play-INF basketball
‘I was the right size **to play** basketball’ (a3mc)¹

¹ The code in parentheses identifies the specific speaker from the Caracas corpus from whose interview the token was extracted. The first letter refers to the speaker’s age (a = 14-29, b = 30-45, c = 46-60, d = 61 or older), the number represents social class (1 = upper class, 2 = upper middle class, 3 = middle class, 4 = lower middle class, 5 = lower class), the following letter indicates gender (m = male, f = female), and the last letter identifies each speaker within his or her own group.

(6) subjunctive

hay que orar pa² que la gente se cure
there is that pray-INF for that the people REFL cure-3SG-PRES-SUBJ
'One has to pray in order for people **to be cured**' (a3mc)

Factor group 2: This factor group coded for the individual speakers (n=40) from whose interviews the tokens were extracted, in order to determine whether or not individual variation plays a role in the use of the infinitive and the subjunctive after *para*. While there is nothing in the literature to suggest such a role, its possibility should not be excluded a priori.

Factor group 3: Gender of the speaker. Factors: i) male, ii) female.

Factor group 4: Age of the speaker. Factors: i) 14-29 years old, ii) 30-45 years old, iii) 46-60 years old, iv) 61 years old and older.

Factor group 5: Social class. Factors: i) upper class, ii) upper middle class, iii) middle class, iv) lower middle class, v) lower class.

Factor group 6: This factor group addresses the type of subject in the *para* clause. Factors: i) null subject, ii) noun, iii) pronoun.

(7) null subject

cogimos el terreno pa' Ø jugar beisbol
take-1PL-PAST the piece of land for play-INF baseball
'We took the piece of land to play baseball (on it)' (d5mc)

(8) noun

Para que las personas sientan tú tienes que darles sentimientos
for that the people feel-3SG-PRES-SUBJ you have-2SG-PRES that give-INF.them feelings
'In order for **people** to feel you have to give them feelings' (a2fa)

(9) pronoun

para yo salir de noche es un rollo
for I go out-INF of night be-INF an affair
'It's a pain for **me** to go out at night' (a2fa)

Factor group 7: This factor group indicates the person of the subject of the *para* clause. Factors: i) first person, ii) second person, iii) third person, iv) impersonal v) unknown vi) multiple subjects (more than one verb after *para*, each with a subject of a different person). The first four of these are illustrated below.

(10) first person

tiene que ser que me hagan algo
have-3SG-PRES that be-INF that me do-3PL-PRES-SUBJ something
para yo molestarte ¿ves?
for I bother-INF.REFL see-2SG-PRES

² The preposition *para* is sometimes reduced to *pa'* in spoken Spanish.

‘They have to do something to me in order for **me** to get annoyed, you see?’ (c4fb)

- (11) second person

tú eres caraqueño y para que (tú) te
 you be-2SG-PRES from Caracas and for that (you) REFL
adaptes a un medio de esos, te
 adapt-2SG-PRES-SUBJ to an environment of those you
cuesta bastante
 cost-3SG-PRES quite a lot
 ‘You are from Caracas and for **you** to adapt to an environment like that, it’s really difficult’ (b3mb)

- (12) third person

Y mi niño pelea pa' ver ese canal
 and my boy-DIM fight-3SG-PRES for see-INF that channel
 ‘And **my little boy** fights to watch that channel’ (b5fb)

- (13) impersonal

se le levanta a uno el ánimo para que uno no
 REFL him raise-3SG-PRES to one the spirit for that one not
se ... se deprima todo
 REFL REFL depress-3SG-PRES-SUBJ all
 ‘One’s spirits are raised so that **one** doesn’t get completely depressed’ (a4fa)

Factor group 8: This factor groups addresses the presence or absence of coreferentiality between the subjects of the *para* clause and the matrix clause. Factors are the same as those used in Bentivoglio (1987a): i) same reference (SR) - the subjects of the matrix and subordinate *para* clause are identical, ii) different reference 1 (DR1) - the subject of the *para* clause is coreferential with some other part of the matrix clause, iii) different reference 2 (DR2): the subject of the *para* clause is not coreferential with any constituent of the matrix clause, iv) unknown v) multiple subjects (more than one verb after *para*, each with a subject of a different coreferentiality). The first three of these factors are illustrated below.

- (14) same reference (SR)

No podíamos abrir la boca para hablar
 no poder-1PL-PAST open-INF the mouth for speak-INF
 ‘We couldn’t open our mouth to speak’ (c5fb)

- (15) different reference 1 (DR1)

por eso te vuelve Dios al mundo, para que
 for that you return-3SG-PRES God to.the world for that
vuelvas a surgir
 return-2SG-PRES-SUBJ to emerge-INF

‘That’s why God sends you back to the world, in order for you to emerge again’ (a5ma)

- (16) different reference 2 (DR2)

yo_i tenía_i que ir_i en diciembre a recoger_i naranjas
I have-1SG-PAST that go-INF in December to harvest-INF oranges
¿ves? para que pueda_j venir_j el niño
see-2SG-PRES for that can-3SG-PRES-SUBJ come-INF the child
Jesús_j
Jesus
‘I had to go in December to harvest oranges, you see? so that baby Jesus can come’ (a3mc)

Factor group 9: This factor group addresses the transitivity of the verb in the *para* clause. Factors: i) transitive, ii) intransitive, iii) copula (e.g. *ser/estar* ‘to be’) iv) multiple verbs (more than one verb after *para*, each with a different value for transitivity). The first three of these are illustrated below.

- (17) transitive

*ellos diseñan ... los gráficos para **vender** los productos*
they design the graphics for sell-INF the products
‘They design the graphics in order to **sell** the products’ (a1fb)

- (18) intransitive

*yo estaba contentísima, esperando ese día para **ir** para esa*
I be-1SG-PAST very content wait-GER that day for go-INF for that
fiesta
party
‘I was very content waiting for that day to **go** to that party’ (a5fb)

- (19) copula

*nosotros siempre estamos pendientes para **estar** unidos*
we always be-1SG-PRES attentive for be-INF united
‘We’re always paying attention to make sure we **are** united’ (c5fb)

Factor group 10: This factor group takes into account the theme of the discourse string within which the *para* clause is located. Factors are based on those that appear in Serrano (2004): i) argumentative, ii) narrative, iii) expository/descriptive. Although Serrano (2004) coded expository and descriptive themes as separate categories, they were combined for the current analysis due to the difficulty of distinguishing between them in many cases.

- (20) argumentative

hay que llevar todo tranquilito pa’ poder vivir bien
there is that carry-INF everything calm-DIM for can-INF live-INF well
‘One has to bear everything calmly in order to be able to live well’ (d5fc)

(21) narrative

nos fuimos a comer una cosa para no entrar a clase
REF go-1PL-PAST to eat-INF a thing for not enter-INF to class
'We went to eat something in order not to go to class' (a1mb)

(22) expository/descriptive

yo soy malísima para que me entrevisten
I be-1SG-PRES very bad for that me interview-3PL-PRES-SUBJ
'I'm really bad at being interviewed' (b1fd)

Factor group 11: This factor group indicates the type of information provided in the *para* clause. Factors: i) old (known) information, ii) new information.

(23) old (known) information

En Valencia para verlos ... hay que ir a los
in Valencia for see-INF.them there is that go-INF to the
barríos para verlos ¿no?
neighborhoods for see-INF.them no
'In Valencia to see them ... one has to go the neighborhoods to see them, right?' (a3mc)

(24) new information

Caracas ya no es un lugar para vivir
Caracas already not be-3SG-PRES a place for live-INF
'Caracas is no longer a place where you can live' (c3mc)

3. Results and Discussion

3.1 Research Question 1

The first research question was the following: Does the use of the infinitive versus the subjunctive following *para* vary significantly according to gender, age, and social class? The *Goldvarb* analysis indicated that factor groups 3 (gender), 4 (age), and 5 (social class) were not significant. Thus, the answer to Research Question 1 is no: the use of the infinitive versus the subjunctive following *para* does not vary significantly according to the extralinguistic (social) variables examined in this study. This agrees with the findings of a number of studies cited in Silva Corvalán (2001:133), which have shown very little or no effect of external social variables on syntactic variation (including variable subject expression and variable word order) in Spanish.

3.2 Research Question 2

The second research question was stated as follows: Which linguistic variables are the most significant predictors of this variation when considered together in a single study design? The factor groups in Table 1 on page 10 were selected by *Goldvarb* as

significant predictors of the use of the subjunctive versus the infinitive in *para* clauses in the Caracas corpus. The factor groups appear in the table in the order of their relative significance, with the most significant appearing first. Four of the five factor groups represent linguistic variables, and the other factor group is the individual speaker. It will be noted that factor groups 9 (transitivity) and 11 (information) do not appear in the table; this is because they were not selected by *Goldvarb* as being significant. It will also be noted that although the total number of tokens included in the analysis was 450 (110 cases of the subjunctive and 340 cases of the infinitive), the totals for factor groups 2 (individual speaker), 7 (person of subject), and 8 (coreferentiality) do not add up to 450. This is because each of these factor groups included one or more knockouts (factors for which there was no variation in the dependent variable), and knockouts must be eliminated before performing a *Goldvarb* analysis. In all cases of knockouts, categorical use of the infinitive was observed.

The subjunctive was selected as the application value for this analysis, due to the fact that it appears less frequently in this corpus than the infinitive. Thus, probability weights above 0.50 indicate a factor that favors the subjunctive, while those below 0.50 indicate a factor that disfavors it.

Table 1: Factor groups selected by *Goldvarb* as significant predictors of subjunctive versus infinitive after *para*

Factor Group	Factors	Cases of subjunctive / total cases	%	Probability weight
8: Coreferentiality	Different reference 1	53/70	75	0.981
	Different reference 2	50/135	37	0.892
	Unknown	1/9	11	0.336
	Same reference	6/235	2	0.086
7: Person of subject	Second person	5/9	55	0.995
	Third person	63/147	42	0.815
	Impersonal	23/152	15	0.319
	First person	19/138	13	0.251
6: Type of subject	Noun	22/24	91	0.918
	Pronoun	22/47	46	0.677
	Null subject	66/379	17	0.439
2: Individual speaker	Too many to list (36 speakers)	110/406	27	Weights for individual speakers range from 0.951 to 0.060
10: Discourse theme	Narrative	67/224	29	0.617
	Descriptive/expository	18/113	15	0.542
	Argumentative	25/113	22	0.248
Input 0.070 Significance 0.021 Cases of subjunctive: 110 Cases of infinitive: 340				

Factor Group 8: Coreferentiality. Coreferentiality was selected by *Goldvarb* as the most significant factor group influencing the choice between the infinitive and the subjunctive after *para*. The factors favoring subjunctive use in this group were the two types of non-coreferentiality between the matrix and *para* clause subjects, DR1 and DR2, with probability weights of 0.981 and 0.892, respectively. In cases of coreferentiality between the subjects of the matrix and *para* clause (SR), subjunctive use was strongly disfavored, with a probability weight of 0.086, and thus the infinitive was strongly favored. The factor *unknown* represents one case where coreferentiality could not be determined by the researcher due to the ambiguity of the discourse context, and had a factor weight of 0.336. The factor multiple subjects was identified as a knockout and was eliminated before running the *Goldvarb* analysis.

The findings of the current study for SR agree with Bentivoglio's (1987a) results, in which *para* Ø inf (*para* + infinitive without an overt lexical subject) was observed in 92% of SR cases. The result obtained here, i.e. DR1 favors subjunctive use more than DR2, also agrees with Bentivoglio's (1987a) findings, in which *para* Ø infinitive was used in 66% of cases of DR2 but in only 32% of cases of DR1, about half as often.

The above results for coreferentiality are not surprising since they are consistent with the rules of prescriptive Spanish grammar mentioned in the introduction (Bello 1921, Butt & Benjamin 2000, Dominicis & Reynolds 2007), which state that the infinitive should be used when the matrix and subordinate clause subjects are coreferential, while the subjunctive should be used when they are non-coreferential. However, the empirical evidence is more convincing in cases of SR than DR (especially DR2, where the subjunctive is used only 37% of the time in the corpus examined here).

Factor Group 7: Person of subject. This factor group was selected by *Goldvarb* as the second most important variable determining infinitive versus subjunctive use. Second person subjects most highly favor the use of the subjunctive within this factor group, with a probability weight of 0.995. Third person subjects also favor the subjunctive, with a probability weight of 0.815. The categories of impersonal and first person subjects both disfavor the use of the subjunctive, with probability weights of 0.319 and 0.251, respectively. The factors unknown and multiple subjects were knockouts and were eliminated from the analysis.

No explanation is immediately apparent for the extremely high probability weight assigned by *Goldvarb* to second person subjects, especially in light of the fact that only 55% of these subjects (5/9) appeared with a subjunctive verb form. The scarcity of tokens of second person subjects means that caution should be exercised in interpreting these results. The favoring effect of second person subjects on the use of the subjunctive should be considered tentative until a corpus containing a greater number of such tokens is analyzed.

The fact that third person subjects favor the use of the subjunctive in *para* clauses makes sense, in light of the fact that Bentivoglio (1987a) found that use of lexical subjects with the subjunctive can have a disambiguating function. Disambiguation is generally more of an issue in the third person than in other persons due to the greater number of possible referents (he, she, it, etc.) that can be associated with third person verb forms.

The disfavoring effect of first person subjects on subjunctive use makes sense as well in light of the fact that speakers were often asked questions to get them talking about

themselves during the Caracas interviews. Since use of the infinitive after *para* is associated with coreferentiality between the matrix and *para* clause subjects, and a *para* clause with a first person subject would be more likely to be coreferential with the matrix clause than other types of subjects, we would expect subjunctive use to be disfavored in this context. The disfavoring effect of impersonal subjects can be explained in a similar fashion, since the Caracas corpus participants would also frequently make comments about people in general, and thus an impersonal subject in the *para* clause would also be likely to be coreferential and appear with the infinitive. The current study's finding that impersonal subjects were used with the subjunctive only 15% of the time is similar to Serrano's (2004) finding that impersonal referents were only used 5% of the time with the subjunctive³.

Factor Group 6: Type of subject. The type of subject of the *para* clause was the factor group selected by *Goldvarb* as the third most significant. The highest probability weight within this group was 0.918 for noun subjects, followed by 0.677 for pronouns, both of which favor the subjunctive. The probability weight for null subjects was 0.439, indicating that the subjunctive is slightly disfavored, and the infinitive slightly favored. This finding agrees with the rules of prescriptive Spanish grammar previously discussed (Bello 1921, Butt & Benjamin 2000, Dominicis & Reynolds 2007). The infinitive is the prescribed usage after *para* when the subjects of the matrix clause and the subordinate clause are coreferential, and in these cases, the subject tends not to be expressed (Bentivoglio 1987a), as in example (1), repeated as (25) below:

- (25) *Tengo_i que regresar_i a la casa para estudiar_i.*
 have-1SG-PRES that return-INF to the house for study-INF
 'I have to return home (in order) to study.'

It should be noted, however, that the use of the infinitive with null subjects was not categorical in this corpus, since the subjunctive was used in 17% of cases.

The fact that noun subjects and pronouns favor the use of the subjunctive agrees with prescriptive Spanish grammar as well, since the subject of a *para* clause is often overtly expressed when it is not coreferential with the matrix clause, and the subjunctive is the prescribed usage in these cases, as in example (2), repeated as (26) below:

- (26) *Tengo_i que regresar_i a la casa para que mi esposa_j estudie_j.*
 have-1SG-PRES that return-INF to the house for that my wife
estudie_j.
 study-3SG-PRES-SUBJ
 'I have to return home so that my wife can study.'

Factor Group 2: Individual speaker. The factor group selected by *Goldvarb* as the fourth most significant predictor of choice of verb form (infinitive versus subjunctive) after *para* was the individual speaker. Four speakers used the infinitive categorically after *para* and were thus eliminated from the analysis as knockouts. Probability weights for the

³ A frequency count of impersonal, first, second, and third person subjects in the Caracas corpus as a whole would be useful in order to confirm or invalidate the interpretation of the results for person of subject offered here. Such a frequency count, however, is beyond the scope of the current investigation.

use of the subjunctive by the remaining 36 speakers ranged from a high of 0.951 to a low of 0.060. This wide disparity between subjects should probably not be assigned too much importance, since the average number of *para* clauses produced per speaker (11.28) was relatively small. Perhaps with a larger sample of linguistic production from each participant, the range of probability weights would not have been so great, and this factor group may not have turned out to be significant.

Factor Group 10: Discourse theme. Discourse theme was the least significant of the five factor groups selected by *Goldvarb*. Narrative discourse strings favor the use of the subjunctive, with a probability weight of 0.617, and descriptive/expository discourse slightly favors this form, with a weight of 0.542. Argumentative discourse disfavors the use of the subjunctive, and thus favors the infinitive, with a probability weight of 0.248.

The results for this factor group partially agree with those obtained by Serrano (2004), who found that both narrative and argumentative themes favored the subjunctive. The discrepancy in the findings for argumentative discourse may be due to the fact that Serrano's (2004) corpus included oral data from a variety of sources in addition to interviews. Future studies should examine other types of corpora (including written data) in order to allow for stronger conclusions regarding the influence of this linguistic factor to be made.

The results of the current study can be used to address the ongoing theoretical debate concerning the difficulty of extending the sociolinguistic variable to the syntactic level (Labov 1978, Lavandera 1978, Romaine 1984). Such an extension is problematic for at least four reasons, summarized by Silva-Corvalán (2001:129-130) in her discussion of the issue. First, syntactic variables tend to have fewer variant forms than phonological variables. Second, the variants of a syntactic variable tend to occur less frequently in spontaneous speech than those of a phonological variable, making it more difficult to obtain a sufficient number of tokens for a quantitative analysis. Third, the context of occurrence of a syntactic variable tends to be harder to identify and define than in the case of a phonological variable. Finally, and most importantly, while it is usually clear that variants of a phonological variable represent two or more ways of saying the same thing (Labov 1972:271, cited in Silva-Corvalán 2001:130), such is not necessarily the case with syntactic variables, as different semantic, discursive, and/or pragmatic meanings may be associated with each variant. Silva-Corvalán (2001:130) gives the example of variable word order in Spanish, in which the order of the components of a sentence often correlates with their informational status, with new information appearing at the end and old information appearing earlier. Thus a given sentence component does not vary freely in terms of its position (final vs. non-final) in the same discourse context, since there is a different semantic/discursive meaning applied in the case of each position.

The results of the current study suggest that the variation between the infinitive and the subjunctive after *para* is less problematic than some other syntactic variables in relation to the issues discussed in the preceding paragraph. As indicated by the results reported in Table 1, a sufficient number of tokens of both variants of this variable were available in the Caracas corpus to permit a quantitative statistical analysis to be performed. In addition, the context of occurrence of the variable was easily defined as all clauses introduced by *para*. Finally, contexts of categorical use of one of the forms (in all cases the infinitive) were few and were eliminated before running the *Goldvarb* analysis,

meaning that variation between the two forms was observed with all of the remaining factors composing the various factor groups.

The next question to be asked is whether, in the absence of categorical usage, any of the factor groups provide compelling evidence of different semantic, discursive, and/or pragmatic meanings being strongly associated with the subjunctive and the infinitive after *para*. Clearly this is not the case for discourse theme, where the probability weights for the subjunctive range from 0.617 for narrative discourse to 0.248 for argumentative discourse, and where the percentages of subjunctive use range from 29% for narrative discourse to 15% for descriptive/expository discourse. In the case of person of subject, although the probability weight for second person subjects (0.995) indicates that they strongly favor the subjunctive, the small number of tokens involved (n=9) coupled with the fact that only 55% of these represent cases of the subjunctive indicate that this result should be considered tentative pending the examination of a larger number of tokens. While third person subjects favor the subjunctive with a probability weight of 0.815, this still represents only 42% subjunctive use. The other two factors, impersonal and first person subjects, while clearly disfavoring the subjunctive (with probability weights of 0.319 and 0.251, respectively) show enough variation (15-13% subjunctive use) to warrant hesitation in suggesting that they are strongly associated with the infinitive.

For the factor group type of subject, there is clear association between noun subjects and the subjunctive. This factor has a probability weight of 0.918 and shows 91% subjunctive use. However, the association between null subjects and the infinitive is not quite as strong, as the factor has a probability weight of 0.439 and shows 17% subjunctive use. In addition, pronoun subjects are not clearly associated with the infinitive or the subjunctive, as indicated by the corresponding probability weight (0.677) and percentage of subjunctive use (46%).

The factor group coreferentiality provides the strongest evidence of possible association of different meanings with the infinitive and the subjunctive after *para*. The infinitive is used 98% of the time in cases of SR, and the probability weight indicates that this factor strongly disfavors the subjunctive (0.086). The probability weight of 0.981 for DR1 indicates a correspondingly strong favoring of the subjunctive, which is used 75% of the time. With DR2, although the probability weight (0.892) indicates that it strongly favors the subjunctive as well, there was a much more even distribution in the use of the two forms, with the subjunctive occurring only 37% of the time.

The results for coreferentiality suggest the possibility that different discursive meanings may be associated with the infinitive and the subjunctive. The infinitive seems to be associated with coreferential matrix and *para* clause subjects, and thus with discourse continuity, and the subjunctive, to a lesser extent, with non-coreferential subjects, and thus with lack of continuity. However, there is still too much variation between these two structures observed in this factor group, especially in the case of DR2, to allow for a simple declaration of equivalence between verb form and discourse meaning. Thus it would appear that overall, this study provides evidence that the sociolinguistic variable can be felicitously extended to the syntactic level in the case of *para* +infinitive/subjunctive. This of course does not imply that use of the two verb forms will necessarily vary according to social factors, and indeed the results reported in section 3.1 indicate that it does not.

4. Conclusions and Future Research

The current study has presented a variationist analysis of the choice of the infinitive versus the subjunctive after *para*. First, the study addresses the extent to which social stratification correlates with this case of synchronic variation in syntactic structure by examining three extralinguistic (social) variables: social class, gender, and age. The findings indicate that social stratification does not correlate with the variation between the subjunctive and the infinitive after *para* in Caracas Spanish, thus providing a negative answer to the first research question. As far as the second research question is concerned, this study is the first to determine which of the linguistic variables influencing the choice of infinitive or subjunctive after *para* are significant when considered together in a single study design. Coreferentiality was determined by *Goldvarb* to be the most significant variable, followed by person of the *para* clause subject, the type of subject of the *para* clause, and discourse theme.

This study has contributed to the debate on the difficulty of extending the sociolinguistic variable to the syntactic level (Labov 1978, Lavandera 1978, Romaine 1984) by providing evidence that the variation between the infinitive and the subjunctive after *para* is perhaps less problematic than many other syntactic variables as a locus for such an extension. Future research, however, is needed to come to more definitive conclusions regarding this issue, as the factor group coreferentiality did provide evidence of a strong (though not categorical) association of each verb form with a different discursive meaning. One way to avoid this complication would be to carry out an analysis where coreferentiality is held constant – for example, by examining only cases of DR, where in fact most of the variation between the two verb forms was observed.

Additional suggestions for future studies include examining a larger number of tokens and including additional linguistic variables such as number of the subject (singular versus plural), which has been shown to contribute to another case of morphosyntactic variation in Spanish, that between null and overt subject pronouns (Bentivoglio 1987b, Cameron 1992, Flores-Ferrán 2002). Future investigations should also include data from countries other than Venezuela and from other types of corpora (including written corpora) in order to shed additional light on the use of the infinitive and the subjunctive after *para* in Spanish.

References

- Bello, Andrés. 1921. *Gramática de la lengua castellana destinada al uso de los americanos*. Madrid: Imp. de A. Marzo.
- Bentivoglio, Paola. 1987a. Clauses introduced by the preposition *para* in spoken Spanish: an analysis from a discourse viewpoint. *Studies in Romance languages*, ed. by Carol Neidle and Rafael Núñez Cedeño, 1-14. Dordrecht: Foris.
- Bentivoglio, Paola. 1987b. *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.
- Bentivoglio, Paola and Mercedes Sedano. 1993. Investigación sociolingüística: sus métodos aplicados a una experiencia venezolana. *Boletín de Lingüística* 8.3-35.
- Butt, John and Carmen Benjamin. 2000. *A new reference grammar of modern Spanish*. 3rd edn. New York: John Wiley and Sons.
- Cameron, Richard. 1992. Pronominal and null subject variation in Spanish: constraints, dialects, and functional compensation. Unpublished PhD dissertation, University of Pennsylvania.
- De Mello, George. 1995. Preposición + Sujeto + Infinitivo: "Para yo hacerlo". *Hispania* 78.825-836.
- Dominicis, María Canteli and John Reynolds. 2007. *Repase y escriba: curso avanzado de gramática y composición*. 5th edn. New York: John Wiley and Sons, Inc.
- Flores-Ferrán, Nydia. 2002. *Subject personal pronouns in Spanish narratives of Puerto Ricans in New York City: a sociolinguistic perspective*. München: Lincom Europa.
- Kirschner, Carl. 1992. The Spanish subjunctive and the Spanish-English bilingual: a semantically-motivated functional shift. *Hispanic Linguistics* 5.89-108.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. 1978. Where does the sociolinguistic variable stop? a response to Beatriz Lavandera. *Working Papers in Sociolinguistics*, vol. 44. Austin: Southwest Educational Development Laboratory.
- Lavandera, Beatriz. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7.171-183.
- Morales, Amparo. 1989. Some considerations on the subjunctive-infinitive alternation in constructions with *para*. *Nueva revista de filología hispánica* 37.27-42.
- Murillo Medrano, Jorge. 2003. La marcación del sujeto en las oraciones finales introducidas por *para*+infinitivo en el habla culta costarricense. *Revista de filología y lingüística de la Universidad de Costa Rica* 29.229-240.
- Romaine, Suzanne. 1984. On the problem of syntactic variation and pragmatic meaning in sociolinguistic theory. *Folia Lingüística* XVIII.409-437.
- Serrano, María Jose. 2004. Between grammar and discourse: completives with *para* + infinitive/subjunctive in a socio-communicative context. *Estudios de sociolingüística* 5.129-150.
- Silva-Corvalán, Carmen. 2001. *Sociolingüística y pragmática del español*. Washington, D.C.: Georgetown University Press.
- Suñer, Margarita. 1986. Lexical subjects of infinitives in Caribbean Spanish. *Studies in Romance linguistics*, ed. by Osvaldo Jaeggli and Carmen Silva-Corvalán, 189-203. Dordrecht: Foris.