Null and overt subjects in a variable system: The case of Dominican Spanish

Cristina Martínez-Sanz

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements

For the PhD degree in Spanish

Thesis Supervisor: Juana M. Liceras

Department of Modern Languages and Literatures

Faculty of Arts

University of Ottawa

© Cristina Martínez-Sanz, Ottawa, Canada, 2011

Abstract

This dissertation investigates subject expression patterns in Dominican Spanish (DS). In this variety, the null subject constructions associated with Non-Caribbean Spanish co-exist with the widespread use of overt subjects, which are found in specific constructions that are either rare or unattested in other Spanish varieties. Interestingly, these structures co-exist in the Dominican grammar with the null subject constructions associated with Non-Caribbean Spanish.

While subject expression has been studied in a number of Spanish dialects within the generative and the variationist paradigms, monolingual Dominican Spanish, to the best of my knowledge, has not been investigated in previous variationist work. This study covers this gap by examining a large corpus of spontaneous speech (N=6005) gathered in the capital city of Santo Domingo and a rural area in the northwestern Cibao region. Furthermore, in line with the cohesive approach to syntactic variation developed in recent work (Adger and Smith 2005), theoretical implications are drawn from quantitative results.

The results obtained in this study show that null and overt subject patterns in DS are regulated by the same constraints that have been found relevant in previous variationist work, i.e. discourse-related factor groups and *Person* (Otheguy, Zentella and Livert, 2007). These results depart from previous work in that evidence for language change in progress has been found in subject position patterns, rather than in null and overt subject distribution. When this phenomenon is examined, urban, young, high-middle class and female speakers arise as the social groups leading grammatical restructuring.

Quantitative and qualitative evidence is taken into account for testing previous syntactic-theoretical proposals on DS. Taking the cartographic approach to syntactic structure (Rizzi 1997) as a point of departure, it will proposed that multiple specifier positions are available within the TP and CP fields to host strong and weak subjects. This proposal, in turn, makes it possible to account for the Null Subject Parameter profile displayed by synchronic DS without resorting to competing grammars in the minds of the speakers.

Table of Contents.

Abstract ii

Index of Tables. vii

Acknowledgments xiv

Chapter One. Introduction. Dialectal variation and diachronic change: Null and overt subject patterns in Dominican Spanish. 1

1.1. Language variation and change in the generative tradition. 8

1.2. Language variation and change in the variationist paradigm. 14

1.3. The present study: Dominican Spanish in the context of language variation, language change and formal grammars. 18

1.3.1. The nature and limits of dialectal variation. 20

1.3.2. Language change and the Dominican variable pattern of subject expression. 40

1.3.3. Formal grammars and variable systems. 42

1.4. Organization. 44

Chapter Two. Theoretical Background: The Null Subject Parameter. 46

2.1. Introduction: Null and overt subjects in linguistic theory. 46

2.2. The Government and Binding Model. 49

2.2.1. The Null Subject Parameter 49

2.2.2. On the definition of rich agreement. 54

2.2.3. Non-standard null and overt subject patterns: Null Expletive and Partial Null Subject languages. 59

2.2.4. Dialectal variation in Western Romance. 63

2.3. Language variation in the Minimalist Program: Features and microparameters. 66

2.3.1. Back to basics: Sound and meaning. 67

2.3.2. Parameters in the Minimalist Program: on the locus of language variation and change. 71

2.3.3. The Minimalist view on the Null Subject Parameter(s). 73

2.3.4. Non-NSL and NSL Spanish. 78

2.3.5. On the feature composition of Spanish pronouns. 80

2.3.5.1. Null Subject Spanish. 80

2.3.6. Dialectal variation and pronominal feature assembly: Dominican Spanish. 86

2.3.6.1. Toribio (2000, 2001). 86

2.3.6.2. Martínez Sanz (2007). 88

2.3.6.3. Camacho (2008). 91

2.4. On the structural position of subjects in Spanish. 95

2.4.1. Overt preverbal subjects belong in Spec, CP. 96

2.4.2. Overt preverbal subjects belong in Spec, CP *and* Spec, IP. 99

2.4.3. On 'free' inversion: Subject positions and information structure. 104

2.4.3.1. Wide focus in Non-Caribbean Spanish. 107
2.4.3.2. Narrow focus in Non-Caribbean Spanish. 109
2.4.3.3. Focus in Dominican Spanish. 114
2.4.4. Subject licensing in interrogative constructions. 117
2.4.4.1. Wh-interrogatives in Non-Caribbean Spanish. 120
2.4.4.2. Wh-interrogatives in Caribbean Spanish. 124
2.4.5. Subject licensing in non-finite structures. 130
2.4.5.1. Personal infinitives in Non-Caribbean Spanish. 132
2.4.5.2. Infinitives in Caribbean Spanish. 136
2.5. Summary. 139
Chapter Three. Null and overt subject patterns in diachronic change: The case of Brazilian Portuguese. 143
3.1. Introduction: Approaches to the subject licensing grammar of Brazilian Portuguese. 143
3.2. Defining the variable in Brazilian Portuguese: Kato and Taralho (1984). 144
3.3. The diachronic view on null and overt subjects in Brazilian Portuguese. 146
3.4. The synchronic view on Brazilian Portuguese. 154
3.4.1. Root null subjects. 158
3.4.2. Embedded null subjects. 160
3.4.3. BP as a Topic-prominent language. 163
3.4.4. Brazilian Portuguese as a partial Null Subject Language. 166
3.5. Summary 173
Chapter Four. Variationist views on Spanish subject expression. 174
4.1. The quantitative paradigm and Spanish subject expression. 174
4.1.1. Discourse-related variables. 180
4.1.2. Morphosyntactic variables. 188
4.1.3. Lexico- semantic variables. 190
4.2. Subject expression in Caribbean Spanish. 194
4.2.1. Subject expression: Overall rates of expression and factor group rankings. 194
4.2.2. On the site of cross-dialectal differences in subject expression: *Person* constraint rankings. 200
4.2.2.1. *Person* constraint rankings and *Verb Ambiguity*. 210
4.2.2.2. Morphological ambiguity and contextual ambiguity: On the role of *Co-reference* in Caribbean Spanish subject expression. 214
4.2.2.3. Structural constraints on Spanish subject expression. 223
4.2.2.3.1. Word order patterns in interrogative constructions: pronouns versus DP subjects. 223
4.2.2.3.2. Clause Type. 228
4.2.3. Language external fac tors and grammatical restructuring in Caribbean dialects. 232
4.2.3.1. *Gender* and *Age*. 234
4.2.3.2. Interpreting distributions in apparent time: on the nature of grammatical restructuring. 241

4.3. Summary. Bridging paradigms: where variationist methodology and theoretical concerns meet. 244

Chapter Five. Methodology. Probing into a restructured subject expression grammar. 247

5.1. Introduction. Restructured grammars and the intersection between language variation and language change. 247

5.2. Research questions and hypotheses. 250

5.2.1. Internal factors. 251

5.2.2. External factors. 258

5.2.3. Grammatical restructuring and formal grammars. 262

5.3. The speaker sample. 265

5.4. Data collection: fieldwork protocol and materials. 270

5.5. The envelope of variation. Transcribing and coding procedure. 273

5.6. Summary 284

Chapter Six. Results: The configuration of a restructured subject expression grammar. Null and overt subject distribution and subject position patterns in Dominican Spanish. 285

6.1. Introduction. Null and overt subject distribution and subject position patterns. 285

6.2. Null and overt subject patterns: Overt pronominal and lexical subjects versus null subjects. 286

6.2.1. Pronominal and lexical subjects versus null subjects. Variables, variable ranking and constraint ranking: Discussion. 297

6.2.2. On the role of other language-external variables: *Social class* and *Education*. 300

6.2.3. On the relative role of contextual and morphological ambiguity. 311

6.2.4. Interrogative and non-interrogative clauses: Differences. 324

6.3. Null and overt subject patterns: Overt pronominal subjects versus null subjects. 329

6.3.1. Overt pronominal subjects versus null subjects: Significant factor groups, factor group ranking and constraint ranking. 333

6.3.2. Language-external variables: *Social class* and *Education*. 338

6.3.3. On the relative role of contextual and morphological ambiguity. 345

6.3.4. Interrogative and non-interrogative clauses: Distribution of pronominal and null subjects. 356

6.4. Null and overt subject patterns: Lexical subjects versus null subjects. 360

6.4.1. Variables, variable ranking and constraint ranking 361

6.4.2. Null and overt subject patterns in Dominican Spanish: Summary. 364

6.5. Subject position patterns: On the distribution of preverbal and postverbal subjects. 365

6.6. Null and overt subject distribution and subject position patterns in Dominican Spanish. Summary. 374

Chapter Seven. Theoretical implications: Formal and variationist contributions to a theory on subject expression in Dominican Spanish. 377

7.1. Introduction. 377

7.2. Null and overt subjects: Patterns of distribution. 379

7.2.1. Rates and factor group rankings: on the role of language-internal factor groups in Dominican subject expression. 379
7.2.2. Structural constraints on Dominican subject expression. 393
7.2.3. Language-external factor groups and language change. 399
7.3. Grammatical restructuring and formal grammars. 404
7.3.1. On the quantitative expression of parameters. 404
7.3.2. Qualitative evidence and the description of grammatical systems. 410
7.3.2.1. Overt expletive constructions. 411
7.3.2.2. Personal infinitives. 414
7.3.2.3. Clefts and Focus: Affirmative and interrogative clauses. 416
7.4. On the formal definition of the Dominican subject expression grammar. 419
7.4.1. Case assignment in Dominican Spanish. 422
7.4.2. On the nature of null and overt pronominal subjects in Dominican Spanish. 427
7.4.3. On syntactic change. 435
7.4.4. Personal infinitives in diachrony. 438
7.4.5. Null subjects in Mainland and Insular Scandinavian. 440
7.4.6. On the pragmatic functions of overt pronominalization in Caribbean and Non-Caribbean Spanish. 443
7.5. (Micro) parametric variation in Romance: On the NSP profile of a restructured subject grammar. 455
7.6. Summary. 457
Chapter Eight. The cohesive approach to cross-dialectal variation and language change: contributions and suggestions for further research. 458
8.1. Language variation and language change: The Dominican grammar of subject expression. 458
8.2. Dialectal variation and formal grammars: the issue of 'optionality'. 461
Appendix A. Demographic Information. All Speakers. 464
References 466

Index of Tables.

Table 3.1. Reduction of inflectional endings in BP (1845-1992)..... 148
Table 3.2. Rates of overt subjects in BP (1845-1992)..... 148
Table 3.3. Overt subject rates in synchronic BP 149
Table 3.4. Strategies for indetermination in synchronic EP and BP. 150
Table 4.1. Spanish subject expression: Factor groups and constraints..... 193
Table 4.2 Dialectal variation in overall rates of subject expression. 195
Table 4.3. Newcomers' pronoun rate by national origin (Otheguy et al. 2007:785)..... 197
Table 4.4. Variable hierarchies in overt pronoun expression: Newcomers (Otheguy, Zentella and Livert 2007:789) 198
Table 4.5. Two independent variable rule analyses of the contribution of factors selected as significant to the probability of expressed subjects in NM and Colombian data. (Travis 2007:115). 199
Table 4.6. Overt subject rates for specific pronouns: San Juan, Puerto Rico, Madrid, Spain, and Buenos Aires, Argentina (Morales 1997:155)..... 201
Table 4.7. Rates of Subject Pronoun Expression in San Juan, Puerto Rico and Madrid, Spain (Cameron 1992: 233)..... 202
Table 4.8. Overt pronominal subjects for person and number: San Juan and NYC (Flores-Ferrán 2004:59)..... 204
Table 4.9. Constraint rankings in Person: San Juan versus Madrid (Cameron 1992:239)..... 204
Table 4.10. Constraint ranking within the Person variable (Otheguy, Zentella and Livert 2007:790). 205
Table 4.11. Frequencies of types of overt non-specific (=arbitrary) *ellos* 209
(Lapidus and Otheguy 2005:170) 209
Table 4.12. Rates of /s/ deletion in Puerto Rican Spanish: Specific versus nonspecific *tú* (Cameron 1992:292)..... 210
Table 4.13. Overall rates of /s/ deletion: males versus females (Cameron 1992:298) 211
Table 4.14. Overall rates of /s/ deletion and chi-square values: occupational class crosstabulated with males versus females (Cameron 1992:302). 211
Table 4.15. Pronominal expression crosstabulated with +/- inflection: specific *tú* in male versus female speakers (Cameron 1992:315). 212
Table 4.16. Rates of pronominal expression associated with *Verb Class* in San Juan and Madrid (Cameron 1996:317). 213
Table 4.17. Rates of pronoun expression for Co-reference (Cameron 1992:177) 215

Table 4.18. Rates of subject expression by plurality (Cameron 1992:241). 215

Table 4.19. *Co-reference* crosstabulated with singular and plural subjects (Cameron 1992: 178). 216

Table 4.20. Set to elements saliency: San Juan versus Madrid (Cameron 1992:249)..... 217

Table 4.21. Variable rule program weights for set-to-elements saliency: San Juan versus Madrid (Cameron 1992:250). 217

Table 4.22. *Co-reference* in singular subjects only: San Juan versus Madrid. (Cameron 1992: 244). 218

Table 4.23. Variable Rule Program weights for *Co-reference* (singular subjects only) San Juan versus Madrid (Cameron 1996: 244). 218

Table 4.24. *Verb class* crosstabulated with *(Non-)Co-reference*: San Juan versus Madrid (Cameron 1992: 268). 219

Table 4.25. Variable rule program weights associated with *Verb Ambiguity* and *Co-reference* in San Juan and Madrid (Singular subjects) (Cameron 1992:270). 219

Table 4.26. Comparison of Puerto Rican and Mexican-American Spanish rates of subject-verb inversion (Cameron 1996:322). 222

Table 4.27. The effect of argumentality on overt subjects (Lizardi 1993:49)..... 224

Table 4.28. The interaction of *Argumentality* and *Subject Type* (Lizardi 1993:52). 224

Table 4.29. Distribution of preverbal, postverbal and null variants among individual pronouns (Lizardi 1993:65) 226

Table 4.30. Distribution of preverbal, postverbal and null variants: Before and after the exclusion of *yo*, *tú* and *uno* (Lizardi 1993:66). 227

Table 4.31. The effect of argumentality on pronouns (b) (Lizardi 1993:67) 227

Table 4.32. *Clause Type* crosstabulated with subject position in San Juan, Madrid and Buenos Aires (Morales 1997:156-158) 229

Table 4.33. Preverbal subjects in Relative Clauses: San Juan, Madrid and Buenos Aires (Morales 1997:161)..... 229

Table 4.34. Differences in *Clause Type*: Nominals and demonstratives (Lizardi 1993:69)..... 231

Table 4.35. Overall gender differentiation among Puerto Rican speakers (Lizardi 1993:78)..... 234

Table 4.36. Overall gender differentiation among Puerto Rican speakers, preverbal occurrences: Pronouns versus nominals (Lizardi 1993:79)..... 234

Table 4.37. Overall age differentiation among Puerto Rican speakers (Lizardi 1993:80)..... 235

Table 4.38. Overall age differentiation among Puerto Rican speakers: Percentages for pronouns and nominals (Lizardi 1993:81) 235

Table 4.39 Distribution of Personal Pronouns in Puerto Rican Spanish (Ávila-Jiménez 1996:97) 236

Table 4.40 Pronominal distribution by age group in Puerto Rico (Ávila-Jiménez 1996:98) 237

Table 4.41. Pronominal distribution by age group (Morales 1986) 237

Table 4.42. Uniform grouping of *Gender* according to the cut-off point of 50 years of age (Lizardi 1993:83) 238

Table 4.43. Apparent time redistribution for women (Lizardi 1993:84) 238

Table 4.44. Fine age distribution by *Gender* (Lizardi 1993:85) 239

Table 4.45 Pronominal distribution based on educational level (Ávila-Jiménez 1996:113) 239

Table 4.46. Pronominal distribution based on educational level and age (Lizardi 1996:115) 240

Table 4.47. Pronominal distribution by occupation (Ávila-Jiménez 1996:117) 240

Table 5.1. The present study: Speakers. (N=34; 6006 tokens) 266

Table 5.2. The present study: selected factor groups and associated constraints. 283

Table 6.1. Dominican Spanish. Overall percentages of null and overt subject use: persons of the paradigm. 287

Table 6.2. Dominican Spanish. *Distance from previous mention*: Overall rates of null and overt subjects (N=6005). 289

Table 6.3. Dominican Spanish. *TMA* : Overall rates of null and overt subjects. 290

(N=6005) 290

Table 6.4. Dominican Spanish. Null subjects versus pronominal and lexical subjects: Patterns of variation. 296

Table 6.5. Dominican Spanish. Asymmetries in language external factor groups: Social class and Education (N=3606). 300

Table 6.6. Dominican Spanish. Social class and Education in Santo Domingo. Patterns of variation for null versus overt (pronominal and lexical) subjects (N=3606). 301

Table 6.7. Dominican Spanish. Working class in Santo Domingo and El Cibao: Patterns of variation for null versus overt (pronominal and lexical) subjects (N=3851). 303

Table 6.8. Dominican Spanish. *Verb* Patterns of variability: Urban and Rural speakers and Social Class. 304

Table 6.9. Dominican Spanish. Patterns of variability, discourse-related factor groups: Urban and Rural speakers and Social Class. 305

Table 6.10. Dominican Spanish. TMA Patterns of variability: Urban and Rural speakers and Social Class. 305

Table 6.11. Dominican Spanish. Cross-tabulation of Person and Clause Type. All speakers (N=6005)..... 307

Table 6.12. Dominican Spanish. Cross-tabulation of *Person* and *Clause Type*. Only urban speakers (N=3606). 308

Table 6.13. Dominican Spanish. Cross-tabulation of *Person* and *Clause Type*. Working class speakers in Santo Domingo and El Cibao (N=3781). 309

Table 6.14. Dominican Spanish. Chi-squares of the differences in Clause Type. Urban speakers versus all speakers and working class urban and rural speakers versus urban speakers. 310

Table 6.15. Dominican Spanish. Null versus pronominal and lexical subjects: TMA patterns of variation. 312

Table 6.16. Dominican Spanish. Null versus pronominal and lexical subjects: patterns of variation in discourse-related factor groups. 314

Table 6.17. Dominican Spanish. Null versus pronominal and lexical subjects: TMA and discourse-related factors patterns of variation. 315

Table 6.18. Dominican Spanish. Null versus pronominal and lexical subjects: First and second person singular *TMA* patterns of variation. 317

Table 6.19. Dominican Spanish. Null versus pronominal and lexical subjects: First and third person singular *TMA* patterns of variation. 318

Table 6.20. Dominican Spanish. Null versus pronominal and lexical subjects: patterns of variation in discourse-related factor groups, first and second person singular. 319

Table 6.21. Dominican Spanish. Pronominal and lexical subjects versus null subjects: patterns of variation in discourse-related factor groups, first and third person singular. 320

Table 6.22. Dominican Spanish. Null versus pronominal and lexical subjects: patterns of variation in TMA and discourse-related factor groups, first and second person singular. 321

Table 6.23. Dominican Spanish. Null versus pronominal and lexical subjects: patterns of variation in *TMA* and discourse-related factor groups, first and third person singular. 322

Table 6.24. Dominican Spanish. TMA and discourse-related factor groups: variable ranking in first and second, first and third person singular. 323

Table 6.25. Dominican Spanish. Cross-tabulation of *Interrogative* and *Distance from previous mention*. (N=6005)..... 325

Table 6.26. Dominican Spanish. Cross-tabulation of *Interrogative* and *Co-reference* (N=6005). 325

Table 6.27. Dominican Spanish. Cross-tabulation of *Interrogative* and *Discourse Function*. (N=6005)..... 326

Table 6.28. Dominican Spanish. Cross-tabulation of *Interrogative* and *TMA of the previous verb*. (N=6005) 326

Table 6.29. Dominican Spanish. Cross-tabulation of *Interrogative* and *Person*. (N=6005) 327

Table 6.30. Dominican Spanish. Cross-tabulation of *Interrogative* and *Clause Type*. 327

Table 6.31. Dominican Spanish. Cross-tabulation of *Interrogative* and *Origin*. (N=6005) 328

Table 6.32. Dominican Spanish. Cross-tabulation of *Interrogative* and *Gender*. (N=6005) 328

Table 6.33. Dominican Spanish. Cross-tabulation of *Interrogative* and *Age*. (N=6005) ... 328

Table 6.34. Dominican Spanish. Cross-tabulation of *Interrogative* and *Class*. (N=6005) 329

Table 6.35. Dominican Spanish. Comparison with Person overall rates San Juan, Puerto Rico (Cameron 1992: 233). 330

Table 6.36. Dominican Spanish. Pronominal subjects versus null subjects: Patterns of variation. 332

Table 6.37. Dominican Spanish. Factor group ranking in lexical and pronominal subjects (N=6005) versus null subjects and pronominal subjects versus null subjects (N=4830). 333

Table 6.38. Comparison of factor group rankings: The present study, Otheguy et al. (2007) and Orozco and Guy (2008) 335

Table 6.39. Dominican Spanish. Constraint rankings in null versus lexical and pronominal subjects (N=6005) and pronominal subjects versus null subjects (N=4830). 337

Table 6.40. Urban Dominican Spanish. Cross-tabulation of *Person* and *Clause Type*: Pronominal subjects versus null subjects (N=4830). 338

Table 6.41. Dominican Spanish. *Social class* and *Education* in Santo Domingo. Patterns of variation for overt pronominal subjects versus null subjects. 340

Table 6.42. Dominican Spanish. Working class and Santo Domingo and El Cibao. Patterns of variation for overt pronominal subjects versus null subjects. 342

Table 6.43. Dominican Spanish. *Verb* Patterns of variability: Urban and Rural speakers and Social Class. Pronominal versus null subjects 343

Table 6.44. Dominican Spanish. Patterns of variability, discourse-related factor groups: Urban and Rural speakers and Social Class: Pronominal versus null subjects. 344

Table 6.45. Dominican Spanish. TMA Patterns of variability: Urban and Rural speakers and Social Class. Pronominal versus null subjects. 345

Table 6.46. Dominican Spanish. Pronominal subjects versus null subjects, singular paradigm: *TMA* patterns of variation. 345

Table 6.47. Dominican Spanish. Pronominal subjects versus null subjects, singular paradigm: Discourse- related factors patterns of variation. 346

Table 6.48. Dominican Spanish. Pronominal subjects versus null subjects, singular paradigm: *TMA* and discourse-related factor groups. 347

Table 6.49. Dominican Spanish.Comparison of variable rankings in TMA and discourse-related factors. Pronominal versus null subjects, all singular paradigm. 348

Table 6.50. Dominican Spanish. Pronominal versus null subjects: First and second person singular *TMA* patterns of variation. 349

Table 6.51. Dominican Spanish. Pronominal versus null subjects: First and third person singular *TMA* patterns of variation. 350

Table 6.52. Dominican Spanish. Pronominal subjects versus null subjects: Patterns of variation in discourse-related factor groups, first and second person singular. 351

Table 6.53. Dominican Spanish. Pronominal subjects versus null subjects: Patterns of variation in discourse-related factor groups, first and third person singular. 352

Table 6.54. Dominican Spanish. Pronominal subjects versus null subjects: Patterns of variation in TMA and discourse-related factor groups, first and second person singular. 353

Table 6.55. Dominican Spanish. Pronominal versus null subjects: Patterns of variation in TMA and discourse-related factor groups, first and third person singular. 354

Table 6.56. Dominican Spanish.Comparison of variable rankings in TMA and discourse-related factors. Pronominal versus null subjects, all singular paradigm. 355

Table 6.57. Dominican Spanish. Cross-tabulation of *Interrogative* and *Distance from previous mention*, pronominal subjects (N=4830). 356

Table 6.58. Dominican Spanish. Cross-tabulation of *Interrogative* and *Co-reference*, pronominal subjects (N=4830). 357

Table 6.59. Dominican Spanish. Cross-tabulation of *Interrogative* and *Discourse function*, pronominal subjects (N=4830). 357

Table 6.60. Dominican Spanish. Cross-tabulation of *Interrogative* and *TMA of the previous verb*, pronominal subjects (N=4830). 358

Table 6.61. Dominican Spanish. Cross-tabulation of *Interrogative* and *Person*, pronominal subjects (N=4830). 358

Table 6.62. Dominican Spanish. Crosstabulation of *Interrogative* and *Clause Type*, pronominal subjects (N=4830). 359

Table 6.63. Dominican Spanish. Cross-tabulation of *Interrogative* and *Origin*, pronominal subjects (N=4830). 359

Table 6.64. Dominican Spanish. Cross-tabulation of *Interrogative* and *Gender*, pronominal subjects (N=4830). 359

Table 6.65. Dominican Spanish. Cross-tabulation of *Interrogative* and *Age*, pronominal subjects (N=4830). 360

Table 6.66. Dominican Spanish. Cross-tabulation of *Interrogative* and *Social class*, pronominal subject (N=4830). 360

Table 6.67. Dominican Spanish. Lexical subjects versus null subjects: Patterns of variation. 361

Table 6.68. Dominican Spanish. Cross-tabulation of Verb and Person, lexical subjects (N=3371). 364

Table 6.69. Dominican Spanish. Distribution of preverbal (versus postverbal) subjects (N=3344). 367

Table 6.70. Dominican Spanish. Subject position patterns. Cross-tabulation of Subject Form and Verb (N=3344). 369

Table 6.71. Dominican Spanish. Subject position patterns. Cross-tabulation of Subject Form and Clause Type (N=3344). 370

Table 6.72. Dominican Spanish. Distribution of preverbal and postverbal subjects in Santo Domingo (N=2188). 371

Table 6.73. Dominican Spanish. Distribution of preverbal and postverbal subjects in urban and rural working classes (N=2308). 372

Table 6.74. Dominican Spanish. Subject position patterns. Cross-tabulation of *Subject form* and *Interrogative* (N=2308). 373

Table 6.75. Dominican Spanish. Subject position patterns. Comparison of factor group rankings for all speakers, urban speakers and working class speakers. 374

Table 6.76.Dominican Spanish. Null and overt subject patterns and subject position patterns. Linguistic factor groups. 375

Table 6.77.Dominican Spanish. Null and overt subject patterns and subject position patterns. Language-external factor groups 375

Table 7.1. Dominican Spanish. Gender and Age null/overt and preverbal/postverbal subject patterns. 385

Table 7.2. Dominican Spanish. Effects of TMA and discourse related factors on pronominal and lexical subjects; pronominal subjects; lexical subjects. 389

Table 7.3. Dominican Spanish. Rates of overt subject use across clause types. First, second, and third person singular subjects. 392

Table 7.4. Focus and word order in Spanish. 446

Table 7.5. Null subjects in Dominican Spanish and Brazilian Portuguese. 457

Acknowledgments

In reading dissertation acknowlegments overt the years I have found that 'it takes a village' is the classic conclusion one arrives at upon finishing this type of work. My dissertation is no different in this respect, except perhaps that stating that 'it took a village' would be an understatement. It has taken at least two 'villages', in the form of the two communities that participated in data collection, the guidance of an exceptional group of scholars, and the huge suport of family and some really good friends for this project to come into existence. It is my pleasure to open this dissertation by thanking them all.

First and foremost I want to deeply thank my dissertation advisor, Professor Juana M. Liceras, for her support over the years. Juana and I met in Madrid in 2002 and have been working together ever since. Anyone who has done linguistics in Europe or North America has for sure come across Juana's work; however, her professional expertise, as impressive as it might be, does not even compare to who Juana is as a person. She is someone that you can count on to walk with you the toughest parts of your path, and once she is with you in something, she will be with you until the very end, no matter how difficult things get. I am fully aware of how lucky I have been to meet Juana and all the linguistics and life learning that has come out of our partnership. I also have to thank Juana for always trusting that I would complete this project, particularly when I didn't think I would make it myself. ¡Muchas, muchas gracias por todo Juana!

Professor A. Jacqueline Toribio has also had a big role in the development and completion of this study. Her work inspired this dissertation, and my data would have been very different without her very generous help. My fieldwork trip to *el campo* meant the world to me for reasons well beyond the linguistic analysis. Jacqueline and Professor Barbara Bullock made me feel at home in *Los Compos*, where everyone was very welcoming and a lot of fun to be with. ¡Mil gracias Jacqueline! I am also grateful to Gerard Van Herk, who taught me how to Goldvarb and whose enthusiasm in the beginning of this project helped in making this dissertation a reality. I want to

also thank Eric Mathieu, who was a part of this project for a long time and guided the syntactic, theoretical aspects of this dissertation. Very many thanks as well to John Lipski and Elena Valenzuela, who were a part of this dissertation's committee and provided great insights and very relevant suggestions for further research.

The Dominican speakers that participated in this study deserve my warmest acknowledgment. The biggest privilege of doing this work has been, no doubt, the time I spent listening to people's stories while doing fieldwork. To this day I am amazed at how people will so generously share their life stories and views when you are genuinely interested in what they have to say. Fieldwork was an incredibly fun, tough, intense, life changing experience, ¡un millón de gracias a todos! I had access to the Santo Domingo speakers thanks to my Dominican family; Abuela Josefina, my uncles, aunts and cousins all helped me greatly and granted me the privilege to reconnect with them in my two months in the Dominican Republic. Son demasiados para nombrarlos a cada uno por su nombre, pero todos ustedes saben quiénes son, mil gracias a todos, ¡sin ustedes nunca lo hubiera conseguido!

The Deparment of Modern Languages at the University of Ottawa has also been very supportive over the years. Working at the Language Acquisition Laboratory has been a great learning experience at many different levels, and has allowed me to work with a great group of people always ready to contribute to each other's work. I have also been offered the opportunity to teach every single Spanish course offered in the department, and I loved my teaching experience and my students in Ottawa. Many thanks also to the Part-Time Professors Association, the Faculty of Postdoctoral Studies and the Association of Graduate Students. Without their generous contributions, I would have not been able to attend conferences and meetings that have been fundamental in the development of this project. I am also very grateful to the academic assistants at the department, particularly to Catherine Bernard and Manon Lavallée-Pratte for all their help and patience. Thanks so much also to Flavia Brun del Re, who did a great job in coordinating all matters related to my defense.

In my years at the department I have met some great people I am happy to count as my friends. Thank you very, very much to Anahí Alba de la Fuente, Andrés Arteaga, Susie Cruess, Rocío Pérez-Tattam, Cristina Senn and Alberto Villamandos for all the good times and support. You guys made all the hardships bearable! Outside of the department, I want to thank my amazing Ottawa friends Fouad Matar and Paula Gettler, thanks so much for being the greatest friends ever! You made leaving Ottawa hard, and I still miss you tons.

It was also thanks to a group of friends that I was able to finish transcribing and coding my data. My rockstar coders, Joanne Markle, Laura Segundo, Carlos Soler Montes and Maribel Usuga, you don't even understand how much you helped me! Carlos and Laura are also part of my Madrid group of friends that I love dearly. Along with them, my warmest thanks to the rest of 'los amigos de Peter', Javier Bahamonde, Khadija Ibrahim, Irene López, Miguel Lucía and Carolina Villanueva (& Manu): sois los mejores. ¡Un millón de gracias por todo el apoyo, las horas de skype, las visitas a Ottawa y todo lo demás! Each of my Madrid friends has provided me with their unique and very special love and support when I needed it the most, y una de las cosas que más me gustan de mi vida son nuestros más de 20 años (and counting!) de amistad. Que sean otros 20, y otros 20 más.

In the last year, two people have been a part of this work as much as me: Guillermo Lafuente and María José Villahoz. María José, muchas gracias por caminar conmigo todo este tiempo y por ayudarme siempre, siempre a sacar lo mejor de mí. Guiller, mi Guiller, gracias infinitas por todo. Volver a Madrid para conocerte ha sido lo mejor que me ha pasado en la vida.

Finally, and above all, a huge thank you to my family. I literally don't know where I would be today without their limitless, unconditional love and support. Gracias infinitas a mi padre, mi madre, Ricardo y Mariaje, cada uno de vosotros ha sido fundamental para terminar este trabajo. My mom is the most intelligent person I know, and I have no doubt that if I have made it through grad school is because she passed on to me some of her talents. However, this is not the most important thing she has done for me: from my mother I have learned the value of always doing good work, whether you are doing some fancy intellectual work or the most mundane task; she has also given

me the opportunity of finding the joy of working with language. It is because how she brought up my brother and me that I discovered the pleasure of reading and learning, ¡muchas, muchas, muchas gracias, mamá! My brother Ricardo has been the best big brother a little sister can hope for; I deeply know that since the day I was born there hasn't been a day that he hasn't thought about and cared for me. From Ricardo and Mariaje I have only received unconditional love over the years. ¡Un millón de gracias!

My dad has taught me the greatest life lessons I know. Most importantly, he has done that not by talking, but just living his life in such a way that he has provided the best model I could ever find of genuine kindness, generosity, honesty and hard work. I have seen my father get out of his way to help other people more times that I can count, and of course he has done it all and then some more for me as well. You will never meet two half-Dominicans who do not 'look' or speak Dominican more proud of their Caribbean heritage than my brother and me. I have no doubt that this is just a function of how proud we are of being our father's children. Although I could dedicate this thesis to all my family, because they all have been fundamental in the completion of this project, I think everyone will excuse me as I dedicate it specially to my father, who ultimately inspired this work. Un millón de gracias, papá, por todo. Te dedico esta tesis como un pequeño regalo para agradecerte todo lo que has hecho por mí, todo este tiempo.

Chapter One. Introduction. Dialectal variation and diachronic change: Null and overt subject patterns in Dominican Spanish.

In this dissertation I approach the issue of the extent and nature of variation within dialects of a language. Specifically, I investigate the variable null and overt subject pattern displayed by Dominican Spanish (henceforth DS). In recent years the study of dialectal variation, which had been traditionally carried out within the field of sociolinguistics, has emerged as an instrumental tool for evaluating generative theories on the nature of language variation and language change. By way of an example, theoretical studies of the syntactic phenomena associated with the expression of subjects in DS (Toribio 1993, 2000, 2001, Suñer 2003, Camacho 2006, 2008, Ordóñez and Olarrea 2006) have shown that dialectal data are relevant for disentangling questions related to word order patterns and the structural position of subjects in Spanish, and more generally for addressing issues concerned with the nature of null subject phenomena cross-linguistically. The present investigation aims to build on this research; however, it departs from previous work in that theoretical implications will be drawn from the analysis of a large corpus of data gathered following standard variationist methodology. Consequently, this study also stands in the context of a line of work developed recently that explores the potential of bridging quantitative and formal paradigms for the explanation of syntactic variation (Adger and Smith 2003, 2005, Cornips and Corrigan 2005a, b, among others). In this view, quantitative results are not only considered relevant in that they may lend support to structural analysis, but they also act as a powerful tool for defining the limits of cross-linguistic variation and for deepening our general understanding of the language faculty.

DS has been described as a variety that departs significantly from Non-Caribbean Spanish in its grammar of subject expression. Specifically, in DS the rates of overt pronominal subjects are much higher that in Non-Caribbean Spanish, and these expressed pronominals arise in a specific set of syntactic environments. Furthermore, some of the constructions that call for an overt subject in

DS -overt subjects with a Topic interpretation, or interrogatives with no subject-verb inversion, for instance- are shared by other Caribbean varieties of Spanish, such as Puerto Rican, Cuban, or coastal Colombian, whereas some others have been described as exclusively Dominican, the existence of overt expletive constructions being a case in point. Interestingly, these overt subject constructions coexist in the speakers' grammars with the null subject structures that constitute the unmarked option in Non-Caribbean dialects.

The DS facts pose a challenge for formal theories on language variation. This is particularly so in the case of the first version of the Principles and Parameters theory formulated under the Government and Binding framework (Chomksy 1981, 1986): the variation displayed by DS would reflect two possible settings of the Null Subject Parameter, a totally unexpected result in a theory that conceives parameters as devices that are set in one direction or the other in the process of first language acquisition. In addition, it is also unexpected for high overt subject rates and specific constructions bearing an overt subject to arise in a dialect of a language such as Spanish, which has been traditionally described as a prototypical agreement-rich, null subject language. Lastly, in this framework, dialectal differences are not expected to mirror cross-linguistic variation. The proposed explanation for this puzzle in the previous literature (Toribio 2000; Martínez Sanz 2007) is that DS is undergoing grammatical restructuring and as a result a null subject grammar and a non-null subject grammar coexist in the minds of the speakers, in line with Kroch's (1989, 1994) *Competing Grammars Hypothesis*. This hypothesis presents some shortcomings, however, which will be discussed below.

The most recent incarnation of the Principles and Parameters theory, formulated in the framework of the Minimalist Program (Chomksy 1995, 2000, 2001, 2004), offers a context in which the DS facts, in principle, could be accommodated without resorting to *ad hoc* assumptions. In the Minimalist framework, language variation is located in the (functional) Lexicon, specifically in the feature composition of the functional items that enter the derivation. Therefore cross-

linguistic and cross-dialectal variation are expected to present similar patterns, since the elements subject to variation will always be defined in terms of the features of functional projections.

The Minimalist view on language variation implies also a shift in the conception of parameters; while Government and Binding parameters were conceived as large scale devices, and the setting of a parameter entailed the fixing of a whole compartment of a language-specific grammar, in Minimalism parametric differences in feature composition are expected to have more local consequences (Kayne 2005; Gianollo, Guardiano and Longobardi 2009). For instance, the Null Subject Parameter has been deconstructed in recent linguistic theory into a number of Null Subject Parameters (Holmberg 2005, Sheehan 2006). The formulation of smaller-sized parameters makes it possible to account for null and overt subject profiles that have been found over the years to be a possibility among natural languages, such as the so-called partial *pro*-drop patterns of Finnish and Hebrew (Vainikka and Levy 1999, Holmberg 2005), the discourse *pro*-drop properties of East Asian languages as first described in Huang (1984), or the (micro)variation found among Romance dialects (Tortora 2003 for Italian; Modesto 2000, 2009; Sheehan 2006 for Brazilian Portuguese).

The study of the DS facts from a Minimalist perspective would need to address two main issues: on the one hand, the specific formal definition of the null subject profile instantiated by DS, in terms of which the features -or lack thereof- are responsible for overt subject insertion in environments that call for a null subject in Non-Caribbean dialects; on the other, the explanation of the variable use of null and overt subjects in DS. While it might be the case that ongoing language change is the source of the apparent 'optionality' in the selection of null and overt subjects, the specific constraints involved in subject selection in each of the Dominican overt subject constructions are awaiting explanation. In other words, DS in its synchronic state, as an instantiation of a 'possible grammar', should be accountable for within the limits offered by formal grammars.

The investigation of the null and overt subject patterns displayed by Caribbean Spanish varieties has also been relevant for the study of cross-linguistic and cross-dialectal subject expression carried out within the variationist paradigm (Labov 1982, 1984, 1994). Within this sociolinguistic framework the main object of study is precisely the variability, syntactic and otherwise, that arises in natural languages. Variationist studies have shown that the variability that is found in the different components of a speaker's grammar is not 'free' or 'optional', but rather constrained by a myriad of factors, both language-internal and language-external. Two main findings emerge from the research on subject expression across dialects of Spanish: first, subject expression constitutes a variable phenomenon in all Spanish varieties, and not only in the varieties that are described as undergoing grammatical restructuring, such as the Caribbean dialects; second, subject expression is constrained by the same core set of factors across varieties, while showing differences in the relative relevance of those factors in specific dialects or language contact situations.

Two main sets of factors have been found to be relevant in Spanish subject expression: discourse-related variables such as the ones concerning the informational structure of the clause (Enríquez 1984; Bentivoglio 1987; Silva-Corvalán 1982, 1994; Bayley and Pease-Álvarez 1996; Flores-Ferrán 2002; Otheguy, Zentella and Livert 2007), and morpho-syntactic variables such as the grammatical person of the subject (Bentivoglio 1987, Cameron 1992, Morales 1997, Travis 2006, 2007, among others). Language-external variables are generally not significant for subject expression, except in situations of language contact (Otheguy, Zentella and Livert 2007; de Prada 2009), where they become very relevant for explaining patterns of subject use.

Variationist studies of Latin American Spanish further show significant contrasts in subject expression across dialects, which essentially distinguish Caribbean from Mainland speakers (Otheguy; Cameron 1992; Otheguy, Zentella and Livert 2007; Guy and Orozco 2007; Orozco 2008). Caribbean speakers show the highest overt subject rates; a perhaps more interesting finding is that they also show different rankings with respect to the factors involved in subject expression,

such as person of the paradigm. Furthermore, person constraints have been convincingly proven to not support the *Functional Compensation Hypothesis* (Hochberg 1986), according to which overt subjects are expected to arise when inflectional morphology fails to identify subjects (Cameron 1992, 1996). This unexpected result suggests that a principled explanation of the restructuring of subject expression systems in Spanish might reside in the investigation of constraints beyond the morphological licensing of subjects.

The null and overt subject patterns of a number of Caribbean varieties, either in isolation or in contact with English in the United States, have been explored in the literature: Bentivoglio (1987) studies subject expression in Caracas, Venezuela; Cameron (1992, 1993, 1996), Morales (1982, 1996) and Flores-Ferrán (2002, 2004, 2007) focus on Puerto Rican Spanish; Guy and Orozco (2007) and Orozco (2008) explore subject expression in Coastal Colombian Spanish. However, to the best of my knowledge the only study carried out within the quantitative paradigm that investigates to some extent the Dominican variable subject pattern is Otheguy, Zentella and Livert (2007), which compares the speech of Mainland and Caribbean Spanish speakers in New York City, in order to study the impact on subject expression of both language contact and dialect leveling across different generations of speakers. In this study, Dominican newcomers to the city arise as the speech community that displays a higher rate of overt subjects, and depart from Mainland as well as from other Caribbean speakers in the specific constraints and constraint rankings associated with subject expression, particularly in the case of morpho-syntactic factors. These results suggest that the Dominican system of overt and null subjects, while framed in the context of Caribbean varieties, presents relevant differences that are worth of in-depth exploration. The present study of the Dominican pattern of subject use, carried out in a monolingual context where factors such as bilingualism and dialect leveling are not present, could contribute to the understanding of the sources of this difference. If these were to be found in language change in progress, this study might also shed light on the direction of language change.

The above discussion of null and overt subjects cross-linguistically shows the specific ways in which generative and variationist approaches have traditionally differed in their approach to linguistic facts. In the quantitative paradigm, variation is considered an inherent property of language, and hence the focus is on the description of the constraints that regulate variable natural-language patterns. In generative studies, the aim is to explain the nature and limits of possible grammars, and the focus is thus on finding strong, categorical restrictions on these attested grammatical systems. These diverging conceptions have also had as a consequence that these theories have been largely articulated by reference to different types of data: the tendency in variationist sociolinguistics is to study local or non-standard varieties, whereas generative research has traditionally investigated the limits of cross-linguistic variation. In turn, the differing conceptions of grammar and the choice of different objects of study has led to a mainstream view in linguistic research according to which these approaches to language variation are complementary at best, if not irreconcilable.

However, in recent years a growing body of research on dialectal variation and language change is developing an innovative and more cohesive approach to syntactic variation that integrates insights from both the generative and the variationist traditions (Adger and Smith 2003, 2005, Cornips and Corrigan 2005a,b,c), among others. This shift in view has been motivated by different causes:

First, in the context of purely theoretical studies, there has been in the last decade a growing interest in studying dialectal variation and the specific instantiations that parameters can take in different dialects and non-standard varieties. This renewed interest in dialectal studies stems from the conception of language variation advocated for in the Minimalist Program, in which parameters are defined as devices that trigger mostly local consequences in the speaker's grammar. The study of closely related varieties in which this sort of local 'microvariation' is to be found has proven to be instrumental in determining the limits of syntactic variation (Tortora 2003, Cardinaletti 2004).

Second, already within the Government and Binding framework a number of studies investigated dialectal parametric variation. Outside the Romance domain, Henry (1995, 1997), Wilson and Henry (1998) already showed in their work on variable agreement patterns in Belfast English that differences within the dialects of a language can mirror cross-linguistic differences, and that this fact suggested that dialectal variation should be constrained within the parametric space defined by formal grammars. Inside the Romance linguistics field, the seminal work by Kato and Taralho (1984) on the Brazilian Portuguese pronominal system, in which theoretical implications were drawn by reference to quantitative data gathered with a variationist methodology, opened the possibility of investigating corpus data from a standpoint in which dialectal studies are conceived not as merely descriptive, but also as theoretically insightful.

The ideas set forth in Henry (1995, 1997) and Kato and Taralho (1984) have been reviewed and developed within the context of the Minimalist Program in the last few years. In these studies (Meechan and Foley 1995, Cornips 1998, Adger and Smith 2003, 2005, Henry 2002, Muysken 2005) it is proposed that the generative and the quantitative paradigm, far from being mutually exclusive, can contribute to each other in relevant ways. On the one hand, as pointed out by Meechan and Foley (1995:82), "by availing of a variationist methodology, one can truly catch a glimpse of grammatical structure", from which theoretical implications can be formulated; in addition, quantitative data can shed light on the categorical or variable behavior of the linguistic elements subject to study, which in turn provides linguistic theory with information on the nature and extent of variation afforded by natural-language grammars. On the other hand, linguistic theory can contribute to variationist studies in the definition of variables, which is a pre-requisite to decisions on what variants to include in quantitative investigations. By resorting to linguistic theory variationist sociolinguistics might find effective measures for deciding which variants are syntactically related. The accurate definition of variables is also an essential element for disentangling the issues related to the locus of language variation in the speakers' grammars.

In this dissertation, I will build on this recent body of research that attempts a cohesive view of syntactic variation by investigating the variable patterns of null and overt pronominal subject use in DS. The goal of this study is to give an accurate description of the Dominican variable null and overt subject pattern, which in turn aims to contribute to our formal understanding of language variation and the nature of possible grammars. In order to determine the specific contexts of overt subject use in this variety, quantitative data collected with a standard variationist methodology have been gathered to carry out this study. In selecting the variants for analysis as well as the coding procedures that precede the statistical analysis, the insights emerging from both the variationist and the generative literature on subject expression and subject licensing in Caribbean Spanish have been taken into account. The goal behind exploring factors that have been deemed relevant by linguistic theory but usually ignored in variationist studies on subject expression is to provide an accurate account of the Dominican null and overt subject system in its synchronic state. In addition, the specific ways in which the DS variable pattern departs from other Spanish dialects in the constraints associated with subject selection has been attributed to diachronic change in progress, as discussed above. For this reason, this dialect arises as a very suitable venue for probing into questions regarding the relationship between language variation and language change.

The next issue to be addressed is the specific way in which the generative and the variationist approach to language variation can be integrated in the study of the Dominican variable subject pattern. In the next two subsections I discuss the views on language variation and change taken in each of these two approaches. In doing this, I will show how the insights of both approaches can contribute to our understanding of attested patterns of variation in subject expression.

1.1. Language variation and change in the generative tradition.

Generative grammar sets as its ultimate goal the understanding of the nature of the 'possible grammars' produced by the innate language device that is part of the genetic endowment shared by all humans. As a cognitive theory of grammar, it focuses mainly on the individual, and hence the

methodological practice associated with these studies is the gathering of data in controlled experimental settings, or the grammaticality judgments that probe into the competence of native speakers.

From the onset of the generative enterprise, the theory has attempted to resolve the tension between accounting for the variation found across grammars while explaining the underlying principles that are shared by all natural languages. This is achieved by imposing very strong restrictions on the form of possible grammars, which are formulated in the form of parameters that are set in the process of first language acquisition on the basis of the input available to language acquirers. The understanding of language acquisition is therefore central to this theory. Acquisition is also central in language change processes. As summarized by Roberts (2001), in generative models of language change (Lightfoot 1979, 1991, 1998, 2002, Roberts 1993, Roberts and Roussou 2003), syntactic change is viewed as shift in the parametric values specified for a given language. In other words, parameter values can change as a function of time, and parameter change is an instantiation of parameter setting. In this sense, in the generative paradigm the study of language acquisition and language change model and theory building are conceived as the higher goals of the enterprise.

In standard generative theories, diachronic change proceeds as follows: a change is initiated when (a population of) speakers converge on a grammatical system internalized by the speakers whose linguistic behavior provides the input to learners. It is assumed that the change is carried through the speech communities, and that phenomena that happen at the community level are involved in the transmission of changes. However, from the perspective of linguistic theory the attempt is to abstract away from these factors and aim, as far as the theory permits, to focus on change purely as a relation between grammatical systems.

Two concerns emerge as fundamental for generative theories on diachronic change: the determination of the shifts in the input to language acquirers that trigger the changes in parameter

settings, and the form and nature of the grammars that arise in the process and as a result of language change.

A standard solution for the former issue (Lightfoot 1991, Roberts 1999) is to propose that parameters are expressed morphologically; this entails that when independent morphological or phonological changes conspire to 'obscure' the morphological expression of a given parameter, a change might take place. The explanations provided in the generative literature for the processes of diachronic change that result in a change of value in the Null Subject Parameter –the French shift to from null to overt subject language, or the restructuring of the Brazilian Portuguese grammar, for instance- reflect this view. In these accounts, the loss of the null subject property is attributed to the impoverishment of the agreement morphology that licenses and identifies null subjects (Roberts 1993, Duarte 1995).

This model is not entirely problem-free, however. The research on diverging null and overt subject systems has convincingly shown over the years that while it might be the case that inflectional morphology is key in licensing null subjects in prototypical null subject languages like Spanish or Italian, natural-language grammars offer mechanisms of subject licensing that go well beyond the morphological identification of subjects (Huang 1984, Vainikka 1999, Holmberg 2005); recently, it has also been shown that non-morphological strategies for subject licensing are available even in synchronic or diachronic varieties of prototypical, Romance null subject systems. This is the case of Brazilian Portuguese, which will be discussed at large in Chapter Three. These findings suggest that not all diachronic shifts in subject licensing are tied to morphological expression. Furthermore, the morphological view on diachronic parameter setting bears the burden of explaining what are the specific morphological triggers for the setting of a given parameter value in a such a way that, when not present, the parameter is set to the opposite value. In Roberts and Roussou's (2003:15) words, the theory needs to formally define what constitutes "robust parameter expression". A number of solutions have been proposed for this problem; essentially, they either locate triggers for change in the input for language acquirers -Lightfoot's (1991) theory on 'cues'

for parameter setting, for instance- or they place those triggers in the interaction between the input and the learner's acquisition device. This is the view in Clark and Roberts' (1993), Roberts (1999, 2007) theory on the computationally 'conservative' nature of the learning device which, in case of ambiguity in the morphological triggers for language acquisition has a built-in preference for relatively simpler representations, preferring for instance representations that involve base-generating structure to structure-building via movement.

The second crucial concern for generative models of language change is accounting for the grammars that arise in language change processes. In a model that poses strong, categorical restrictions on linguistic representations, the coexistence in a given synchronic grammar of structures that reflect two different parameter values constitutes a challenge. Kroch's (1989, 1994) *Competing Grammars Hypothesis* deals with the so-called 'optionality' issue of diachronic change by essentially proposing that a single grammar cannot sanction 'optionality'. Rather, in processes of language change the speakers have access to two different grammars, one that represents the 'old' parameter value, and another that reflects the parameter value that is making its way into the language. These two grammars coexist in the speakers' minds until one of the grammars wins over the other, i.e. until the shifts in the input for language acquirers are significant enough to trigger the definite (re)-setting of a given parameter value. Again, this has been the solution adopted in the investigations on the Null Subject Parameter in diachrony. Mixed null and overt subject profiles like the Dominican variable system studied in this dissertation have been accounted for by resourcing to Kroch's *Competing Grammars Hypothesis* (Toribio 2000, 2001, Martínez Sanz 2007).

However, the explanatory power of Kroch's theory has been questioned recently, both on pre-theoretical (Koopman and van der Wurff 2000, Biberauer and Richards 2005) and empirical grounds (Sorace 2000, 2005). Sorace points out that 'optionality', defined as the coexistence within an individual grammar of two or more variants of a given construction that make use of the same lexical resources and converge the same meaning, is well attested in grammatical systems. She explains that 'optionality' is not only found in the grammars of native speakers, but it is an essential

element in the developmental grammars that correspond to transitional stages of first and second language acquisition and in the ultimate attainment, end-state grammars of non-native speakers. In fact, 'optionality' has been identified as one of the core elements configuring interlanguage grammars, as shown by the extant research on the issue carried out in second language acquisition studies (Bley-Vroman 1989, Eubank 1994, Epstein, Flynn and Martohardjono 1996, Schwartz 1998, Schwartz and Sprouse 1994, 1997, Liceras 1998, 2007, Zobl and Liceras 2005, Liceras et al. 2006, Sorace 1993, 1999, Robertson and Sorace 1999, among many others). Language acquisition research suggests that 'single' developmental grammars yield systems that incorporate 'optionality' as a stabilized, rather than transitional, feature, and second, that 'optionality' must be accounted for within the limits of formal grammars, particularly in a model that conceives diachronic change as crucially driven by language acquisition.

Koopman and van der Wurff (2000) further contend that resorting to competing grammars in the minds of the speakers basically sidesteps the issue of reconciling the attested cross-linguistic variability with the deterministic generative model of grammar. In this model it is assumed that in structure-building operations there is competition for well-formedness among possible derivations, in such a way that optional syntactic operations are ruled out according to an evaluation metric based on economy principles. Attempts at explaining 'optionality' in developmental or end-state grammars within the generative framework have followed two main lines of reasoning: either it is proposed that 'true' optionality does not exist (Grimshaw and Samek-Ludovici 1995, Adger 1996), or it is hypothesized that optionality is to be relegated to domains other than the grammar proper (Adger 1996). That is to say, either optional forms involve different semantic representations and therefore they do not belong to the same candidate set, or discourse-pragmatic constraints regulate the distribution of these optional forms.

Kroch's (1989, 1994) *Competing Grammars Hypothesis* follows the line of reasoning according to which 'optionality' is more apparent than real, and therefore different surface structures are to be attributed to different underlying parameter settings. However, as pointed out by

Koopman and van der Wurff (2000), this approach has a number of shortcomings: first, one can wonder why only two grammars are ever in competition. Given that up to *n* binary parameters can in principle be contained in Universal Grammar, we would need to potentially allow for 2^n competing grammars, which in turn would be very difficult to account for in computational terms. Second, the authors suggest that Kroch's theory essentially implies replacing 'optional' operations in the speakers' minds with 'optional' grammars, which questions the overall elegance of this theory for dealing with diachronic variable systems. Finally, several studies show that in fact grammatical systems allow for some amount of semantically-vacuous optionality, i.e., instances in which the two alternatives for a given construction cannot be ascribed either to different Numerations or to different interpretive effects at the interfaces with the computational module, and are therefore equal in terms of economy of derivation. Fukui (1993) shows that this is the case with certain types of movement operations, such as rightward movement in English or scrambling in Japanese. Biberauer and Richards (2005) report similar facts related to optional movement operations for subject and expletive licensing in Germanic, which leads them to conclude that 'true optionality' should be accommodated within a single grammatical system. In this respect, they propose that the grammatical, uninterpretable features that trigger computational operations like movement, such as the EPP-features involved in subject licensing, only need to be checked and valued in the course of the derivation; the manner in which EPP-checking proceeds is not relevant for the computational system, as long as these features are valued before the derivation is sent to the interfaces. In other words, optionality arises not from the presence or absence of a given grammatical feature, but from the way this feature is satisfied. This difference in *how* the feature is checked is what yields semantically vacuous variation[1].

To sum up, syntactic 'optionality' is a phenomenon that has always constituted one of the objects of study in developmental theories of grammar, and is receiving increasing attention in recent linguistic theory. This suggests that a formal theory of grammar should include an

[1] Biberauer and Richards' (2005) theory will be discussed in relationship with subject-licensing cross-linguistically in Chapter Six.

explanatory account of 'optionality'. Recent Minimalist generative theories such as Biberauer and Richards (2005) outline specific ways in which 'optionality' can be captured by a formal theory of grammar.

The description and explanation of this sort of intra-linguistic variability has traditionally been the goal of language variation and change investigations carried out within the variationist sociolinguistics paradigm. In the next section I discuss the sources and foundations of this approach, as well as the insights on language-internal variability that arise from the variationist paradigm and are of interest for the study of variable syntactic patterns such as the one represented by the Dominican null and overt subject system.

1.2. Language variation and change in the variationist paradigm.

The variationist model of language variation (Labov 1966, 1969, 1972, 1984, 1994) is rooted in two main ideas, namely that 'optionality' –in Labov's terms, 'variation'- is inherent to natural language (Weinrich, Labov and Herzog 1968), and that the workings of grammar may have a quantitative and non-categorical component. In addition, as a sociolinguistic model, it aims to probe into the pervasive social meaning of language, and therefore the community rather than the individual has traditionally constituted the object of study in variationist research.

As far as the first core idea is concerned, Labov (1969) proposed the concept of 'variable rule' as an extension of the notion of 'optional rule' that was part of early generative models (Chomsky 1957, 1965). In these transformational models, natural-language grammars contained formal rules that could be applied obligatorily or optionally. Formal rules were conceived as transformations that connected 'deep' with 'surface' structures. For example, optional rules derived passive or negative structures from declarative sentences. Labov's variable rule was a conventional optional rule, redefined to include the social and stylistic -i.e. language-external- dimensions of language, which were to be investigated in sociolinguistic studies along with linguistic –i.e. language-internal- dimensions. Consequently, early theoretical linguistics work interfaced relatively easily with sociolinguistic research. However, the conception of the language faculty in subsequent

generative models (Chomsky 1981, 1986, 1995) as a system of principles that allows for a very limited range of cross-linguistic variation had as a consequence that these two approaches to language variation followed their own avenues, with generativists probing into the explanatory power of categorical rules and variationists studying the sort of "structured heterogeneity" (Labov 1972:36) inherent to natural language that variable rules account for.

Labov (1972) defines the linguistic variable as a structural unit that consists of a class of variants, which are in practice "alternative ways of saying the same thing". Crucially, the output of a linguistic rule thus conceived can be probabilistic rather than discrete, and a given linguistic constraint can have a quantitative rather than a deterministic effect on the outcome of a variable rule. For this reason, as stressed by Labov (1984), an analysis of the linguistic variable that is 'accountable' to the data is an essential step in sociolinguistic research. Accurate analyses of linguistic variables are driven then by the 'principle of accountability', which states that every variant that is part of the variable context, whether the variants are realized or unrealized elements in the system, must be taken into account. In this way it guarantees that the entire range of variability present in the data will be dealt with.

The accurate definition of syntactic linguistic variables and the differences between this type of variable and phonological and morphological variables has been extensively discussed in the variationist literature (Lavandera 1978, Romaine 1984, Cheshire 1987, Winford 1996, D. Sankoff 1988). The discussion stems from the idea assumed in variationist studies that variable rules are part of a single though heterogeneous system: if linguistic variables are structural units that amount to "alternative ways of saying the same thing", they must have semantic equivalence and underlying identity of structure. Full semantic equivalence is not always found among the variants of a syntactic variable, as pragmatic considerations are sometimes at the source of syntactic variation. Romaine (1984) points out in this respect that this raises questions on the nature of syntactic variation, as well as on the motivations for it. In any case, as noted by Winford (1996:177), whether the constraints on syntactic variation are motivated by pragmatic factors alone

or they serve other purposes, if the condition on semantic and structural identity is reformulated as a "condition of functional comparability" the quantitative study of syntactic variants can be carried out similarly to the quantitative study of morpho-phonological variables[2].

In short, the variationist theory features a view of language variation according to which the linguistic elements that make up a given speech community, whether syntactic, morphological, phonological, lexical or discursive may vary in a systematic way.

As pointed out by Young and Bayley (1996), the behavior of a speech community is best modeled by availing of quantitative methodology, since the core goal of the variationist approach is to discover not individual occurrences of a given linguistic item, but patterns of variability. To achieve this goal, sociolinguist studies carried out within this framework employ multivariate analysis to describe and explain the linguistic phenomena under investigation. The advantage of this sort of analysis is that it allows for capturing subtle grammatical tendencies and regularities in a given data set. Furthermore, it makes possible to assess the relative strength and significance of these tendencies and regularities when all possible factors operating on them are treated simultaneously.

The variationist methodology has proved particularly fruitful in dealing with dialectal differences (Wolfram 1999, 2001) or in isolating systems in language contact (Poplack and Meehan 1995). As already pointed out by Labov (1980), this methodology is also ideal for tracking historical connections between varieties of a given language, as the constraints operating on a given set of synchronic linguistic features are, at least to some extent, the reflection of diachronic patterns. It is precisely in this way that the variationist approach brings together sociolinguistics, dialectology, historical linguistics and language contact studies. This view, coupled with the recognition of systematic variation in all of its language-internal and language-external dimensions,

[2] The syntactic variable studied in this dissertation is a well-defined variable, and for this reason I do not consider relevant to discuss further the nature of syntactic variation. For thorough discussion of these issues, the reader is referred to Winford (1996), D. Sankoff (1988) and references therein.

yields a perspective in which the question of the relationship between language variation and language change can be addressed in a principled way.

In this model, language change is conceived as a gradual process in which the frequency of application of a given variable rule is key in determining the spread of a given linguistic feature across potential contexts of application; these potential sites for the spreading of a linguistic rule can be defined in a multivariate analysis, which can model the complex constraints and relative weights of the numerous factors that operate simultaneously on a linguistic variable. The successive diachronic shifts in the application of a variable rule in linguistic environments should be reflected in a series of shifts in the relative weight of the factors that constrain the application of that rule. As expressed by Labov (1982:75), linguistic change proceeds as an "ordered set of shifts in the frequency of application of [a given] rule in each enviroment". As pointed out by D.Sankoff (1988), the locus of diachronic change is to be found precisely in these variable aspects of the grammar of a given language. However, as explained by G. Sankoff (2009), the relationship between language variation and language change is not biconditional: since variation as understood by Labov is an inherent property of natural language, it is expected to arise in all areas of a given grammar, whether language change is effectively occurring or not. In brief, language change implies language variation but not vice versa.

The variationist model of language variation was designed by Labov (1972, 1984) to accommodate both inter- and intra-speaker variation. In this respect G. Sankoff (2009) points out that so far sociolinguistics research has laid out several patterns of possible relationships between language change and variation at the level of the speech community, leaving aside the role of the individual in these natural language phenomena. However, individual patterns or strategies for rule application add up to community-level instances of language variation and change in a way that needs to be precisely defined. This is the only logical possibility if we assume diffusion of language change from individuals or smaller groups to the speech community as a whole, and it involves probing into questions related to the relationship between community and individual grammars,

particularly with respect to the impact of both type grammars on each other in ongoing change scenarios. It is G. Sankoff's (2009:659) contention that "the reintegration of the individual into the overall matrix of the speech community and the evolving languages represents the greatest challenge and the greatest scope for advancement in the research of the next decade".

In sum, while differing in their approach to linguistic facts, the generative and the variationist approach share at their core the assumption that language variation is not random or unconstrained. Furthermore, as pointed out by Cornips and Corrigan (2005c), patterns of language variation and change, as much as they may be driven by a myriad of language external and language internal factors, are constrained at one level by the nature of possible grammars. Therefore both inter- and intra-speaker syntactic variation should be accounted for within the limits afforded by formal theories. This is the idea behind the cohesive approach to syntactic variation developed in recent years and adopted in this work. In the following section I explain how a study carried out within this approach could contribute to our understanding of the sources of the Dominican syntactic variation in the expression of subjects. I will also present the main DS facts as described in previous research, and outline the research questions that articulate this work.

1.3. The present study: Dominican Spanish in the context of language variation, language change and formal grammars.

The Dominican mixed pattern of null and overt subjects presents itself as a suitable venue for exploring the possibilities of a cohesive approach to syntactic variation for dealing with dialectal differences: Spanish subject expression constitutes a clear case of syntactic variation; in fact, quantitative studies carried out in the last two decades within the variationist paradigm have convincingly proven that the selection of subjects in this language is not always constrained by categorical rules, but rather is a variable phenomenon across varieties of Spanish. Furthermore, these studies also show that the same set of factors, which include syntactic, morphological, lexical and pragmatic constraints, regulate the output of subjects across dialects, while the relevance of specific factors varies from dialect to dialect. Theoretical studies on DS syntax have described this

variety as possessing a subject expression system that significantly departs from other Spanish dialects, so much so that this variety has been proposed to be undergoing change toward becoming a non-null subject language. I propose that a quantitative analysis of the Dominican pattern of subject use is best equipped to confirm or reject the theory that this variety is in the process of diachronic change. The reasoning behind this claim is as follows: if DS were in the process of becoming a non-null subject language, the effects of this ongoing change should be reflected in the factors that constrain subject use in synchronic DS; in addition, the variationist methodology allows for multivariate analysis, with which the relative weight of lexical, morphological, syntactic and discursive factors in the shifts in subject expression in DS can be captured in detail. In other words, this kind of analysis allows one to capture whether restructuring is located in specific areas of the speaker's grammar. Having available adequate theoretical observations and a detailed structural analysis of the Dominican constructions that are infrequent or unattested in other dialects allows one to undertake a solid variationist study of the DS null and overt subject pattern. In turn, the level of descriptive adequacy in capturing syntactic patterns of variation provided by quantitative work will make it possible to explore how the Dominican null and overt subject system could be accounted for by a formal theory of grammar.

Taking these ideas as a point of departure, three broad questions articulate the present study: What are the specific factors that constraint subject expression in DS? Is the output of subjects in DS regulated by the same factors that operate in other Spanish varieties, or does the Dominican subject expression system display significant differences that must be interpreted as derived from ongoing language change? How can a formal grammar account for the specific subject expression system displayed by DS in its synchronic state? These questions, and the contributions to our understanding of linguistic variation that the present study will make in addressing them are developed and substantiated in the following paragraphs.

1.3.1. The nature and limits of dialectal variation.

The Dominican dialect stands in the context of 'Caribbean Spanish', a label often used to refer to the Spanish varieties spoken in Cuba, Puerto Rico and the coastal regions of Colombia, Venezuela and Ecuador. DS has been described as bearing many similarities with these other varieties, while also significantly departing from them in that it displays a number of morpho-phonological and syntactic properties either infrequent or unattested elsewhere in the Caribbean. Notable among these properties are specific phonological reduction processes, some of which derive in the restructuring of morpho-phonological patterns of verbal and nominal inflection, and a number of syntactic phenomena, among them variation in null subject expression. In the following paragraphs I will present the main facts related to subject expression as they have been described in previous work, and point out the specific ways in which they speak to an overall restructuring of the subject licensing grammar. It will be shown that although the widespread availability of overt subject pronouns is the single linguistic trait invariably noted in syntactic studies on Dominican Spanish, relevant questions related to the null and overt subject system displayed by this variety await accurate description and explanation. In order to set this discussion in perspective, I will first offer a brief outline of the socio-historic circumstances that might have influenced the particular configuration of the Dominican dialect and that have rendered this variety a unique source of empirical data for the study of the limits afforded by dialectal variation.

As explained by Lipski (1994a), the socio-historic origins of Latin American Spanish varieties are tied to four main factors:

(i) The distribution of pre-Columbian indigenous populations across Latin American territories;

(ii) The social and geographical origins of the European settlers that populated Mainland and Caribbean territories, as the socio-demographic profiles of settlers significantly vary from one area to another;

(iii) The influence of languages oher than Spanish, mostly represented in the colonial Caribbean by the African languages of slave workers; and

(iv) The relative chronology of successive settling processes, which is intimately related to the degree of isolation or social, economic and politic relevance of each of the colonies.

The history of DS reflects this pattern of mixed causes. The eastern part of the island of Hispaniola, labeled Santo Domingo in 1496, was the first European colony in the New World. Before the arrival of the Spanish, Taíno people populated the island. As noted by Lipski (1994b), the influence of the Arawak languages spoken by these groups in contemporary DS is limited to lexical items used nowadays elsewhere in the Caribbean and throughout Latin America, as 80 to 90% of this population was devastated by diseases and killings in the first thirty years of the colonial period (Moya Pons 1996). From that point on, two historical circumstances have fundamentally determined the profile of the Dominican dialect: the marginality of the Spanish colony of Santo Domingo, and the French and English presence in the western side of the island.

During the colonial years, linguistic innovation was brought to the Americas via the main ports, in turn connected with the southern Spanish ports in Andalusia. The port of Santo Domingo, however, only enjoyed predominant status during the first three decades of colonization. As soon as the gold mines in Hispaniola were exhausted and the Taíno population mostly disappeared, the colonizers found the riches of Mexico and Peru, and Santo Domingo rapidly declined and turned into a colonial backwater. The colonists that remained in Santo Domingo after the first few decades of settling engaged in sugar cultivation. The exploitation of sugar plantations triggered the massive movement of African slave workers throughout the Caribbean; this was the case in the western side of Hispaniola, the French Saint-Domingue. In Saint-Domingue the black population grew astronomically as a consequence of the growth of the sugar economy (Lipski 1994b). The situation was radically different in Spanish Santo Domingo, where the economic marginality derived from the Spanish monopolistic system, under which the colonies were not allowed to trade with foreign

countries, dictated the patterns of slave importations. As a result, the proportion of blacks in Santo Domingo never came close to the figures found in French Saint-Domingue. The economic marginalization of the Spanish colony in Hispaniola extended well into the 18th century, when an effort was made to repopulate it with settlers from the Spanish Canary Islands and from other Spanish colonies, especially Cuba and Puerto Rico. One of the reasons behind this repopulation was the attempt to resist the French incursions into Santo Domingo, which had been frequent since the 17th century. However, Santo Domingo was never able to fully prosper. The colony did not benefit either from the sugar plantation boom that took place at the turn of the 19th century as a result of the termination of the monopolistic system and the opening of Spanish Caribbean ports to foreign ships. In other Antilles like Cuba and Puerto Rico, opening the ports entailed the massive importation of African slaves, a very significant growth in the economy, and the urbanization of many coastal regions. However, in Santo Domingo the consequences of the Haitian revolt, which started in 1791 and ended in 1804 with Haiti's declaration of independence, were too close at hand. Santo Domingo then refused to import great numbers of African slave workers for fear of duplicating the French system that had just been overthrown in Saint-Domingue.

As noted by Lipski (1994b), these circumstances are key to understanding why, as opposed to what is the case in other Antilles, vestigial Afro-Hispanic is rarely attested in the Dominican Republic. He also notes that with the notable exception of Megenney (1990), the Afro-Hispanic influence in the Dominican dialect has not been investigated in depth. However, and even if exact numbers on the demographic proportions of Africans in Santo Domingo do not exist for the colonial period, the major trends shown by the figures available (Larrázabal Blanco 1975, Moya Pons 1986) suggest that a long-lasting African contribution to the Spanish spoken in Santo Domingo was not viable: by the end of the 17th century, the slave population had dropped to around 15%; consequently, the slaves born in Africa and speaking Spanish as a weak second language or *bozales* were a minority, and the portion of Santo Domingo's population with African roots was mainly *criollo*, i.e. black or *mulatto* population that was locally-born and lived in close contact with

Spanish. While African-born *bozales* were still found in the Dominican Republic in the 19th century, they were proportionally few in comparison with the *criollo* blacks that acquired Spanish as their first language[3].

Lipski (1994b) concludes that in Santo Domingo the greatest non-hispanic linguistic influence is to be attributed not to African languages, but to Haitian Creole, a language that was in close contact to Spanish practically since its inception towards the end of the 17th century. In fact, he argues that even if the Dominican Republic and Haiti share a history of mutual hostilities that extends to present day, the contact between at least marginal varieties of DS and Haitian Creole has been wide and long lasting: after being under French and Spanish control between 1809 and 1821, Santo Domingo declared its independence, only to be occupied by Haiti in 1822 and remain under Haitian rule until 1844, when the Dominican Republic became finally independent. In subsequent years, the Dominican Republic tried to annex alternatively to Spain and the United States, and in fact became a Spanish protectorate between 1861 and 1865, when it received large numbers of Spanish settlers as a result. The country has enjoyed independent status ever since, with the exception of the period 1916-1924, when it was invaded by the United States. Lipski (1994b) submits that within the past century, the impact of Haitian Creole on DS has been confined to the rural border region, on the one hand, and to the sugar plantations or *bateyes* throughout the country, where hundreds of thousands of Haitian migrant laborers have been recruited over the years, on the

[3] As noted by Lipski (1994b), the geographical distribution of Afro-Hispanic texts closely reflects the profile of the African slave trade in Latin America, and for this reason the *bozal* language is the centerpiece of the studies on Afro-Hispanic varieties. *Bozal* is first attested at the turn of the 16th century in the Iberian Peninsula, where it is documented through the middle of the 18th century. Latin American *bozal* Spanish was first acknowledged by 17th century writers like Sor Juana Inés de la Cruz, who reproduced the language used by black slaves in Mexico, Peru, Bolivia and Central America. Most *bozal* representations, however, date from the 19th century. In this respect, the speech of Afro-Cubans, for instance, has been taken as evidence for the existence of a stable Afro-Hispanic creole in Cuba and elsewhere in the Caribbean (Megenney 1984, Otheguy 1973). Other authors sustain that while it might well be the case that a Spanish creole was formed in an early stage of the colonial period, evidence for its existence might not be gathered from 19th century *bozal* representations, as these the texts offer a series of configurations which sometimes are indistinguishable from traces of *bozal* or pidgnized Spanish, or from post-creole remnants which signal an earlier stage in which Spanish was creolized in Afro-Spanish communities. From a lengthy discussion on the issue of the "missing" Spanish creoles, see McWhorter (2001).

other. The extent of the influence of Haitian Creole, if any, outside these two enclaves remains unexplored, to the best of my knowledge.

Both Lipski (1994b) and Bullock and Toribio (2008) point out that contemporary DS interfaces not only with Haitian Creole, but also with American English, as a result of the successive waves of Dominican immigration to the United States that started in the 1960s, following the termination of the dictatorial regime lead by Rafael Leónidas Trujillo. Most of the Dominicans that entered the United States as transient immigrants in the 1960s stayed in the country; however, as noted by Toribio (2000b), the patterns of return migration to the Dominican Republic in the last few decades, and the fact that typically Dominican enclaves in the United States replicate the cultural and social institutions of the island, has led to the formation of a dense transnational community with a Dominican/American identity.

In sum, Haitian Creole and more recently American English constitute the main non-hispanic languages that have influenced the Dominican dialect. The effects of language contact in the United States, particularly with respect to language maintenance among Dominicans in the diaspora and regarding the language attitudes of Dominicans towards their dialect have been examined to some extent in the previous literature (Toribio 2000, 2006, Suaréz-Budenbender 2009). In these studies Dominicans emerge as very language retentive, while displaying the complex linguistic behavior that has been described for stigmatized varieties (Labov 1972) and that also characterizes Dominicans in their homeland (Pérez Guerra 1989), in which linguistic insecurity coexists with the covert prestige of certain speech forms. Toribio (2000b) finds that Dominicans are particularly conscious of the non-standard phonological features in their speech, while most of the lexical and syntactic properties that are usually associated with the Dominican variety in the linguistic literature go unnoticed by speakers. On the contrary, with the exception of Lipski (1994b), it is only in the last few years that the role of Haitian Creole in the configuration of rural Dominican varieties has been acknowledged in the literature (Bullock and Toribio 2007, Ortiz López 2007, 2009). Instead, the theoretical studies carried out to date on DS usually focus on either

the phonological or the syntactic features that distinguish the Dominican variety from its Caribbean and Mainland Latin American counterparts.

Investigations of Dominican phonology have centered on phenomena such as the deletion of postvocalic /s/ without aspiration, or the distribution of postvocalic nasals and glides characteristic of the Cibao region (Bullock and Toribio, 2008). Theoretical studies on Dominican syntax have focused on three main phenomena: the widespread use of overt subject pronouns (Toribio 1993a, 2000, 2001, Martínez Sanz 2007, Cabrera-Puche 2008); the availability of overt expletive constructions (Toribio 2000, Hinzelin and Kaiser 2007, Martínez and Toribio 2008); and the existence of specific structures related to the expression of phrasal and clausal focus, namely focalizing *ser* constructions and structures with clause final negation and affirmation (Bosque 1999, Toribio 2002, Camacho, 2008).

The use of overt subject pronouns in contexts where non-Caribbean varieties of Spanish would call for a null pronoun and the availability of overt expletive constructions was already noted in traditional studies on DS such as Henríquez Ureña (1940) and Jiménez Sabater (1975). These investigations constitute general descriptions of the language, in which the main morpho-phonological and syntactic properties of the Dominican dialect are identified. In them, DS is presented as a variety in which relic forms unattested elsewhere in the Caribbean coexist with apparent linguistic innovation. In the domain of syntax, such innovation is attributed to the widespread use of overt subject pronouns, and to the shifts in word order patterns in the direction of a fixed subject-verb word order in contexts in which other varieties require postverbal subjects.

Studies carried out within modern theoretical frameworks that have further examined the Dominican null and overt subject system also point to shifts in pronoun use that distinguish DS from its Caribbean counterparts. Although Caribbean Spanish generally displays more overt subject pronouns than the rest of the Spanish varieties, DS has been proposed to be "at the forefront of linguistic innovation" (Toribio 2000:317) in this respect. Dominican overt pronominal subjects are documented with every person of the paradigm, and they seem to bear a Topic reading in most

cases. As such, they are akin to default elements that do not add any pragmatic force to the sentence, in clear contrast to Non-Caribbean Spanish overt subject pronouns, which are generally inserted for focus or emphasis reasons. The example in (1) illustrates the insertion of first and third person singular pronouns that would be redundant in other varieties, as the referent of the pronoun is well established in the discourse. Overt pronominal subjects are in italics[4]:

(1) Cuando yo vine a la ciudad, fue porque mi tía me rescató de allá como quien dice. Yo era chiquitico, yo tenía aproximadamente ocho años, y ella habló con mi papá, y ø le dijo que ø quería como que yo estuviera con ella. Ella le gustaba mucho los niños chiquitos, pero siempre ella iba a los campos a buscarlos, porque ella decía que los de la ciudad tenían los ojos muy abiertos, entonces ella quería adaptarlos a su manera, no dizque ya agarrar uno cualquiera que ya tuviera experiencia, ella siempre le gustaban los muchachitos chiquitos de campo.

(Mauro, 25, UW)

'When I came to the city, it was because my aunt rescued me from back there. I was teeny-tiny, I was about 8, and she talked to my dad, and told him that she wanted me to like, stay with her. She liked little children a lot, but she always went to look for them in the *campos*, because she said that the city children had their eyes too open, and she wanted to adapt them to her ways, not like, getting a random one who already had experience. She always liked the little *campo* guys.'

The example in (1) also shows the mixed nature of the DS null and overt subject system: while pronominal subjects tend to be overt even when the referent is well established in the discourse and there is no shift in reference from one sentence to the next (cf. *ella iba, ella decía,*

[4] In what follows, the examples offered to illustrate facts on Dominican Spanish have been drawn from the interviews that constitute the corpus for the present study, if possible; where this is not the case, sources for specific examples are provided in the text. In the examples culled from my fieldwork, speakers will be identified with their pseudonym and their basic socio-demographic information (age, geographical origin – U=urban, R=rural-, and social status -W=working class, H=middle/high class-). In examples containing answer-question pairs, S stands for speaker, and I for interviewer. The grouping of speakers with respect to these categories and the reasons behind it will be detailed in Chapter Five.

ella quería), instances of null subjects also arise in the speech of the same speaker, which shows that the use of overt pronominal subjects is not categorical (cf. *Ella habló con mi papa y ø le dijo como que ø quería que yo estuviera con ella*).

Overt pronominal subjects are inserted not only in contexts where they would be pragmatically odd in other varieties, but also in environments in which Non-Caribbean Spanish would categorically select a null subject. This is the case of instances with an overt non-specific third person plural pronoun, as in (2) below, which alternate with the familiar constructions with a null pronoun. Similar effects have been found by Lapidus and Otheguy (2005, 2009) in the speech of Caribbean New Yorkers.

(2) Me hicieron cartas de recomendación algunos ortodoncistas, y tuve suerte que me *ø* aceptaron [en una universidad argentina para hacer un postgrado]. Porque *ellos* aceptan diez por promoción; las promociones entran cada dos años, entonces *ellos* aceptan a diez, y de esos diez siete son argentinos y tres extranjeros, y *ø* aplican de todos lados, o sea que tuve suerte.

'Some orthodontists wrote me recommendation letters, and I was lucky they accepted me [as a graduate student in an Argentinian university]. Because they accept ten students per class; new classes enter the program every two years, and then they accept ten students, among those ten students seven are Argentinian and three are foreign, and they get applications from everywhere, so I was lucky.'

(Martín, 26, UH)

The Spanish grammar allows for a number of options for the expression of non-specificity, aside from null subject insertion. One of these options, very productive in other Spanish varieties, involves an impersonal construction with a null subject and the clitic *se*. Bullock and Toribio (to appear) note that in Dominican speech these impersonal constructions are often replaced with a construction with the overt pronoun *uno*, as depicted in (3a) below. As shown in (3a), the impersonal construction with *se* is still available to some speakers, however. This construction is

widespread in the data set analyzed in this dissertation as well. Furthermore, in some cases this pronoun seems to shift its referential meaning and become equivalent to a first person singular pronoun. This is illustrated in (3b); note that the speaker alternatively employs *yo* and *uno* to refer to himself:

(3) a. En Enriquillo, que era el sitio que yo te decía desolado, en que había una carretera excavada en la roca, cuando *uno* iba para allá, había que avisarle a la guardia que *uno* iba, para que no dejaran a alguien venir de allá para acá, porque no *se podía* dar la vuelta.

'In Enriquillo, which was the desolated place I was telling you about, where there was a road dug out on the rock, when one went there, one had to tell the guard that one was going, so he would stop people coming the opposite way, because it was impossible to do a u-turn.'

(Federico, 81, UH)

b. S: Porque tú sabes, yo entro a las siete y salgo a las cuatro, entonces de cuatro en adelante hasta las diez de la noche voy a la universidad en las materias que tengo.

Because, you know, I get in at seven and get off at four, so from four o'clock on and until ten at night I go to university to the classes I'm taking.

I: Pero duermes poco tú.

So you sleep very little.

S: A veces sí, a veces *uno* duerme solamente desde las 12 hasta las 5 y media que *uno* se levanta para prepararse para venir para el trabajo.

Sometimes I do, sometimes one (=I) sleeps only from midnight to five thirty, when one (=I) gets up to get ready to come to work.

(Chago, 39, UW)

These constructions speak to a restructuring of the Dominican grammar in the direction of displaying the overt subject-verb fixed word order already pointed out by Jiménez Sabater (1975), which extends to contexts in which Non-Caribbean Spanish would resort either to a null subject construction or to an impersonal construction. These overt subject-verb constructions have spread as well to structures in which the subject of the sentence is an inanimate referent, again another context in which other Spanish varieties categorically select a null pronoun:

(4) a. *Ella* [la comunidad] es grande, no tan grandes, un poco pequeña…
Lit. The community is big, well, not that big, actually *she* rather is rather small.'

b. Pero yo no sé qué le pasó (a la camioneta) porque ella tiene gasolina y ella estaba caminando bien.
Lit. 'But I do not know what happened (to the truck) because *she* has gas and *she* was working well'.

c. Pero *ella* (la laguna) antes estaba llena sí.
Lit. But *she* (the lagoon) was indeed full.
(Bullock and Toribio, to appear:11)

Another instance of the same phenomenon is the availability of overt pronominal subjects not only with personal forms of the verb, but also with infinitives and gerunds:

(5) a. Así como nosotros estudiamos, como tú estudias, como yo estudio, hay un libro que nos ayuda a nosotros a conocer la carrera y entenderla y hacer los pasos que debemos hacer para *nosotros* aplicarla en la carrera, asimismo es en la Biblia. En la Biblia es la que nos enseña las cosas que Dios dejó para *nosotros* hacer y para lo que no debemos hacer. Entonces para *tú* conocer más a Dios, más a Jesucristo, y más a fondo lo que es la iglesia, solamente tienes que leer la Biblia.

'The same way we study, like you study, like I study, there is a book that helps us to know our career and understand it and take the steps we need to take for us apply them in our career, exactly like that is in the Bible. In the Bible it teaches us the things God left for us to do and not to do. So in order for you to know God better, to know Jesus Christ better, and the church better, you only have to read the Bible.'

(Eliser, 35, UW)

b. Cuando estaba en Enriquillo, que dormía ahí, optaba por juntarme con algunos amigos y me tomaba unos tragos y llegaba prendío pa' ø acostarme, porque si no tenía que acostarme antes pa' *yo* dormirme porque con los ronquidos de ellos no podía dormir.

'When I was in Enriquillo, where I slept, I opted to join some friends and have a few drinks, so I would get to bed drunk, if I didn't do that I had to go bed before the others so I could fall asleep, because with their snores I couldn't fall asleep.'

(Federico, 81, UH)

In addition to personal pronouns, overt expletive constructions are also documented in DS. Overt expletives are possible in existential constructions (6a), impersonal constructions, as in (6b), with metereological verbs, (6c), and with unaccusative verbs, (6d); examples (6c) and (6d) are from Bullock and Toribio (2008:11):

(6) a. Vamos ahí que ello hay sillas (Valderrama, 73, RW)

go.1PPL there that EXP is chairs

'Let's go there where there are chairs'

b. Ello tiene que haber otro paso (Ramphis, 40, RW)

EXP has that is.INF another path

'There has to be another way to get through.'

c. Ello no está lloviendo aquí pero allá sí.

EXP NEG is raining here but there yes

'It is not raining here but it is raining there.'

d. Ello vienen haitianos aquí

EXP come.3PPL Haitians here

'Haitians come here.'

All these structures would categorically require a null subject in Non-Caribbean Spanish. In fact, the categorical use of null expletives, as will be discussed in Chapter Two, was one of the core properties associated with the availability of null subjects in a language (Rizzi 1982, 1986). For this reason, these constructions have received some attention in the theoretical literature. A number of studies (Silva-Villar 1998, Hinzelin and Kaiser 2007) align Dominican *ello* with the overt expletives found in other Romance languages, where expletives are discourse-related elements, and as such are not merged in the canonical subject position for subjects. Discourse-related expletives were available in earlier stages of Non-Caribbean Spanish as well, as noted in traditional descriptions of DS (Henríquez Ureña 1940), in which Dominican overt expletive constructions are interpreted as relic structures characteristic of the historically more isolated rural Cibao region. This view, however, has been disputed in the recent literature (Martínez Sanz 2007, Martínez Sanz and Toribio 2008); Dominican *ello* does not seem to comply with the properties associated with discourse-related expletives, which in these studies is taken as evidence that it is in fact merged in the canonical subject position, and that its availability is related to the general restructuring of the subject expression grammar. Furthermore, in the corpus gathered for this study overt expletives arise not only in the speech of rural speakers, but also in the urban vernacular. A detailed discussion of these structures is provided in Chapter Four.

Also of interest is the fact that Dominican and generally Caribbean Spanish subjects retain their preverbal position in interrogative clauses. In these structures, Mainland Spanish varieties would call either for a null subject, or for an overt postverbal subject:

(7) a. ¿Y qué *tú haces* ahí?

And what you do.2PSG there

'And what are you doing over there?'

(Danilo, 52, UH)

b. ¿A qué hora *ustedes cierran*?

at what time you.2PPL close.3PPL

'At what time do you guys close?'

(José, 48 UH)

There is disagreement in the literature with respect to whether these constructions are only available with pronominal subjects. While it is the stand taken by Ordóñez and Olarrea (2006) that only pronominal elements can be non-inverted in questions, Toribio (2000) documents interrogatives with an NP subject that do not display inversion. Even if marginal, they are also attested in the corpus gathered for this dissertation, as illustrated in (8a-b) with a proper noun and a full NP subject respectively:

(8) a. ¿Qué *Dios* quiere para con nosotros, lo que yo haga?

What God wants for with us, CL.ACC what I do.1STPSG.SUBJ

'What does God want for us, what does he wants me to do?'

b. ¿Cómo los curanderos curan?

How the witch doctors cure.3PPL

How do the witch doctors cure?

A detailed look at the data in the corpus gathered for this study shows that interrogatives with non-inverted pronominals are clearly the default option for Dominican speakers. However, a small number of Non-Caribbean interrogatives with inversion are also found in the data set:

(9) Entonces Trujillo lo mandó llamar: "¿qué *hace usted* visitando la casa de don Fulano de tal?"

'And then Trujillo called for him: "what are you doing visiting this person's house?'

Note that in all the constructions depicted in (1)-(8) overt pronominal subjects arise in a preverbal position. This fact speaks to a restructuring of the Dominican grammar, as in Non-Caribbean Spanish the position of subjects is flexible, and crucially related to the lexical properties of verbs (unaccusatives usually bear postverbal subjects, for instance), and to the informational structure of the clause –overt subjects are used to signal focus. Dominican word order patterns have been central to syntactic-theoretical proposals on the structural projections that host and license pronominal and NP subjects in Romance (Toribio 1993, 2001, Suñer 2003, Camacho 2006, 2008, Lunn 2002, Martínez Sanz 2007). Specifically, it has been argued that overt subjects in Non-Caribbean and Dominican Spanish do not have the same status, and that while Non-Caribbean overts are 'strong' pronouns in the sense of Cardinalleti and Starke (1999), Dominican expressed pronouns constitute 'weak' pronominals. If this were the case, two positions for subjects should be available in the structure (Suñer 2003); this in turn has consequences for a general theory of subject expression and word order in Romance, as it is relevant for determining the linguistic material pertaining to different structural projections, and for questions related to the nature of null and overt subjects crosslinguistically. These issues will be approached in Chapters Two and Seven.

As pointed out above, aside from the constructions that display overt pronominal subjects, the investigations of Dominican syntax have also looked into the strategies available in this variety for the expression of phrasal and sentential focus. Some of these constructions are relevant for the present study, as they interface with the restructuring of the Dominican subject grammar. Two structures have received particular attention in the literature: the so-called focalizing *ser* constructions, illustrated in (10), which involve phrasal focus, and the structures with final negation, affirmation, or verb *ser* ('to be') shown in (11), which signal sentential focus:

(10) a. No, porque ese es un sitio que va *es* la gente que hace camping y vainas, y lo alquilan y que sé yo. (Marigabi, 19, UH)
'Lit. No, because that is a place to which the people that go *is* for camping and stuff, and they rent it and things like that.'

b. Uno hace eso es para enfatizar algo. (Toribio 2002:130)

'Lit. One does that *is* to emphasize something'

(11) a. Yo no me pienso bañar ahora no.

'Lit. I am not bathing now no.' (Néstor, 56, UH)

b. Yo salí de viaje hace tiempo fue. (Magdalis, 34, UW)

'Lit. I went on a trip long time ago (it) was.'

c. Nosotros hablamos ingles sí. (Toribio 2002:130)

'Lit. We speak English yes/ We do indeed speak English'

These constructions will not be dealt with in this dissertation. In this investigation the concern is with the strategies available in DS for the focusing of subjects, therefore the sentential focus phenomena represented in (11) will be set aside. The constructions in (10), even if they involve phrasal focus, are not fundamental for this investigation either. Focalizing *ser* structures involve two clauses, one of them headed by a copular verb; the focused constituent, which can be an object, an adverb or a postverbal subject, is placed in front of the copular verb. These constructions are widely available also in Colombian and Venezuelan Spanish (Sedano 1988, 1995, 2001), and have been accounted for as deriving from pseudocleft structures (Toribio 2002), or as being closer to prosodic focus constructions (Camacho 2008)[5]. In any case, as noted by Camacho (2008), focalizing *ser* structures in which the focused constituent is a subject, as opposed to what is the case in other Caribbean dialects, are not attested in DS. He attributes this to the fact that postverbal subjects are rare in this variety. In the data set analyzed in the present study no single instance of a focalizing *ser* construction with a postverbal subject was found, which supports Camacho's claim.

More relevant for the present study are the strategies that DS has available for the focusing of subjects. The grammar of Spanish offers two basic options for focusing constituents, namely

[5] The availability of these constructions has been considered relevant for the formulation of a general theory of focus in Romance. For a detailed discussion on the nature of these constructions and the relationship between sentential and phrasal focus in DS the reader is referred to Toribio (2002), Camacho (2008).

prosodic and syntactic focus, illustrated in (12) and (13) respectively, taken from Camacho (2008:1). Focused constituents are in brackets:

(12) a. Fue {Marta} la que trajo los regalos. ['true' cleft structure]
was Marta cl.acc that brought the gifts.

b. Quien trajo los regalos fue {Marta}. [pseudocleft structure]
who brought the gifts was Marta.

c. {Marta} fue quien trajo los regalos. [inverted pseudocleft structure]
Marta was who brought the gifts
'It was Marta who brought the gifts.'

(13) Los regalos los trajo {Marta}.
'{Marta} brought gifts.'

It would be reasonable to think that in a variety like DS, in which overt subjects are widespread and tend to bear a Topic reading, prosodic focus constructions should be rather infrequent, more so given that prosodic focus in Spanish entails inserting a subject in postverbal position. The extent to which these constructions are available in DS has not been explored in previous work; however, it has been pointed out (Toribio 2000, 2001) that clefting is a very productive strategy in this dialect, even in interrogative constructions, as in (14). In these clefted questions subjects are often preverbal, although there is some degree of variability, as illustrated in (14b):

(14) a. ¿Qué es lo que *yo* tengo que hacer?
what is CL.ACC that I have.1PSG that do
'What is it that I have to do?'

b. ¿Cómo es que se llaman *esas maticas*, Maria Amelia?
How is that CL call.3PPL those little.plants Maria Amelia
'How is it that those little plants are called, Maria Amelia?'

Affirmative sentences in which subjects are focalized through clefting are widespread in the data set gathered for this study. Inverted pseudoclefts, in which the subject appears in a preverbal position (15a,b) alternate in the speech of most speakers with the pseudoclefts that bear a postverbal subject (15c,d); a small number of speakers also use the 'true' clefts in (12a) above, which also require a postverbal subject (15e):

(15) a. Yo la vi en el Banco Popular, y yo digo, yo como que conozco a esta vieja… ella me miró a mí, [ella fue que me dijo mi nombre.]
'I saw her at the bank, and I thought, I think I know this old lady…
it was her who looked at me, it was her who called my name.'
(Néstor, 56, UH)

b. Entonces como yo era pitcher, yo dije en mi mente, [esa es que dio a mi hermana.]
'And then, since I was the pitcher, I said in my mind, that is the one who hit my sister earlier.'
(Ramona, 28, RW)

c. Yo no me siento mal, al revés, yo me siento bien, porque [fui yo que le dije]
'I don't feel bad, on the contrary, I feel good, because it was me who told him.'
(Bruno, 52, UW)

d. Cuando decían El Jefe nadie se equivocaba, [era Trujillo que llamaba.]
'When they said El Jefe nobody got it wrong, it was Trujillo who was calling.'
(Federico, 81, UH)

e. Espectacular me pasó con una artista, que teníamos una producción en la cual [la que se fajó fui yo.]

'Spectacular is what happened to me with an artist, that we had a production together in which the one that did all the hard work was me.'

(Atilia, 50, UW)

The quantitative study carried out in this dissertation might contribute to an account of the focus strategies available in DS. Specifically, a relevant inquiry is whether these subject-focusing structures are part of a system of focus assignment that reflects a significant restructuring of the Dominican grammar, or whether the discursive constraints that regulate the organization of the informational content in syntactic structures are applied in DS as well.

The widespread availability of overt pronominal and NP subjects in DS, as well as the shifts in word orders attested in this variety, suggest that the Dominican grammar of subject licensing departs in some respects from its Caribbean and Mainland counterparts. It also points to the emergence of novel strategies for the expression of subjecthood. It is important to note that relevant differences seem to exist among the constructions discussed so far: for instance, some constructions seem to be part and parcel of the general Caribbean Spanish grammar, such as personal infinitives, attested also in Puerto Rican Spanish (Morales 1997). This is also the case for interrogatives with no subject-verb inversion, which are also documented in Cuba and Puerto Rico, and that according to Granda (1991) were already available in DS in the 18th century. However, while in these other Caribbean dialects non-inverted interrogatives are typically found only with pronominal elements (Cameron 1992), in DS these constructions are documented with NP subjects as well, as shown above. Facts of this sort have led Toribio (2000) to propose that DS has attained a further level of restructuring in its subject expression system, compared to other Caribbean Spanish varieties. The presence of constructions involving overt subjects that are exclusive to DS, such as overt expletive constructions, has been interpreted as confirming this claim. It is proposed here that a quantitative study of the DS patterns of overt subject use is well suited to test the hypotheses put forward in theoretical studies, as it will allow for an accurate description of each of the contexts in which overt subjects are attested.

First, a study of this sort will show if in any of these contexts the use of overt subjects is categorical, or whether subject expression is generally variable; given that variable subject expression has been found in all Spanish varieties, if it were the case that Dominican overt subjects are regulated by categorical constraints in any of the contexts studied, a major restructuring of its subject licensing system would be confirmed. Second, this investigation will let us determine whether the fixed subject-verb word order affects also NP subjects and to what extent, and whether there are significant differences in pronominal and NP subject expression generally. Third, the constraints involved in subject expression in Spanish, as will be discussed in detail in Chapter Four, rank from morpho-phonological to lexical, syntactic and pragmatic constraints; a study that takes into account all these constraints simultaneously will inform us of the relative significance of each of them in DS. For instance, in previous variationist studies constraints related to the informational structure of the clause, such as switch reference or focus, arise as fundamental in subject selection across dialects of Spanish. An issue not approached in previous work is how discursive constraints for subject selection are managed in a grammar in which overt subjects are so widespread and do not seem to add pragmatic force. The expectation, if theoretical studies are on the right track, is that discursive constraints should not be as relevant in DS as they are in other varieties, and that strategies other than overt subject insertion are available in this variety for the expression of non-presupposed informational content; however, only a detailed quantitative study can confirm this prediction.

Finally, it has been proposed that Dominican and Non-Caribbean Spanish pronominal subjects are of a different nature and as such are hosted by different structural projections. However, since the focus in these studies has been on the structures that display overt pronouns that are pragmatically odd or ungrammatical in Non-Caribbean Spanish, no attention has been devoted to the specific contexts in which null subjects are still available in this variety. The assumption in these studies is that Dominican null subjects are equivalent to Non-Caribbean Spanish null subjects in all respects.

Interestingly, this is not what has been found in the investigations of Brazilian Portuguese, another Romance variety whose high rates of overt pronominal subjects have been attributed to diachronic change in progress (Duarte 1993, 1995, Modesto 2000, 2009, Sheehan 2006, among others). Since in earlier Brazilian and contemporary European Portuguese subjects are identified by the 'rich' agreement morphology displayed by verbs, the shifts in the Brazilian subject system are attributed to the loss of inflectional distinctions undergone by this variety over time. Quantitative studies of Brazilian Portuguese have shown that overt subject use in this variety is almost categorical, around 90% for the younger generations. In spite of this and even in the absence of 'rich' inflectional morphology, third person singular null subjects are still licensed in a well-defined set of contexts. The theoretical studies that approach the structural analysis of these null subject constructions have provided evidence that when inflection is lost, other syntactic strategies are available for subject licensing. Brazilian Portuguese facts will be presented in detail in Chapter Three. For now it will suffice to say that delimiting the specific contexts in which null subjects are licensed has made it possible to account for the null and overt subject profile of this variety, characterizing it either as a partial null subject system (Sheehan 2006), or as a discourse null subject language (Modesto 2009). The quantitative analysis of DS null subjects aims to shed light on the nature of the null subject constructions available in this variety, in order to provide relevant information for the formal definition of the Dominican subject expression system in its synchronic state. As discussed in the following paragraphs, former views on contemporary DS have explained its overt and subject pattern as the result of the availability of more than one grammar in the minds of the speakers, a profile proposed to be typical in instances of diachronic change, rather than attempting the definition of its subject licensing system within the limits of a single, stable grammar that allows for structured variability.

1.3.2. Language change and the Dominican variable pattern of subject expression.

Toribio (2000, 2001) proposes that DS is undergoing a diachronic change in progress with respect to the value that this language assigns to the Null Subject Parameter. She assumes Kroch's (1989) *Competing Grammars Hypothesis* discussed above, and submits that Dominicans are in this sense 'bilingual in their own language' (Toribio 2000: 339). As a result, they are able to display in their mental grammars the two possible values of the null subject parameter, one akin to Non-Caribbean Spanish that allows for null subjects, and one in which pronominal subjects are overt and the set of specific constructions with overt subjects mentioned above arise. In my previous work, (Martínez Sanz 2007) I have put forward an account of the Dominican overt and null subject pattern along the same lines, proposing that the main difference between the Non-Caribbean and the vernacular grammars available to speakers is the availability in the former of the verb-movement operation, responsible for null subject licensing. In these theories it is more or less explicitly assumed that the loss of null subject properties is associated, at least to some extent, with the loss of inflectional distinctions in DS. Final /s/ is elided across Dominican regional and social dialects, which has consequences in the verbal system: since {-s} is the morpheme that identifies the second person in the singular paradigm, second and third person converge in most tenses and moods; furthermore, the first, second and third person have become homophonous in the imperfect and the conditional.

The claim that null subjects in Romance languages are in some way linked to the presence of rich agreement is a long standing one, and it is empirically supported by synchronic facts on standard Romance null subject languages like Non-Caribbean Spanish, Italian, or European Portuguese. Therefore, it might well be the case that the impoverishment of agreement morphology over time triggered by the loss of phonological distinctions derive in the restructuring of null subject systems. The shift from the Old French null subject profile to the contemporary non-null subject properties of Modern French confirms this idea, according to Roberts (1993a). However, the

investigation of the loss of null subject properties in other Romance languages shows that factors other than the blurring of agreement patterns are involved in such change processes. In the case of Brazilian Portuguese, quantitative studies (Duarte 1993, 1995, 2000) have shown that in addition to the impoverishment of inflection, the simplification of the Brazilian pronominal system has also been fundamental in the path of restructuring followed by this variety.

The idea that the appearance of overt subject pronouns is linked to the loss of final consonants has been also explored in the variationist literature on Caribbean Spanish. Cameron (1992, 1993) tests Hochberg's (1986) *Functional Compensation Hypothesis* against Puerto Rican Spanish data. His results show that the insertion of overt pronominals do not confirm this hypothesis, according to which the highest rates of overt subjects should be found in the second and third person, where verbal forms are homophonous.

Furthermore, as will be discussed in Chapter Two, the examination of diverse null and overt subject systems in the last three decades has shown that natural languages allow for a number of strategies for licensing null subjects other than identification by the morphological features of the verb. This is the case for discourse null subject systems like the ones found in East Asian languages, or the partial null subject patterns that have described for non-typologically related languages such as Hebrew or Finnish. Recent work on Brazilian Portuguese, as discussed in the previous section, has shown that even within Romance in the absence of 'rich' agreement morphology the speakers can resort to alternative ways for the licensing of null pronouns in specific contexts.

The investigation of Brazilian Portuguese has shown that it might be the case that a language employs different strategies for the identification of null pronominals over time. It has also shown that diachronic shifts can yield grammars that do not correspond to any of the values typically associated with a given parameter: the null subject system exhibited by contemporary Brazilian Portuguese does not correspond to the one displayed by European Portuguese, nor to the full-fledged non-null subject system of a language like English. Note that this result is not surprising if we look at it from a developmental point of view. It has been convincingly proven over

the years that neither the grammars of L1 acquisition in its early stages nor L2 acquisition interlanguage grammars conform to target language systems. Given that language acquisition has a fundamental role in language change processes, it is expected to find grammars that do not conform to a given parameter value in ongoing diachronic change.

The next question that is relevant for the present study is how a formal theory of grammar could account for the systems that arise from grammatical restructuring. Since these are 'possible languages', they should fall within the limits afforded by UG, whether or not they correspond to established parameter values.

1.3.3. Formal grammars and variable systems.

Within the Minimalist Program the loci of parametric variation are sharply delimited: either variation is to be attributed to the feature composition of lexical items, or it derives from interface conditions. Therefore, all cross-linguistic differences in subject licensing strategies should stem from these two sources, or from independent properties of specific languages. In recent years a number of studies have explored how parametric differences proposed in previous frameworks can be captured by minimalist analysis (Holmberg 2005, Sheehan 2006, among others). In these investigations, as extensively discussed in Chapter Two, it is claimed that Romance null and overt subject systems vary across three different dimensions:

(i) The feature composition of Inflection, the functional category related to null subject licensing, in which features related not only to the morpho-phonological expression of subjects but also features associated with referentiality are valued;

(ii) The specific assembly of inflectional features determines as well which structural positions are projected to host subjects in different varieties;

(iii) Independent properties of specific languages are relevant for explaining cross-linguistic variation in subject licensing, and are related to the differing nature of both null and overt subject pronouns in different Romance varieties.

The study of the Dominican null and overt subject pattern offers an opportunity to explore the power of these theories for explaining the limits of syntactic variation within the dialects of Spanish. My aim is to build on formal theories on DS by accurately defining variable subject expression in this dialect. In previous theoretical studies, DS has been described as a system in which the use of null or overt subjects is 'optional' to speakers. The study of the contexts in which these null and overt pronominals respectively arise is oriented in this dissertation towards determining the sources of this 'optionality'; specifically, I will explore whether Dominican 'optional' forms involve different semantic representations and therefore 'optionality' is more apparent than real, or whether syntactic variation in this variety is actually an instance of the 'true optionality' studied by Biberauer and Richards (2005) and discussed above.

In sum, the main contributions of the present study will be as follows: first, it will constitute a quantitative analysis of the null and overt subject patterns displayed by Dominicans born and raised in the island. While there are studies of such variation in Dominican communities in the diaspora (Otheguy and Zentella 2007), an accurate description of the level of restructuring of the Dominican subject expression grammar in the island is not available. It has been claimed that DS is a variety that significantly departs from its Caribbean and Mainland counterparts in its patterns of subject use; a study of this sort could allow us to confirm confirm or reject this hypothesis. Second, if it is the case that subject expression is constrained in DS by factors that do not operate in other Spanish varieties, this study could contribute to the explanation of the patterns of diachronic change followed by DS; in particular, this investigation will probe into the relative relevance of lexical, morphological and syntactic constraints in a variety of Romance that is undergoing diachronic shift, thus informing linguistic theory on the direction of language change. Third, at the theoretical level this study represents a contribution to the body of research on non-standard null subject languages. In recent years, the study of these varieties has significantly added to our understanding of the processes of diachronic change and the limits of parametric variation among and within languages. A detailed investigation of the Dominican null and overt subject system could shed light on

fundamental issues related to subject expression in Romance, such as the precise nature of 'rich' agreement morphology able to license null subjects or the position in the structure of preverbal and postverbal subjects. It is the stand taken in this work that while recent linguistic theory offers the tools to explain the differences found among closely related grammatical systems, gathering quantitative data, which allow for descriptive adequacy, makes it possible to provide a the fine-grained structural analysis of such microvariation.

1.4. Organization.

This dissertation is organized as follows: Chapters Two and Three will establish the theoretical background for the present study. The discussion in Chapter Two focuses on the theories that have been proposed over the years to account for attested null and overt subject patterns. I will show how the examination of several number of non-typologically related languages within the generative paradigm led to the discovery that diverging cross-linguistic subject expression systems cannot be reduced to the reflex of a single parameter. This discovery in turn was fundamental for the significant revision of theories related to the nature of linguistic variation carried out within the Minimalist Program (Chomsky 1995, 2000, 2004). This model offers a novel view of parametrization, which allows for an examination of previously problematic facts. Specifically, it makes it possible to define the sort of variation that arises in closely related grammatical systems. Chapter Three is devoted to the discussion of these issues by reference to Brazilian Portuguese, a language that displays a null and overt subject pattern close to the DS system. In this chapter the advantages of the cohesive approach to syntactic variation adopted in the present study are also presented. In the study of Brazilian Portuguese, while quantitative investigations have identified the sources of syntactic variation in the expression of subjects, theoretical studies that attend to the results obtained in quantitative work have offered a fine-grained structural analysis of the Brazilian Portuguese null subject profile in its synchronic state.

Variationist approaches to subject expression are discussed in Chapter Four. In Chapter Five I present the methodology used in the quantitative investigation carried out in this study. I will

explain how the findings of both quantitative and theoretical research on Caribbean Spanish varieties have been incorporated into the formulation of the research questions that articulate this dissertation.

In Chapter Six I present the results of the statistical analyses. Results show that while subject expression in DS is subject to constraints that regulate it in other Spanish varieties, differences in factor and constraint rankings suggest a restructuring of the Dominican grammar of subject expression. Chapter Seven discusses the theoretical implications of the results presented in Chapter Six. Chapter Eight contains conclusions and suggestions for further research.

Chapter Two. Theoretical Background: The Null Subject Parameter.

2.1. Introduction: Null and overt subjects in linguistic theory.

A central assumption in generative linguistic theory is that syntactic variation is parametric in nature. Consequently, pursuing an explanation of the form that parameters take in the language faculty has been a fundamental concern both in the Government and Binding Theory (Chomsky 1981, 1986) and in the Minimalist Program (Chomsky 1995, 2000, 2001, 2004). The Null Subject Parameter, as one of the parameters studied in most depth within linguistic theory, reflects in its successive reformulations the shifts undergone by the generative view on language variation in the last three decades. As it will become clear in the ensuing discussion, while early theories of parametrization were mainly formulated by reference to data related to cross-linguistic variation, in recent years the investigation of the (micro)variation found between closely related grammatical systems, such as the dialects within a language, has proven particularly fruitful in disentangling questions on the form and nature of parameters.

The aim of this chapter is twofold: on the one hand, to present an overview of the null and overt subject systems that have been discovered in theoretical linguistics research over the years, as well as the shifts in parametric theory that the study of these grammars has triggered; on the other, to frame the facts on dialectal variation in Spanish that are the focus of this study, as they are relevant for a theory of language variation and change. Laying out this theoretical background will pave the way for the discussion carried out in Chapter Three, which will center on the contributions that the cohesive approach to language variation taken in the present study can make to our understanding of cross-linguistic and cross-dialectal variation patterns.

The initial formulation of the NSP within the Government and Binding framework (Chomsky 1981, Rizzi 1982, 1986) was grounded in two core assumptions: first, the correspondence between morphology and syntax, which has as a consequence that the availability of null subjects in a given language is essentially attributed to its morphological makeup; and

second, the idea of parameter clustering, i.e., that the fixing of a given parameter entails that a number of properties cluster together in all the languages that set that parameter according to the same value.

This early formulation of the NSP triggered in the eighties and early nineties the investigation of a myriad of facts related to the possibilities attested in natural languages for the expression of subjecthood. As the description of typologically different languages was made available, it was clear that cross-linguistic and cross-dialectal subject licensing is a complex phenomenon that cannot simply be viewed as the reflex of a single parameter. Specifically, different species of null subjects exist cross-linguistically, and they may derive from quite different constraints from language to language (Sigurdsson 1993; Holmberg 2009). Findings of this sort were key in the revision of parametrization theory that has been carried out within the Minimalist Program for linguistic theory (Chomksy 1995, 2000, 2001, 2004, 2009). This theory features a view on parametrization that allows accounting for null and overt subject profiles that posed problems for previous theories.

It is particularly relevant for the purposes of this study that minimalist theory is well suited to deal with issues related to dialectal variation and language change: in this model, languages are not the immediate locus of parametric variation; rather, all variation is attributed to differences in the features that make up the items contained in the Lexicon, on the one hand, and on legibility conditions imposed by the mental components that interface with the grammar, on the other. Under this view, the fact that cross-dialectal differences may mirror cross-linguistic variation, evidenced by the Dominican data presented in Chapter One, can be accommodated in pre-theoretical terms. Furthermore, the sort of fine-grained analysis granted by minimalism makes possible to account for these facts also at an empirical level.

For instance, variation in the feature assembly of functional items will determine the differing mechanisms available in specific grammars for subject licensing and nominative Case assignment, and these differences constitute the microparameters that languages set to a specific

value. In other words, a (macro)parameter such as the NSP can be deconstructed in a Case assignment (micro)parameter, a null subject licensing (micro)parameter, and so on. Fragmentation in smaller-scale parameters entails that subject licensing systems can vary in a number of ways, yielding different types of mixed patterns, depending on the language-specific choices in the setting of these microparameters and the local consequences that each of these choices has. This was not the case in Government and Binding theories, in which the setting of a parameter had global consequences in the grammar (i.e. parameter clustering). Note, however, that macroparametric differences can also be derived in a minimalist model, for instance by way of the combination of smaller-scale parameters.

The chapter is organized as follows: In sections 2.2. I present the initial formulations of the NSP, proposed under the Government and Binding framework. This explanation is followed in §2.2.1. and §2.2.2. by the discussion related to the NSP profiles found cross-linguistically that posed a challenge for this initial formulation, namely East-Asian Topic-drop languages, Null expletive languages, such as German, Icelandic and a number of Creole languages, and specific Romance languages with mixed pro-*drop* patterns, such as DS. Next, the Minimalist framework is outlined in §2.3. The basic assumptions on the language faculty put forward in this theory are explained in §2.3.1, while the minimalist view on parametric variation and the different NSP theories that stem from this theory are presented in §2.3.2. and §2.3.3. respectively. The empirical coverage and explanatory adequacy of the main accounts presented to deal with subject licensing cross-linguistically are discussed in this section as well. Section 2.4. focuses on subject licensing in Caribbean and Non-Caribbean Spanish; in this section, the main facts on DS are presented, and the main issues that the Dominican pattern of null and overt subjects raises are discussed: the question of the morphological makeup of pronouns in Non-Caribbean and DS respectively is approached in 2.4.1.; issues related to the structural position of subjects in this variety are dealt with in section 2.4.2., while the questions related to subject positions and information structure in Spanish are discussed in section 2.4.3. Finally, the specific constructions that significantly differ in Non-

Caribbean and DS, namely interrogative and infinitive structures, are presented in sections 2.4.4. and 2.4.5. respectively. Section 2.5. contains a summary.

2.2. The Government and Binding Model.

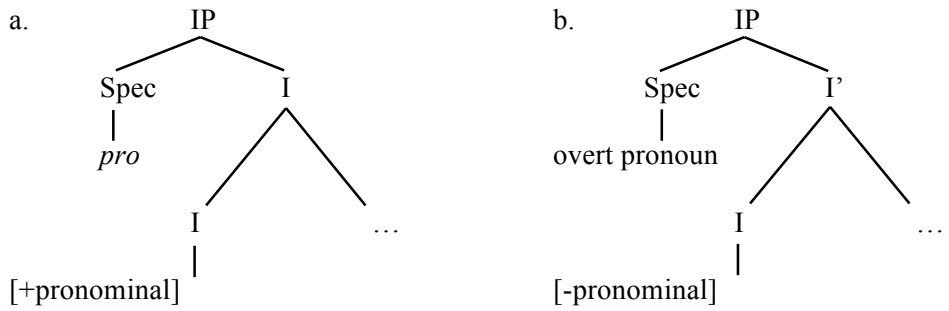
In this first theory on language variation, parameters were conceived roughly as binary options that can be fixed only in one direction or the other in the process of language acquisition. Each of the two values of any given parameter was associated with a set of properties and with the availability of a set of specific constructions in a language. Accordingly, the fixing of a parameter in the language acquisition process was assumed to trigger the fixing of a whole compartment of a specific language grammar. Baker (2005) labels this view of parametrization the "macroparametric" approach to language variation, and he notes that support for this view on language variation first came from facts on the Null Subject Parameter (henceforth NSP) (Chomsky 1981, Rizzi 1982, 1986) concerning the differences between languages like Italian or Spanish and the Germanic languages.

2.2.1. The Null Subject Parameter

Building up on previous work by Perlmutter (1970), Taraldsen (1978), the initial formulation of the NSP attempted to distinguish null subject languages (henceforth NSLs) from non-null subject languages (henceforth non-NSLs) on the basis of one principle and one parameter: the Extended Projection Principle, which requires all clauses in all languages to project a subject, and the presence or absence of morphological agreement. Languages like Spanish or Italian, with 'rich' verbal inflection, allow for the empty category subject to be identified whereas languages like English do not:

(1) *Null Subject Parameter* (cf. Rizzi 1982, 1986)

Subjects can be *pro*, where *pro* is a [+pronominal, -anaphoric] entity which is licensed by X_y, where X_y is a subset of the heads in a given language.



Infl licenses *pro* Infl does not licenses *pro*

Chomsky (1981), Rizzi (1982) noted how null subject languages like Italian or Spanish manifest the following cluster of properties:

(i) Possibility of null referential subjects. As opposed to what happens in a language like English, referential subjects can be null in Spanish, as the following contrast shows:

(2) a. Juan duerme

'John sleeps.'

b. duerme

sleeps

'He sleeps.'

(3) a. John is sleeping.

b. *(He) is sleeping.

(ii) Possibility of 'free' inversion in simple clauses. Rizzi (1982) proposed, taking as a point of departure facts from Italian, that the availability of referential null subjects in a given language correlates with the possibility in that language for "free inversion", which has as a consequence that the referential subject appears post-verbally. Spanish replicates this fact, as exemplified in (4a-b)[6]:

[6] While Rizzi (1982) labeled this inversion as 'free', as noted by López (2008) subsequent investigations have pointed out that inversion carries important semantic-pragmatic information on the status of the subject as given or new information in the discourse. This issue will be discussed in detail in light of Spanish facts in section 2.4.1 For a summary of relevant work on the relationship between subject position and the informational content of the clause, also see López (2008) and references therein.

(4) a. Juan ha llamado.

Juan has telephoned

'Juan has called.'

b. Ha llamado Juan.

has telephoned Juan

'Juan has called.'

(iii) Impossibility of overt expletive pronouns. While in a language like English, as in (5) an expletive subject is obligatory in sentences with weather predicates, as well as in impersonal and existential constructions, in languages like Spanish, in (6), which allow for referential null subjects, overt expletives are ungrammatical. In these languages a null subject is obligatory in these constructions:

(5) a. *(It) thunders. [weather predicate]

b. *(It) is true that I am leaving. [impersonal construction]

c. *(There) are many students in the party. [existential construction]

(6) a. {*pro* / *EXPL} relampaguea. [weather predicate]

thunders

'It is thundering.'

b. {*pro* / * EXPL} es cierto que me voy. [impersonal construction]

is true that me leave-1SG

'It is true that I am leaving.'

c. {*pro* / *EXPL} hay muchos estudiantes en la fiesta.

there-is many students in the party

'There are many students in the party.' [existential construction]

(iv) Absence of complementizer-trace effects. While in a language like English it is not possible to extract a subject across a *that*-type complementizer, which would yield the sequence

complementizer-trace of *wh*-movement, in languages that allow null referential subjects, like Spanish or Italian, this sequence is allowed in the grammar:

(7) a. Who do you think (*that) is calling?

b. ¿Quién$_i$ piensas que t$_i$ está llamando?

Who think-2SG that telephones

'Who do you think is calling?'

(v) 'Long wh-movement' of subjects. According to Rizzi (1982), languages like Italian or Spanish allow for 'long movement' of the subject in interrogative constructions:

(8) El hombre que me pregunto a quién había visto.

The man that to-me ask-1SG.PRES to who see-3SG.PAST

'The man I ask myself who has he seen.'

(vi) Empty resumptive pronouns in embedded clauses. Finally, Chomsky (1981) noted as well how null subject languages typically allow for empty resumptive pronouns in embedded clauses:

(9) Esta es la chica que me pregunto que puede hacer.

This is the girl that to-me ask-1SG.PRES what be able-3SG.PRES to do

'This is the girl I ask myself what is she able to do.'

As the examples in (2) through (9) show, the fact that Spanish allows for referential null subjects has specific consequences in other areas of the grammar, as expected by the Government and Binding theory on parametrization. Furthermore, the crosslinguistic variation found between languages like English, which does not allow for null subjects, and null subject languages like Spanish is accounted for by this theory as well.

This formulation of the NSP allowed accounting for a good deal of the variation in subject expression found cross-linguistically. It also triggered an extant amount of research on the nature of different languages and language varieties with respect to their null subject status, the acquisition of the null subject parameter and related properties (Hyams 1992, Pierce 1992, Clahsen et al. 1996,

Meisel 1994, among many others) and on the processes of diachronic change undergone through specific languages with respect to the NSP (for instance Vance (1989), Roberts (1993) on the shifts in Old French).

Note that the Government and Binding formalization of the NSP in (1) above assumes a very close relationship between morphology and syntax. This notion gets translated in the pivotal role of 'rich' agreement morphology in licensing null subjects. In other words, morphological variation was expected to correlate with syntactic variation. It is also expected that in the process of first language acquisition the learning of inflectional paradigms entailed the acquisition of syntactic structures, so throughout the Government and Binding period first language acquisition studies on the NSP focused mainly on the relationship between the acquisition of inflectional paradigms and parameter settings, on the acquisition of the clustering of properties related to the NSP, and whether NSLs and non-NSLs follow different acquisition paths, among other issues. In a theoretical model in which language change is guided by language acquisition (Lightfoot 1979, 1991, 1998), parameter change constitutes an aspect of the process of parameter setting, and therefore shifts in the properties of Agreement are also expected to drive diachronic change in null and overt subject patterns (Roberts 1993).

Parameters within the Government and Binding framework provided then a very powerful theory of language variation that practically reconciled synchronic and diachronic concerns in theoretical linguistics. However, as a wider range of languages was explored, this first formulation of the NSP, which relied entirely on the differences in inflectional paradigms, was proven to fail on empirical grounds. As pointed out by Holmberg (2005), when the NSP was proposed, the concept of inflectional 'richness' was only roughly defined, taking 'rich' agreement to mean that verbs in a given language carry enough morphology to provide non-ambiguous information on the person and number of the subject. Testing the specific nature of 'rich' agreement against the data from different languages proved this notion very hard to quantify. Work as early as Gilligan (1987), who compared synchronic null and overt subject patterns in over a hundred languages, showed that a

relevant number of morphologically 'rich' languages, such as Icelandic or Russian, are not NSLs. Findings of this sort triggered the exploration of what specifically is 'rich' agreement, and a number of proposals were put forward to accurately define the nature of an inflectional paradigm able to license null subjects. The main proposals on this issue are presented in the following section. In doing this, the issues that comparative syntax studies raised with regards to the other core idea contained in Government and Binding theories on the NSP, i.e. parameter clustering, will be discussed as well.

2.2.2. On the definition of rich agreement.

East Asian languages such as Chinese, Japanese, Thai or Korean, which bear no agreement morphology at all and yet license null subjects, represent a significant challenge to rich agreement theories:

(10) *e* lai-le (Chinese)
come.PERF
'He came'

The existence of these languages therefore questions that inflectional morphology alone is responsible for licensing null subjects cross-linguistically. Huang (1984) proposed a solution to this puzzle, arguing that these languages are Topic-oriented, and that what is licensed in null argument constructions is not a null subject/object, but rather a null discourse topic.

The evidence provided by Huang on the possibility of UG of allowing null subjects without 'rich' agreement in Chinese and other East Asian languages led Jaeggli & Safir (1989) to reformulate the NSP, adding to it the *Morphological Uniformity Condition* in (11):

(11) a. The Null Subject Parameter
Null subjects are permitted in all and only languages with morphologically uniform inflectional paradigms.

b. *Morphological Uniformity*

An inflectional Paradigm P in a language L is morphologically uniform iff P has either only underived inflectional forms or only derived inflectional forms.

This reformulation of the NSP allowed accounting for he fact that languages such as Chinese, with virtually no inflectional morphology at all, allow for null arguments. Moreover, the notion that null subjects are linked to rich agreement morphology in Romance-type languages seems to be empirically supported.

In fact, the role of inflectional morphology in Romance null subject licensing seems uncontroversial. Consequently, classical accounts of language change in NSP properties, such as Roberts (1993) submit that historically, when Romance NSLs have lost a substantial part of their inflectional morphology, null subject properties have disappeared from the languages in question. This is what Roberts found in his study of the shifts from NSP to non-NSL undergone by French. Based on these facts, he puts forward a formalization of rich agreement that stems from his comparison between Old and Modern French inflectional paradigms:

(12)

Old French	Modern French
chant	chante
chantes	chante(s)
chante(t)	chante
chantons	chantons
chantez	chantez
chantent	chantent

As showcased in (12), while Old French had six distinct agreement morphemes, Modern French, in spite of orthographical appearances, has only three or four, depending on the presence of liaison. Therefore the lost of null referential subjects in French seems to correlate with the loss of distinctions in inflectional morphology. Roberts (1993) definition of rich agreement goes as

follows: he considers a paradigm inflectionally rich when it has a complete set of person-number features. Old French would comply with this condition, given that Roberts regards zero endings as encoding inflectional richness, and that he claims that a complete set of inflectional features tolerates up to one syncretism.

Nonetheless, as soon as non-Romance languages enter the picture, exceptions to this definition are found. A case in point would be Icelandic. The Icelandic paradigm is presented in (13) alongside a paradigm from a consistent Romance NSL like Romanian:

(13)

Romanian	Icelandic
cînt	tel
cînti	telur
cîntã	telur
cîntãm	teljum
cîntati	teljið
cîntã	telja

The data in (13) show that Icelandic has an inflectional paradigm as rich as the Romanian one, with five distinct forms and one synchretism, which according to Roberts definition of rich agreement should suffice to license null referential subjects. However Icelandic, as opposed to Romanian, does not allow for null referential subjects. Roberts attributes the non-NSL status of Icelandic to the interaction of the V2 property in this language with the NSP.

Rohrbacher (1994, 1999) constitutes an attempt to provide a definition of rich agreement with wide empirical coverage that overcomes Roberts' (1993) problems. His definition of rich agreement is given in (14):

(14) Inflectional features are strong iff, for at least one number, the forms of the verb in the most common verb class maintain the following morphological distinctions:

a. [1st person] ≠ [2nd person]

b. i. [1st person] ≠ [3rd person]

ii. [1st person] ≠ [infinitive]

c. i. [2nd person] ≠ [3rd person]

ii. [2nd person] ≠ [infinitive]

Rohrbacher's idea is that rich agreement correlates with both verb movement and null subject licensing. In his account, 'strong' agreement affixes move to the Agr head in a Split-IP (Pollock 1989), allowing for null subject licensing. Pollock first proposed that IP is further split in two functional projections, Agreement and Tense. This Split-IP allows for a landing site for the moved verb in languages that license verb movement, and to account for the word order differences found between French and other Romance languages, on the one hand, and English, on the other.

Rohrbacher extends Pollock's account to subject licensing, and elaborates a proposal that relies on the differing nature of first and second persons versus the third person. He resources to the idea originally posited by Benveniste (1946) and taken up by Beard (1991) according to which third person is the unmarked person, i.e., the lack of any person specification, as opposed to first and second. In languages where there is any zero marked finite form of the verb, this null-marked form corresponds to a third person form. Left only with first and second person features, in Rohrbacher's account a given language will license both V-to-I movement and null subjects if and only if its regular verbs distinguish the forms for first and second person in at least one number of one tense from each other as well as from any third person forms, and from the form of the infinitive.

Rohrbacher's theory is able to account for a wide range of verb movement phenomena in the Germanic languages: on the one hand, Standard Mainland Scandinavian dialects bear no agreement morphology, and as predicted by the definition in (14), they display neither verb movement nor null referential subjects. On the other hand, Icelandic, which posited a problem for

Roberts (1993), distinguishes the first and second person plural form of regular verbs (cf. (11)), and has verb movement in non-V2 environments and allows for non-thematic null subjects. Finally, non-standard dialects of Swedish, which distinctively mark first and second person in the plural, also display verb movement, while in non-standard Norwegian dialects, which only make a number distinction, the verb stays in its base-generated position.

However, the definition in (14) does not have the same power in covering empirical grounds with respect to the distribution of null subjects. French represents a significant problem that Rohrbacher's theory needs to deal with. According to (14), French agreement is not distinctively marked, so it would be predicted to allow neither verb movement nor null subjects. Since French undisputably displays verb movement (Pollock 1989), Rohrbacher turns to adopt the idea that subjects in French are really agreement clitics, first put forward in Roberge (1990) and later developed by Sportiche (1995).

In sum, the notion of 'rich' agreement has proved very difficult to define without resourcing to language-specific stipulations, and in spite of several attempts, no theory has been able to account for all the possibilities displayed by different languages.

Furthermore, 'rich' agreement theories leave unexplained other possible null and overt subject patterns. Cases in point would be the so-called Null Expletive Languages such as German or a number of Creole languages. These grammars ban null referential subjects, while allowing for null expletive subjects in specific conditions. In addition, further cross-linguistic research on attested NSP profiles showed that Partial Null Subject Languages such as Finnish or Hebrew (Holmberg 2005, Vainikka 1999), also allow only for a subset of the null subjects licensed in NSLs like Spanish or Italian. These phenomena cast doubts on parameter clustering, (i.e. the set of properties that are taken to derive from the setting of the NSP outlined in (2)-(9) above), which along with 'richness' of morphology constitutes a central notion in traditional, Government and Binding, NSP theories, as pointed out above.

2.2.3. Non-standard null and overt subject patterns: Null Expletive and Partial Null Subject languages.

German, as well as some varieties of Dutch and Afrikaans, allows for expletive null subjects:

(15) a. Gestern wurde (*es) getanzt.
Yesterday was (it) danced
'Yesterday there was dancing'.

b. Gestern war *(es) geschlossen.
Yesterday was (it) closed
'Yesterday it was closed'.

In (15a), the expletive pronoun *es* cannot be realized, while the same pronoun in identical syntactic position, but with a referential interpretation, must be overt for the sentence to be grammatical. Rizzi (1982) already noted this asymmetry between German and NSLs of the Romance type, and submitted that within the NSP there are actually two related, however autonomous, parameters: one regulating whether null pronouns are allowed at all, and one regulating whether referential pronouns need to be expressed. English would set these two parameters as negative, while German would sets the first parameter as positive, and the second as negative. In Rizzi's account, the relationship between the two parameters is implicational; the presence in a language of null referential subjects implies the presence of null expletive subjects, but not the other way around. This seems to be the case for prototypical Romance NSLs, such as (Non-Caribbean) Spanish.

Research on a range of Creole languages has shown that some of these languages present a somewhat similar pattern. Null expletive subjects have been found in Cape Verdean, Berbice Dutch, Kriyol, Papiamentu, Palenquero, Saramaccan (Nicolis 2005, 2008), in a number of Creoles spoken in the Philippines (Lipski 1999), and in Haitian and Jamaican Creoles (Roberts 2007). The lack of inflectional morphology has been identified as one of the most common traits of Creole languages

(Bickerton 1981, Roberts 1999). In this sense, Creoles constitute a good testing ground for the Government and Binding hypothesis that person and number morphology (or the lack thereof) is a crucial factor in subject licensing. The fact that expletive null subjects are licensed in a number of these languages has been taken to suggest (Pratas 2002) that there are mechanisms other than inflectional richness that may license null subjects. Additional evidence for this claim might be found in the interesting fact that null expletive subjects have been documented both in Creoles that take Romance NSLs as their lexifiers (Papiamentu, Palenquero, Capeverdean creole, which have Spanish and Portuguese as their lexifiers), as well as in Romance and non-Romance lexified creoles whose lexifiers are non-NSLs (Haitian Creole, Saramaccan, Kriyol, which have French and English respectively as their lexifiers). The example in (16) makes the case for Papiamentu:

(16) *pro* (tawata) parse ku Maria ta(wata) malo. (Nicolis 2008:10)
PAST seem COMP Maria be.PAST ill
'It seems/seemed that Mary is/was ill.'

Partial Null Subject Languages also allow for null pronouns only in a subset of the contexts in which unexpressed subjects would appear in consistent NSLs. In Finnish and Hebrew, for instance, null subjects arise only in some persons of the paradigm. This phenomenon is illustrated in (17)-(18) with Finnish examples:

(17)

a.	(Minä) puhun	ruotsia.	(Me) puhumme ruotsia.
	I	speak.SG Swedish.	we speak-1PL Swedish
b.	(Sinä) puhut	ruotsia.	(Te) puhutte ruotsia.
	you	speak.2SG ruotsia.	you-PL speak.2PL Swedish
c.	*(Hä) puhuu	ruotsia.	*(He) puhuvat ruotsia.
	he/she	speak.3SG Swedish	they speak.3PL Swedish

(18) Pekka väittää [että (hän) puhuu ruotsia]
Pekka claims that he speaks Swedish

As shown by the contrast between (17) and (18), while Finnish first and second person pronouns can be null in any environment, a third person pronoun can only be null in certain contexts, specifically when it is bound by a higher subject. As noted by Holmberg (2005), these examples seem to support the 'rich' agreement theory, since while agreement morphology is heavier in this language in the first and second persons of the paradigm, third person agreement is not as inflectionally rich, since the third person singular is null in the past tense and the conditional mood.

However, Finnish also presents a further challenge to traditional, Government and Binding theories on the NSP: in this language, the availability of null referential subjects co-exists with the possibility of overt expletive constructions, so in this language parameter clustering does not obtain. Verb-initial sentences are generally not allowed in Finnish, as exemplified in (19a-c). To avoid a verb in the first position, either an expletive is merged, or an adverbial or argument is fronted, as illustrated in (20) to (22):

(19) a. *Sattui minulle onnettomuus.
Happened to-me accident
'I had an accident.'

b. *Meni nyt hullusti
went now awry
'Now things went wrong.'

c. *Viihtyy saunassa.
feels-good in-sauna

(20) a. Minulle sattui onnettomuus.
to-me happened accident

b. *Sitä* sattui minulle onnettomuus.
EXPL happened to-me accident
'I had an accident.'

(21) a. Nyt meni hullusti.

now went awry

b. *Sitä* meni nyt hullusti.

EXPL went now awry

'Now things went wrong.'

(22) a. Saunassa viihtyy.

in-sauna feels-good

b. *Sitä* viihtyy saunassa.

EXPL feels-good in-sauna

'One feels good in the sauna.'

The set of facts depicted in (19)-(22) is not predicted to be possible by Government and Binding NSP theories. Furthermore, the issue of the availability of overt expletive constructions in (partial or consistent) NSLs extends to the Romance domain. Overt expletive constructions have been found in a number of Romance diachronic and synchronic varieties. The existence of these constructions in languages like Spanish raises a number of issues for traditional NSP theories: first, a theory akin to Rizzi (1982) does not predict at all that null referential subjects and overt expletives may co-exist in consistent NSLs; second, Romance languages with overt expletives show that the set of properties typically attributed to the NSP do not necessarily cluster together; finally, as was already pointed out in Chapter One (cf. §1.3.1.), the co-existence of overt expletives and null referential subjects within the dialects of Spanish shows that cross-dialectal differences can mirror the variation found among languages, again a non-welcome result in a theory that reserves core syntax differences to cross-linguistic variation. The relevant facts on Romance overt expletives, along with the general issues that dialectal variation in NSP patterns raises for classical parametrization theories are presented in the next section.

2.2.4. Dialectal variation in Western Romance.

Uriagereka (1995) and Silva-Villar (1998) note that overt expletive constructions are attested in a number of Western Romance varieties. These constructions have been documented in both full-fledged NSLs, such as Galician or European Portuguese, as well as in languages that present a variable null and overt subject pattern, like Dominican Spanish. Overt expletives appear in a variety of contexts, among them declarative and interrogative existential constructions, such as (23a-b) and (27); impersonal constructions, depicted in (26); constructions with meteorological verbs, as in (25); and constructions with unaccusative verbs (23c):

(23) Dominican Spanish

a. *Ello* hay mujeres bonitas. (Silva-Villar 1998:249)
EXP there-is pretty women
'There are pretty women.'

b. ¿*Ello* hay dulce? (Patín Maceo 1947:76)
EXP there-is dessert
'¿Is there any dessert?'

c. *Ello* llegan guaguas hasta allá. (Toribio 2000:321)
'There arrive buses there.'

(24) Occitan

O pléut (Silva Villar 1998: 296)
EXP rains
'It rains.'

(25) Leonese

Ellu foi que nun llegarun a casasi.
EXP was that not ended-up.3PL to marry-SE
'It (just so) happened that they never ended up getting married.'

(26) Basso Polesano (Silva-Villar 1998: 298)

A pare que Nane vegna qua.

EXP seems that John comes here

'It seems that John comes here.'

(27) Galician

Il hai cecais outro problema. (Silva-Villar 1998:248)

EXP there-is perhaps another problem

'There is perhaps another problem.'

Both Uriagereka (1995) and Silva-Villar (1998) attribute Romance expletive constructions to discourse-related factors and submit that as a consequence, they belong in a position akin to Spec, CP. This proposal entails that that two subject positions must be available in Overt Expletive Romance: Spec, TP, which hosts the thematic subject, and a discourse-related functional position above TP that would be reserved for overt expletives. The attested Romance Transitive Expletive Constructions (TECs) would constitute evidence for these two distinct subject positions. The constructions in (28a) and (28b) exemplify TECS in synchronic Portuguese and 17th Century Spanish respectively:

(28) a. *Ele* os lobos andan con fame. (Silva-Villar 1998:269)

EXP the wolves go.3PL with hunger

'Wolves are hungry.'

b. *Ello* yo no sé por qué mi padre no me

EXP I not know why my father not me

llamó la torda o la papagaya.

called.3SG the thrush or the parrot

'I don't know why I was not called either the thrush or the parrot by my father.'

While Uriagereka (1995) and Silva-Villar (1998) submit that their proposal accounts for all instances of overt expletives in Romance, in previous work (Martínez-Sanz 2007, Martínez-Sanz and Toribio 2008), we have proposed that a detailed look at the different types of overt expletive constructions displayed by the different Romance varieties reveals that there is (micro)parametric variation with respect to the specific conditions that trigger the merging of an expletive. Specifically, we submitted that Dominican Spanish departs from other Romance languages, and that in this language overt expletives do not occupy a discourse-related position above TP; rather, in DS the merging of overt expletives does not obey discourse-related factors, and they are merged in Spec, TP, the same position that hosts overt pronominal subjects in this variety. Consequently, TECs are not grammatical in DS, as shown in (29):

(29) *Ello yo no sé por qué mi papá me puso Almeida. (Toribio, p.c)
EXP I not know why my dad me-called.3SG Almeida
'I don't know why my dad named me Almeida.'

This fact suggests that there is not a relationship between merging an expletive and triggering a discourse effect in Dominican Spanish, and that expletives in this variety belong in a Spec, TP position, similarly in this sense to a language like Finnish. As outlined in Chapter One, it also seems to be case that referential overt subjects do not trigger either a discourse-related effect in this variety (Toribio 2000). Therefore, the DS facts point to an instantiation of the NSP that differs from other Romance varieties that allow for overt expletive constructions.

As discussed in §1.3.1., DS represents an additional challenge to classical NSP theories, since overt expletive and overt pronominal structures are displayed in this variety along the Non-Caribbean null subject constructions. As was pointed out, this fact in itself poses a problem from the Government and Binding view on dialectal variation, in which cross-dialectal variation is not expected to target core syntax. Note also that the Dominican facts conflate the two set of problems that have been discussed in §2.2.1.: on the one hand, this variety does not show the expected parameter clustering derived from a positive setting of the NSP; on the other, the wide availability

of overt subjects questions whether agreement is the key factor in licensing null subjects in DS. For these reasons, arriving at an accurate explanation of the Dominican null and overt subject patterns has drawn some attention in theoretical studies. In what remains of this chapter, the existing views on the Dominican subject system will be presented. However, since these views are mainly framed in the Minimalist Program, before moving on to the detailed description in DS as presented in the previous literature, a brief description the foundations of minimalism will be provided.

2.3. Language variation in the Minimalist Program: Features and microparameters.

Under the Government and Binding theory parameters were expected to meet the following standards: descriptive simplicity, binarity, smallness of number, implicational organization and rich deductive structure. Since the comparative study of a number of languages showed that null subjects stem from differing grammatical constraints cross-linguistically, it seems to be the case that more than one parameter is involved in determining the null and overt subject profile of a given language. This result was not necessarily expected in Chomsky's (1981, 1986) theory of language variation. If there is not a single Null Subject Parameter but rather Null Subject Parameters, which in turn tend to be more construction-specific, parameters are at risk of turning into merely descriptive devices, and not the primitives or linguistic theory they were initially conceived as. The fragmentation in smaller, less general parameters undergone by the NSP described in this chapter also affected other parameters, such as word order, movement or configuration parameters[7]. Issues of this sort triggered the reconsideration of the very nature of linguistic variation carried out in the Minimalist Program.

The Minimalist Program emerges as a framework designed to critically evaluate the previous generative view on the nature of the language faculty. As far as language variation is concerned, Minimalism carries over from Government and Binding Theory the fundamental assumption that major syntactic variation among grammatical systems is to be attributed to parametric differences. However, as Chomsky (2005) notes, the complex descriptive technology

[7] For a lengthy discussion on the fragmentation of parameters during the Government and Binding era, see Baker (2005) and Kayne (2009).

developed in the eighties within the Government and Binding framework had as a result the postulation of relatively unconstrained parameters. The goal of Minimalism then would be to construct a theory of parameters that places substantive restrictions on their form and function, while maintaining their descriptive power. As we will see, this view implies restricting the locus of language variation within the language faculty, as well as reducing to the minimum the Government and Binding theoretical machinery. Before we move on to discuss Minimalist ideas on the nature of parameters, the locus of parametric variation, and the distribution of parameters in space and time, we present in the next section the view on the language faculty proposed in this model.

2.3.1. Back to basics: Sound and meaning.

The shift to the Minimalist Program represents a 'back to basics' in theoretical linguistics research in that it is an attempt to explore how much of the language faculty can be accounted for appealing only to *a priori* necessary conditions. Language, or the result of mental linguistic computation, is essentially sound and meaning. Therefore the computational device must interface with two mental components outside of the language module, the one responsible for sound and the one responsible for meaning, which are the Sensory-Motor system and the Conceptual-Intentional system respectively. Following this logic, the only *a priori* necessary conditions on language are legibility by the interfaces of spelled-out structures generated by the computational device. These interfaces are Phonological Form (PF) and Logical Form (LF). This is explicitly stated in the Strong Minimalist Thesis.

(30) The Strong Minimalist Thesis (SMT):

Language is an optimal solution to legibility conditions (Chomsky 2000:97).

The language faculty is then conceived as a system comprised by a Lexicon from which units are taken to build up derivations, a computational device that generates structures with these units taken from the Lexicon, and the interfaces with the Sensory-Motor and the Conceptual-Intentional systems. The lexical items that are part of a particular language Lexicon are proposed to be no more than bundles of three types of features: phonological features, legible to the PF

component, semantic features, legible to the LF component, and formal features, which drive structure-building by the computational system.

Since language is an 'optimal' or perfect solution for interface conditions, as stated by the Strong Minimalist Thesis above, it is assumed to be maximally economical. Notions of Economy, both at the abstract pre-theoretical level and at the level of computation (i.e. eliminating as much Government and Binding technical machinery as possible) are taken to be crucial in the Minimalist inquiry. This also entails that the principles of UG get reduced basically to: basic operations for building grammatical sentences, conditions on Spell Out of structures, and global economy metrics.

As far as basic operations go, according to Chomsky (1995), combination of phonological, formal and semantic features to form lexical items in the Lexicon; Numeration of lexical items –selecting a specific set of lexical items from the Lexicon to build up a given structure-; and Merge –the operation that puts together lexical items in the computational system- are of virtual conceptual necessity. Any system would require basic operations that take specific units and combine them. With regards to the features that make up lexical items, not all formal features that drive the computation are legible to the interfaces, and therefore must be deleted before reaching PF and LF. Features not readable in LF are labeled 'uninterpretable' –as opposed to 'interpretable' features, legible to the semantic component[8]. For this reason, the system requires an operation that allows for deletion of these uninterpretable features before Spell Out to the interfaces. This operation is labeled Agree, and consists in the local, Spec-head relation established between two items with matching formal features; one of the items will carry an uninterpretable feature, which will be matched and deleted –in minimalist terms, checked or valued (Chomsky 1995, 2000)- against the corresponding interpretable version of that feature carried by the other item entering Agree. What we know about specific languages shows that it appears that some uninterpretable features, for reasons not fully understood, do not get deleted locally via Agree and trigger movement instead. Move then arises as a basic operation in the computation of lexical items as well.

[8] According to standard minimalist assumptions (Chomsky 2000) the set of formal features and the set of semantic features intersect, while the set of phonological features does not intersect with the two other sets.

An example of a formal feature is the [Person] feature. With [Number], it makes up the set of the so-called phi-features, involved in the licensing of sentential (NP, pronominal, null) subjects. As we will see below, a Case feature and an EPP or 'subject' feature are involved in subject licensing. [Person] is interpretable on DPs, and uninterpretable on verbs (Pesetsky and Torrego 2001):

(31) Du komm-st (German) (Zeijlstra 2009:148)

you come-2SG

'You come'

In (31), the information that the sentential subject is second person singular is encoded twice in the morphosyntax, by the subject *du*, and by the marker *–st* on the verb. Since there is only one second person singular subject in the semantics of this sentence, the subject marker of the verb is meaningless, or in minimalist terms, uninterpretable in LF. In order to check and delete this uninterpretable feature, *du* and *kommst* enter into an Agree relationship, so the verb's uninterpretable person feature gets deleted against the subject's interpretable counterpart of [Person]:

(32) $[_{CP}$ $Du_{[i2SG]}$ $kommst_{[u2SG]}$

The purpose of Move is also to delete uninterpretable features before Spell Out, and in that sense it is actually a superfunction of Agree:

(33) Move = Agree+Pied-pipping+Merge

For instance, *wh*-fronting is triggered by an uninterpretable *wh*-feature on C. The *wh*-word, which carries an interpretable *wh*-feature, moves then to Spec, CP, where the uninterpretable *wh*-feature on C can be valued and deleted against its interpretable version in the *wh*-word:

(34) a. Who have you seen?

b. $[_{CP}Who_{[iWHj]}$ $[_{C'}$ $[_{C}$ [have $_{[uWH]}$ $[_{TP}$ you seen $t_j]]]]]$

According to the representation in (34), Move occurs in the familiar, Government-and-Binding-style: the *wh*-element moves to its landing site, Spec, CP, leaving a trace in its base-generated position in VP. However, as Chomsky (1995) notes, the existence of traces, or of any other empty category for that matter, is not justified in minimalist syntax. On the one hand, they are not drawn from the Lexicon but rather created later in the computation, and by minimalist assumptions, the only elements that enter into the syntax are lexical items selected from the Lexicon; on the other, empty categories are not legible by the interfaces. Chomsky proposes then the Copy Theory of Movement, according to which traces are simply copies of a lexical item contained in the Numeration for a given structure. The representation of (33a) above would then be (35):

(35) b. $[_{CP}\text{Who}_{[iWHj]}\ [_{C'}\ [_{C}\ [\text{have}\ _{[uWH]}\ [_{TP}\ \text{you seen}\quad \text{who}_{j}]]]]]$

This way Move is actually conceived as a composite operation of two primitive operations, Copy and Merge:

(36) a. Move = Copy+Merge

b. Move α =

(i) Copy $\alpha = \alpha'$, and

(ii) Merge (α', ß)

To sum up, *Select* (items from the Lexicon); *Merge* (items to build structure), *Agree* and *Move/Copy* (to eliminate uninterpretable features); and *Spell Out* to PF and LF are the basic operations engaged in linguistic computation.

Regarding the elimination of uninterpretable features, while Chomsky (1995) proposes that features are *checked* and deleted, Chomsky (2001) explains that the uninterpretable features that enter the operations Agree or Move actually get assigned a *value* through these operations. In other words, *feature checking* is reformulated into *feature valuation*. The idea that sustains this reformulation is that there is a formal difference between interpretable features and their

uninterpretable counterparts. The latter enter the derivation unspecified and are assigned values as part of the process of structure building. The proposal in Chomsky (2001) gives in this way formal expression to the intuition that agreement is directional, i.e., in a structure like (31) above, the verb agrees with the subject, and not vice versa.

According to the minimalist view, basic operations, conditions on Spell Out of structures and global economy metrics are the only instruments that the theory has to deal with language variation, and to answer the familiar questions of parametric linguistics, namely:

(i) What are the actual pararameters of UG?

(ii) What is the form of parameters?

(iii) How are parameters distributed in space and time?

As laid out above, in the minimalist model the language faculty is conceived as containing a Lexicon with items made up of bundles of features, a computational system and two interface components. Operations carried out by the computational device and Spell-Out conditions are not language specific but universal; thus, we are left with two logical possibilities for approaching language variation: the first one is to attribute cross-linguistic differences to the feature specification of lexical items; the second is to attribute them to interface conditions. The standard Minimalist view on parametrization locates language variation in the feature specification of functional lexical items, i.e., in the features that trigger feature valuation through the familiar operations of Agree and Copy.

2.3.2. Parameters in the Minimalist Program: on the locus of language variation and change.

The notion that parameters might in fact be associated with lexical items is a long-standing one, first proposed in the context of the Government and Binding framework as the *Lexical Parametrization Hypothesis* (Borer 1984), recently relabeled by Baker (2005) the *Borer-Chomsky Conjecture*:

(37) *The Borer-Chomsky Conjecture*

All parameters of variation are attributable to differences in the features of particular items (e.g., the functional heads) in the lexicon.

The Conjecture implies strong restrictions on cross-linguistic differences[9]: if lexical items constitute the sole focus of parametric variation, acquiring these will, in effect, amount to acquiring the language as a whole. In other words, learning "vocabulary" entails acquiring the parametric structure of a language.

The next question that arises is what exactly constitutes the set of functional items that get projected in the structure, or rather what is the specific set of formal features that are able to project, and whether this set of projections is UG-determined. This is not a novel issue in theoretical linguistics either: since Pollock's (1989) *Split-IP Hypothesis*, the question of which are the specific functional projections in the clausal structure of a given language has been approached from various points of view. Subsequent minimalist work has further investigated the set of functional heads that are present in each clausal domain; for instance, Rizzi (1997) studies the set of projections contained in the CP domain, Zanuttini (1997) the projections involved in Negation phenomena, and Cinque (1999) looks at the IP field, adding a number of functional projections to the original AgrP and TP proposed by Pollock. Rizzi's approach has become known as the *cartographic approach* to clausal structure and has been further developed in Cinque (2002, 2006), Rizzi (2004) and Belletti (2004). Under his view, the set of functional projections is taken to be UG-determined. Some recent approaches to functional categories (Zeijlstra 2009, *Flexible Formal Feature Hypothesis*), on the contrary, explore the idea that UG does not contain any set of formal features; rather, the set of formal features of a given language, and consequently its set of functional projections, is created during the process of language acquisition. A particular feature might be analyzed as a formal

[9] There has been a substantial amount of discussion regarding to which extent all parameters are candidates for reduction in terms of the BCC. For discussion on these issues, the reader is referred to Kayne (2009) and references therein.

feature able to project a functional projection if and only if there is substantial evidence for it in the input for acquisition.

Questions related to the number and nature of functional projections are central to subject licensing, as this phenomenon is related to the differing mechanisms for overt subject insertion and subject-drop available in different languages, the position of subjects in the structure of the clause and their semantic nature. With these specific concerns in mind, I now turn to outline the prevalent Minimalist views on the NSP, as they are relevant for the study carried out in this dissertation. In doing this, it will also be shown how the specific null and overt subject patterns found in cross-linguistic and cross-dialectal studies that used to pose problems for previous models are accounted for within this view on language variation.

2.3.3. The Minimalist view on the Null Subject Parameter(s).

The initial Government and Binding formulation of the NSP relied on Chomsky's (1981) Extended Projection Principle (EPP). The intuitive basis for this Principle is the concept of predication, and that it makes explicit a potentially universal property of natural language, namely that all sentences require subjects for structural reasons. Evidence for postulating this principle was initially gathered from the existence of overt expletives in languages like English, in which the only function of overt expletive subjects is to realize the subject position of a sentence, since overt expletives lack semantic content. This principle translates in formal terms in the requirement for the specifier of IP to be projected in all structures, as this is the canonical position for sentential subjects. In other words, in languages like English, if there is not a category such as the thematic subject to fulfill the EPP, the grammar turns to other means, such as the merging of overt expletives, in order to have the Spec, IP position filled. For NSLs such as (Non-Caribbean) Spanish, which do not have overt expletives and allow for null referential subjects, the Spec, IP position was proposed to be filled by the empty category *pro*. Importantly, the existence of *pro* allowed maintaining the idea that the EPP is a principle of UG.

Minimalist views on the NSP also take the Extended Projection Principle as the point of departure for the explanation of subject licensing conditions cross-linguistically: in early minimalism, Chomsky (1995) captured the structural requirement for subjects in all clauses by postulating the need for a feature in T that participates in subject licensing[10]. He labeled this feature EPP in reference to the Principle proposed in the eighties, since this is a 'subject feature' of sorts. This EPP feature is assumed to be universal, automatically merged in T and uninterpretable, and as such must be valued and deleted before the derivation is sent to the interfaces. He further proposes that the EPP reduces to a morphological property of T: strong or weak D features (Chomsky 1995:199). In other words, the requirement that all clauses must have subjects is reduced to the presence of an uninterpretable nominal feature in the TP domain.

Chomsky proposes that there are two possible ways to check and delete this feature, depicted in (38), each of them corresponding to the structures in (39a-b):

(38) EPP checking (Chomsky 1995):

(i) Merge XP

(ii) Move X

(39) a. There arrived a man.

b. A man arrived.

In (39a) the EPP feature in T is checked via *Agree* by merging an expletive. In (39b) EPP checking is obtained via movement. Recall from the discussion on feature checking in § 2.3.1. that T, in addition to the EPP feature, also contains a set of phi-features, and a Case feature. In (39b), for the phi-features in T there is only one choice to match these features under *Agree*, the phi-feature set of *A man*. This same operation could check and delete the Case feature in T. However, since there is an EPP feature in T that needs to be checked as well, *A man* moves to Spec, TP. This *Move*

[10] Since Pollock (1989) proposal on the split of IP in AgrP and TP, the canonical subject position is assumed to be Spec, TP. It has been discussed, particularly with respect to Romance (Toribio 2000), whether this is parametrized and Spec, AgrP and not Spec, TP is the canonical position for subjects in some languages. For the sake of clarity, unless specified in the following discussion I will take Spec, TP to be the canonical subject position.

(=*Agree* + *Merge*) operation eliminates all uninterpretable features in T, namely phi-features, Case and EPP, and the structure in (39b) is grammatical.

The next relevant issue related to the study carried out in this dissertation is whether EPP checking proceeds in the same way in (Romance) NSLs. In this respect, Roberts and Holmberg (2009) explain how in principle Rizzi's (1982, 1986) analysis of consistent NSLs translates well into minimalist terms, as he derives all of the correlations associated with the availability of null subjects in a given language from the parametrized properties of agreement. Therefore, framing the NSP for consistent NSLs follows naturally in the Minimalist view, in which parametric differences are attributed to the feature content of lexical items. Specifically, two lexical requirements must be met for a language to be a consistent NSL: this given language must have a specific type of functional category, i.e. a T with person and number features, and a particular type of pronominal, *pro*.

The first lexical requirement poses no problem for Minimalism. However, as explained by Holmberg (2005), this is not the case with the second lexical requirement, since within the Minimalist framework the existence of empty categories is not justified, as they lack both PF and LF properties[11]. If there is no category to merge in Spec, TP it may be the case that this specifier does not need to be projected at all in consistent NSLs. If correct, this would be the case for most Romance languages, which allow for null referential subjects and lack overt expletives.

However, Holmberg (2005) points out that Rizzi's theory on the licensing and identification of *pro* is incompatible with a theory of feature valuation like Chomsky (2001), in which uninterpretable features enter the derivation unspecified, and are assigned values as part of the process of deleting them from the derivation via Agree. Chomsky submits that he phi-features on T are uninterpretable, and are valued by entering in an Agree relationship with the subject DP. Argumental DPs are fully specified for interpretable phi-features, and therefore able to value and delete the uninterpretable features on T. The empty category *pro* is incompatible with this approach

[11] Recall that this was the same problem that other empty categories, such as the traces of movement, posed for Minimalism.

because "within this theory of agreement, it is obviously not possible for an inherently unspecified pronoun to be specified by the phi-features if T, as those features are themselves inherently unspecified" (Holmberg 2005:537). In other words, without specification verbal inflectional features can play no role in recovering *pro*'s grammatical content.

The question is then what is the nature of the EPP in NSLs and how this 'subject feature' is valued and deleted. Two solutions for this puzzle have been explored in the recent literature:

The first one is instantiated in Alexiadou & Anagnostopoulou (1998), and further developed in Manzini and Savoia (2002), Platzack (2003, 2004), among others. Alexiadou & Anagnostopoulou study consistent NSLs like Greek and (Non-Caribbean) Spanish and propose that EPP is parametrized: while in English-type languages the EPP is valued and deleted via subject insertion in Spec, TP**,** in consistent NSLs agreement morphology is strong enough to check the EPP feature, which has as a consequence that subject movement is not needed and Spec, TP is not projected:

(40) Parametrized mode of EPP checking: Move/Merge XP versus Move/Merge X^0.

The second solution, elaborated in Roberts (2004), Holmberg (2005) and Sheehan (2006) takes as its main idea that the possibility of having a zero pronoun is a PF matter: EPP checking would proceed as in English-type languages, and the subject pronoun is not spelled out. In a nutshell, the two options presented by each account are that either *pro* does not exist, or that *pro* is an ordinary pronoun without PF features.

Holmberg (2005) summarizes these two possibilities for dealing with the EPP in consistent NSLs in terms of the notion of interpretable and interpretable features. He presents them thus:

(41) a. Hypothesis A

There is no *pro* at all in null subject constructions. Instead, Agreement (the set of phi-features of I) is itself interpretable; Agreement is a referential, definite pronoun, albeit a pronoun phonologically expressed as an affix. As such, Agreement is also assigned a subject theta-role.

b. Hypothesis B

> The null subject is specified for interpretable phi-features. It values the uninterpretable features of Agreement, and moves to Spec, TP, just like any other subject. This implies that its nullness is a phonological matter: the null subject is a pronoun that is not pronounced.
>
> (Holmberg 2005:537-38)

Holmberg points out that in pre-Minimalist NSP theories, the empty category *pro* licensed and identified by verbal morphology was responsible for carrying the subject theta-role, bearing nominative Case, and satisfying the Extended Projection Principle. On the contrary, following *Hypothesis A*, if Agreement is interpretable and therefore referential, it can check the EPP and carry itself the subject theta-role. If Agreement also absorbs nominative Case, there is no need for the empty category *pro*. This in principle seems like an elegant solution for EPP licensing in consistent NSLs.

However, there are other null and overt subject profiles that a theory on the NSP should be able to account for. In this respect, Holmberg notes that if there were a partial NSL in which null thematic subjects and overt expletives co-existed, *Hypothesis A* could not deal with it. Recall from the discussion in §2.2.2.1 above that Finnish is such a language. This leads Holmberg to propose that *Hypothesis B* arises as the correct formulation of the NSP, at least for partial NSLs.

The gist of Holmberg's analysis of Finnish can be summarized as follows: he claims that expletives are merged overtly in Finnish as a result of a language-specific EPP formulation. Specifically, if the sentence contains a referential category, Spec, TP must be filled overtly either by a referential category, or by an expletive pronoun (cf. (20)-(22) above). The fact that adverbials and arguments can be merged in Spec, TP shows that this position in Finnish is not a subject position *per se*, but it constitutes essentially a Topic position, as it can be filled by those referential categories (Holmberg and Nikanne 2002).

Holmberg's minimalist NSP theory makes two main contributions to our understanding of subject licensing cross-linguistically: first, it constitutes an account of partial null subject phenomena that were difficult to deal with under and Government and Binding approach to parametrization; second, it shows how theories under the approach depicted by *Hypothesis A*, such as Alexiadou and Anagnostopoulou (1998), in which agreement is proposed to be interpretable, make the wrong predictions for languages with partial null subject profiles.

Holmberg notes, however, that it might be the case that *Hypothesis A* could accurately account for consistent null subject phenomena, if agreement were parametrized, and thus uninterpretable in some languages and interpretable in others. In that case, the null subjects in languages like Finnish and the ones found in consistent NSLs would stem from different sources, EPP-checking via subject movement to T in the former and strong, interpretable agreement morphology in the latter. Both strategies would yield the same surface effect, in the form of a null subject.

NSP variation in Spanish presents a suitable venue for testing the power of *Hypothesis A* and *Hypothesis B* respectively to deal with consistent and non-consistent NSLs. As it will become clear in the discussion below, accounts along the lines of both hypotheses have been put forward in the recent literature on Spanish syntax. The study of dialectal variation, and particularly the study of varieties with mixed null subject patterns like DS, has been crucial to evaluate the empirical coverage of *Hypothesis A* and *Hypothesis B* accounts.

2.3.4. Non-NSL and NSL Spanish.

In general, Minimalist accounts of null and overt subject licensing in Romance maintain Chomsky's (1995) idea that the Extended Projection Principle reduces to the morphological properties of Inflection, and that the licensing of null subjects in consistent NSLs such as Spanish is –at least to some extent- a function of 'rich' agreement. However, the differing approaches taken for explaining the nature of agreement in each of these accounts make very different predictions with respect to the structural positions of overt subjects, as well as with regards to the word orders

that are signaled as marked or unmarked in Spanish: In the accounts on Spanish subjects developed along the lines of *Hypothesis A* above (Contreras 1991; Olarrea, 1996; Ordóñez 1997; Alexiadou and Anagnostopoulou 1998; Ordóñez and Treviño 1999; Kato 1999; Barbosa 1995, 2009), which propose that Spec, TP is not necessarily projected in Null Subject Spanish, overt subjects are proposed to belong in Spec, CP. Note that placing the locus of overt subjects in Spec, CP crucially entails that the unmarked word order in consistent Romance NSLs is not SVO, as in traditional NSP theories such as Rizzi (1982, 1986), but VSO, and that overt subjects in NSLs and non-NSLs are of a different nature. On the other hand, accounts along the lines of *Hypothesis B* above (Belletti 1998; Zubizarreta 1999; Goodall 2001; Ortega-Santos 2008) take Spec, IP to be the structural position of Spanish overt subjects and propose that overt preverbal subjects are essentially of the same nature in NSLs and non-NSLs, and that SVO constitutes the unmarked word order in Spanish.

As explained by Rizzi (1997), CP and TP are the locus of two distinct clausal phenomena: while the morphosyntactic features of the verb and the propositional content of the clause belong in IP, the CP field acts as the interface between this propositional content in the IP and the superordinate structure and/or the previous discourse. Consequently, locating the default position of subjects in each of these domains has important consequences for a general theory of subject licensing in Spanish.

In addition to characterizing the structural position of subjects, both *Hypothesis A* and *Hypothesis B* accounts need to deal with the fact that constraints beyond morpho-syntax have a role in null and overt subject licensing in Spanish. It is well known that null and overt subject insertion is also governed in this language by information structure constraints at the syntax-pragmatics interface, such as Focus or Topichood, as well as by constraints related to the lexical properties of subjects, such as the type of predicate involved in a given structure. For instance, unaccusative predicates are more likely to bear postverbal subjects (Burzio 1986), and certain psychological predicates, such as *gustar* 'to like' typically bear postverbal (logical) subjects.

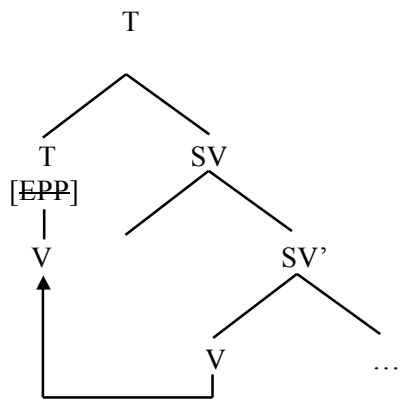
To sum up, any theory on subject licensing in Spanish is required to provide answers for the following questions: what is/are the mechanism(s) for null subject licensing? What are the structural positions for null and overt subjects respectively? In which syntactic position belong overt preverbal and overt postverbal subjects and what is their nature with respect to semantic/information structure? Ideally, a theory on subject licensing should be also able to deal with the attested cross-dialectal differences in null and overt subject insertion. In fact, the study of varieties like DS has become key in evaluating the empirical coverage of the theories on subject licensing recently proposed in the recent literature. In Minimalist theory, syntactic operations crucially rely on the feature composition of the lexical items that enter the computation; for this reason, the analysis of the feature assembly of both overt and null pronouns constitutes the point of departure of both *Hypothesis A* and *Hypothesis B* accounts of Spanish subjects.

2.3.5. On the feature composition of Spanish pronouns.

2.3.5.1. Null Subject Spanish.

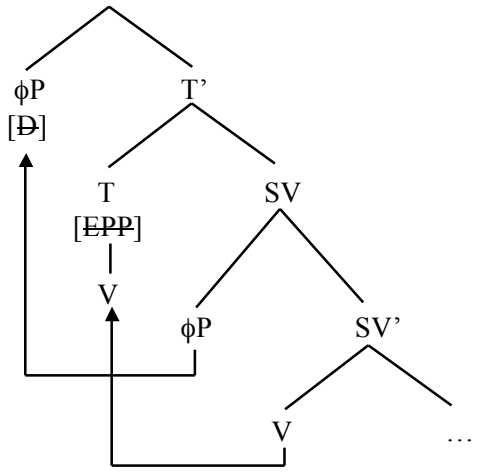
In a *Hypothesis A* theory such as Alexiadou & Anagnostopoulou (1998) verbal morphology is attributed pronominal nature. Consequently, Romance NSLs include in their Inflection a nominal element [+D], interpretable phi-features, and Case. In other words, agreement is not only interpretable in these languages, as pointed out above, but it also has the same status in terms of feature composition and semantics that pronouns have in a non-NSL like English, and is able to check the EPP feature on T. The operation by which agreement satisfies the EPP in Null Subject Romance is verb movement, as represented in (42):

(42)



On the contrary, in *Hypothesis B* theories along the lines of Holmberg (2005), in which agreement morphology is uninterpretable, null pronouns are taken to have different properties in terms of their feature composition and semantics. In Holmberg's view, for instance, null subjects are 'phi-phrases', ϕPs in his notation. These ϕPs are bundles of interpretable phi-features with no PF form that serve to value the uninterpretable (agreement) phi-features on T. The null subject needs to raise to Spec, TP because while bearing phi-features, it lacks the deictic/referential [D] feature and hence it cannot refer. This movement operation allows the null subject to incorporate referential value. The relevant representation is provided in (43):

(43)



Sheehan (2006), contra Alexiadou and Anagnostopoulou (1998), submits that Holmberg's theory is best suited to explain null subject phenomena not only in Partial NSLs, but also consistent Romance NSLs. It is her contention that if the position of subjects is the same in NSLs and in non-

NSLs, null pronouns value the uninterpretable features in T essentially the same way overt pronominals do in non-NSLs like English, and the fact that they are not spelled out is a PF matter. As PF deletion is an operation required in the system for independent reasons, and it is relevant for instance in VP ellipsis phenomena or in copy deletion after movement, it is theoretically justified.

As far as the specific way in which PF deletion of potentially null subjects takes place, according to Sheehan pronouns are not spelled out due to the universal PF economy principle proposed in Roberts (2004), which avoids the duplicate spelling out of identical feature bundles in multiple sites. Roberts argues that subject pronouns in Spec, TP are deleted when the featural content of the pronoun is identical to that of T and the pronoun Agrees with the T head:

(44) Deletion under feature-identity: (Roberts 2004:2)

α deletes under identity of features with ß only if ß Agrees with α.

This Agree relationship also checks the pronoun's [Nominative] feature, which results in featural identity between the specifier and its head. Thus, referential pronouns have the option of deletion when T contains a [uD] feature and it agrees with its specifier in phi-features. The formalization in (44) predicts that pronouns can fail to be spelled out as long as the EPP feature of T is deleted. It also predicts that null subjects occupy a preverbal position. For a non-NSL such as French the proposal would be that T lacks an uninterpretable D feature and so featural identity between the subject and T does not obtain, despite the fact that agreement between T and the DP takes place, and that the verb, which displays agreement, raises to T at PF. The NSP is then formalized thus:

(45) The Null Subject Parameter: T is +/– uninterpretable (Roberts 2004:2)

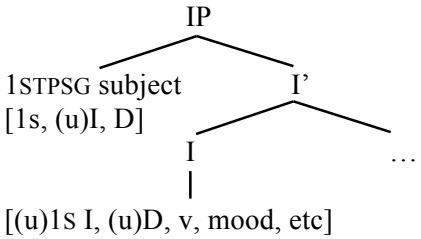
A deletion account like the one proposed in Roberts (2004) has the advantage that it strongly constrains syntactic variation between languages, while accounting for the diversity found in Romance. However, Roberts (2004) crucially relies on the notion of 'rich agreement', since the [uD] in T is in his theory is in fact the minimalist formal reflection for the notion 'strong' or 'rich' inflectional morphology contained in pre-mininalist NSP theories. For this reason his account could

be rejected arguing that this is a concept that has resisted a formal definition within generative linguistic theory since the eighties (cf. §2.2.1.). Holmberg (2005, 2009) presents a possibility to overcome this problem: he proposes that the D feature present in the agreement morphology of 'rich' agreement languages cannot be the result of the specific PF content of said inflectional morphemes. Recall that under Chomsky's (2001) theory on feature checking as feature valuation, uninterpretable features enter the derivation unspecified and get assigned a value as part of the feature checking process. This has the direct implication that in this theory of feature checking morphology happens after syntax. Therefore it is not clear that the phonetic representation of rich agreement could have any effect on syntactic operations. Holmberg's idea is that it might be the case that agreement morphology is actually a surface effect of a [uD] feature in I, and in that sense not the trigger for acquisition of syntax, but rather an aid to it. Holmberg's view on feature checking theory is very relevant, since it leaves open the possibility that there might be no one-to-one correlation between morpho-phonological richness and the [uD] feature itself[12].

Sheehan (2006) follows Holmberg (2005) and departs from Roberts (2004) in this respect, and she dissociates the uninterpretable D feature on T from the richness of morphological agreement, avoiding in this way the problem of the elusive nature of 'rich' inflectional morphology. She assumes Robert's definition of *featural identity*, that she relabels *feature non-distinctness* and to which she incorporates Holmberg's view on agreement morphology. The structure in which feature non-distinctness obtains is illustrated in (46), in which the overt subject is a first person pronoun. In her theory, it is the specifier that deletes rather than the head, as the head contains more information:

[12] Note that Holmberg's (2009) idea on the possible non-correlation of the presence of a [uD] feature with the presence of 'strong' agreement morphology has very relevant implications for accounts of diachronic change in null and overt subject patterns in Romance. Traditionally, the shift of a given language from NSL to non-NSL has been taken as a direct consequence of loss of inflectional morpholoy (Roberts' 1993 study on French for instance). Holmberg's view opens the possibility of an explanation of the shifts in NSP patterns in a given language that does not take changes in inflectional morphology as the only trigger for changes in NSP profiles. This issue will be discussed in detail in subsequent chapters.

(46) Feature non-distinctness (Sheehan 2006:224)



In sum, the theory on feature identity/feature non-distinctness formalized in (44) and (46) accounts for Romance null referential subjects, while also deriving SVO as the default word order in consistent NSLs. Nonetheless, the theory has two other crucial cases to deal with: on the one hand, the cases in which null non-thematic pronouns are compulsory, and the contexts in which overt pronominal referential subjects are equally categorical in Romance NSLs, on the other. The first case is instantiated by null expletive structures, and the second by Focus constructions.

Sheehan (2006) accounts for focalized mandatory overt pronouns proposing that Spell Out of a pronominal subject will be forced where pronouns bear an interpretable [Focus] feature, or where deletion would yield ambiguity. This is predicted by the rule on *featural identity/feature non-distinctiness* in (46), since the overt pronouns resulting from Focus bear additional features not matched on T. Consequently, all instances of overt pronouns in Romance NSLs are attributable to the failure of deletion because of lack of non-distinctness.

As far as NSL expletive constructions are concerned, Sheehan proposes expletive deletion in Romance is also subject to feature non-distinctness as formalized in (46), i.e., null expletives are deleted under non-distinctness with T, just like null referential pronouns. This is so because the PF-features of expletives are a subset of those in T, which with it agrees in person only. To account for the null expression of expletives in Romance versus their overt expression in non-NSLs like English, Sheehan proposes a parametrized PF-interpretability condition on T. This parametrized

condition regulates the requirement or lack thereof for Spec, TP to be spelled out at PF. Sheehan formulates it as variant of Roberts and Roussou's (2001) * parametrized property[13]:

(47) PF interpretability parameter

T is +/- * , where * means that Spec, TP must be spelled out at PF.

Deletion of all expletives is therefore permitted in languages lacking a PF-interpretability condition on T. Although expletives have an LF interpretation not shared by T, at PF they realize only (optionally) Case and third person phi-features.

Consider that as expletives lack the D feature contained in null referential subjects, expletive deletion is predicted to be possible even in languages that lack the uninterpretable [uD] feature on T. That is to say, this theory predicts the existence of languages that lack referential null subjects while allowing for null expletives. This is the case of the so-called 'null expletive languages' discussed in § 2.2.1. Among these languages are German, Icelandic, and a wide range of Creole languages. This is a welcome result, since these languages represented an issue for pre-minimalist NSP theories[14].

Sheehan's (2006) formulation of the NSP(s) would therefore be as in (48):

(48) The Null Subject Parameter:

(i) *The rich agreement parameter*

T lacks/bears a [uD] feature.

(ii) *The PF-interpretability parameter*

T lacks/bears a *, where * requires the specifier of I to be spelled out at PF.

The *Hypothesis B* theory in (48) can account both for consistent NSLs and for Null Expletive languages. It also shows in which specific ways Minimalist NSP theories allow for wide

[13] * is conceived in Roberts and Roussou's (1997, 2001) as a "parametrization operator" which forces the morphophonological realization of functional material via Move or Merge. * is just a randomly choosed diacritic to represent this parametrization operator. For the full development of this idea, see Roberts and Roussou (1997).

[14] For a description of the (micro)variation found among null expletive languages with respect to the specific types of expletives they allow for and the interaction of null expletive constructions with other language-specific factors, see Sheehan (2006) for German and Icelandic, and Sheehan (2006), Nicolis (2004) for Creoles.

empirical coverage, at the cost of formulating small-scale parameters. As will become clear in the ensuing discussion on dialectal variation in Spanish, Minimalist theory, with its fine-grained distinction of the specific features that drive syntactic computation, is also particularly well suited to account for variation among closely related grammatical systems. In fact, differences in feature assembly are at the core of recent accounts of subject deletion and insertion in DS; accounts along the lines of both *Hypothesis A* (Martínez Sanz 2007) and *Hypothesis B* (Toribio 2000, 2001; Camacho 2008) have been put forward to explain the Dominican subject grammar. In what follows we present the advantages and shortcomings of each of them.

2.3.6. Dialectal variation and pronominal feature assembly: Dominican Spanish.

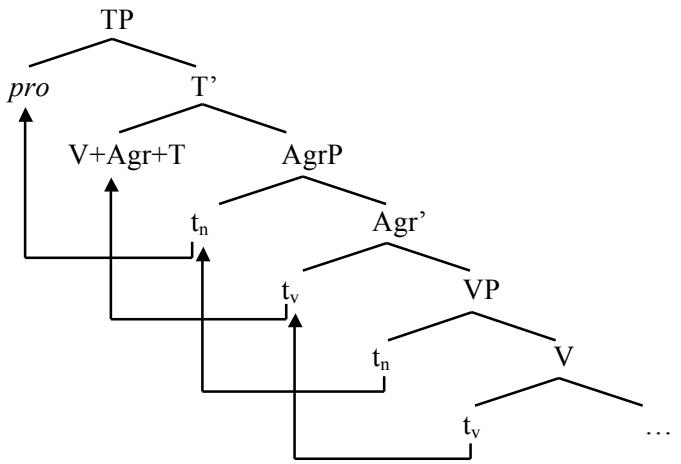
2.3.6.1. Toribio (2000, 2001).

Toribio (2000, 2001) takes as a point of departure Henry's (1995) idea that dialectal variation can be accounted for within parametric limits, and that consequently cross-dialectal differences may mirror cross-linguistic variation. Also building up on Kroch's (1989, 1994) *Competing Grammars Hypothesis*, Toribio puts forward a proposal to explain the co-existence of NSL and non-NSL patterns in DS. In her theory, the differences between DS and Mainland Latin American Spanish are derived from the differences in the feature composition of the functional nodes involved in subject licensing, namely Tense (T) and Agreement (Agr). In standard minimalist theory, while T checks the nominative Case feature on the subject DP and the tense on the verb, Agr checks the phi-features on the verb and its arguments. As a theory that is framed in early minimalism (Chomsky 1995), cross-dialectal variation is attributed to the 'strength' of the morpho-syntactic features of the relevant functional nodes. In Chomsky (1995), while 'weak' morphological features are readable by PF, 'strong' morphological features must be checked and deleted in the course of the derivation for it to converge.

Toribio's proposal is that in Mainland Latin American Spanish verbal T and Agr features are strong; as far as nominal features in these projections are concerned, while nominal Agr features

are strong, nominal T features are specified as weak. Null subject licensing in Mainland Latin American Spanish is depicted in (49):

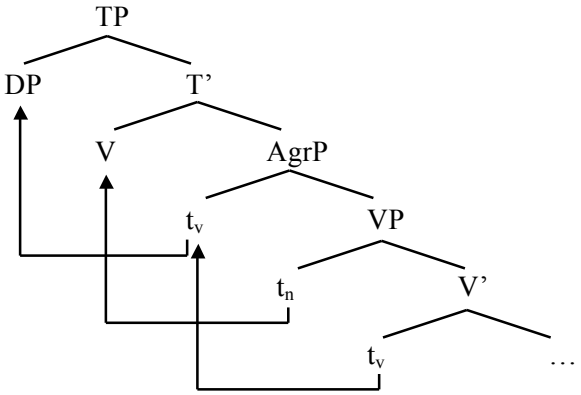
(49) Mainland Latin American Spanish, with *pro* subject:



In (49), the verb moves in order to check and delete the strong verbal features of T and Agr. Since the nominal features of Agr are also strong, verb movement produces the [Tense+Agr] complex required for the licensing of null subjects. The null subject in turn moves to the specifier of Agr, and with this operation the strong nominal features of Agr get checked and deleted; from this position the subject moves to Spec, TP; this movement is not related to feature checking and deletion, since nominal T features are specified as weak, but it occurs, according to Toribio, as a 'last resort' strategy to comply with 'the *pro* theorem' (Toribio 2000:335).

In DS, however, two possible feature specifications are possible, each of them corresponding to each of the two competing grammars available in the mind of the Dominican speaker: grammar A would parallel the Mainland Latin American Spanish grammar, while in grammar B Agr may bear weak nominal features, and T may bear strong nominal features. The feature specification of Agr in grammar B entails that, on the one hand, the DP subject does not raise to Spec, AgrP, and, on the other, that null subjects cannot be licensed. Furthermore, the strong nominal features of T trigger subject DP raising to Spec,T. Finally, verbal Tense and Agreement features are always strong, as is the case in grammar A, which entails that main verbs in this variety also raise to Agreement and Tense before Spell-Out. This is represented in (50):

(50) Dominican Spanish, with weak nominal Agreement features and strong nominal Tense features:



While this theory is able to explain structurally the two possible null and overt subject patterns displayed by DS, it also presents the same shortcomings that *Competing Grammars* theories generally show and that were discussed in Chapter One; in short, while it definitely seems to be the case that when language change is taking place different structures that convey the same meaning are available to speakers, it is not as clear whether this necessarily entails that they should correspond to differing grammars in the mind of speakers; it might be the case that these competing structures correspond to a single, and for a given period of time even stable, grammar. A different proposal that aims to overcome this challenge is provided in Martínez-Sanz (2007).

2.3.6.2. Martínez Sanz (2007).

This hypothesis takes as a point of departure the fact that overt expletive constructions are available in DS. As mentioned above, Urigereka (1995) and Silva-Villar (1998) already pointed out that these constructions are available in a number of Western Romance varieties (cf. examples (23)-(27)). As explained above, while these authors propose that all Romance overt expletives are inserted for discourse-related reasons, in my previous and collaborative work (Martínez-Sanz 2007; Martínez-Sanz and Toribio 2008), substantial evidence has been provided to show that Dominican expletives belong in the core syntax and are probably merged in the same position as overt non-expletive pronouns. As already noted above, Transitive Expletive Constructions (TECs) are not a possibility in DS (cf.(29)); the fact that an expletive pronoun and an overt thematic subject cannot

co-occur in DS strongly suggests that Dominican expletives are not merged in a discourse-related position like Spec, CP, but rather in a position akin to Spec, TP. The contrast between other Romance varieties and DS with respect to TECs is shown in (51a-b), which repeat for ease of discussion the examples provided in (28a) and (29) above:

(51) a. *Ele* os lobos andan con fame. (Galician; Silva-Villar 1998:269)
EXP the wolves go.3PL with hunger
'Wolves are hungry.'

b. *Ello yo no sé por qué mi papá me puso Almeida. (DS; Toribio, p.c)
EXP I not know why my dad me-called.3SG Almeida
'I don't know why my dad named me Almeida.'

In addition, overt expletives in DS are also displayed in embedded contexts, while they are banned in this same position in other Western Romance varieties, which again strongly suggests that Dominican expletives are not merged in Spec, CP:

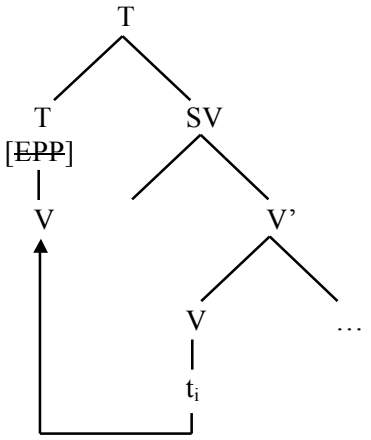
(52) a. *Xa sei que el chove. (Galician; Silva-Villar 1998:236)
Of course know.1SG that it rains
'Of course I know that it rains.'

b. Dicen que ello está lloviendo. (DS; Martínez-Sanz and Toribio 2008:2)
say.3PL that EXP is raining.
'They say that it's raining.'

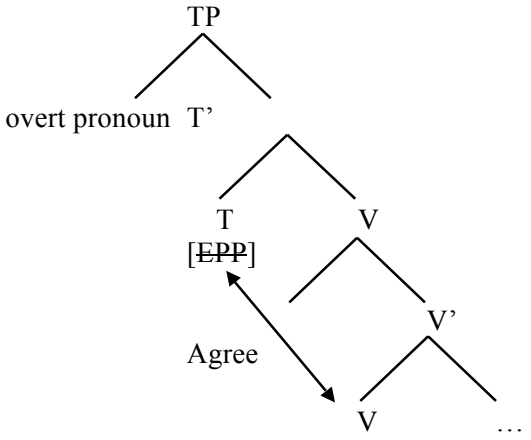
Therefore, it does not seem to be the case that merging an expletive triggers any type of discourse effect in DS. It is the contention in Martínez-Sanz (2007) that the merging of both expletives and personal overt pronouns stems from the particular instantiation of the NSP represented by this variety. Building up on Alexiadou and Anagnostopoulou (1998), it was argued that the differences between these two varieties could be explained as follows: in Non-Caribbean Spanish it is possible for the pronominal features of the verb (i.e. the strong D-features on Agr) to check the EPP on T via V-raising. In other words, the V-raising operation is crucially related to EPP

checking in this language. In DS, where V to T movement is no longer an option, the EPP feature on T remains unchecked by the time TP is merged. For this reason, an expletive is merged in Spec, TP in this variety as a 'last resort' strategy to avoid the derivation to crash. It was assumed that an Agree relation that checks the uninterpretable features of V and T can proceed independently of EPP. The differences between Non-Caribbean and Dominican Spanish are illustrated in (53):

(53) Non-Caribbean Spanish



Dominican Spanish



This theory accounts for the Topic-like nature of overt subjects in DS. It also correctly predicts that if overt pronominal subjects must be preverbal in Dominican Spanish, specific word order patterns should arise to comply with this condition. This prediction is borne out, as DS is usually described as displaying word orders in interrogative and infinitive constructions in which subjects are typically preverbal, as opposed to what happens in other European and Latin American varieties of Spanish, as thoroughly discussed in Chapter One. This account, however, also presents some shortcomings: while the claim that V-to-T does not occur in DS can be sustained to some extent in the fact that the Dominican inflectional paradigms are substantially less 'rich' than their counterparts in Non-Caribbean Spanish, the existence of languages in which null subjects are licensed and verb raising does not occur could pose a problem for this theory. In addition, in a language like French main verbs do raise (Pollock 1989) and null subjects are not possible. In sum, relative positioning of adverbs and negation in DS would need to be thoroughly in order to confirm whether V-to-T is an option in DS.

Neither Toribio (2000, 2001) nor Martínez-Sanz (2007) elaborate on the possible sources of the DS pattern of overt subject insertion and deletion. Recently, Camacho (2008) has taken on this issue. In doing this he has provided an explanation on the Dominican subject system in which the insertion of overt pronouns is attributed also to their particular feature assembly, but with the difference that he does not resort to grammars in competition to explain the DS facts. The specifics of his account are presented in the following section.

2.3.6.3. Camacho (2008).

Camacho (2008) compares DS and Brazilian Portuguese (in what follows, BP) facts. He notes that in BP studies (Kato and Taralho 1984, Taralho 1986, Duarte 1992, 1993, 2000, Kato and Negrão 2000) the widespread use of overt pronominal subjects is not only attributed to an erosion of the inflectional paradigm; rather, a reorganization of the pronominal paradigm is also taken as critical in the extension of overt subjects to the contexts that would feature null pronouns in consistent NSLs. Specifically, it is claimed in these works that it is the reduction of the set of personal pronouns what has triggered the loss of functional distinctions that have resulted in high rates of overt pronouns, and not the other way around.

Camacho builds up on this idea and suggests that the frequency of use of certain pronouns might have had an impact in the transformation of the pronominal system in varieties like BP and DS. He further proposes that they key feature specifications that have shifted in these pronominal systems are related to the feature [referentiality], on the one hand, and to the feature related to the nature of a given pronoun, on the other. In this latter respect, Camacho follows Cardinaletti and Starke (1999), who establish a tripartite typology of pronouns in natural languages: strong, weak and clitic pronouns. Both strong and weak pronouns may appear as subjects in a given language (the Italian pronouns *lui* and *egli*, strong and weak respectively, being a case in point), and they are syntactically, semantically and prosodically distinct. Basically, weak pronouns are 'defective', in a

sense that needs to be specified, as compared to their strong counterparts[15]. Camacho's idea is that frequency of use and the erosion of the inflectional paradigm have contributed to DS and BP pronouns to be specified as [weak]. Bearing a [weak] feature implies a requirement for the pronoun to be attached to a head, and a syntactic requirement that this head must be Inflection.

Subject pronouns in consistent NSLs such as Null Subject Spanish are strong, and consequently specified as [+referential]; they bear no [weak] feature, as represented in (54) for a null subject pronoun and an overt third person singular pronoun:

(54) Null Subject Spanish

a. *pro* {α person, α number α ref }

b. *ella* {3rd person, sg, +ref }

If an overt strong pronoun entered the derivation in Non-Caribbean Spanish, two [+referential] features would be present in the computation: a [+referential] feature in Inflection and a [+referential] feature in the pronominal; since the operation Agree must be established between [+] and [–] versions of a feature, it would not be possible in this context. For this reason, the pronoun appears in its null version and the derivation contains a single [+referential] feature. For an overt pronominal to be felicitous in Null Subject Spanish, it would need to carry additional features, related with the information structure of the clause, such as contrast or emphasis.

Overt pronouns in BP and DS, on the contrary, are specified as [-referential] and [weak]. Bearing a [-referential] feature makes the pronoun a potential semantic variable, distinct from a [+referential] item, which is potentially anaphoric. However, [+anaphoric] pronouns are also available in DS, since null subject Non-Caribbean Spanish-like structures are available in this variety as well. This leads Camacho to propose that in DS there are two independent lexical entries for pronouns, one with the [-referential] feature and one without it, as represented in (55):

[15] The issue of the distinction between strong and weak pronouns will be revisited in §2.3.4.4. when discussing interrogative structures in Spanish, and relevant examples of each type of pronouns will be provided. For a thorough discussion of the differing properties of weak and strong subject pronouns in Italian and other Romance languages, see Cardinaletti and Starke (1999), Cardinaletti (2004). For the present discussion it suffices to point out that weak pronouns are 'defective' as compared with strong pronouns and that this is reflected in their feature specification.

(55) Dominican Spanish

a. *pro* {α person, α number -ref }

b. *ella* {3rd person, sg, weak, +ref }

c. *ella* {3rd person, sg, weak, -ref }

Postulating two different lexical entries for pronouns is theoretically justified on Cardinaletti and Starke's (1999) tripartite classification of pronouns. Furthermore, it is empirically supported by Spanish facts: note that the availability of two lexical entries for pronouns is also found in Non-Caribbean Spanish; specifically, third person plural pronouns can be specified either as [+referential], in specific readings, or as [-referential], in arbitrary interpretations that convey an impersonal meaning. Relevant examples and the feature specification of the pronouns are represented in (56)-(57):

(56) a. (Ellos) trajeron el vino a la fiesta.
they brought.3PPL.SPECIFIC the wine to the party
'They brought the wine to the party.'

b. *-n* {3rd person, pl, +ref }

(57) a. Dicen que beber mucha agua es bueno para la salud.
they say 3PPL.ARBITRARY that to drink a lot of water is good for the health
'They say that drinking a lot of water is good for your health.'

b. *-n* {3rd person, pl, -ref }

According to Camacho, in a variety like DS once the pronominal loses its exclusively [+referential] interpretation and a separate, [-referential] lexical is created, overt pronouns are able to be merged in Spec, TP and satisfy the EPP like in English-type languages. He further argues that once the weak pronominal is able to check the EPP in T, the verb's inflection loses its pronominal nature. This shift is represented in (58):

(58) higher frequency → [± referential pronoun → pronoun merged in Spec, TP → pronoun satisfies EPP, becomes weak, Inflection becomes [-ref].

Camacho also notes that there are relevant interpretative differences among overt pronouns that bear a [+referential] feature, which makes them potentially anaphoric, and [-referential] pronouns. Montalbetti (1984) first formulated these differences in the *Overt Pronoun Constraint*. This *Constraint* states that whenever there is a contrast between overt and null pronouns, the overt pronoun cannot be bound as a variable:

(59) a. Todo estudiante$_i$ cree que *pro*$_i$ es inteligente. (Non-Caribbean Spanish)
every student thinks that *pro* is intelligent
'Every student$_i$ thinks that he$_i$ is intelligent.'

b. Todo estudiante$_i$ cree que él$_{*i/j}$ es inteligente.
every student thinks that he is intelligent
'Every student$_i$ thinks that he$_j$ is intelligent.'

In the varieties in which pronouns are not exclusively linked to [+referential] readings, such as BP and DS, we would expect for the Avoid Pronoun Constraint to not apply. This is in fact the case; in embedded clauses, pronouns that are coreferential with matrix subjects can be expressed overtly in BP, as noted by Duarte (2000) and exemplified in (60a), and in DS, as pointed out by Toribio (2000); these examples are also found in the corpus gathered for this study, as illustrated in (60b):

(60) a. Você, no Canadá, você pode ser o que você quiser. (Duarte 2000:191)

b. Tio Papi$_i$ murió tan feliz que *pro*$_i$ no sabe que él$_i$ murió.
Tio Papi died.3PSG so happy that *pro* not know.3PSG that he died
'Tio Papi died so peacefully that he doesn't even know that he's dead.'
(Evangelita, UFH1)

The proposal in Camacho (2008) has the significant advantage that it dispenses with postulating two grammars in competition in the minds of DS and BP speakers. The co-existence of null and overt subject structures is accounted for by availing of two separate lexical entries for at least a subset of the pronouns in the paradigm, a possibility that is already attested in the grammar

of Spanish for third person plural pronouns, and in other Romance languages like Italian, in which distinct weak and strong pronouns show up as phonetically distinct (*lui* versus *esso*).

Camacho (2008) addresses only in passing the issue of what is the assigned position in the structure of the clause for different types of pronouns, proposing that weak Spanish pronouns are merged in Spec, TP. He leaves open the question of which structural position is the merging and/or landing site for overt preverbal pronouns in Non-Caribbean Spanish. In his proposal, however, it is implicit that overt pronouns in Caribbean and Non-Caribbean Spanish should belong in different subject positions in the clausal structure, as a logical consequence of these pronouns bearing different features in each variety.

The issue of the structural positions of subjects in Spanish is to this day the object of intense debate in the theoretical syntax literature. Questions related to this issue are for instance, what is the position of overt preverbal and overt postverbal subjects, or what is the nature and assigned position of null subjects. The latter question has been discussed in this section. In what follows, the proposals put forward in the recent literature regarding the position of overt preverbal and postverbal subjects in Caribbean and Non-Caribbean Spanish are presented.

2.4. On the structural position of subjects in Spanish.

As is the case for other Romance languages, overt Spanish subjects have been proposed to belong either in Spec, CP (Contreras 1991; Olarrea, 1996; Ordóñez 1997; Ordóñez and Treviño 1999; Barbosa 1995, 2009, among others), or in Spec, TP (Belletti 1998; Zubizarreta 1999; Goodall 2001; Ortega-Santos 2008, among others). As noted by Camacho (2008), several questions are conflated in the debate on the default structural position of subjects in Spanish, among them whether there is a uniquely determined position for preverbal subjects, and whether this position is uniquely associated with specific informational content, such as Topichood.

In recent years, specific facts from both Caribbean and Non-Caribbean Spanish have been proposed to suggest that Spanish overt subjects can present differing properties, in terms of their A or A-bar properties and the informational content they carry. This is turn suggests that more than

one subject position must be available in the Spanish grammar, or even that there are landing positions available for subjects in both the CP and the TP domain (Sheehan 2006, Suñer 2003, Camacho 2006, among others).

2.4.1. Overt preverbal subjects belong in Spec, CP.

Proponents of the subjects-in-CP theory take as a point of departure the parametrized EPP proposal in Alexiadou and Anagnostopoulou (1998) account of Romance null subject licensing. Recall that in this theory verbal agreement was attributed pronominal nature, which gives this proposal the advantage of dispensing with the theoretically objectionable category *pro*, which has no interface value at PF or LF. This theory also makes two very strong empirical predictions: on the one hand, if the EPP is valued and deleted via verb movement and subject movement to Spec, TP is precluded, VSO is predicted to be the unmarked word order in NSLs. On the other, the canonical, A-position for subjects Spec, TP does not need to be projected, as there is no *pro*. Overt pronominal subjects, when they are merged, must occupy a different, A-bar position above TP. In other words, if this theory is correct, all Spanish overt subjects should have A-bar properties and they should be merged in Spec, CP.

Alexiadou and Anagnostopoulou (1998) submit that these two predictions are met. Their proposal is that in consistent NSLs the apparent SVO constructions are actually Clitic Left Dislocated structures in which the preverbal subject is merged in an A-bar position within the CP field. They find evidence for this claim in distributional and interpretational evidence facts involving preverbal subjects in consistent NSLs like Greek and Spanish.

Clitic Left Dislocated constructions (henceforth CLLD) involve a fronted phrase that is doubled by a clitic within the core sentence (Cinque 1990). CLLDed elements bring to the foreground old or given information in the universe of the discourse; the CLLDed constituent is set as a point of reference against which the main predication of the sentence is then construed (Raposo 1996). These constructions are frequent in consistent NSLs like Spanish or Portuguese with direct and indirect objects. When fronted, these objects are dislocated to a left-peripheral position in the

structure. This position has been argued to be within CP (Rizzi 1997), or adjoined to IP (Barbosa 1995). An example of an object CLLD construction in Spanish is provided in (61):

(61) A la hermana de Susana, la vi el otro día. (Camacho 2006:1)
to the sister of Susana, CL saw.1SG the other day
'Susana's sister, I saw her the other day'.

As far as the distribution of preverbal subjects is concerned, Alexiadou and Anagnostopoulou argue that in Greek and Spanish SVO orders there is not Spec, head relation between the subject and the finite verb, since adverbs may intervene between the preverbal subject and the verb. This would constitute evidence that the subject is not moved to Spec, TP:

(62) a. Greek
O Petros xtes meta apo poles prosphaties sinandise ti Maria.
Petros yesterday after from many efforts met Maria
'After many efforts, Peter met Mary yesterday.'

b. Spanish
Pedro ayer inesperadamente conoció a su profesor.
Pedro yesterday unexpectedly met to his professor
'Pedro met his professor unexpectedly yesterday.'

Interpretational effects involving Quantifier Phrases and Indefinites also suggest that preverbal subjects are CLLDed constituents in Greek and Romance NSLs. As shown in (63), CLLD objects have unambiguous wide scope:

(63) Kapjo pedito to eksetase kathe kathigitis. (A&A 1998:505)
some child CL-ACC examined every professor
'There is some child such that he was examined by every professor.'

Alexiadou & Anagnostopoulou claim that NSL quantificational elements and indefinites in a preverbal subject position also have unambiguous wide scope. Their reasoning goes as follows: postverbal subjects can have wide or narrow scope interpretations in consistent NSLs, as illustrated

in Greek (64b). The scope properties of a quantifier are preserved when it is moved to an A-position (May 1985). Therefore, if preverbal subjects were raised to a Spec, TP A-position, they would preserve their narrow scope interpretation. This is what happens in a language like English, in which subjects undergo movement to Spec, TP and consequently they have ambiguous scope, as shown by the gloss to (64a). However, in the Greek (64a) the indefinite *some student* in a preverbal position has necessarily wide scope over the universally quantified NP in object position. The subject in (64a) has the same scope properties as the CLLDed object in (61), which Alexiadou and Anagnostopoulou take as evidence that preverbal subjects are akin to CLLDed constructions in Greek:

(64) a. Kapios fititis stihiothetise kathe arthro. (A&A 1998:505)
some student filed every article
b. stihiothetise kapios fititis kathe arthro.

Preverbal Spanish subjects also have unambiguous wide scope readings:

(65) a. Algún/un estudiante archivó todos los artículos.
some student filed every the articles
'Some student filed every article.'

Another interpretational property typically ascribed to CLLD constructions is their failure to reconstruct (Cinque 1990). According to Alexiadou and Anagnostopoulou, indefinite preverbal subjects also pattern with CLLD constructions in this respect: if an indefinite preverbal subject is in an A-bar position, it cannot reconstruct and therefore it should have scope over a modal or negation. This is precisely the case, as illustrated with negation in (66), where the only possible interpretation is *many* > *not*:

(66) Poli andres dhen eroteftikan ti Maria.
Many men not fell-in-love Maria
'Many men did not fell in love with Maria.

Spanish replicates this same effect:

(67) Muchos hombres no se enamoraron de María.

Many men not CL fell.in.love of María

'Many men did not fall in love with María.'

To sum up, under Alexiadou and Anagnostopulou's approach, SVO orders in consistent NSLs are actually CLLD structures. Note that for this account to be empirically accurate it must be the case that *all* preverbal subjects are CLLDed. The evidence provided in recent work shows that this does not seem to be the case (Suñer 2003, Cardinaletti 2004, Camacho 2006, Sheehan 2006). Specifically, the comparative study of closely-related grammatical systems such as Non-Caribbean Spanish and DS has shown that while some Spanish subjects *can* appear in the Comp domain, it is not the case that they *must*.

2.4.2. Overt preverbal subjects belong in Spec, CP *and* Spec, IP.

Sheehan (2006) investigates null and overt subjects in Romance, with the aim of testing Alexiadou and Anagnostopoulou's (1998) theory against a number of data from consistent NSLs such as Non-Caribbean Spanish, European Portuguese and Italian. Her analysis of Romance data leads her to conclude that, while a number of preverbal subjects in these languages are CLLDed, at least some preverbal subjects seem to occupy an A-position in all of the languages she discusses. Sheehan finds three types of evidence that support her claim that some Romance subjects are in an A-position: pragmatic facts; syntactic facts; and facts regarding the nature of *pro*.

As far as pragmatic phenomena are concerned, she argues that preverbal subjects and CLLD constructions do not have the same informational value. If all preverbal subjects were CLLDed as argued by Alexiadou and Anagnostopoulou, they should have the same role in the discourse. However, Sheehan (2006) argues, while CLLD structures always bring to the foreground old or given information in the universe of the discourse, this is not the case for Romance preverbal subjects. Furthermore, null subjects in Romance are generally constrained by pragmatic factors: when ambiguity arises and the null subject referent cannot be identified by the inflectional

morphology in the verb, an overt preverbal subject is inserted. These preverbal subjects have no special discourse value as Topics or Foci. Rather, if there is ambiguity, the speaker looks for a salient discourse Topic, and if it is not found, the structure requires an overt subject. Again, a CLLD approach to overt preverbal subjects fails to allow for the presence of overt preverbal subjects that do not receive a Topic reading, like the ones in (68):

(68) a. María y yo llegamos a casa. (Yo) encontré las llaves…
María and I arrived.1PL at home. 1SG found the keys…

b. María y yo llegamos a casa. *(Yo/ella) tenía las llaves…
María and I arrived.1PL at home had.1SG/3SG the keys

Syntactic facts also show that preverbal subjects in Spanish are not always fronted topics (Goodall 2001). In Spanish, clauses with fronted topics are islands for extraction, whereas clauses with preverbal subjects are not:

(69) *¿A quién crees [que el premio se lo dieron]?
to whom you-think that the price DAT-CL ACC-CL they-gave
(Lit.'Who do you think that the price they gave to?'

(70) ¿A quién crees que Juan le dio el premio?
to whom you-think that Juan DAT-CL gave the prize
'Who do you think that Juan gave the prize to?'

It is not clear how a CLLD approach to preverbal subjects would explain this difference. On the contrary, an approach that locates the preverbal subject in an A-position does not face this problem. If the preverbal subject raised to an A-position then the embedded sentence in (69) has a different status from that in (70). Assuming that it is the presence of an A-bar element in the left periphery of (69) which blocks A-bar extraction of another, more embedded element, it follows that the preverbal subject in (69), as an A-element will not block A-bar extraction in the same way.

Finally, Sheehan (2006) notes that if as suggested by Alexiadou and Anagnostopoulou (1998) the unmarked word order in consistent NSLs is VSO, the properties of the null subject

should align with those of postverbal subjects. This does not seem to be the case. The investigation on pronoun typology in Cardinaletti and Starke (1999) shows that *pro* patterns with preverbal subjects. Recall that in this theory three types of pronouns are distinguished, strong, weak and clitics, each of them with different syntactic, semantic and prosodic properties. Taking as point of departure Italian facts, they describe *pro* as a weak pronoun. As such, it patterns in its distribution and semantics with defective pronouns, rather than with strong ones. Thus, *pro*, like weak pronouns, can refer to non-human referents, and it can act in specific contexts as an expletive, quasi-referential subject, or as an impersonal subject. In Spanish, evidence for the fact that *pro* patterns with preverbal subjects comes from the fact that floating quantifiers are possible with preverbal subjects or *pro*, but not with postverbal subjects:

(71) a. [Todos los chicos] se han comprado un coche.
All the boys CL-SELF have bought a car

b. [Los chicos] se han comprado todos un coche.
The boys CL-SELF have bought all a car

c. Se han comprado un coche [todos los chicos]
CL-SELF have bought a car all the boys

d. *Se han comprado todos un coche [los chicos]
CL-SELF have bought all a car

e. Se han comprado todos un coche
CL-SELF have bought all a car
'All the boys have bought themselves a car.'
(Sheehan 2006:91)

All these findings lead Sheehan to discard the idea that *all* preverbal subjects in Non-Caribbean Spanish are CLLDed and therefore occupy an A-bar position; while preverbal subjects and CLLD structures pattern alike with respect to interpretational effects related to narrow focus and quantifier phrases, other evidence seems to align them with elements that belong in an A-

position like Spec, TP. Therefore, if Spanish subjects have mixed A and A-bar properties, it can be sustained that they can occupy either of these positions.

Dialectal variation further supports this claim. Suñer (2003) specifically suggests that given that overt pronouns in Caribbean varieties such as DS are defective as compared to Non-Caribbean Spanish pronouns, probably as a result of the loss of morphological distinctions. She characterizes DS pronouns as weak, and submits that as a consequence they are limited in their distribution: while Non-Caribben strong subject pronouns like *él*/*ella* can be separated from the verb by a parenthetical, the Dominican overt expletive pronoun *ello* cannot.

(72) a. Juan/Él, a mi parecer, es muy simpatico. (Suñer 2003:5)
Juan/he, according to me is very nice
'Juan/he according to me is very nice.'

b. Ello (*a mi parecer) no sería malo estudiar
it according to me not would be bad to study
'It (according to me) wouldn't hurt to study.'

This suggests, according to Suñer, that *ello* is structurally closer to the verb than strong pronouns and lexical NPs, and therefore it is merged lower in the structure, as represented in (73):

(73) $[_{YP}$ {Juan/él} $[_{XP}$ parenthetical $[_{ZP}$ {ello} V]]]

Camacho (2006) finds further evidence to support the claim that Dominican pronouns are merged lower than strong pronouns in subject adjacency effects attested elsewhere in the Spanish grammar: in addition to the adjacency effects seen in (72) with the Dominican pronoun *ello*, preverbal negative quantifier subjects in both DS and Non-Caribbean Spanish must also be adjacent to the verb (74a-b), by contrast to lexical subjects, postverbal quantifiers, and CLLDed items (cf.75a-c):

(74) a. Nadie pudo avanzar tres metros
no-one could advance three meters
'No one could advance three meters.'

b. *Nadie casi pudo avanzar tres metros.

no-one almost could advance three meters

*Ninguno de ellos casi pudo avanzar tres metros

none of them almost could advance three meters

(75) a. La tortuga casi pudo avanzar tres metros

the turtle almost could advance three meters

'The turtle could almost advance three meters.'

b. Casi no pudo avanzar tres metros nadie

almost not could advance three meters no-one

'No one could almost advance three meters.'

c. A ninguno de los culpables, casi lo descubrieron.

to none of the guilty almost cl discovered

'None of the guilty, were almost discovered.'

To conclude, the evidence presented thus far suggest that while it seems to be the case that some preverbal subjects in Spanish are merged in Spec, CP, it cannot not be maintained that all of them are, as showed by pragmatic, syntactic and distributional evidence. In addition, dialectal variation also supports the claim that a Spec, TP position should be available for the landing of subjects in varieties like DS.

So far, I have discussed questions related to the position of Spanish subjects in default, SVO word orders. I now turn to discuss subject placing as related to information structure, since the informational content encoded by the subject also constitutes a very powerful constraint for subjects to surface pre- or postverbally. Consequently, notions related to the syntax-pragmatics interface such as Focus or Topichood largely determine subject position. Furthermore, informational content interacts with constraints at the lexicon-syntax interface, responsible for the structural properties of specific types of predicates.

2.4.3. On 'free' inversion: Subject positions and information structure.

'Free' inversion, in classical, Government and Binding NSP theories, (Rizzi 1982) was defined as one of the core properties associated with the availability of null referential subjects in a given language (cf.§2.2.1.). It was proposed that this property allowed referential subjects to appear either preverbally or postverbally.

However, the cross-linguistic research carried out over the years has convincingly shown, as summarized in Hulk and Pollock (2001), that while it seems to be the case that some inverted structures are typically associated with Romance NSLs, it is not the case that inversion is really 'free': on the one hand, there is not only a single type of inversion; rather, a number of distinct syntactic phenomena trigger VS word orders. On the other, specific types of inverted constructions are tightly related with information structure matters, and therefore the choice of a preverbal or postverbal subject is not optional or 'free'.

There is general consensus in the literature that Romance Inversion[16] comprises at least three distinct inverted structures (Hulk and Pollock 2001,Casielles-Suárez 2004, Sheehan 2006, López 2009, among others):

(i) The so-called Stylistic Inversion, responsible for subject-verb inversion in interrogative constructions and constrained by purely syntactic factors.

(ii) Locative inversion. These constructions, also syntactically constrained, feature a P(repositional) P(hrase) in preverbal position and the word order PPVS.

(iii) The so-called 'Free' Inversion, characteristic of Romance NSLs and constrained by the information structure of the clause.

[16] Hulk and Pollock (2001) explain that Romance Inversion is distinct from Germanic Inversion. The latter is seen in English questions and French subject-clitic inversion:

(i) Have you been feeling ill lately?

(ii) A-t-il dit cela?
has him.CL said.3PSG that?
'Has he said that?'

Germanic Inversion can be derived straightforwardly by I-to-C movement (Rizzi and Roberts 1989, Rizzi 1996), as is usually a root phenomenon and the verb is in complementary distribution with overt complementizers. In Romance, inversion does not display root/embedded asymmetries, which suggests that it does not clearly involve the CP domain.

The three phenomena are exemplified in (76), (77), and (78):

(76) ¿Qué dijo María?

what said.3PSG María

'What did María say?'

(77) En este hotel han trabajado muchas mujeres extranjeras.

In this hotel have worked.3PPL many women foreign

'Many foreign women have worked in this hotel.'

(78) a. Juan ha llamado.

Juan has telephoned

'Juan has called.'

b. Ha llamado Juan.

has telephoned Juan

'Juan has called.'

In all the structures in (76)-(78) the subject appears in the right periphery of the sentence, or preceding the object, but obligatorily following participles and infinitives (Hulk and Pollock 2001:3). The interrogative constructions in (i)-(76) have received a good deal of attention in the literature, and they present quite distinct patterns in Non-Caribbean and Caribbean Spanish; for this reason, they will be discussed in detail in section 2.3.4.4. below. The Locative Inversion constructions in (i)-(77) are not in principle of great interest for the work carried out in this dissertation so they will be left aside. In the remainder of this section attention will be devoted to the constructions in (iii)-(78), as they are key for the explanation of a number of issues related to subject placement in Spanish.

Roughly, while Romance preverbal subjects, unless in contexts of contrast or emphasis, bear by default a Topic reading, postverbal subjects signal Focus. However, matters are rather more complex than this, as evidenced by the extant research available on the finer properties of VS structures (see Rizzi (1998, 2000, 2004), Belletti (1998, 2001, 2004, 2005), for Italian; Kayne and

Pollock (2001) for French; Ambar (1992, Costa 2004) for European Portuguese; Zubizarreta (1998), Gutiérrez-Bravo (2002), Casielles-Suárez (2004), López (2009), among others, for Spanish. In this section I will only focus on the attested Spanish word order possibilities that are relevant to the present study. Before moving on to discuss the word order possibilities of Spanish, a clarification of the basic information structure concepts relevant to this discussion is in order.

According to Zubizarreta (1998:1), the Focus is "the non-presupposed part of the sentence". The presupposed part of the sentence is composed by the assumptions shared by speaker and hearer in a given discourse situation. Chomsky (1971), Jackendoff (1972) provided the classic answer/question tests to establish the focus of a proposition. All sentential constituents can be focused; of interest here are the so-called *wide (subject) focus* and *narrow (subject) focus* readings. For an utterance as *John ate the pie*, the question in (79) and the answers to it in (80a-b) and the associated presuppositions in (80a'-b') establish the two possibilities for the focusing of subjects:

(79) What happened?

(80) a. Something happened.

a'. There is a x such as John did x.

b. Someone ate the pie.

b'. There is an x such that x ate the pie.

The answer (80a) receives wide focus, as the whole proposition is new. In the answer (80b), the subject alone will receive narrow focus.

Contrastive focus is the other relevant concept for the discussion on Spanish VS structures. According to Zubizarreta (1998:7), "like emphasis, contrastive focus makes a statement about the truth or correctedness of the assertion introduced by its context statement." In (81), the subject *el sindicato*, 'the union' is contrastively focused:

(81) EL SINDICATO habló contra el gobierno (y no el partido).

The union spoke.3PSG against the government (and not the party)

'The UNION spoke out against the government and not the party.'

Contrastive focus assignment is relatively uncontroversial. It is generally accepted that it is assigned prosodically, and that while it can fall on any constituent *in situ*, there is a preverbal contrastive focus position in languages like Italian, Spanish and French. In the following two subsections the more controversial Spanish word order possibilities for marking wide and narrow focus are presented.

2.4.3.1. Wide focus in Non-Caribbean Spanish.

With transitive verbs, both SVO and VSO word orders are felicitous in wide focus contexts in Spanish (Zubizarreta 1998:125):

(82) a. ¿Qué pasó ayer?

What happened yesterday?

b. Ayer, Juan ganó la lotería.

Yesterday Juan won the lottery.

(83) a. ¿Qué pasó ayer?

b. Ayer ganó Juan la lotería.

Yesterday won Juan the lottery.

As far as intransitive verbs are concerned, while the VS order is more likely to appear with unaccusative predicates, as in (84), VS with a wide focus reading is also available in Spanish with intranstive verbs in which the subject is not assumed to be an underlying object, as noted by Zubizarreta (1998, Barbosa 2009) and exemplified in (85):

(84) a. ¿Qué pasó?

b. Llegó Juan.

(85) a. ¿Qué pasó?

b. Juan se rió.

b'. Se rió Juan.

However, it is not the case that Spanish allows VS order without restrictions. Rather, as proposed by Suñer (1982), the availability of this word order is constrained by the semantics of the

utterance: if intransitive predicates appear in presentational contexts, as in (84)-(85), VS is felicitous; however, if the semantics of the utterance is not presentational, VS is infelicitous in out-of-the-blue contexts, as in (86):

(86) *?Llegó tarde el tren.
arrived.3PSG late the train
'The train arrived late.'

Sheehan (2006) notes that the fact that unmarked VSO order is available in Spanish to convey wide focus might suggest that an analysis of NSL along the lines of Alexiadou and Anagnostopoulou (1998) is correct; if verb movement to I satisfied the EPP in SVO structures, the subject would stay *in situ* and Spec, TP would not be projected. However, as suggested by Zubizarreta (1998) and Ordóñez (2000), unmarked VSO in Spanish is available only where another XP or a clitic surfaces before the finite verb*:*

(87) a. ¿Qué pasó ayer?
b. Ayer ganó Juan la lotería.
b.' #Ganó Juan la lotería.

The examples in (87) suggest that on the one hand, that the Spec, TP position is projected in the structure, and on the other, that elements other than the subject can satisfy the EPP in Spanish. Sheehan (2006) proposes in this respect that the EPP, rather than the D-feature in Chomsky (1995), has the qualities of an edge feature, as in Chomsky (2001), i.e., a feature that manifests as a requirement for the specifier of a head to be filled. Her account of the Spanish data in (87) goes as follows: if the head T bears a (Topic or Focus) discourse feature, the EPP associates itself with this discourse feature, triggering movement of a Topic or Focus element to Spec, TP. If there is no discourse feature in T, the EPP associates itself with phi-features, and attracts the subject to Spec, TP. This contrast is illustrated in (88)-(89):

(88) a. ¿Qué pasó?
b. Juan ganó la lotería.

(89) a. ¿Qué pasó ayer?

b. Ayer ganó Juan la lotería.

In (88b) there is no Topic, and the whole sentence receives wide focus as new information; the EPP is satisfied by the subject, which agrees in phi-features with T and has its case feature valued. In (89b), on the contrary, a Topic is present. The EPP is therefore construed as an edge feature and triggers movement of the adjunct *ayer* to Spec, IP. Agreement and Case checking between the subject and T proceeds as in (88b), the only difference being that no subject movement takes place. If the Topic element occupies Spec, TP, the next question that arises is what is the position of the subject in this structure. In this respect Sheehan (2006) suggests that the most minimal analysis is to assume that the subject remains *in situ* in Spec, vP, and that Case is valued and deleted via Agree. Furthermore, the fact that elements other than the subject can occupy Spec, TP would explain why the literature on Spanish subjects, as mentioned above, has pervasively shown that Spec, TP in Spanish has mixed A and A-bar properties.

In sum, both SVO and VSO can convey wide focus readings in Spanish, and the choice of either word order is determined of the type of predicate involved, on the one hand, and the presence of Topic-like constituents in the structure, on the other.

2.4.3.2. Narrow focus in Non-Caribbean Spanish.

Narrow-focused subjects, as opposed to wide-focused elements, must surface postverbally in Spanish. Ordónez (1998, 1999, 2000) and Zubizarreta (1998) have both shown that only post-verbal subjects are felicitous in answer to *wh*-questions in Spanish. This is the case with all verb types:

(90) a. ¿Quién se rió?

Who himself laughed

'Who laughed?'

b. Se rió Juan.

Himself laughed Juan

'Juan laughed.'

b'. *Juan se rió.

Juan himself laughed

'Juan laughed.'

(91) a. ¿Quién llegó?

'Who arrived?'

b. Llegó Juan.

Arrived Juan

'Juan arrived.'

b'. *Juan llegó.

'Juan arrived.'

Furthermore, with intransitive verbs the subject must not only be post-verbal, but in fact must be sentence final. For this reason, VOS but not VSO (or SVO) orders are felicitous in structures featuring intransitive predicates:

(92) a. ¿Quién te regaló la botella de vino?

who CL.YOU gave the bottle of wine

'Who gave you the bottle of wine?'

b. Me regaló la botella de vino María.

CL.ME gave the bottle of wine María

'María gave me the bottle of wine.'

b'. *Me regaló María la botella de vino.

CL.ME gave María the bottle of wine

'María gave me the bottle of wine.'

b''. *<u>María</u> me regaló la botella de vino.

Maria CL.ME gave the bottle of wine

'<u>María</u> gave me the bottle of wine.'

The classical analysis of Spanish narrow focus is Zubizarreta (1998), who derives word order effects from constraints at the prosody/syntax interface. Her analysis goes as follows: she submits that all natural languages are subject to a *Focus Prominence Rule*:

(93) The Focus Prominence Rule (FPR)

Given two sister categories C_i (marked [+Foucs]) and C_j (marked [-Focus]), C_i is more prominent that C_j.

(Zubizarreta 1998:21)

In other words, the rule in (93) states that focused constituents are more intonationally prominent than unfocused constituents. Zubizarreta further claims that in both Romance and Germanic stress, i.e., intonational prominence, is assigned to focused constituents via a *Nuclear Stress Rule* (NSR):

(94) The Nuclear Stress Rule (NSR)

Constituent driven NSR- Given two sister categories C_i and C_j, the one lower in the asymmetric c-command ordering is more prominent.

(Zubizarreta 1998:19)

In Zubizarreta's analysis prosody corresponds directly to information structure in Spanish, which entails that nuclear stress encodes narrow focus. According to (94), in an utterance which represents the answer to the question 'what happened?', nuclear stress will fall on the 'lowest' constituent in the structure, as the whole sentence is focused. In (95), this constituent would be the past participle:

(95) [[Un [niño] ha [<u>bailado</u>]]]

a boy has danced

'A boy has danced.'

(96) *Un <u>niño</u> ha bailado

'A boy has danced.'

In (96) the subject is not lowest in asymmetric c-commanding order; therefore it can neither receive nuclear stress via the NSR, nor can it receive narrow focus, according to the FPR.

The NSR and the FPR apply in both Romance and Germanic languages. The differences in word orders displayed between both types of languages are attributed by Zubizarreta to a 'prosodic invisibility' constraint that applies in the latter but not in the former: in Germanic languages and French defocalized constituents are invisible to the NSR, whereas in Spanish and Italian they are not. For this reason nuclear stress can be assigned to preverbal subjects in English and French, but not in Spanish or Italian. The examples in (97)-(100), all answers to the question 'who ate the cake'?, and in which metrically invisible constituents are in italics, show the differences in metrical (in)visibility between French and Spanish:

(97) <u>Marie</u> ate the cake.

(98) <u>Marie</u> a mangé le gateau.

Marie has eaten the cake

'<u>Marie</u> has eaten the cake.'

(99) *<u>María</u> ha comido el pastel.

María has eaten the cake

'<u>María</u> has eaten the cake.'

(100) Ha comido el pastel <u>María</u>.

Has eaten the cake María.

'<u>María</u> has eaten the cake.'

Zubizarreta further proposes that if there is a mismatch between the NSR and the FPR, the alternative solution in Spanish is what she labels P-(rosodic) Movement. This operation occurs in the phonological component and triggers movement of the object over the subject so the NSR can apply and assign nuclear stress, and narrow focus, to the low subject.

According to her, P-movement is responsible for the fact that VOS Spanish is triggered only where the subject receives narrow focus. Therefore, In VOS word orders, the object cannot receive contrastive focus. Recall that contrastive focus can be assigned to any constituent *in situ*, as well as to elements in the preverbal contrastive focus position. In VOS, however, the object is p-moved over the subject so nuclear stress can fall on the subject; it is not possible, then, to move the object and stress the object rather than the subject, as exemplified by (101), which repeats (81) above, and (102)-(103):

(101) EL SINDICATO habló contra el gobierno (y no el partido).
The union spoke against the government and not the party
'The UNION spoke out against the government and not the party.'

(102) La camarera del hotel puso la VALIJA sobre la cama (y no el maletín).
The attendant of-the hotel put the suitcase on the bed and not the briefcase
'The hotel attendant put the suitcase on the bed, not the briefcase.'

(103) *Habló contra el GOBIERNO el sindicato (y no contra el partido).
spoke against the government the union and not against the party
'The union spoke out against the GOVERNMENT and not the party.'

Zubizarreta's approach is generally accepted in the literature as it presents several advantages: it captures the close parallels between prosody and syntax displayed by Spanish, and in addition it correctly accounts for the specialization of certain word orders in conveying wide and narrow focus. For her account to work, she proposes that the language faculty includes an intermediate level of ∑-structure (i.e., Assertion Structure) prior to LF and PF, where operations such as p-movement take place.

Recently, Sheehan (2006) has offered a reformulation of Zubizarreta's account in a more restrictive model, without a level of assertion structure. In Sheehan's account, p-movement takes place in the phonological component. Given that according to standard minimalist assumptions movement operations entail copy deletion, p-movement is reformulated as PF movement; the

difference between syntactic movement and p-movement would be that the latter involves sending a lower copy rather than a higher one to the PF interface. The structure sent to the phonological component after p-movement would look as in (104):

(104) $[_{IP}$ S Agr T $[_{vP}$ O S v $[_{VP}$ V O]

When the subject receives narrow focus, the derivation proceeds as follows in Italian and Spanish:

(105) $[_{IP}$ ~~S~~ **[V+v+T+Agr]** ~~T~~ $[_{vP}$ O S ~~v~~ $[_{VP}$ ~~V O~~]

In (105), the object is not spelled-out low because it would violate the NSR. If the subject were not to receive narrow focus, subject and object would be spelled out in their higher positions, to minimize computation. In this structure, however, prosodic interface conditions require a lower copy of the subject to be spelled out.

To sum up, narrow focus is assigned in Spanish to sentence-final constituents, and it seems clear that operations at the syntax-phonology interface are responsible for VSO and VOS word orders.

A theory on focus assignment in Spanish should be able to account as well for dialectal variation in word orders. Wide focus interpretations can be conveyed in Spanish by both SVO and VSO word orders; however, one may wonder whether in a variety like DS, in which SVO is the default word order, VSO and VOS orders that convey narrow focus are available, or whether in such a restructured grammar other strategies are available for signaling focus. As discussed in the next section, while general issues related to focus assignment in DS have been approached in previous work (Toribio 2002, Camacho 2008b), the question of how subjects get narrow-focused in this variety has not received much attention in the relevant literature.

2.4.3.3. Focus in Dominican Spanish.

As outlined in Chapter One, focus can be assigned in Spanish not only via prosody or prosody-syntax interface operations, but also through purely syntactic means, as clefted structures

also convey focus. Three types of clefted structures are available in Non-Caribbean Spanish (cf. (12a-c) in Chapter One, repeated here for ease of discussion as (106a-c):

(106) a. Fue Marta la que trajo los regalos. [cleft]
was.3PSG Marta CL.ACC that brought. 3PSG the gifts

b. Quien trajo los regalos fue Marta. [pseudocleft]
who brought.3PSG the gifts was.3PSG Marta

c. Marta fue quien trajo los regalos. [inverted pseudocleft]
Marta was.3PSG who brought.3PSG the gifts
'It was Marta who brought the gifts.'

In Caribbean varieties of Spanish, another structure is available for signaling phrasal focus, the so-called focalizing *ser* construction. This structure involves two clauses, one of them headed by a copular verb; the focused constituent, which can be an object, an adverb or a postverbal subject, is placed in front of the copular verb, as illustrated in (107a-c); as shown by (107d), preverbal subjects are not possible in focalized *ser* structures:

(107)

a. El gobierno le cedió los terrenos fue a las multinacionales
the government CL.3PSG granted.3PSG land was.3PSG to the transnationals
'It was transnational corporations that the government granted land to.'

b. Llegamos fue ayer
arrived.1PPL was.3PSG yesterday
'It was yesterday that we arrived.'

c. Compró papas fue Juan.
bought.3PSG potatoes was.3PSG Juan

d. * Juan fue compró papas.

Juan was.3PSG bought.3PSG potatoes.

'It was Juan who bought potatoes.'

(Camacho 2008b:2)

Camacho notes, with Bosque (1999), that the copular verb cannot modify preverbal subjects because generally it cannot modify any category higher than vP. He further points out that *ser* focalizing structures are not attested in DS for the focusing of subjects, probably because postverbal subjects are rare in this variety. Note that the Non-Caribbean Spanish structures in (104a-b) also feature a postverbal subject. We would expect then that the preferred construction in DS for the expression of syntactic focus is one along the lines of (106c). In fact, the use of clefted structures to convey focus is widespread in the corpus of spontaneous speech gathered for the present study, as shown by the examples in (108a-d), in which as is usually the case in spontaneous speech, the relative 'que' is chosen over 'quien', associated with more formal registers:

(108)

a. Yo fui que la crié.

I was.1sg that her.CL raised.1PSG

'It was me who raised her.

(Evangelita, 70, UHF)

b. Ella fue que me dijo mi nombre.

She was.3PSG that me.CL said.3PSG my name

'It was her who said my name.'

(Néstor, 56 UHM)

c. Yo no me siento mal, al revés, yo me siento bien

I not me.CL feel.1PSG bad, on the contrary, I feel.1PSG good

porque fui yo que le dije

because was.1PSG I that cl.3PSG said.1PSG

'I don't feel bad, on the contrary, I feel, good, because it was me who told him.'

(Bruno, 52, UWM)

d. Cuando decían El Jefe nadie se equivocaba

When said.3PPL El Jefe no-one se.cl took-wrong.3PSG

era Trujillo que llamaba.

was.3PSG Trujillo that called.3PSG

'When they said El Jefe nobody took it wrong, it was Trujillo who was calling.'

(Federico, 81, UHM)

As shown in (108), clefted structures tend to feature preverbal subjects, but postverbal subjects are also found (108c). Only a fine-grained study of the attested DS focus structures could tear apart the constructions available to Dominican speakers to convey narrow subject focus. While this is not the main goal of the study carried out in this dissertation, the quantitative analysis on subject positions in main and embedded structures carried out in Chapter Six will contribute to the understanding of these focus strategies.

Having presented the main findings in the previous literature on focus assignment in Caribbean and Non-Caribbean Spanish, I now turn to discuss interrogative constructions.

2.4.4. Subject licensing in interrogative constructions.

Dominican, and more generally Caribbean wh-questions have been described as generally featuring SVO word order in matrix and embedded questions, in sharp contrast with Non-Caribbean Spanish, in which the sequence whSV yields ungrammaticality:

(109) a. ¿Qué comes (tú)? (Non-Caribbean Spanish)

what eat.2PSG (you)

'What are you eating?'

a'. *¿Qué tú comes?

b. ¿Qué tú comes? (Caribbean Spanish)

what you eat.2PSG

'What are you eating?'

b'. ¿Qué comes (tú)?

(110) a. No sé a quién ha llamado (él). (Non-Caribbean Spanish)

not know.1PSG to who has called.3PSG (he)

'I don't know who did he call.'

a'. * No sé a quién él ha llamado.

b. No sé a quién él ha llamado. (Caribbean Spanish)

not know.1psg to who he has called

'I don't know who did he call.'

b. No sé a quién ha llamado (él).

Torrego (1984), Ordóñez and Treviño (1999), Zubizarreta (2000), Ordóñez and Olarrea (2006) have pointed out that two amendments are to be made to the claim that whSV is banned in Non-Caribbean Spanish:

First, while bare wh-words (*qué*, 'what', *quién*, 'who', *cuándo*, 'when', *dónde*, 'where') must be inverted and the wh-word is necessarily adjacent to the verb, non inversion is possible with complex wh-words, as shown by the contrast between (111a) and (111b), where wh-words are signaled in italics:

(111) a. *¿A *quién* tu hermana visitó? (Ordóñez and Olarrea 2006:67)

to who your sister visited.3PSG

'Who did your sister visit?

b. ¿A *cuál de estas chicas* tu hermana (la) había visitado en Sicilia?
to which of these.FEM girls your sister (her.CL) had visited.3PSG in Sicily
'Which of these girls had your sister visited in Sicily?'

Second, a [+/- argumental] asymmetry also applies: while thematic arguments obligatorily trigger VS order, non-thematic elements allow for non-inversion, as showcased in (110a), which features the wh-word *por qué*, 'why'. Gallego (2006) further notes that the bare wh-word *cómo*, 'how' can surface preverbally where emphatic intonation is added, as illustrated in (112b):

(112)

a. ¿Por qué Juan quiere salir de clase antes que los demás?
Why Juan wants get out.INFIN of class before that the others
'Why Juan wants to get out of the class before the others?'

a'. No sé por qué Juan quiere salir de clase antes que los demás.
not know.1PSG why Juan wants get out.INFIN of class before that the others
'I don't know why Juan wants to get out of the class before the others.'

b. ¿Cómo Juan pudo hacer eso?
How Juan was able.3PSG do.INFIN that
'How on earth was Juan able to do something like that?'

b.' No sé como Juan pudo hacer eso.
not know.1PSG how Juan was able.3PSG do.infin that
'I don't know how on earth Juan was able to do something like that'.

As far as Caribbean Spanish is concerned, the availability of non-inverted wh-questions was noted already in early descriptions of these varieties (Navarro Tomás (1948) for Puerto Rican Spanish, Henríquez Ureña (1940), Jiménez Sabater (1975) for Dominican Spanish). In more recent work, it is generally agreed in both theoretical (Ordóñez and Olarrea 2006) and quantitative work (Lizardi 1993) that even if Caribbean speakers accept wh-interrogatives with VS order as grammatical, in production non-inverted interrogatives constitute the default option for the

speakers. However, there has been strong disagreement in the literature with respect to the type of subject constituent that can appear in preverbal position: while some authors have claimed that only pronominal elements can surface as non-inverted (Davis 1971, Lantolf 1980, Núñez Cedeño 1983), in more recent work (Lizardi 1993, Toribio 1993), as well as in the corpus gathered for this dissertation, non-inverted interrogatives with full DP subjects are also attested:

(113) a. ¿Cómo los padres pueden ayudar a sus hijos? (Lizardi 1993:51)

b. ¿Cuánto un médico consume en un mes? (Toribio 1993:26)
how much a doctor uses in one month
'How much a doctor uses in one month?'

c. ¿Cómo los curanderos curan? (Magdalis, RFW)
How the witch doctors cure.3PPL
How do the witch doctors cure?

Ordóñez and Olarrea (2006) suggest that the diverging patterns in the possibility of inverting different subject constituents described in the literature might be a function of intra-Caribbean dialectal variation, while Lizardi (1993) points out that the availability of non-inverted full DP subjects might be due to the extension of SVO word order in Caribbean varieties as a result of grammatical restructuring. In both pieces of work it is pointed out that even if full DP non-inverted subjects are possible in Caribbean dialects, subject pronouns are non-inverted in much higher numbers. As will be discussed in Chapter Six, a similar pattern is also found in the data set gathered for the present study.

A number of accounts have been put forward in the previous literature to explain the distribution of pronominal and full DP subjects in Caribbean and Non-Caribbean interrogatives. I present them in turn in the following two subsections.

2.4.4.1. Wh-interrogatives in Non-Caribbean Spanish.

There are at least two logical possibilities for the required adjacency between the wh-phrase and the finite verb in Non-Caribbean Spanish, and both of them have been explored in previous

work: either the verb moves to C (Torrego 1984, Goodall 1993, Rizzi 1996), or the wh-phrase moves to Spec, IP (Vallduví 1990, Zubizarreta 1998, Barbosa 2001, Sheehan 2006). Ordóñez and Olarrea (2006), developing ideas in Kayne and Pollock (2001), Pollock and Poletto (2004), propose a third type of analysis, in which subject-verb inversion is derived via remnant movement of the IP to a Spec position within CP.

Analyses along the lines of Rizzi (1996), which posit V-to-C in Spanish, take as a point of departure the similarity between Romance interrogative constructions with subject-verb inversion and Germanic inversion phenomena such as Subject-Auxiliary inversion in English (cf. for instance *Who must John visit?* or V2. Rizzi (1996) formulates the Wh-Criterion, a well-formedness condition that is responsible for the interpretation of wh operators at LF:

(114) *Wh-Criterion* (Rizzi 1996)

a. A wh-operator must be in a Spec-head configuration with an X [+wh]

b. An X [+wh] must be in a Spec-head configuration with a wh-operator.

Rizzi also proposes that Inflection can be specified as [+wh]; consequently, when I-to-C is available in a given language, this movement operation will take the [+wh] feature of I to a position where the Wh-Criterion can be satisfied. The Criterion would derive then the ungrammaticality of structures with no subject-verb inversion in Non-Caribbean Spanish (cf. **¿Qué tú comes?* in (107) above): the [+wh] feature on Inflection is not in a Spec-head configuration with a wh-operator in this sentence, since the subject *tú* intervenes between the wh-word *qué* and the verb *quieres*.

As explained by Ordóñez and Olarrea (2006), Rizzi-style analyses of Non-Caribbean interrogatives face a number of challenges:

First, as the examples in (109)-(110) show, there is no main/embedded asymmetry in Spanish with respect to subject-verb inversion phenomena, which suggests that movement to C is not an option in this language. Second, as showcased by (111), the complexity of the wh-word has a role in whether non-inversion is available in Spanish, an effect that is unattested in Germanic

languages. Finally, Rizzi's theory could not account for the preverbal position of subjects in Caribbean interrogatives[17].

Ordóñez and Olarrea (2006) propose an alternative analysis to overcome these challenges. In their theory, inversion consists of movement of a remnant IP to a Specificer position in CP located above the position in which the subject lands[18]. Following the development of Rizzi's (1997) cartographic approach in Kayne and Pollock (2004), they adopt the structure in (115) for CP in Spanish:

(115) [Op 2 [Force [Ground P [TopP [Op1]]]]

In (115), the extended projection of the CP domain contains the following functional projections: Force Phrase, which is a reflex of the classic CP, i.e., the functional node that expresses clause type and that is the ultimate site of complementizers; Ground Phrase, the locus of presupposed information and the landing site of various movement operations; Topic Phrase, where Topicalized constituents lands; and two Operator Phrases, for the landing of wh-words. The postulation of two different Operator Phrases is sustained on the availability of wh-moved and wh-in situ structures cross-linguistically.

The derivation of a wh-structure with subject-verb inversion such as (116) would proceed as represented in (117), according to Ordóñez and Olarrea (2006):

(116) ¿En qué fiesta comió curry José?

in what party ate.3PSG curry José

'At which party did José eat curry?'

[17] Ordóñez and Olarrea (2006) also point out that adopting Rizzi's (1996) account would imply to assume an innecesarily complex theory of clitic placement in Spanish; for details about this counter-argument to Rizzi (1996), see Ordóñez and Olarrea (2006).

[18] A remnant constituent is defined as a constituent from which material has already been extracted. Moving such a constituent has as a consequence that traces of earlier movements can be carried along by remnant movement to positions where they are no longer c-commanded by their antecedents. While this configuration was banned in previous generative models, it is allowed by minimalist theory. Remnant movement theories have been applied to different syntactic phenomena in the last few years, as they allow to maintain bottom-up structure building (proposed as the only structure building mechanism in Chomsky 1995) and to maintain as well the notion that all movement is leftward (Kayne 1994).

(117) a. Movement of the wh-word to an operator position →

[$_{OP1}$ En qué fiesta OP10 [José ha comido curry t]]

b. Movement of the subject to Top P →

… [$_{TopP}$ José Top^{0} [$_{OP1}$ En qué fiesta OP1^{0} [$_{IP}$ José comió curry t]]]

c. Preposing of IP above TopP →

[$_{GroundP}$ [$_{IP}$ comió curry] G^{0} [$_{TopP}$ José [$_{OP1}$ En qué fiesta OP1^{0}]]]

d. Movement of the wh-word to Op2 →

[$_{OP2}$ En qué fiesta [[$_{IP}$ comió curry] José [OP1]]]

The position to which the remnant IP moves is identified, following Kayne and Pollock (2004), as Ground Phrase. Before the IP moves to Spec, GroundP, however, the wh-word has moved to an operator position OP1 above IP to check its features, (117a). This movement is followed by movement of the subject to the Topic-like position located above OP1, TopP (117b). The next step is for the remnant IP, from which the subject has just been extracted in the previous step, to move over TopP into Spec, GroundP. Finally, the wh-word moves to OP2, deriving the final output for this structure.

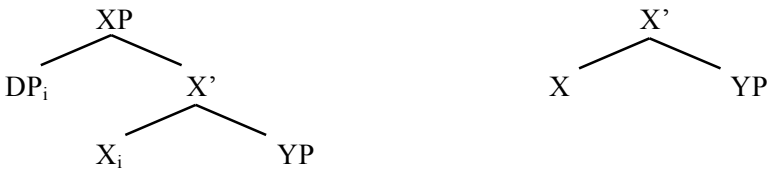
Ordóñez and Olarrea (2006) correctly derive the word order in Non-Caribbean Spanish inverted interrogatives. Sheehan (2006), on the contrary, presents an account in which VS order is also correctly derived without resourcing to remnant movement: in her account, the wh-phrase moves to Spec, IP and the verb remains in situ in vP, yielding subject-verb inversion. Recall that according to the analysis put forward by Sheehan for VS orders in declaratives discussed in §2.3.4.3.1., in Spanish VS the subject can remain in Spec, vP as long as another element lands in Spec, IP and is able to satisfy the EPP. Case and Agreement are checked locally by the subject in vP, under Agree.

In light of VS facts in declarative and interrogative structures, Sheehan then presents a view on Case assignment along the lines of Koopman and Sportiche (1991). According to these authors, who construed their account under the Government and Binding framework, two possibilities are

available cross-linguistically for Case assignment: while in some languages Case is necessarily assigned under government, in other languages it can be assigned via Specifier-Head agreement, as represented in (118):

(118) Case assignment mechanisms (Koopman and Sportiche 1991):

a. Spec-Head agreement Case b. Government Case



Sheehan translates Koopman and Sportiche's account in minimalist terms as follows: in some languages nominative Case is specified as [strong], which entails that it requires a local, Spec, Head configuration with the agreeing head I in order for the unintepretable feature [Nominative] to delete. This would be the case in a language like English, for instance, in which subjects surface necessarily in Spec, IP. In some other languages, Non-Caribbean Spanish being a case in point, nominative Case is [weak], and a DP specified as [nominative] can remain in situ, as long as the EPP is otherwise satisfied.

To sum up, while V-to-C accounts of subject-verb inversion in interrogatives face a number of challenges, remnant IP movement accounts (Ordóñez and Olarrea 2006), and accounts proposing wh-movement to Spec, IP (Sheehan 2006) correctly derive the Non-Caribbean facts. In the following section the power of these accounts to deal with Caribbean Spanish interrogatives is discussed, along with the specific accounts proposed for these varieties put forward in the previous literature.

2.4.4.2. Wh-interrogatives in Caribbean Spanish.

The available accounts of Caribbean wh-interrogatives generally frame non-inversion patterns in the global restructuring of Caribbean subject licensing grammars. Three main theories have been put forward in the literature, which focus on the shifts undergone by Caribbean agreement morphology, the mechanisms available for Case licensing in Spanish, and the

informational content of overt subjects, respectively: the *Cliticization Hypothesis* (Lipski 1977, Contreras 1989, Heap 1990, Benedicto 1993); The *Case Hypothesis* (Toribio 1993, 1996); and the *Theme-Rheme Hypothesis* (Suñer 1982, Lizardi and Suñer,1992, Lizardi 1993).

In addition, in the recent work by Ordóñez and Olarrea (2006) discussed in the previous section a fourth proposal reminiscent in some respects of the *Cliticization Hypothesis* is put forward to deal with the attested dialectal variation in inversion patterns. Before I move on to outline the specific ways in which Ordóñez and Olarrea's account deals with Caribbean interrogatives, the other three hypotheses are presented in the following paragraphs.

Proponents of the *Cliticization Hypothesis* build up on Lipski's (1977) early ideas on the consequences that impoverishment of inflectional morphology in Caribbean Spanish has for their distribution in declarative and interrogative constructions. Lipski proposed that as a result of weakened agreement, Caribbean pronouns behave essentially like clitics, and therefore need to be attached to the verb.

However, as noted by Suñer (1982), Lizardi (1993), Caribbean pronouns can be separated from the verb by negation, which tears them apart from clitics:

(119) ¿Por qué tú no se lo dijiste? (Lizardi 1993:56)
why you not CL.DAT.3PSG CL.ACC.3PSG said.2PSG
'Why did you not tell him/her/them?'

Furthermore, the *Cliticization Hypothesis* cannot deal with the fact that SVO word orders in interrogatives are also available with full DP subjects.

The *Case Hypothesis* (Toribio 1993a,b, 1996) posits a difference in Case assignment mechanisms in Non-Caribbean and Caribbean Spanish. Developing the ideas in Koopman and Sportiche (1991) discussed in the previous section, according to which Case can be assigned under Government or via a Spec-Head relation, Toribio (1993b) proposes that as a result of the different nature of the Inflection head in Non-Caribbean and Caribbean Spanish, in the former Case is assigned under government, while in the latter it is assigned under Specifier-Head agreement. That

Case is assigned via Specifier-Head agreement implies that the subject has to necessarily land in Spec, IP. In the languages in which Case is assigned under Government, the subject can remain in VP and the word order VOS is possible. As explained above, while in Caribbean Spanish subjects surface in preverbal position, VSO and VOS are possible word orders in Non-Caribbean Spanish.

Note the striking similarity between the theory put forward by Toribio (1993b) to deal with the differences between Non-Caribbean and Dominican Spanish and the account sustained in Sheehan (2006) to explain the differences between consistent NSLs and English discussed in §2.3.4. It might be worthwhile to point out in this respect that since cross-dialectal differences may mirror cross-linguistic differences, similar accounts can be posited to deal with both types of language variation.

The *Case Hypothesis* has the considerable advantage over the *Cliticization Hypothesis* that it can account successfully for both pronominal and full DP non-inverted subjects. Lizardi (1993), however, points out that while the *Case Hypothesis* correctly derives the facts on interrogative constructions with preverbal subjects, it bans postverbal subjects, and rather what is found in Caribbean Spanish varieties is a three-way alternation: while overt preverbal subjects are the default option, structures with null and postverbal subjects are accepted, and produced in small numbers, by the speakers. Recall that in this respect Toribio (1993a) claimed that Dominican speakers, in accordance with Kroch's (1989, 1994) *Competing Grammars Hypothesis*, would have available also the Non-Caribbean Spanish Case assignment mechanism.

In order to account for the Caribbean three-way alternation Lizardi and Suñer (1992), Lizardi (1993) put forward the *Theme/Rheme Hypothesis*. This hypothesis is built on two ideas: first, generally in Spanish, in the absence of emphatic stress or intonation, preverbal subjects are thematized, i.e., the preverbal position is reserved to presupposed information; in the same unmarked circumstances, the postverbal position is mainly used to introduce rhematic, asserted elements. Second, they point out that while pronouns are used when identification of the referent is not crucial for carrying the communication forward, full DPs are used when identification of the

referent is necessary different distribution of pronouns. This would explain, in their view, why pronominals surface as non-inverted much more frequently than full DP subjects. With these two notions in mind, the authors suggest that the postverbal position in Caribbean interrogatives is increasingly acquiring the role of introducing elements under the scope of assertion, in contrast with Non-Caribbean Spanish, in which the unmarked, default position for subjects in interrogatives is postverbal. This theory, along with the *Case Hypothesis*, overcomes the problem of the availability of full DP non-inverted subjects in wh-structures. However, without a detailed study of the informational content of Caribbean interrogative postverbal subjects, it is not possible to assess the validity of this proposal. Furthermore, since Caribbean varieties tend to maintain the SVO word order in declaratives and interrogatives, it is questionable why a strategy that places subjects postverbally would be developed as an innovation.

Ordóñez and Olarrea (2006), rather than focusing on the informational content of subjects or on how nominative Case assignment takes place in interrogative constructions, base their account on the different nature that pronominals have in Caribbean and Non-Caribbean Spanish. In this sense, their account bears some similiarities with the *Cliticization Hypothesis* in Lipski (1977). Their theory, build with reference to Dominican data, departs from Lipski's early account in that they take as a point of departure the more recent ideas in Cardinaletti and Starke (1999) on the tripartite classification of pronouns cross-linguistically. In line with accounts on Caribbean pronouns such as Camacho (2008), presented in §2.3.4, they suggest that subject pronouns in Caribbean varieties are specified as [weak], as opposed to Non-Caribbean pronouns, which are specified as [strong]. Also in line with Camacho, Ordóñez and Olarrea explain the availability of inverted interrogatives along with non-inverted ones in DS by availing to the existence of two series of homophonous subject pronouns, constituted respectively by strong and weak pronominals (cf.§2.3.4.3).

Ordoñez and Olarrea find evidence for the [weak] status of Dominican pronouns in their distributional properties:

First, similarly to the Italian weak pronouns analyzed in Cardinaletti and Starke (1999), Cardinaletti (2004), Dominican weak pronouns cannot be coordinated, do not allow modification, as illustrated in (120a-b), as opposed to the strong, postverbal pronouns in (120a'-b'):

(120) a. *¿Qué tú mismo comes? (Ordóñez and Olarrea 2006:74)
what you yourself eat.2PSG
'What do you eat yourself?'

a'. ¿Qué comes tú mismo?
what eat.2PSG you yourself
'What do you eat yourself?'

b. *¿Qué tú y él comieron?
what you and he ate.2PPL
'What do you and he ate?'

b.' ¿Qué comieron tú y él?
what ate.2PPL you and he
'What did you and he ate?'

Second, weak Dominican pronouns cannot be focused and they cannot precede a left-dislocated constituent:

(121) a. *¿Qué TÚ quieres?
what YOU want
'What do YOU want?'

b. *¿Tú, esa carta la echaste al correo?
you that letter CL.ACC dropped in-the mail
'You, that letter you dropped in the mail?'

The authors take these differences between strong postverbal pronouns and weak preverbal ones also as evidence that they occupy different positions in the clausal structure; specifically, they claim that weak pronouns are merged lower in the structure than strong pronouns, again in line with

Camacho (2008) and previous work such as Suñer (2004). For this reason, in the derivation of a Caribbean wh-structure, the weak pronoun is part of the IP that moves into a Specifier position in the CP field, and appears in preverbal position. Recall that in Non-Caribbean Spanish the subject pronoun 'escaped' the IP before remnant movement of this whole constituent took place (cf.117b). The steps involved in a non-inverted Caribbean interrogative such as *¿Qué tú quieres?* are represented in (122):

(122) a. Movement of the wh-word to an operator position →

$[_{OP1}$ Qué OP^0 $[_{IP}$ tú quieres]]

b. Movement of IP to GroundP→

$[_{GroundP}$ $[_{IP}$ tú quieres] $Ground^0$ $[_{OP1}$ qué $OP1^0]]$

c. Movement of the wh-word to Op2 →

$[_{OP2}$ Qué $OP2^0$ $[_{GroundP}$ $[_{IP}$ tú quieres] $Ground^0$ [OP1]]

The account in Ordóñez and Olarrea (2006) successfully derives the dialectal variation observed in Spanish interrogatives. It also provides further evidence to the hypothesis that Caribbean and Non-Caribbean pronouns bear different feature specifications, an idea that has found ample support in the literature on Romance mixed *pro*-drop patterns (Toribio (2000), Camacho (2008), for Dominican Spanish, Kato (1999) for Brazilian Portuguese). However, since their hypothesis is formulated to deal only with non-inverted pronominals, it faces the challenge, like the early *Cliticization Hypothesis*, that it cannot account for non-inverted full DP subjects. A possible solution for this problem could be to sustain that non-inversion with pronominal and full DPs is derived from different grammatical constraints, but it is doubtful that such a theory could be maintained against minimalist Economy restrictions.

Caribbean non-finite structures with nominative preverbal subjects have been described as posing similar challenges to theoretical accounts on subject licensing and subject placement. Likewise to wh-structures, Caribbean infinitive constructions can bear a preverbal subject. Also, a three way alternation is available to Caribbean speakers, in that structures with preverbal,

postverbal and null subjects are attested. Finally, not only pronominal but also full DP elements can be featured in these constructions. The ensuing discussion describes these issues.

2.4.5. Subject licensing in non-finite structures.

Canonical Romance infinitive constructions license PRO, a null, Caseless pronoun, which represents the understood subject of an infinitive complement:

(123) La niña quiere PRO salir.

'The girl wants PRO to go out.'

In these constructions, the subject or object of the matrix verb controls the subject of the infinitive, as represented in (124):

(124) [TP subject T_{MATRIX} object [TP PRO $T_{INFINITIVE}$]]

These constructions alternate in Spanish with finite embedded clauses that feature a verb in the subjunctive. The finite structure arises when the subjects in the matrix and the embedded clause are disjoint in reference:

(125) La niña quiere que salga su madre.

The girl wants that go-out.3PSG.PRES.SUBJ her mother

'The girl wants her mother to go out.'

Two Romance constructions have been documented that deviate from this pattern: on the one hand, Portuguese inflected infinitives, as in (126a), in which the non-finite form surfaces with tense and agreement morphology, and personal infinitives, as in (126b) found in Spanish, Catalan and Sardinian, among other Romance languages, which feature a postverbal overt subject.

(126) a. Eu penso terem os deputados trabalhado pouco.

I think.1PSG have.INFIN.3PL the deputies.NOM work.PART little

'I think that the deputies worked little.'

(Raposo 1987:87)

b. Para poder una lograr esa meta,

For be able.INFIN one.FEM.NOM to meet that goal

hay que tener visión.

have-to.3PSG have vision

'For one to be able to meet that goal, one must have vision.'

(Suñer 1986:2)

Inflected and personal infinitive constructions have in common that they do not license PRO, but a referential subject. This fact, in addition to posing a challenge for the theory of Control, raises the question of how nominative Case is licensed in non-finite constructions.

Personal infinitives are also attested in Caribbean Spanish, in Venezuela (Suñer 1986), Puerto Rico (Lizardi 1993), and the Dominican Republic (Toribio 1993, 2000). Lipski (1991) has documented these constructions also in coastal Colombia, Cuba, the Canary Islands, and in the vestigial Spanish spoken in Trinidad, the Philippines, and Louisiana in the United States (Lousiana *Isleño* Spanish). In these varieties, the infinitive can appear in preverbal position, as exemplified in (127) with Venezuelan examples offered by Suñer (1986); note that personal infinitives with an overt preverbal subject are possible in Caribbean Spanish not only when the matrix and embedded clause share the same subject (127a) but also when they are disjoint in reference (127b):

(127)

a. …pero nunca llegué a pensar que los tuviera que dejar

but never had-to.3PSG think.INFIN that CL.3PPL had-to leave.INFIN

para yo ir fuera.

For I.NOM go.iNFIN abroad.

'But I never had to think that I had to leave them (my children) to go abroad.'

b. Ya está asentado ese té, pero está muy caliente para yo tomarlo.

already is steeped that tea, but is very hot for I.NOM drink.INFIN

'That tea is already steeped, but it is too hot for me to drink it.'

Since these constructions, both in Non-Caribbean and Caribbean Spanish versions, represent a challenge to generative theories on subject licensing in which Case is assigned by Agreement, they have received some attention in the previous literature (Fernández Lagunilla and Anula (1994), Rigau (1995), Torrego (1998), Sitaridou (2002), for Non-Caribbean Spanish, Suñer (1986) for Venezuelan Spanish, Morales (1988) for Puerto Rican Spanish, Lipski (1985, 1991) for a number of varieties of Latin American Spanish, Toribio (2000), Pérez-Leroux (1999), for Dominican Spanish, among others). These studies have shown that the distribution of overt subjects in personal infinitives is limited in all varieties of Spanish. As far as Caribbean Spanish is concerned, the preverbal position of the subject in these varieties is attributed to the strong preference for maintaining the SVO word order, which in turned is linked, with the exception of Pérez-Leroux (1999), to the syntactic restructuring undergone by the subject licensing system in these varieties. The fine-grained properties of these constructions, in both Caribbean and Non-Caribbean Spanish are discussed in the following two sections.

2.4.5.1. Personal infinitives in Non-Caribbean Spanish.

Fernández Lagunilla and Anula (1994), Rigau (1995) and Torrego (1998) all point out that distribution of personal infinitives is by no means free. First, they can only occur in non-L-related, adjunct, positions like the ones exemplified in (128) -(132), or in subject position, as in (133):

(128) Antes de actuar Caballé, el público estaba expectante.
before of perform.INFIN Caballé the audience was.3PSG expectant
'Before Caballé performed, the audience was expectant.'

(129) Después de actuar Caballé, cantó Carreras.
after of perform.INFIN Caballé sang.3SG Carreras
'After Caballé perform, Carreras sang.'

(130) Se casaron justo al acabar la guerra.
they.refl married.3PPL just in end.INFIN the war
'They got married just when the war ended.'

(131) Para corregirlo yo, tenía que trabajar toda la noche.
for correct.INFIN-it I.NOM had-to.1SG work all the night
'In order to correct it, I had to work all night.'

(132) Sin corregirlo yo, todo funciona mal.
without correct.INFIN-it, I.NOM everything work.3PSG bad
'Without me correcting it, everything works badly.'
(Rigau 1995: 280)

(133) Cantar yo la Traviata sería interesante.
sing.INFIN I.NOM the Traviata be.cond.3SG interesting
'It would be interesting for me to sing the Traviata.'
(Yoon and Bonet-Farran 1991:353)

As noted by Torrego (1998), the personal infinitive cannot surface as a complement of a matrix verb:

(134) *Se queja de trabajar la familia demasiado.
he.REFLEX complain.3SG of work.INFIN the family too much
'S/he complains about her family working too much.'
(Torrego 1998:210)

Additionally, the adverbial clause needs to appear before the main clause:

(135) *Después de Caballé actuar, cantó Carreras.
after of Caballé perform.inf sang.3PSG Carreras
'After Caballé performed, Carreras sang.'

There is no doubt that the case of the subject is Nominative in these structures. When the personal infinitive is introduced by a preposition, it is clear that the preposition does not license the Case on the subject because the latter does not surface in accusative Case, as shown by the fact that when substituted by pronoun forms which how oblique and non-oblique Case distinctions (cf. yo 'I' vs mí 'me'), the result is ungrammatical:

(136) *Para mí corregirlo, tenía que trabajar toda la noche.

for acc.I correct.INFIN-it had-to.1sg work.INFIN al l the night

'In order to correct it, I had to work all night.'

The ungrammaticality of (136) is evidence that Case is licensed by the infinitive itself and not by the matrix verb.

With regard to the position of the subject, if no emphasis is added, the postverbal position is the obligatory position and a preverbal subject gives rise to ungrammaticality. Torrego (1998) and Uriagereka (1995) note that for some speakers preverbal subjects are accepted when the word order carries a contrastive focus interpretation for the subject:

(137) Al ELLA tener que pagar, y no los otros, María gritó

On she.NOM have-to.INFIN pay and not the others Maria shouted

'Since she had to pay and not the others, Maria shouted'

(Uriagereka 1995:216)

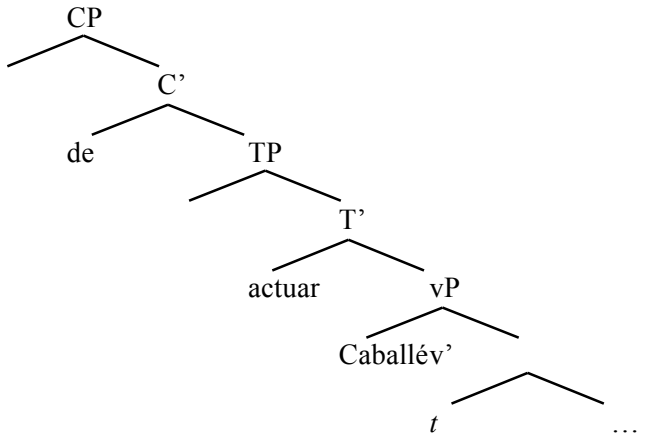
Two main views on personal infinitives have been put forward in the literature: some researchers argue that these structures are a peripheral or marked phenomenon (Rizzi 1987, Rigau 1995, Torrego 1998). The opposite claim is sustained by Sitaridou (2002), who derives the possibility of nominative subjects with non-finite constructions from a configurational property made available by UG. While in the former accounts it is suggested that (abstract) Agreement assigns Case to the subject of personal infinitive constructions, Sitaridou proposes that it is Tense and not Agreement the functional category responsible for Case assignment in these structures:

Taking as point of departure the patterns shown by Portuguese inflected infinitives, which display agreement morphology, Rigau (1995), among others, proposes that Abstract Agreement licenses Case in personal infinitives. Pires (2000) founds further support for this claim in Brazilian Portuguese. Brazilian 'inflected' infinitives have lost their agreement morphology, but have maintained their ability to license nominative subjects, a fact that he attributes to the presence of Abstract Agreement.

The other logical possibility, developed by Sitaridou (2002), is that in the absence of (overt) Agreement, nominative is licensed by the Tense of the infinitive. Trivially, she points out, the mere existence of personal infinitives suggests that Case and Agreement must be dissociated. She argues, in line with early minimalist assumptions (Chomsky 1995), that in both finite and non-finite sentences nominative is the structural Case assigned to the subject, rather than the instantiation of the agreement relation between the subject and the verb. In other words, the licensing of the nominative is a configurational property, and as such is not always necessarily linked to Agreement.

Moreover, since the personal infinitive can appear in an adjunct position, it must have an independent temporal frame, which in turn entails that it has the ability to project a TP. T in embedded infinitives is defective as compared to matrix T, since they are different in their ability for independent sentencehood. The derivation of a sentence like (128) above, *Antes de actuar Caballé, el público estaba expectante*, would proceed as follows: according to minimalist assumptions, the matrix and the adjunct clause represent two different Phases (Chomsky 2001); the matrix clause would be built first and sent to PF; when matrix TP1 is built, adjunct TP2 starts out as a separate derivation. The preposition *antes* selects a CP headed by the preposition de. TP2 is phi-incomplete, and for this reason, when the infinitive is attracted to T, there is no strong EPP to attract the subject to Spec, TP, so the subject remains in Spec,vP. This does not preclude T from checking the Case of the subject, because Agree can apply locally and [nominative] is valued and deleted. The derivation converges because *de* is at the head of the CP phrase, and this makes impossible for T to raise to C, in accordance to the Phase Impenetrability Condition (PIC) of Chomsky (2001), which makes elements below the head of a phase invisible to operations from outside said phase:

(138)



Therefore, in Sitaridou’s account, Agreement is not responsible for licensing nominative Case, but rather its absence triggers that the nominative subject of infinitives in Non-Caribbean Spanish surfaces in postverbal position.

Sitaridou correctly derives the properties of Non-Caribbean Spanish personal infinitives. However, her account crucially relies on the postverbal position of subjects in these structures, and personal infinitives with a preverbal subject are widely attested in Caribbean Spanish. While Caribbean Spanish facts also suggest that nominative Case assignment and presence of Agreement do not correlate, Sitaridou’s theory of T as a Case-assigning head leaves unexplained how a preverbal subject is derived in these varieties.

2.4.5.2. Infinitives in Caribbean Spanish.

The available studies on personal infinitives in Caribbean Spanish (Suñer 1986, Lipski 1991, Pérez-Leroux 1999) all describe them as having essentially the same distribution as in Non-Caribbean Spanish, with the relevant difference that the preverbal position for the subject is also allowed without added emphasis or contrast. Also, even if pronominal elements are far more frequently featured in a preverbal position in Caribbean Spanish, (Suñer 1986) full DP subjecs are also a possibility:

(139)

a. Para el gobernador tomar eso en consideración
for the governor.NOM take.INFIN that in consideration
debe ser presentado por escrito.
must be presented in writing
'For the governor to take that in consideration it must be presented in writing.

b. Antes de Juan conocerte a ti, se la pasaba vagando.
before of Juan know.INFIN.CL.DAT CL.DAT.3PSG CL.ACC.3PSG waste time
'Before Juan met you, he used to waste his time.'

Suñer (1986) presents an account of Caribbean non-finite structures somewhat similar to the *Case Hypothesis* (Toribio 1993b) explanations put forward to account for wh-structures. She departs from *Case Hypothesis* proposals in that her theory is based on the attested weakening of agreement morphology in these dialects. Suñer claims that for finite constructions, this shift in agreement patterns has two consequences: on the one hand, it precludes postverbal subjects from being licensed; on the other, Case is assigned in Caribbean Spanish not through Government, like in Non-Caribbean dialects, but in purely structural terms. This Case assignment mechanism is extended, in her account, to non-finite structures, as in (140):

(140) …prep [S' [S' NP INF…
[+lexical]
[nominative]

Toribio's (1993b) *Case Hypothesis*, in which nominative Case is assigned via Specifier-Head agreement in Caribbean dialects, also accounts for non-finite constructions. Toribio proposes that the availability of this mechanism for Case assignment is derived from the properties of the Inflection head. Specifically, building up on Van Riemsdijk (1990) and Grimshaw (1992), she suggests that while Inflection qualifies as a lexical head in Non-Caribbean Spanish, it constitutes a non-lexical pure, functional category in Caribbean Spanish.

Grimshaw's (1992) theory on extended projections captures the relation between functional heads and their lexical complements proposing that a functional projection is an extension of a lexical projection. Abney (1987) had already formulated such a relation in terms of dual headedness, stating that a functional projection is headed by its own functional projection and by the head of its lexical complement. Van Riemsdijk (1990), in turn, proposes that such a relation is expressed categorially, and that a functional projection is categorially identical to its lexical complement. In Toribio's account, the functional head Inflection would be categorially [-N, +V, ±F], where [±F] refers to functional status, and languages are free to choose either option.

Lipski (1991) also studies non-finite structures with nominative subjects. His comparison of these constructions in a number of Caribbean, vestigial, and non-standard varieties allows him to probe into the historical origin of this construction. He suggests that infinitives with preverbal subjects might have arisen through a process of ambiguity resolution: he notes that the preposition *para*, 'for' is usually chosen in Spanish to introduce infinitival clauses, not only when the following constituent is nominative, but also when it bears oblique Case:

(141) Es difícil para mí hacer eso. (Lipski 1991:205)
is difficult for me.CL.DAT do.INFIN that
'It is difficult for me to do that.'

The sentence in (141) involves a pivotal ambiguity, which would trigger, in a subsequent stage, the use of the infinitive with "objective subject Case" (Lipski 1991:206), like in the Louisiana Isleño example in (142):

(142) Para ti tené un bote tienes que ser sosedano americano.
for you.CL.DAT have.INFIN a boat have-to be citizen American
'For you to have a boat you have to be an American citizen.'
(Lipski 1991:202).

When examples like (142) arise, a grammatical tension is set between the requirement for oblique Case of the preposition, and the reanalysis of the objective pronoun as nominative. Since,

according to Lipski nominative is the unmarked Case in Spanish, the tension is resolved in favor of the nominative and the reanalysis can be extended to other prepositions:

(143) a. replace PRO by [NP, +lexical]; or, equivalently:

b. adjoin [NP, +lexical] to any Infinitive.

The explanations in Toribio (1993b) and Lipski (1991) do not crucially rely on a process related to the weakening of agreement morphology in Caribbean Spanish: while Toribio builds her theory with reference to the categorial nature of the Inflection head, Lipski derives synchronic infinitive constructions from language change driven by maximally unmarked processes. This is a considerable advantage, given that the very concept of morphological richness or lack thereof, as discussed above, has proven to be particularly difficult to define. Furthermore, as shown by the discussion on 'free inversion' in 2.3.4.4, subject insertion in Spanish is not constrained by inflectional morphology alone, which might suggest that shifts in areas of the grammar other than the syntactic component might be involved in the restructuring of the Caribbean subject licensing grammar. This possibility is explored in the studies reviewed in the following two chapters, in which theoretical are gathered from quantitative data.

2.5. Summary.

This chapter has centered on the discussion of NSP theories, as they are relevant for the study carried out in this dissertation. In the first part of the chapter the complex NSP profiles that posed problems for early NSP formulations (Partial NSLs, Null Expletive languages and dialectal variation in Western Romance) were presented. The remainder of the chapter has been devoted to discussing the two types of accounts that have been submitted within the Minimalist framework to overcome these empirical shortcomings, generally labeled for ease of discussion *Hypothesis A* and *Hypothesis B*. *Hypothesis A* accounts (Alexiadou and Anagnostopoulou 1998) take EPP to be parametrized: while in non-NSLs the EPP is licensed via subject movement to Spec, TP, in consistent NSLs agreement morphology is pronominal and therefore referential, which precludes the Spec, TP position to be projected. *Hypothesis B* explanations (Holmberg 2005, Sheehan 2006)

take agreement morphology to be always uninterpretable. EPP checking would then proceed similarly in non-NSLs and consistent NSLs, via subject movement to Spec, TP, which is projected in both types of languages.

Hypothesis A and *Hypothesis B* theories converge in that they pursue a view on parametrization in which differing null and overt subject licensing systems vary along three different dimensions:

(i) Feature assembly in the TP domain;

(ii) The nature of (null and overt) subject pronouns; and

(iii) Projection or lack thereof of the Spec, TP position.

These three possible sites of variation determine the smaller-scale parameters in which the NSP can be reformulated in minimalist terms. For instance, Sheehan (2006) proposes that two microparameters are relevant in subject licensing: *the rich agreement parameter* and *the PF interpretability parameter* (cf. (48) above). Therefore, under the minimalist view, the properties of Agreement are not exclusively responsible for subject licensing, as was the case in Government and Binding theories (Rizzi 1982, Jaeggli and Safir 1989, Rochbacher 1994, Roberts 1993).

Explanations along the lines of both *Hypothesis A* and *Hypothesis B* have been put forward in previous work to account for the differences in subject grammar displayed by Non-Caribbean and Caribbean Spanish (Suñer 1986, 2003; Toribio 1993a,b, 2000; Martínez Sanz 2007; Camacho 2006, 2008a,b; Ordóñez and Olarrea, 2006, among others). The discussion of these accounts in section 2.3.4. has shown that the study of Caribbean facts conflates a number of the long-standing questions on subject licensing in Spanish, such as the specific feature assembly of subject pronouns, the mechanisms available for null and overt subject licensing, or the different components of the grammar involved in subject insertion. While explanations of Caribbean subject grammars generally share the assumption that the phenomena usually associated with these dialects all stem from the general restructuring of their subject expression grammar, they differ in their identification of possible sources for this diachronic shift. Since it seems uncontroversial that agreement

morphology has a role in licensing null subjects in Non-Caribbean Spanish a number of authors (Suñer 1986, 2003, among others) submit that the attested weakening of agreement morphology is related to overt subject insertion and the preverbal positions of subjects in these varieties. However, recent research (Holmberg 2005) has shown that mechanisms other than richness of agreement are available cross-linguistically for the licensing of subjects. Furthermore, null subject structures are still available to Caribbean speakers. The co-existence of overt and null subject, NSL-like constructions in varieties like Dominican Spanish has been explained in previous work (Toribio 2000, 2001) by availing of the *Competing Grammars Hypothesis* (Kroch 1989, 1994). While this hypothesis accounts for the attested facts, it is not an ideal solution under minimalist economy restrictions. In the work of Camacho (2008b) and Ordóñez and Olarrea (2006) this challenge is overcome by proposing that Caribbean (single) grammars feature a series of weak pronouns, along the Non-Caribbean strong subject pronouns. The former appear overtly and have specific prosodic and distributional properties, while the latter have the properties of NSL subject pronouns. However, it is not clear how weak pronoun theories could deal with the fact that not only pronouns, but also full DPs appear overtly in a preverbal position in declarative, interrogative and non-finite constructions.

In sum, while a number on theories have been put forward in previous work to explain Caribbean facts, a number of issues remain unclear. In addition, as discussed in 2.3.4.4 some phenomena, related mainly with the expression of informational content in varieties that feature SVO as their default word order, still await description.

The discussion in Chapter Three will center on the contributions that a cohesive approach to dialectal variation has made to the understanding of issues of this sort. Specifically, data from Brazilian Portuguese will be discussed. Brazilian facts are very relevant for the work carried out in this dissertation for a number of reasons: on the one hand, this variety displays a system of filled and empty pronouns similar to the one described for Dominican Spanish (Kato 1999). In addition, the quantitative and the theoretical view on syntactic variation have been showed to converge in the

study of Brazilian Portuguese. While diachronic approaches that analyze quantitative data have offered an accurate description of the BP NSP profile, synchronic approaches have explored the possibilities of formal grammars to account for attested null and overt subject patterns.

Chapter Three. Null and overt subject patterns in diachronic change: The case of Brazilian Portuguese.

3.1. Introduction: Approaches to the subject licensing grammar of Brazilian Portuguese.

The label 'Brazilian Portuguese' (henceforth, BP) usually refers in the literature relevant for this study to the variety spoken in the southeast of the country, i.e. Sao Paulo, Rio and Minas Gerais (Duarte 1993, 1995, 2000, Kato and Negrão 2000). Brazilian facts have been approached in the historical syntax literature, as well as in theoretical investigations that have aimed to explain the null and overt subject pattern displayed by this variety in its synchronic state. The main proposal put forward in diachronic studies is that BP is undergoing language change in the direction of becoming a non-NSL, and that two phenomena are at the source of this change: the shifts in Brazilian pronominal paradigm and the impoverishment of agreement morphology (Duarte 1993, 1995, 2000). However, even if overt subjects have become the unmarked option for subject expression in most syntactic environments, null subjects are still licensed in synchronic BP in specific contexts. Theoretical studies have focused on exploring whether BP null subjects are of the same nature of consistent NSL null subjects, or whether present-day BP can be best characterized as some sort of partial NSL.

Diachronic and synchronic accounts that aim to explain the restructuring of the BP grammar that has resulted in this mixed pattern are presented in the sections to follow. In doing this, it will also be shown that the detailed quantitative work carried out on this variety has afforded a great deal of descriptive adequacy, which in turn has allowed us to draw meaningful theoretical implications on the nature of dialectal variation. The cohesive approach to the variability of subject expression in BP dates back to Kato and Taralho's (1984) work on the Brazilian pronominal paradigm. These authors explored the idea revitalized in recent years that while generative theories on parametrization could provide us with predictions as to what kind of correlations to find in

quantitative data, a well-conducted empirical analysis could reveal new unaccounted for facts and unnoticed patterns.

3.2. Defining the variable in Brazilian Portuguese: Kato and Taralho (1984).

The work carried out by Kato and Taralho (1984) on the semantic shifts occurring in the Brazilian pronominal paradigm was inspired by Laberge and G. Sankoff's (1976) investigation of the use of pronouns in Montreal French. Taking as a point of departure their observation that in this grammar *tu* and *vous* had migrated from the definite to the indefinite system, they study the interaction of this phenomenon with the increased use of the clitic *on* in the speech of Montrealers. They found that *on* was used in discourse with the values traditionally associated with *je*, *tu*, *ils*, and *nous*. In addition, *on* and *tu*/*vous* had become linguistic equivalents. They also showed that variation between indefinite *on* and indefinite *tu*/*vous* was constrained by syntactic and pragmatic factors.

Kato and Taralho (1984) replicated this study and argued for a similar phenomenon of pronominal tension taking place in BP: they found that the elements involved in the expression of indetermination in BP were the pronoun *você*, (*tu*/*vous*, *you*); *eu*, (*I)*; and *nós* (*nous*, *we*). In addition, the pronominalized expression *a gente* (*we folks*) and the null pronoun (∅) can also be used to express indefiniteness in BP. Examples are provided in (1)-(5):

(1) Depois que *você* termine o comércio, *você* vai na área residencial.

'After you are done with the commercial area, you go to the residential area'.

(2) Em primeiro lugar *nós* tenemos identificado claramente uma nova consciência crítica de clase media.

'In the first place we have clearly identified a new critical awareness of the middle class'

(3) E se *a gente* falar que nao tem?

'What if we say there isn't any?'

(4) Então ∅ chega numa rua, não é, a rua é grande.

'Then you get to a street, right, it is a large street'.

(5) E se *eu* pego aquella rua ali, então *eu* chego lá mais rápido.

'And if I (=one) take that street over there, then I (=one) get there faster'.

These authors' results show that *você* is clearly the most frequent form in speech. It is followed by the null form, *a gente*, *nós* and lastly *eu*. There are also register differences among the different variants, and while some of them enjoy standard status (*nós* and *eu*), others (*a gente*) are non-standard.

Kato and Taralho also point out that they did not find a single instance of another (standard) possibility offered by Portuguese for expressing indefiniteness, namely the structure with the impersonal pronoun 'se'. This structure is illustrated in (6):

(6) Fala-SE português no Brasil

'One speaks Portuguese in Brazil'

This structure is widely used in European Portuguese, and other consistent NSLs, as noted by Kato and Taralho (1984) and contrasted against quantitative data a few years later by Duarte (1995). Instead of the use of this construction, in BP what is found instead is either a small number of instances of the structure illustrated in (4) –a construction very similar to the 'se' structure but with the null form[19]- on the one hand, and structures with overt pronominal subjects in the vast majority of cases, on the other.

Kato and Taralho's (1984) contribution to the studies carried out to this day on the shifts in the BP pronominal paradigm is fundamental in a number of respects: first, although not always acknowledged in theoretically-oriented subsequent studies, their use of a large database of

[19] It is important to point out, with Holmberg (2009), that the null subjects in constructions like (4) have a *generic* interpretation, to be distinguished from the *arbitrary* interpretation available for third person plural null pronouns in consistent NSLs. While a generic pronoun is best translated in English as either 'one' or 'you', the arbitrary pronoun translates to English 'they', as in *They speak many different languages in India*. Generic and arbitrary pronouns have different semantic properties: the former denote people in general including the speaker and the addressee, and the latter denote people in general (in some domain), but excluding the speaker and the addressee.

spontaneous data let them look at how the language-specific grammar of BP was shaped with respect to the shifts occurring in the BP pronominal paradigm. Second, their study of the shifts in the referential meaning of different pronouns showed that this restructuring of the grammar was also related to the high rates of overt pronominal subjects found in this variety. This fact constituted a very intriguing phenomenon for a dialect like BP, since Portuguese, like Spanish, constitutes a typical, consistent NSL.

For these reasons, their investigation was followed by a series of studies that attempted a syntactic analysis of the restructuring of the BP grammar. Two types of approaches were taken: a number of studies focused on historical syntax, and in doing so explored the shifts that have occurred within the Brazilian pronominal paradigm over time (Taralho (1986); Berlinck (1989,1996); Duarte (1992, 1993, 2000); Lopes Rossi (1993); Torres Morais (1994); Ribeiro (1994); Cyrino, Duarte & Kato (2000), among others). Other studies focused on the issues related to null and overt subject licensing in synchronic BP (Modesto 2000a, 2000b, 2009; Holmberg (2005), Sheehan (2006)). While the diachronic studies followed the trend initiated in Kato and Taralho (1984) in that they gathered quantitative data for their analyses (Duarte 1993, 1995, 2000), some of the investigations of synchronic BP have been carried out though the traditional methodology used in syntactic research, based on grammaticality judgments from native speakers.

The quantitative and theoretical approaches to BP complement each other: while diachronic studies document the restructuring of the BP grammar and the correlation between loss of inflectional richness and the shift in the pronominal paradigm, theoretical studies contribute to the understanding of NSP phenomena cross-linguistically. The specific ways in which the two perspectives can be bridged in order to deepen our understanding of dialectal variation in Romance is outlined in the following sections.

3.3. The diachronic view on null and overt subjects in Brazilian Portuguese.

Duarte (1993, 1995, 2000) investigates the shifts in the BP pronominal paradigm between 1845 and 1992. The study of three corpora corresponding to this period leads her to conclude that

BP is undergoing diachronic change in the direction of becoming a non-NSL. She grounds this proposal on the fact that overt subject rates are on a constant raise throughout the 150 years covered in her research. She submits that two interacting factors are at the source of the restructuring of the BP grammar: the reduction of forms in the pronominal paradigm, and the impoverishment of agreement morphology. She finds further evidence for her claim that BP is in the process of becoming a non-NSL in the fact that overt subjects do not only appear in higher rates, but they do so in specific syntactic environments: first, overt subjects tend to appear in preverbal position, which entails a sharp decrease in VSO word orders, which are widely attested in European Portuguese and earlier stages of BP. Second, as a result of the restructuring of the Brazilian grammar, new strategies for subject expression that depart from consistent NSL profiles and that are akin to partial NSLs or non-NSLs are found in the speech of Brazilian speakers; these constructions include overt arbitrary subjects, new strategies for building impersonal constructions that do not involve null expletive subjects, or the so-called "double subject constructions", also known as left-dislocated structures, which are found in Romance non-NSLs like French.

Duarte (1993) gathers data from popular plays written in the 19th and 20th century which mirror vernacular speech, and she finds that the reduction in the BP pronominal paradigm began with the replacement of both singular and plural second person pronominals *tu* ('you'), and *vós*, which used to combine with verb forms bearing exclusive inflections, with the former address form *você(s)* ('you'), which combines with third person verb forms. In a second and more recent stage of restructuring, the first person plural *nós* ('we) is being replaced by the pronominal expression *a gente* ('one'), which also requires the third person singular verb form. These processes result in a paradigm that in the past exhibited six different inflectional endings being reduced to four, and even to three in the speech of the younger generation. The impoverishment of BP agreement morphology, as opposed to what seems to be the case for other languages that have changed their NSP profile over time, such as French (Roberts 1993), has been motivated by changes in the set of pronominal subjects, rather than by phonological processes. Duarte finds three different sets of

pronominal paradigms between 1845 and 1992, as illustrated in the following table for the regular verb *amar* (to love):

Table 3.1. Reduction of inflectional endings in BP (1845-1992)

Person & number	Pronoun	Paradigm 1	Paradigm 2	Paradigm 3
1SG	Eu	Amo	Amo	Amo
2SG	Tu	Amas	*	*
2SG	Você	Ama	Ama	Ama
3SG	Ele/Ela	Ama	Ama	Ama
1PL	Nós	Amamos	Amamos	*
1PL	A gente	*	Ama	Ama
2PL	Vós	Amais	*	*
2PL	Vocês	Amam	Amam	Amam
3PL	Eles/Elas	Amam	Amam	Amam

Table 3.2 shows how the three paradigms in Table 1 are distributed over time, as well as the differing rates of overt pronominal subjects in the period of time studied by Duarte. She divides the 1845-1992 time frame in seven periods, ranging from 20 to 35 years each. Subject rates grow from 20% in 1845, when Paradigm 1 was in use, to 74% in 1992, when Paradigms 2 and 3 coexisted in the speech of different generations of speakers:

Table 3.2. Rates of overt subjects in BP (1845-1992)

Paradigm 1	1845	20%
	1882	23%
	1918	25%
Paradigm 2	1937	46%
	1955	50%
Paradigms 2&3	1975	67%
	1992	74%

Table 3.2 shows that rates of overt pronominal subjects are constantly on the rise between 1845 and 1992. However, very significant differences are found between the different persons in the paradigm when they are looked at separately: second person pronouns were the first to raise their overt subject rates, increasing from an average of 18% of null subjects in the first three periods (1845-1918) to 75% in the fourth (1937-1955). This 75% rate is maintained throughout the final period. Overt first person pronouns also increase dramatically, ranging from 31% in 1845 to 82% in 1992. Numbers are quite different for the third person: prior to 1937, the average of overt subjects is 28%, and it only increases to a range of 41% to 48% in the more recent texts analyzed by Duarte.

Therefore, over time it is the third person the one which offers more resistance to an overt expressed subject.

If we were to maintain a NSP theory that fully relied on agreement inflectional morphology for the licensing of Romance null subjects, these results would be difficult to explain, particularly those related to third person singular overt subject rates. In Portuguese, as illustrated in Table 3.1, the ambiguity in inflectional forms arises between the second and the third person in the singular, so it is in these persons that there are fewer chances for subject identification by inflection. However, these are not the persons of the paradigm that show the higher rates of expressed subjects, and the third person shows the lowest rate of overt subjects.

Duarte (1995) suggests a possible explanation for these patterns: once full identification from the inflection is lost, the third person pronoun can rely on an NP in the previous context for identification, particularly in embedded clauses. This would be the reason for the stronger resistance to the implementation of the change towards overt subject pronouns in the third person.

Duarte (2000) further investigates the BP null and overt subject pattern with synchronic data. She gathers a sample of oral data with the aim of testing the hierarchy of overt pronoun use found in her diachronic written sample, and to verify the behaviour of third person pronouns. Oral data were obtained from two sources: the contemporary speech of 12 speakers born in Rio de Janeiro belonging to three age groups, and four hours of radio interviews. The analysis of this data corroborates her previous diachronic findings.

Table 3.3. Overt subject rates in synchronic BP

	1STP	2NDP	3RDP
Group 1 (over 46 years of age)	67%	80%	50%
Group 2 (35 to 46 years of age)	79%	94%	65%
Group 3 (25-35) (25 to 35 years of age)	79%	92%	71%

Table 3.3 confirms the differences between first/second person and the third person found in Duarte (1993): the highest rates of overt subjects are found in the second person for all three age

groups, followed by the first person, then the third person. These results also point to a process of diachronic change, as the speakers in the oldest group exhibit the lowest rates of expressed subjects in all grammatical persons, followed by tmiddle-aged and younger speakers. However, differences among groups do not suggest a constant variation: while the rate of expressed third person subjects for Group 1 approaches the rate found in the 1992 popular play analyzed in Duarte (1993), the rates exhibited for Groups 2 and 3 clearly suggest a change in progress.

Duarte also investigates the syntactic environments in which overt preverbal subjects arise in BP in her synchronic sample. In addition to observing that overt pronominal subjects tend to appear in a preverbal position, she also detects three phenomena that are prevalent in her data set: the filling of arbitrary subjects, the tendency to avoid null expletive constructions, and the appearance of left-dislocated structures.

As far as arbitrary subjects are concerned, her results are very similar to Kato and Taralho's (1984), as shown in Table 3.4, which compares strategies of indetermination in BP and EP. This allows us to observe to which extent the BP expression of arbitrary subjects departs from the pattern akin to a consistent NSL displayed by EP:

Table 3.4. Strategies for indetermination in synchronic EP and BP.

BRAZILIAN PORTUGUESE		EUROPEAN PORTUGUESE	
você	44%	você	6%
nós	2%	nos	14%
a gente	13%	a gente	20%
eles	16%	eles	22%
se	8%	se	38%
zero	17%	zero	__

As was the case in Kato and Taralho (1984), structures with the clitic *se* are the least frequent in Duarte's corpus, and she notes that they only appear in the speech of members of Group 1, the older generation. Note the difference with EP, in which *se* structures account for 38% of the occurrences. In BP the preferred pronominal for indetermination is the second person *você* ('you'), followed by third person *eles* ('they') and first person plural *a gente* ('one'), which together account for 73% of the occurrences. The pronominal *nós* only appears at a rate of 2% in BP, confirming that

it seems to be disappearing from the pronominal system of BP. Duarte also finds in a 17% of the occurrences the structure with a null pronoun also discussed by Kato and Taralho (1984) and provided in (4) above. These structures are unattested in EP. She notes, with Kato and Taralho (1984), that the existence of this construction is intriguing, given that it entails the incorporation of a structure with a null form in to a system that elsewhere progresses in the direction of filling null subjects[20].

The examples in (7)-(10) are taken from Duarte's (2000) corpus and illustrate the core strategies for indetermination in BP:

(7) Você quando você viaja, você passa a ser turista. Então você passa a fazer coisas que você nunca faria no Brasil.

Lit.'You when you travel, you become a tourist. So, you start doing things that you would never do in Brazil.'

(8) Quando eles querem eles fazem. Quando eles querem eles acham dinheiro.

'When they want they do (it). When they want, they find the money.'

(9) Hoje em dia, quando a gente levanta as coisas, é que a gente vê tudo o que aconteceu. Mas na época a gente não podia acreditar. A gente não acreditava nisso, primeiro porque a gente era novo.

'Nowadays, when one investigates things, it is that one sees everything that happened. But at that time one could not believe (that). One would not believe it, in the first place, because one was young'.

(10) Antigamente *pro* punha a mesa pra tomar lanche.

In the old times *pro* set the table.3PSG t o have afternoon snack

'In the old times one used to set the table to have an afternoon snack.'

In Duarte's terms, while EP behaves like a variety that uses Paradigm 1 and therefore is able to license and identify *pro* on the basis of agreement, in BP the unmarked form is becoming the

[20] Synchronic theoretical studies on BP (Modesto 2009, Holmberg 2009, Sheehan and Holmberg 2009) have explored the factors that cause the appearance of these constructions. These factors will be discussed in §3.4.

phonological expression of the subject pronoun. According to her, the defective system of null referential subjects exhibited by contemporary BP is to be interpreted as a stage in the process towards becoming a non-NSL.

For this reason, she claims, structures unknown to EP but available in Romance non-NSLs such as French are starting to appear in Brazilian speech. This is the case of "double subject constructions", also labeled left-dislocated structures, which in the corpus gathered by Duarte (2000) account for 7% of the occurrences of overt referential subjects. As first noted by Rivero (1980), left-dislocated subjects in consistent NSLs like EP or Spanish can only appear in configurations with no syntactic adjacency, and they must be interpreted as bearing focus. This is not the case of BP left-dislocated subjects, which have no focal interpretation and may corefer with NPs and other syntactically adjacent pronouns:

(11) A Clarinha$_i$ ela$_i$ cozinha que é uma maravilha.

'Clarinha she can cook wonderfully.'

(12) Eu acho que o povo brasileiro$_i$ ele$_i$ tem uma grave doença.

'I think that the Brazilian people they (he has) have a serious disease.'

(13) Você$_i$, no Canadá, você$_i$ poder ser o que você quiser.

Lit. 'You, in Canada you can be whatever you wish.'

Finally, BP displays another strategy for filling what in earlier stages were categorical null subjects. These are strategies that arise in impersonal or expletive constructions and that afford avoiding null non-thematic subjects. One of them is replacing an impersonal or existential sentence with *haver/ter* and an overt expletive with structures with arbitrary *você*:

(14) a. EXPL Não há/tem mais clientela no centro da cidade.

'There are no longer customers downtown.'

b. Você não tem mais clientela no centro da cidade.

'You no longer have customers downtown.'

Another strategy affects raising verbs, which appear in two structures with two inflected verbs:

(15) a. EXPL Parece que vocês não pensam a sério na vida.
'(It) seems that you guys don't take life seriously.'

b. Vocês parecem que *pro*/vocês não pensam a sério na vida.

In addition, one-argument verbs such as existentials and unaccusatives can appear with two arguments, the second represented by a demonstrative or a topicalized complement of the internal argument NP:

(16) a. EXPL Era em torno de dez pessoas.
'(It) was around ten people.'

b. Isso era em torno de dez pessoas.
that was around ten people
'There were around ten people.'

Finally, extraposed sentential subjects are often copied by a demonstrative in a typical LD construction, as in (17):

(17) a. Que o partido tomou a medida errada EXPL é evidente.
that the party took the wrong decision, *pro* is evident

b. Que o partido tomou a medida errada, isso é evidente.
that the party took the wrong decision, that is evident
'It is evident that the party took the wrong decision.'

As Duarte (2000) notes, the use of the structures in the (b) examples in (14)-(17) co-occurs with their counterparts in (a) that employ empty non-referential pronouns. The latter, according to Duarte, are still far more frequent in speech.

The contribution of Duarte (1993, 1995, 2000) to the description and understanding of the specific NSP profile instantiated by BP can be summarized as follows: first, her quantitative data,

both diachronic and synchronic, clearly show that there is an asymmetry between first/second and third persons as far as the expression of overt subjects is concerned. Even if Duarte herself entertains the hypothesis that overt subject expression is intimately related to the loss of inflectional endings over time, this asymmetry is not easy to account for in a theory that attributes subject licensing to the strength of inflectional morphology only: as pointed out by Duarte herself, inflectional endings, in the tenses where they converge, do so precisely in the second and third person singular; in an agreement-only approach to subject licensing, these would be the persons expected to fill more referential subjects with overt pronouns. Diachronic and synchronic data clearly show that this is not the case, suggesting that strategies other than 'strong' inflection are at play in licensing BP subjects. Second, working with large-scale corpora allows Duarte to accurately describe the specific syntactic environments in which overt subjects are found, and to detect new structures that are making their way into the Brazilian grammar as a result of the restructuring of the pronominal paradigm undergone by this language.

Recent theoretical syntax studies on synchronic BP (Modesto (2000a, 2000b, 2009) Holmberg (2005), Sheehan (2006), Holmberg and Sheehan (2006), Ferreira (2000, 2002), Rodrigues (2002, 2004), among others) take the findings in Duarte (1993, 1995, 2000) as a point of departure to give an account of the NSP profile of BP. In the next section it will be shown how the application of fine-grained minimalist syntactic theory to the data provided by corpus-based diachronic and synchronic studies on BP has produced very fruitful results in explaining BP null and overt subject pattern, in turn contributing to our global understanding of the nature of cross-linguistic variation and the theory of parameters.

3.4. The synchronic view on Brazilian Portuguese.

Linguists working on BP from a historical perspective focus on explaining the causes of the restructuring of the BP grammar, and take this restructuring to instantiate diachronic change in progress pointed towards BP becoming a non-NSL. This view implies that while diachronic change is taking place, two different NSP grammars are available in the mental grammars of Brazilian

speakers: on the one hand, a NSL grammar, that allows for the licensing of null subjects, and a non-NSL grammar, on the other, responsible for widespread overt subjects. This view on diachronic change is essentially the one put forward in Kroch's (1989, 1994) *Competing Grammars Hypothesis*. The core idea in this hypothesis is that these competing grammars will co-exist until one of them 'wins over' the other, and a given parameter is reset to the value represented by the prevailing grammar in the process of first language acquisition.

Note that within this view parameters are conceived in the traditional, Government and Binding way, i.e., as binary devices that when set trigger the fixing of a whole compartment of a language-specific grammar. However, the extant research of the NSP properties of a wide array of languages carried out during the eighties and nineties, as thoroughly discussed in the previous chapter, has shown that this view of parameters needs to be critically evaluated, as there are a number of possibilities for subject licensing in natural languages other than the NSL and non-NSL profiles proposed in initial NSP theories (Chomsky 1981, Rizzi 1982, 1986).

Theoretical studies on the NSP syntax of synchronic BP incorporate the findings of linguistic theory with respect to the diverging NSP profiles offered by natural languages. When these findings are taken into account, a new possibility opens up for explaining the BP null and overt subject grammar: it might be the case that the complex system for subject expression exhibited by BP is not the result of competing grammars in the minds of the speakers (as proposed in, e.g., Kroch 1989, 1994), but that rather this complex patterns emerges from a single and to some extent stable grammar, in which both referential and non-thematic subjects are licensed in such a way that they can be filled or null in specific syntactic contexts.

Moreover, the perspective taken in theoretical studies complements the view developed in diachronic investigations in that they focus on giving an account of the null and overt subject pattern found in contemporary BP, leaving aside the sources of the Brazilian NSP profile. At the core of this approach is the idea that whether or not present-day Brazilian null and overt subject grammar arises as a result of the impoverishment of the inflectional paradigm, a formal theory of

grammar should be able to account for the diverse null and overt subject patterns that arise cross-linguistically, independently of the specific language change path that has led to said contemporary null and overt subject profiles. In other words, in the same way that an accurate formal theory of grammar needs to be able to deal with consistent NSLs as well as with more complex subject licensing systems, it needs to be able to explain the results of diachronic change in a given point in time.

Within this perspective, the Brazilian null and overt subject system could be defined as follows: BP is a language that lacks unrestricted referential null subjects, but in which null subjects are permitted in certain restricted environments. A particularly interesting question for the studies focusing on the null and overt subject syntax of BP is how these null subjects still available are licensed. Again, this question stems from the research carried out in the last two decades on complex NSP patterns. This research essentially showed that null subjects cross-linguistically might arise as a result of quite different requirements[21]. Therefore, the relevant question for studies on dialectal variation in Romance NSP is whether BP null subjects are of the same nature of those arising in consistent NSLs like EP or Spanish.

Recent explorations of the syntax of subject licensing in BP (Kato 1999; Ferreira 2000, 2002; Rodrigues 2002, 2004; Holmberg 2005, Sheehan 2006) agree that BP null subjects are both semantically and syntactically distinct from those in NSLs. The set of contexts in which null subjects in synchronic BP can be found and that were detected in Duarte (1993, 1995, 2000) is summarized in (18)-(21). Recall that these contexts included null non-thematic subjects (18); null generic pronouns (19); a small number (ranging from 6% to 10% of occurrences in 1992) of null subjects in root clauses (20); and finally, a more significant number of third person null subjects (with a average of 28%), which tend to appear in embedded contexts, as in (21):

[21] Recall for instance that while null subjects in consistent NSL languages such as Spanish stem from a requirement such as the Avoid Pronoun Principle (Chomsky 1981) and rely on 'rich' inflectional morphology, whatever the accurate definition for this notion might be, null subjects in partial NSLs like Finnish, particularly in embedded contexts, arise as a result of binding phenomena and cannot be attributed to this 'rich' inflectional morphology or lack thereof.

(18) Null non-thematic subjects: (Kato 1999:5)

a. Tá chovendo. (weather predicate)

be.3RDPSG raining

'It's raining.'

b. ø Tem novidade. (existential construction)

expl have.3RDPSG news

'There are news.'

c. Parece que vai chover (raising construction)

seem.3RDPSG that go.3RDPSG rain

'It seems that is going to rain.'

(19) Null generic pronouns: (Rodrigues 2004:72)

a. É assim que faz o doce.

is thus that make.3RDPSG the sweet

'This is how one makes dessert/This is how dessert is made.'

b. Nesse hotel não pode entrar na piscina bêbado.

in.this hotel NEG can enter in.the swimming pool drunk

'In this hotel it is not permitted to enter the swimming pool drunk.'

(20) First person null subjects in root clauses: (Sheehan 2006:264)

pro Comprei os livros ontem.

bought.1STPSG the books yesterday

'I bought the books yesterday.'

(21) Third person embedded subjects: (Kato 1999:5)

a. O João$_1$ disse que *pro*$_1$ comprou um carro.

The João$_1$ says that *pro*$_1$bought.3RDPSG a car

'João says that he bought a car.'

b. Ninguém$_1$ acha que *pro*$_1$ é estúpido.

nobody$_1$ think.3RDPSG that is.3RDPSG stupid

'Nobody$_1$ thinks that he$_1$ is stupid.'

The data in (18)-(21) show that null subjects in BP are allowed in both root and embedded contexts. However, while root contexts allow for first person null pronouns, in embedded contexts only third person subjects can be null. Second person subject pronouns are almost categorically overt. Furthermore, a null third person pronoun can be selected in constructions with a generic reading. A number of studies explore the sources of these asymmetries. Most argue that both null root and embedded BP subjects differ from consistent NSLs null subjects. However, differing ideas have been proposed as far as the nature of the mixed null and overt subject pattern displayed by BP is concerned, which I discuss now.

3.4.1. Root null subjects.

Kato (1999), Rodrigues (2002, 2004) and Sheehan (2006) propose that BP lacks deictic null subjects in root clauses, in spite of the grammaticality of examples like (22):

(22) (Eu) falei con João ontem a noite. (Rodrigues 2002:163)

I spoke.1PSG with João yesterday night

'I spoke with João last night.'

As pointed out by Figuereido Silva (1996) these null subjects become ungrammatical in constructions that involve fronting of other constituents. This is illustrated for the third person in (23), where the object is fronted for contrastive focus, and in (24), which features a fronted Topicalized phrase, for the first person:

(23) a. Q: O José vai trazer a salada?

Is José bringing the salad?'

b. A: *Não, O VINHO *pro* vai trazer.

No, the wine *pro* go.3RDPSG to bring

b'. A: Não, O VINHO ele vai trazer.

No, the wine he go.3RDPSG to bring

'No, the wine is what he's bringing.'

(24) *O João, *pro* acho que vai ser promovido.

(cf. O João, *eu* acho que vai ser promovido.)

'As for João, I think he is going to be promoted.'

The null subjects in (23) and (24) would be grammatical in EP, a consistent NSL in which null subjects are deictic. Rodrigues (2002) proposes that the BP null subject in (22) essentially akin to the Topic null pronouns allowed in (otherwise non-referential NSLs) languages such as German, Swedish and Dutch. These constructions were already pointed out in Ross (1982). The sentence in (25) makes the case for German:

(25) a. (ich) hab'ihn schon gesehen

I have-him already seen

I have already seen him.'

The pronoun *ich* cannot be omitted in all configurations, however. Specifically, it can only be dropped when it is in sentence-initial, Topic position. As illustrated in (25), if the Topic position is occupied by another element, in this case the object *ihn*, the subject must remain in the third position and cannot be omitted (cf. the BP example in (24) for the same effect):

(26) Ihn hab'*(ich) schon gesehen

him have-I already seen

'Him, I have already seen'

According to Ross (1982), German null subjects are the result of an operation he labels 'Pronoun Zap'. This operation consists of the deletion of the (topicalized) subject pronoun, located in the highest specifier position of the clause. Rodrigues (2002) and Sheehan (2006) essentially adopt Ross' hypothesis for BP; Sheehan (2006) further proposes that these pronouns belong in the specifier of the TopP projection, above CP. According to her, both null and embedded BP subjects

can be derived in the same way: they all bear a Topic interpretation, and as such raise to Spec, TopP via Topicalization. Before I discuss the details of her account, I present the core data on the complex system of embedded null subject licensing in BP.

3.4.2. Embedded null subjects.

Duarte (1995) found an asymmetry between first and third person pronouns in embedded contexts: only third person pronouns can be null in this environment. As noted by Rodrigues (2002), first person pronouns are banned in BP when the Spec, CP is filled. The examples in (27)-(29) illustrate this effect with a subordinate, an interrogative, and a relative clause respectively:

(27) *_Eu_ não sei [$_{CP}$ quanto tempo [moro lá]] (Rodrigues 2002:164)

(cf. _Eu_ não sei [$_{CP}$ quanto tempo [(_eu_ moro lá]])

'I don't know how long I have been living there'

(28) a. *?Quem _pro_ esqueci de citar? (Rodrigues 2002:164)

(cf. Quem _eu_ esqueci de citar?)

b. *?O que _pro_ fiz?

(cf. O que _eu_ fiz?)

'What did I do?'

c. *? Por que _pro_ fui despedida?

(cf. Por que _eu_ fui despedida?)

'Why was I fired?'

(29) *? _Eu_ comprei aquele vestido [$_{CP}$que _pro_ vi ontem]

(cf. _Eu_ comprei aquele vestido[$_{CP}$ que _eu_ vi ontem])

'I bought that dress that I saw yesterday.'

The facts for third person embedded subjects are different. Duarte (2000) found that embedded third person null subjects are optionally null in BP. In consistent NSLs like EP or Non-Caribbean Spanish, the *Avoid Pronoun Principle* (Chomsky, 1981) would require a null subject in a construction like (30); if an overt pronoun is merged, there is a preference for a non-coreferential

reading (Montalbetti 1984). Rodrigues (2002) notes that this is not the case in BP, where merging an overt third person singular embedded subject does not trigger non-co-referentiality:

(30) Os meninos$_1$ ficavam contentes cuando (eles$_1$) tinham um dia de folga
the children were happy when they have-PST3PL a day of holiday
The children were happy when they had the day off.'

Building upon Duarte's work, Modesto (2000a) points out that the use of a null versus an overt pronoun in constructions like (30) is not truly 'optional'. Rather, there is a contrast between EP and BP as far as null embedded subjects are concerned, which stems from the deictic versus non deictic nature of null pronouns in each of these varieties: whereas in EP a null pronoun can be identified from the pragmatic context, by pointing or looking at a photograph, in BP only linguistic antecedents, and not pragmatic ones, can bind null embedded subjects. This contrast is illustrated in (31):

(31) Mr. A comes back home from the doctor. Mrs A. says: "Tell me what he said".
Spanish: Dime qué (te) dijo.
Tell-me what to you.CL said-3SG
BP: Me diz o que *(ele) falou.
To-me say the what he spoke
'Tell me what he said.'

The nature of BP null subjects has been investigated in Negrão (1997), Modesto (2000a, 2000b, 2009), Rodrigues (2002, 2004) and Sheehan (2006). The following properties have been ascribed to these structures:

Null subjects are not only licensed in embedded CPs when they are bound by a referential matrix subject, but also when they are coreferential with a quantifier phrase (QP) subject:

(32) a. O Pedro disse que pro $_{1/*2}$ ele $_{1/2}$ ganhou na loto.
The Pedro said that he won on-the lottery
'Pedro said that he won on the lottery.'

b. Ninguém disse que pro $_{1/*2}$ / ele $_{*1/2}$ ganhou na loto.

said that he won on-the lottery

'Nobody said that he won on the lottery.'

Interestingly, while normally embedded null subjects need to be bound by an antecedent in the higher clause, a question like 'What about Paulo?' allows for a null subject in the answer to pick up the linguistic antecedent:

(33) a. Q: E o Paulo?

And the Paulo

'What about Paulo?'

b. A: A Maria$_2$ disse1 que EC$_1$ estava doente.

b'. A: *ec*$_1$ está doente.

In the rest of the contexts studied, there is a c-command requirement: as noted by Negrão (1997), only c-commanding antecedents can act as antecedents of embedded null subjects in BP:

(34) A opinião de Zé$_1$ é que *(ele$_1$) vai ganhar (Sheehan 2006:275)

The opinion of Zé is that he goes to-win

'Zé's$_1$ opinion is that he$_1$ is going to win'

In addition, the closest c-commanding DP binder must bind the embedded null subject:

(35) a. *O João$_1$ disse [que a Maria$_2$ acha [que *pro*$_1$ é esperto]]

The João said that the Maria thinks that is smart

'João said that Maria thinks that *he/she is smart.'

b. *O José acha [que os rumores revelam [que *ec* é inteligente]]

The José thinks that the rumours reveal that is intelligent

'José thinks that the rumours reveal that he is intelligent'

It seems that only matrix subjects and not objects can act as binders:

(36) a. O Pedro$_1$ convenceu o João$_2$ que *ec*$_{1/*2}$ tinha que ir embora
Pedro$_1$ convinced the João that had-to go away
'Pedro$_1$ convinced João that he$_1$ had to leave.'

b. O Pedro$_1$ convenceu quem$_2$ que *ec*$_{1/*2}$ tinha que ir embora
The Pedro$_1$ convinced who$_2$ that he$_1$ had-to go away
'Who did Pedro convince that he had to leave?'

Raposo (1986) notes that embedded subject licensing in BP differs both from consistent NSLs and from non-NSLs. In the former embedded null subjects are not required to be c-commanded by the closest linguistic antecedent, and in the latter null pronouns are not available in embedded clauses. Raposo proposes, following Huang (1984), that embedded null subjects in BP are licensed they same way in which dropped Topics are licensed in East Asian languages; i.e., they are actually variables bound by a null topic.

Modesto (2009) builds upon Raposo's proposal and suggests that synchronic BP is a Topic-prominent language, and he argues that as such it shares other properties with languages like Chinese. Rodrigues (2002, 2004) and Sheehan (2006), Holmberg and Sheehan (2009) also ascribe to BP an NSP profile that differs both from the pattern displayed by non-NSLs like English and from the one displayed by consistent NSLs like EP. However, rather than describing BP as a Topic-drop language, they define it as a partial NSL, therefore closer to languages like Finnish.

3.4.3. BP as a Topic-prominent language.

Modesto (2009) takes as a point of departure Holmberg's (2005) proposal on the nature of null subjects in non-consistent NSLs. Recall that Holmberg advocated for a view of the NSP, which he labeled *Hypothesis B*, according to which agreement morphology is always uninterpretable, and therefore it needs to value and delete its uninterpretable features via Agree. Subject raising to Spec, TP makes possible that this Agree relationship is established between T and the null subject, which in his view is a phi-phrase, i.e. a ϕP; ϕPs are bundles of interpretable phi-features with no PF form

that lack the deictic/referential [D] feature and hence they cannot refer. They incorporate the [D] and therefore referential when entering Agree with T.

Modesto (2009) follows Holmberg in defining in BP null embedded subjects as instances of ϕPs. He departs from him in claiming that in BP the ϕP is rendered interpretable not by entering Agree with T via subject raising to Spec, TP, but rather by being bound by a higher A-bar binder in the CP domain. This A-bar binder is the subject in the preceding clause. Note then that in Modesto's account BP subjects are located in an A-bar position higher than Spec, TP. Modesto labels this high, A-bar position FP[22], and proposes that Spec, FP has an EPP feature that requires it to be filled:

(37) $[_{FP}$O Pedro$_1$] $[_{TP}$disse $[_{CP}$que $[_{FP}$ ϕP$_{1/*2}$ $[_{TP}$ganhou na loto]]]].

The Pedro said that *pro* won on-the lottery

'Pedro said that he won on the lottery'

Modesto's proposal is then that synchronic BP is neither a consistent-NSL like EP, nor a partial NSL like Finnish. Rather, it is a Topic-prominent language, closer in this respect to East Asian languages. Consequently, the higher subject in Spec, FP is allowed to surface without a theta-role, since it is a Topic. This subject moves from its based-generated position to Spec, FP – presumably via Spec, TP- where it checks the EPP feature on F. To sustain his proposal, he provides evidence that BP shares some of the properties associated with Topic-prominent languages first described by Li and Thompson (1976), in addition to the core property of controlling co-referential constituent deletion:

(i) There is a "surface coding" for topics in Topic-oriented languages; Topics are coded by a special topic marker (Japanese, Korean) or appear in sentence-initial position (Chinese, BP):

[22] The "F" in FP stands for "Functional", and describes a functional projection above TP that carries the informational content of the clause. It is reminiscent of the F category put forward in Uriagereka (1995) to host a variety of discourse-related elements, such as overt expletives in NSLs that allow for them (Galician, European Portuguese, among others). The role in this position in Caribbean and non-Caribbean Spanish will be discussed in Chapter Four.

(38) a. ?? Pode nadar neste lago (Modesto 2009:404)

may swim.INF in.this lake

b. Neste lago pode nadar

In.this lake may swim.INF

'One may swim in this lake'

(ii) Topic-prominent languages have "double subject" constructions, which are used instead of the null expletive constructions licensed in NSLs:

(39) a. Os professores$_1$ parecem que eles$_1$ gostam da Maria

the teachers seem.3PPL that they like of.the Maria

'The teachers seem to like Maria'

b. Cê$_1$ parece que cê$_1$ sabe a resposta.

you seem that you know the answer

'You seem to know the answer'

(iii) Topic-prominent languages tend to be verb-final. East-Asian languages like Japanese and Korean are typically described as verb-final. Modesto claims that ta similar phenomenon in BP, in the form of marking transitive verbs into intransitives, which make structures verb-final:

(40) O prédio já construiu. (Modesto 2009:404)

the building already constructed.3PSG

'The building has already been constructed'

While Modesto (2009) presents an analysis that captures the main facts related to the BP complex system of embedded null subjects, an alternative account has been put forward in the recent literature, with the intention of accounting not only for null embedded subjects, but also for root null subjects in a principled way.

3.4.4. Brazilian Portuguese as a partial Null Subject Language.

Rodrigues (2002, 2004), Sheehan (2006) and Holmberg and Sheehan (2009) develop an alternative analysis to Modesto's in which BP null subjects are derived not by A-bar binding, but rather by A-movement.

In this theory BP null embedded subjects, like English overt subjects, raise to Spec, TP to check T's uninterpretable features. Successive cyclic movement then lands them in Spec, TopP. Sheehan (2006) further proposes that a PF-interface requirement on the deletion of pronouns is responsible for the null spell out of these subjects. The reasoning behind this analysis is as follows: Sheehan (2006) analyzes the advantages of Modesto's (2000) and Rodrigues' (2002, 2004) analyses, and she concludes that in fact both theories are able to derive BP system of third person embedded null subjects. However, an A-movement analysis of the BP facts affords accounting for these embedded subjects and the root null subjects discussed §3.4.1. in a unified way. According to Sheehan, essentially all BP null subjects arise as a result of a process of Topicalization that lands root and embedded subjects in Spec, TopP.

Sheehan (2006) also points out that an A-movement analysis of the facts is preferable on purely theoretical grounds; building up on Hornstein's (1999, 2003) theory on anaphoric relationships, she states that under the parsimonious requirements of the Minimalist Program, an anaphoric construal relationship should be derived via movement. Hornstein focuses on Obligatory Control and he submits that it is derived via A-movement, being essentially identical to raising, with the only difference being that in Obligatory Control the moved DPs pick up their theta-role through Agree[23]. In this way, many of the properties of Obligatory Control stem from the inherent properties of movement without further stipulation:

[23] It is beyond the scope of this chapter to discuss the benefits and consequences os adopting Hornstein's control-as-movement analysis, where theta-roles can be 'picked up' as features instead of being assigned as 'relations', as Chomsky proposes. For a long, ongoing debate on the issue see Hornstein (1999), Landau (2000, 2004), Manzini & Savoia (2002) and references therein.

(41) a. John promised [<John> to leave]

b. John seems [<John> to be happy]

Taking this theory as a point of departure, Ferreira (2000, 2004), and Rodrigues (2002, 2004) both independently propose a movement account of null subjects in BP. While in English movement is applied to infinitive constructions, in BP it targets finite CPs. As BP appears to license hyper-raising, which is raising from finite CPs, as shown by Duarte (1995), this extension is appealing:

(42) O José parece que está doente (Sheehan 2006:273)

The Jose seems that he is ill

'José seems to be ill'

Ferreira (2004) and Rodrigues (2004) also derive the semantics of BP null subjects. Their bound-variable behavior can be taken as evidence for movement. Recall that Modesto's A-bar binding analysis also predicted that null subjects would be interpreted as variables, and also that only c-commanding antecedents can act as antecedents. Under an A-movement account, null subjects must have a minimally c-commanding antecedent and as such cannot take split antecedents, as showcased in (43) -(44):

(43) *João disse [que a Maria acha [que *ec* é esperto]]

John said [that Mary thinks [that *ec* is smart-MASC]]

'John said that Maria thinks that he is smart.'

(44) *João disse [que a Maria acha [que *ec* são espertos]]

João said [that Mary thinks [that *ec* are smart-PL]]

'João said that Mary thinks that they are smart.'

The A-movement analysis can also capture difference between BP and English/German. In all three languages empty pronouns arise as a result of a process of Topicalization that lands null

subjects in Spec, TopP. In BP, in addition, Topicalization of a subject from an embedded finite clause to the matrix clause is allowed, via successive cyclic movement:

(45) a. Eu$_1$, a Maria disse que *ec*$_1$ comprei um carro muito caro.
I, the Maria said that bought.1PSG a car very expensive
'Me, Maria said that I bought a very expensive car.'

b. A Maria$_1$, o José disse que *ec*$_1$ comprou um carro
The Maria, the José said that bought.3SSG a car
Maria, José said that she bought a car.'

This seems to be good evidence that subject Topicalization, unlike object Topicalization, is derived via *wh*-movement in BP. Consequently, it is sensitive to strong islands (relative clauses and adjuncts) but less so to weak islands (*wh*-islands and factives) (Figuereido Silva 2000):

(46) ?A Maria$_1$, o João$_2$ não sabe [pra quem *ec*$_1$ vai dar o livro]
The Maria, the João not knows for who goes give the book
'Maria, João doesn't know who she'll give the book to.'

(47) ?A Maria$_1$, o João acha uma pena [que *ec*$_1$ vendeu o carro]
The Maria, the João finds a shame that sold the car
'Maria, João finds it a shame that she sold the car.'

(48) *A Maria$_1$, o João$_2$ achou [um carro que *ec*$_1$ tem grana pra comprar]
The Maria, the João found a car that has money for buy
'Maria, João found a car that she has the money to buy.'

(49) *A Maria$_1$, o João$_2$ olha pro pé [toda vez que *ec*$_1$ fala com ele]
Maria, the João looks for-the foot every time she speaks with her
'Maria, João looks at his feet every time that she speaks with him.'
(Figuereido Silva 2000:136)

Similar effects are true of *wh*-movement in BP (Negrão 2000):

(50) Que aluno você não sabe quando vai entregar o trabalho?

What student you neg know when goes hand-in the work

Lit. 'Which student don't you know when will hand in the work?'

(51) *Quem$_1$ o João saiu da festa antes que t_1 tivesse entregado o presente?

Who$_1$ John left of the party before that t_1 had given the present

According to Sheehan (2006), if null Topics are the result of Topicalization, a pronoun might be said to enter the derivation in the normal way and then move to Spec, TopP, where it is deleted by virtue of its high position. It should be explained, however, why embedded subjects and not objects are visible to probes in a higher clause in BP:

(52) *[O João falou [$_{CP}$ que [$_{TOPP}$ <O João> [$_{IP}$ o Pedro viu ~~<O João>~~]]]]

João said that the Pedro saw

João said that Pedro saw him'

It also needs to be explained why in BP, and not in English-type languages, the subject of a finite CP is still 'active', but not able to delete the phi-features of T. Rodrigues (2002, 2004) proposes that the verbal morphology in BP is rich enough to satisfy the EPP in BP, but not to delete the phi-features on T, and attributes this to the loss of inflectional endings documented by Duarte (1995). An overt subject is therefore required to value and delete the phi-features in T. In addition, any DP subject must receive structural (nominative) Case, which is also achieved via subject movement. Rodrigues (2004) formalizes this requirement as follows:

(53) a. A structural Case feature is checked in a Spec-head relation.

b. Agreement in phi-features prompts movement to check Case.

(Rodrigues 2004:123)

As Chomsky (2000) points out, there is no reason why Case and EPP checking need to be one and the same process. In fact, there are many well-known examples where the two are separate

processes. For instance, in English, the expletive 'there' seems to satisfy the EPP but does not absorb Case:

(54) There entered a sinister-looking man. (Sheehan 2006:294)

Sheehan's (2006) thesis is that if null subjects in BP arise as a result of Topicalization, i.e. A-movement, BP is best characterized not as a discourse *pro*-drop language akin to East-Asian languages, as in Modesto (2009), but rather as an instantiation of partial *pro*-drop in Romance.

Holmberg (2009) presents further arguments for characterizing BP as a partial NSL. He points out to the existence of null generic subjects in BP. These constructions were already presented in Kato and Taralho (1984). Recall the example in (19) above, repeated now as (55):

(55) É assim que faz o doce.

is thus that make.3RDPSG the sweet

(cf. EP 'É assim que **se** faz o doce.')

This is how one makes dessert/This is how dessert is made.'

The existence of this null subject construction in a language like BP, which otherwise is shown to move towards extending the use of overt subject pronouns, was seen as puzzling in Duarte (1995). This is particularly so if we take into account that in consistent NSLs like EP or Non-Caribbean Spanish generic meaning is conveyed by use of the overt clitic 'se'.

Holmberg (2009) points out in this respect that null generic pronouns are possible in other partial NSLs, Finnish being a case in point:

(56) Tässä tuolissa istum mukavasti.

this.IN chair.IN sits comfortably

One can sit comfortably in this chair.'

This fact leads him to suggest that the key difference in the feature assembly of T that distinguishes consistent NSLs from partial NSLs might not be the [+referential] feature usually associated with consistent NSL inflection in traditional (Rizzi 1982, 1986) and minimalist (Chomksy 1995) theories, but rather [definiteness]. In other words, while consistent NSLs have a

[definite] phi-feature as part of the feature make up of T, in partial NSLs the Agree relationship established between T and the null subject does not supply the latter with a definiteness value. As a result, what is obtained in partial NSLs is an indefinite, generic, null subject pronoun. Conversely, in consistent NSLs a null definite subject cannot be used to convey a generic meaning, and hence the insertion of the clitic 'se' in EP and Non-Caribbean Spanish[24].

Going back to Sheehan's (2006) theory on BP, she submits that the theory she proposes for Null Subject Romance can also explain the BP null and overt subject pattern. Recall that she provided a NSP theory based on the interaction of two different parameters: on the one hand, the 'rich' agreement parameter, where 'rich' agreement morphology is conceived as a surface of a [uD] feature on T; on the other, the PF-sensitive Spec, TP parameter (cf. (48) in §2.3.4.3.). The former regulates subject deletion, while the latter constrains the presence or absence of expletives in a given language. Evidence for the separate nature of these two parameters is that there are languages which do not license referential *pro*-drop but allow for null expletive subjects, such as German and Icelandic, on the one hand, and a number of Creole languages, on the other. Further cross-linguistic variation in word order can be attributed to differences in prosodic interface conditions, and the way focus is assigned. These two parameters capture the facts in consistent-NSLs and null expletive languages.

The BP facts show that there is a third parameter related to Case assignment that is also involved in subject licensing cross-linguistically. Therefore, Sheehan (2006) concludes that in order to account for partial NSLs three (micro) parameters must be involved in null and overt subject licensing cross-liguistically:

[24] Recall that Holmberg (2009) refers here to generic inclusive null pronouns, i.e., null subjects that denote people in general, including the speaker and the addressee and are translated in English as 'one', and in other languages as second person plural 'you'. Generic exclusive readings, which denote people in general in some domain that do not include the speaker or the addressee and are translated into English 'they', can be conveyed in NSLs like Non-Caribbean Spanish with (third person plural) null pronouns.

(57) *The Null Subject Parameter:*

(i) *The rich agreement parameter*

T lacks/bears a [uD] feature.

(ii) *The PF-interpretability parameter*

T lacks/bears a *, where * requires the specifier of I to be spelled out at PF.

(iii) *The Weak/strong nominative Case parameter*

Nominative Case feature [uI] lacks/bears *, where * requires movement to Spec, TP.

The null and overt subject profile of BP can be defined as follows: BP has a 'negative' setting for the NSP, that is to say, it lacks a [uD] feature on T, meaning that the deletion of referential null subjects under non-distinctness is blocked. However, BP lacks a PF requirement for Spec, TP and so licenses null locatives and expletives. BP also allows movement from finite clauses via a peripheral Topic position. This gives rise to null bound embedded subjects. As such, the null subjects licensed in BP are also the result of deletion. It can thus be maintained that BP has an EPP requirement of the type 'Merge XP', as has been argued for Null Subject Romance. Moreover, in BP, nominative Case is strong, as it is in English, EP, and French, triggering movement of the subject to Spec, TP. Spec TP is only empty at PF where it contains a deleted expletive or locative or the deleted copy of a moved DP.

In sum, the study of non-standard null and overt subject patterns, and particularly the investigation of dialectal variation, has not only been relevant for confirming or rejecting views on the NSP, but has also emerged in recent linguistic theory as tool for defining the nature of parameters, as illustrated by the presentation of the differing views on the BP subject licensing system presented in this chapter.

3.5. Summary

This chapter has focussed on null and overt subject licensing in BP. The Brazilian pattern of subject expression presents a number of similarities with the one described in the literature for DS: in both varieties overt pronominal subjects appear in environments that would require a null pronoun in consistent NSLs like EP or Non-Caribbean Spanish; in BP as well as in DS overt subjects are not only widespread, but they also appear in specific word orders and syntactic structures, some of which are shared by both varieties. DS and BP differ in significant respects as well. For instance, it does not seem to be the case that the DS pronominal paradigm has undergone the shifts documented for BP, and specific BP constructions, such as structures with null generic pronouns, have not been documented for DS.

While the investigations of the syntactic variation attested in Dominican Spanish approach the facts from a formal point of view only, combining formal hypotheses with quantitative data has proven very fruitful in the study of Brazilian Portuguese. In the next chapter, it will be argued that this cohesive approach is best suited to explain the Dominican Spanish null and overt subject system, and to probe into the questions that the Dominican NSP profile rises for recent theories on parametrization.

Chapter Four. Variationist views on Spanish subject expression.

4.1. The quantitative paradigm and Spanish subject expression.

In this chapter I review variationist approaches to subject expression in Spanish. Studies that employ quantitative methodology differ in two basic respects from purely theoretical, syntactic investigations: first, in the variationist paradigm theories are elaborated with reference to naturalistic corpora, and not with reference to data gathered in experimental settings; second, these studies, rather than focusing on a single constraint or area of the grammar involved in the specific nature of a given linguistic phenomenon, examine simultaneously the multiple language-internal and language-external variables that may play a role in the configuration of a grammar.

Theoretical and quantitative approaches to syntactic variation converge on the finding that while the rich morphological system of Spanish is involved in null subject licensing, a number of discourse-pragmatic and lexical constraints are also fundamental in shaping subject expression phenomena. In spite of differences in research agendas, both approaches have faced similar challenges in this respect. Defining the specific nature of 'rich' agreement has proven elusive in theoretical studies (cf. Chapter Two). In variationist investigations, it has been hypothesized that the differing degrees of 'rich' subject-verb agreement that can be observed either in cross-dialectal variation or within the pronominal elements of the paradigm in the same variety are key for explaining differing patterns of subject expression (Hochberg 1986); this hypothesis, however, has been convincingly disproven (Cameron 1992, 1993, 1996). In fact, grammatical properties at the syntax-pragmatic interface have been consistently shown to be among the strongest constraints on subject licensing in all the Spanish varieties studied up to date (Otheguy, Zentella and Livert 2007). As in theoretical investigations, the study of Caribbean restructured varieties that have lost inflectional distinctions over time has emerged as a suitable venue for exploring the role of 'rich' inflectional morphology or lack thereof in null and overt subject insertion.

In this chapter, I discuss variationist accounts of Spanish subject expression in Mainland and Caribbean varieties to point out the contributions that they have made to our understanding of subject licensing phenomena in synchrony and diachrony. I argue that a number of facts examined in variationist studies but usually overlooked in purely theoretical work can shed light on the investigation carried out in this dissertation.

First, variationist studies have consistently shown that subject expression is variable in all Spanish dialects, not only in varieties that have been described as undergoing restructuring in their subject grammars. This implies that specific phenomena related to subject expression are more accurately described in probabilistic rather than in categorical terms. By way of an example, as the discussion in §4.1.1. will make clear, discourse-related constraints on subject expression manifest in environments of low and high variability, rather than in environments in which null or overt subjects are categorically selected.

Second, it will be shown that the variationist framework allows for a detailed comparison of closely related varieties, as it is particularly well suited to modelling the subtle regularities and tendencies in a data set. It is argued throughout this dissertation that this sort of fine-grained description is instrumental for disentangling dialectal differences, which in turn allows for accurate theoretical explanations of restructured varieties like DS.

As explained by Tagliamonte (2002), the quantitative variationist methodology rests on the premise that the phonological, lexical, syntactic and discursive features that shape the grammar of a given speech community vary in a systematic way; more specifically, the central idea is that whenever speakers have a choice available to them in speech, this choice may be constrained by a number of factors, and the role that each factor has in the speakers' selection can be modeled quantitatively (D. Sankoff 1988, Young and Bayley 1996). Consequently, the variationist methodology is designed for the systematic observation of the many different constraints that influence the expression of linguistic phenomena, and for disentangling the relative strength of each.

This systematic observation is achieved through multivariate analysis. This type of analysis measures the impact of a given constraint in variation, relative to other constraints that apply in the same domain. For this reason, more than focusing on individual occurrences or overall rates of occurrence, multivariate analyses focus on patterns of variability, i.e. the 'distribution' of linguistic features, defined as "where a feature occurs in a language, as determined by the relative frequency of the feature across its different contexts of use" (Tagliamonte 2002: 732). In variationist theory, this distribution is taken to reflect underlying grammatical structure. Comparisons across linguistic systems are therefore comparisons of the distribution of linguistic features.

Variationist analysis takes as a point of departure a 'dependent variable', which represents the linguistic phenomenon under study, and any number of 'independent variables' or 'factor groups', which represent the possible constraints on the expression of said linguistic phenomenon. In the study of subject expression, the dependent variable is the form that the subject can take (null, overt pronominal subject and overt DP subject). Independent variables or factor groups are for instance the *Person* of the subject, the *Tense* of the verb that co-occurs with a given subject, or *Topic continuation*, i.e. the discourse status of a subject. Each factor group includes a set of factors or 'constraints'; these factors represent the different contexts within a factor group in which the linguistic feature under study can appear. For instance, for the factor group *Person* first, second and third person, singular and plural, are constraints; for a factor group like *Tense* factors or constraints are all the verb tenses that may be used with a null, pronominal or DP subject; *Topic continuation* includes as factors continuity of Topic, Topic shift and emphasis.

Each factor group selected for a study tests a hypotheses put forward in the literature to explain the distribution of a linguistic feature. By way of an example, it has been found in theoretical studies that while Topic subjects tend to be null, new or focused subjects, as well as emphatic ones, tend to be overt in Spanish. This finding would be tested in the data set at hand with the factor group *Topic continuation*.

Tagliamonte (2002) explains that multivariate analysis of a data set returns three types of measures:

(i) Statistical significance of the analysis,

(ii) Relative strength of each factor group as compared to the rest of factor groups in the application of the dependent variable, and

(iii) Constraint hierarchy, i.e., relative ordering of the constraints within a factor group.

The measures in (i)-(iii), and particularly the examination of the relative strength of factor groups and constraint rankings, allow us to define the regularities and asymmetries in a data set. For this reason, as stressed by Tagliamonte (2002), much of the work carried out in variationist linguistics is comparative, as factor and constraint rankings provide us with the information to assess to what extent two given languages, or a single language in different historical contexts, share an underlying grammar. In other words, the measures in (i)-(iii) allow us to decide on the nature of variability across and within languages.

More specifically, factor groups in the application of a given independent variable are expected to tease apart grammatical systems. Consequently, they are conceived as the fundamental units of language change, and as such diachronic shifts are reflected in "an ordered shift" (Labov 1982:76), over time, in the strength of factor groups. Constraint rankings, on the other hand, are taken to provide diagnostics for comparisons across closely related grammatical systems (Tagliamonte 2006).

Studies of Spanish subject expression confirm this view of cross-linguistic and cross-dialectal variation and change: while hierarchies of factor groups have been found to be stable across varieties, constraint hierarchies have been proven to be heterogeneous in language contact and language change contexts (Otheguy, Zentella and Livert 2007). This effect is particularly visible in the comparison between Mainland and Caribbean Spanish varieties, as the discussion in §4.2. will make clear.

Subject expression in Spanish has been examined in a range of varieties and contexts: Peninsular varieties in Spain (Miró-Vera and Pineda 1982, Enríquez 1984, Ranson 1991) as well as the Spanish of specific regions of Argentina (Barrenechea and Alonso 1977), Chile (Cifuentes 1981), and Mexico (Montes-Miró 1986) have been studied. Among Latin American varieties, Caribbean dialects have garnered a good deal of attention: specific phenomena in Venezuelan Spanish are examined in Bentivoglio (1987), Sedano (1990); Puerto Rican Spanish is studied in Cameron (1992, 1993, 1996), Morales (1980, 1986 1992, 1997), Lizardi (1993), Ávila Jimenez (1996) and Flores-Ferrán (2004); Coastal Colombian dialects are looked at in Orozco (2007), Orozco and Guy (2008), Travis (2005, 2007). Situations of language contact have also been examined: de Prada (2009) studies Spanish-Catalan contact in the island of Minorca, Spain, while an number of investigations probe for the impact of Spanish-English contact in the US (Silva-Corvalán 1984, 1992, Otheguy and Lapidus 2005, 2009, Otheguy, Zentella and Livert 2007, Flores-Ferrán 2002, 2004 among others).

Comparisons across studies are not easy, as stressed by Cameron (1993), Travis (2007) and de Prada (2009): in addition to the difficulty of defining contexts of null and overt subject variation in Spanish, different factor groups are examined across studies. Furthermore, while in a few studies a full-fledged variationist analysis of subject use is carried out, others offer only percentages of null and overt subject use. Even so, some general trends have been found in subject use:

As far as language-external factors such as *Age*, *Gender* or *Occupation* are concerned, most studies do not find them significant (Bayley and Pease-Álvarez 1997, Bentivoglio 1987, Cameron 1994, Enríquez 1984, Flores-Ferrán 2002, Hochberg 1986, Ranson 1991, Silva-Corvalán 1994). Exceptions to this trend have been found in language contact studies (Otheguy and Lapidus 2005, Otheguy, Zentella and Livert 2007, de Prada 2009) and in a few studies on Caribbean dialects (Lizardi 1993, Ávila-Jiménez 1996). Another exception to this general pattern of non-significance for language-external constraints is *Genre*, which distinguishes between conversational and

narrative discourse. This constraint has been found significant in the small number of studies that have examined it (Flores-Ferrán 2002, Travis 2007).

Regarding language-internal factors, all studies converge in finding significant discourse-related constraints on subject use in speech. In short, as noted by Travis (2007), a 'continuity' effect is observed: continuity in subject reference, subject form and tense of the verb essentially trigger null subject use (Cameron 1994, Cameron and Flores-Ferrán 2004, Flores-Ferrán 2002). Among the discourse-related factors associated with continuity, *Coreference*, also labeled *Switch Reference*, has been found significant in practically all varieties examined to date (Ávila-Shah 2000, Bayley and Pease-Álvarez 1997; Bentivoglio 1987, Cameron 1994, 1995, Enríquez 1984, Flores-Ferrán 2002, 2004, Hochberg 1986, Morales 1986, Silva-Corvalán 1982, 1994, Travis 2005, among others). Other discourse-related factors not necessarily associated with continuity effects, such as *Clause Type*, are significant in some studies (Morales 1997), but not others (Cameron 1992).

Lexical factors, such as the semantic nature of the subject, have also been found to constraint subject use (Bentivoglio 1987, Enríquez 1984, Silva-Corvalán 1994, 2007, de Prada 2009). Morphosyntactic factors also have an effect on subject expression. *Person of the subject* has a significant impact on subject use: generally, singular subjects favor overt subjects, while plural persons of the paradigm present the lowest rates of overt subject use in Caribbean and Non-Caribbean dialects (Cameron 1992, Travis 2007). However, the relative role of a closely related variable, *Verb ambiguity*, is highly controversial. In early studies of Spanish subject expression (Hochberg 1986), it was hypothesized that more overt subjects were to be found in the persons of the paradigm that corresponded to morphologically ambiguous verb forms. Accordingly, in Caribbean varieties more overt subjects were expected in second and third person singular forms, since these were the persons of the paradigm that showed more morphological ambiguity as a result of dropping final –s in speech. This is usually referred in the literature as the *Functional Compensation Hypothesis*. While *Verb ambiguity* has been returned as significant in a few studies

(Travis 2005, 2007, for instance), the accuracy of the *Functional Compensation Hypothesis* has been questioned in the relevant literature (Cameron 1993, 1996).

The role of morpho-phonological constraints in Caribbean subject expression will be addressed in §4.2.2. Before discussing this and other factors relevant to the examination of subject use in Dominican Spanish, I outline the different types of variables and constraints that regulate subject use across dialects of Spanish.

4.1.1. Discourse-related variables.

Among discourse-related variables, *Co-reference*, *Topic Continuation* (also referred to as *Discourse Function*), and *Distance from Previous Mention* exert significant effects on subject expression in most investigations. Also related to discourse and discussed in previous work are *Clause Type*, *Tense-Mood-Aspect (TMA) Continuity*, and *Priming*. In what follows I discuss them in turn.

The variationist study of the role of continuity of reference has contributed significantly to our understanding of subject licensing in Spanish: on the one hand, quantitative examination of the role of *Co-reference* has shown that while coreference favors null subjects and non-co-reference disfavors them (Silva-Corvalán 1994, Bayley and Pease-Álvarez 1997, Flores-Ferrán 2004, Otheguy, Zentella and Livert 2007), this is more a tendency than a categorical rule. On the other, the investigation of continuity of reference in speech has shown that constraints beyond the truth-conditional semantics of the sentence are relevant for variable subject expression.

The role of *Co-reference* has been examined in detail in Caribbean dialects, as discussed at length in 4.2.2.2. below. In most studies this is construed as a binary factor group, indicating whether the preceding and target subject share the same referent[25]. Cameron (1992), Flores-Ferrán (2002), however, define it as a scalar variable, taking into account whether a target plural subject partially co-refers with the preceding subject form. In general, whether *Co-reference* is constructed

[25] In the ensuing discussion, 'target subject' refers to the specific instance of a null, pronominal of full DP subject examined in a given context.

as a binary or as scalar variable, it has been found to be a very strong factor in both Caribbean and Non-Caribbean Spanish, as well as in contexts of language contact: for instance, in Los Angeles Spanish (Silva-Corvalán 1982), Venezuelan (Bentivoglio 1987) and Puerto Rican Spanish (Cameron 1992), 70% to 80% of overt subjects correspond to contexts of non-co-reference.

However, even if non-co-reference strongly favours overt subject use, it only accounts for a portion of the overall number of overt subjects: Bentivoglio (1994), Flores-Ferrán (2004) and Travis (2007), among others, have shown that co-referential subjects are expressed overtly at rates of 20% to 40%, depending on which Spanish dialect is looked at (Travis 2007). Among the attested examples of overt co-referential subjects:

(1) a. Ahí estaba *yo*, comenzaba *yo* a hablar inglés. (Travis 2007:107)
there was.1PSG I, started.1PSG I to speak English
'There I was, I started to speak English.'

b. *Yo* no sabía cuánto *yo* quiero a mi hermano.
I not knew.1PSG how much I love.1PSG to my brother
'I didn't know how much I love my brother.'

Recall from the discussion in Chapter Two (cf. §2.3.4.1.2.3) that it was established in theoretical studies that overt and null pronouns have different properties in NSLs and non-NSLs; specifically, NSL pronouns bear a [+referential] feature, which makes them potentially anaphoric. Montalbetti (1984) formalized this property in the *Overt Pronoun Constraint*. This constraint states that whenever there is a contrast between overt and null pronouns, the overt pronoun cannot be bound as a variable. The relevant examples provided in Chapter Two are repeated here:

(2) a. Todo estudiante$_i$ cree que *pro*$_i$ es inteligente. (Non-Caribbean Spanish)
every student thinks that *pro* is intelligent
'Every student$_i$ thinks that he$_i$ is intelligent.'

b. Todo estudiante$_i$ cree que él$_{*i/j}$ es inteligente.

every student thinks that he is intelligent

'Every student$_i$ thinks that he$_j$ is intelligent.'

De Prada (2009) stresses that the finding in variationist work that reference constraints are variable rather than categorical argues in favor of applying quantitative methodology in the study of subject expression. It should be added that applying variationist methods to a large corpus of data would allow for the examination of the shifts that the workings of co-reference have undergone in restructured varieties. This would make it possible to assess whether the differences in numbers of expressed co-referential subjects are derived from the interaction of other constraints, or whether they reflect change in progress that involves changes in the feature specification of pronouns in specific (Caribbean) dialects.

Variationist work has also shown that the role of continuity, in subject form, in discourse function, and in verb form, exceeds the clausal domain. Specifically, null subject insertion has been found to correlate with the general notion of continuity. Factor groups such as *Topic Continuation*, *TMA Continuity*, *Speech Connectedness* and *Priming* lie within this notion.

Topic continuation is designed to examine whether the referent of the target subject has been mentioned in the previous discourse or not. With respect to the continuity of Topic, subjects are either *Topic* (3a), or they represent *Topic Shift* (3b) or *Emphasis* (3c):

(3) a. Terminé el trabajo y me fui a cenar.

finished.1PSG my work and CL.went.1PSG out to dinner.

'I finished my work and went out to dinner'

b. Estábamos cenando cuando llegó ella con su novio.

were.1PPL having dinner when arrived.3PSG she with her boyfriend.

'We were having dinner when she arrived with her boyfriend.

c. YO traje el vino, y ellos las cervezas.

I brought.1PSG the wine, and they the beers.

'I brought the wine and they brought the beers.'

As illustrated in (3a-c), while continuity of *Topic* tends to favor null subjects, *Topic shift* usually entails overt subject use. *Emphasis*, as in (3c), triggers overt subject insertion categorically (Enríquez 1984, Bentivoglio 1987, Silva Corvalán 1982, 1994, Otheguy, Zentella and Livert, 2007, among others).

Note that *Topic continuation* is a factor group that only measures continuity of Topic between the preceding and the target clause. Another variable, *Distance from previous mention*, measures the effect of Topic continuity in discourse. Within this scalar factor group, the distance between the target subject and its previous mention is measured in intervening subjects. If the subject of a clause is co-referential with that of a preceding clause, the distance from the subject is ø, since the referent appears in the preceding clause, as in (3a) above. In (4), if the target subject were (4d), the distance between the preceding and the target subject would be 2, since there are two intervening subject, *Pablo* (4b) and *todos*, 'everyone' in (4c):

(4) a. Terminé el trabajo y me fui a cenar.

finished.1PSG my work and CL.went.1PSG out to dinner.

b. Pablo me llamó y me dijo

Pablo me.CL called.3PSG and me.CL.told.3PSG

c. que estaban todos en su casa.

that be.3PPL everyone at his place

d. Yo estaba cansadísima, pero me apetecía ir.

I was.1PSG very tired, but I CL.felt.1PSG to go.

'I finished work and went out to dinner. Pablo called me and told me everyone was at his place. I was very tired but still felt like going.'

Previous studies find that the further away the referent is, the more probable an overt form (Cameron 1995, Cameron and Flores-Ferrán 2004, Flores-Ferrán 2002, Travis 2005, 2007). Note that as opposed to *Co-reference*, neither *Topic Continuity* nor *Distance from previous mention* take into account whether the target subject and the previous mention of its referent share their syntactic function or not.

The form that the subject takes is also affected by the continuity effect. Continuity in subject form, or *Priming*, is defined as "the unintentional and pragmatically unmotivated tendency to repeat the general syntactic pattern of a utterance" (Bock and Griffin 2000:177). Cameron and Flores-Ferrán (2004) and Travis (2007) have examined the role of priming in subject expression, and have arrived at similar conclusions: in short, null subjects trigger null subject use in subsequent clauses, and the same holds true for overt subjects (Cameron and Flores-Ferrán 2004). Travis (2007) further concludes that there are differences in the strength of *Priming* among genres of speech. In her study she compares the relevance of *Priming* in the Spanish of New Mexico and Coastal Colombian Spanish, and concludes that subject expression is conditioned by a priming effect in both dialects; however, the effect seems to be more short-lived in Coastal Colombian Spanish. Travis (2007) attributes this difference to the interactional nature of her Colombian data set. In her New Mexican Spanish data, which feature narrative speech, priming effects last longer in the discourse.

Continuity in verb tense has also been found to be relevant in subject expression. The factor group *TMA Continuity* indicates whether the tense of the target clause and that of the preceding clause are the same or different. This variable has been returned as significant in a number of studies, among them Bayley and Pease-Álvarez (1997) and Ávila-Shah (2000).

Taking as a point of departure that both *Co-reference* and *TMA Continuity* have a role in variable subject expression, in her study of Brazilian Portuguese Paredes Silva (1993) proposed to merge both variables into a single scalar variable that could probe the relevance to subject use of the relationship or connectedness between preceding and target clauses. This variable is usually referred to as *Discourse* or *Speech Connectedness*, and it measures the effect of referential and

temporal connections among clauses. Paredes Silva hypothesized that in more connected discourse there should be less use of overt pronouns. This hypothesis was borne out in her study, and more recently in Ávila-Shah (2000), Otheguy, Zentella and Livert (2007) and de Prada (2009).

Paredes-Silva (1993) defined six degrees of speech connectedness, which range from most connected discourse, in contexts that feature a preceding and target subject that share both reference and TMA (first degree), to least connected discourse, in environments in which there is no co-reference between preceding and target subjects and there is an intervening referent that is a viable alternative for the target subject (sixth degree). Intermediate degrees or connectedness are represented in her classification by contexts in which subjects share reference but not TMA (second degree); environments in which there is no co-reference and there are one or more intervening subjects between preceding and target clauses (third degree); cases in which there is no co-reference and the preceding subject is not in a subject position (fourth degree); and contexts that feature the introduction of a new discourse topic or narrative section (fifth degree). The examples in (5)-(10) illustrate these six degrees of connectedness:

(5) ø Telefonei procê, ø continuei a copiar o trabalho. ø Fiquei de saco cheio, ø parei de copiar e ø fui juntar-me aos meus familiares na praia.

'I called you up, I went on copying the paper. I got fed up, I stopped copying and I went to the beach to join my relatives.'

(6) Ultimamente eu não tenho tido muitas novidades pra te contar, porque eu estava vivendo em completa alinenação.

'Lately I have not had much news to talk about, because I have been living in complete alienation.'

(7) ø Fiquei estudando ontem até às 3:30 h am. Ainda são 11hs. ø Vou ficando por aqui.

'Yesterday I studied until 3:30 in the morning. Now it's 11 o'clock. I am going to stop here.'

(8) Desculpe-me pela minha reclamação infundada. Eu estava brincando.

'Forgive my unjustified complaing. I was joking.'

(9) ø Vou ter que estudar segunda-feira mesmo. A Vera vai ter que ser rápida e direta, pois eu tenho mais o que fazer.

'I am really going to have to study on Monday. Vera is going to have to be quick and to the point because I have other things to do.'

(10) Diga-me o preço total e envie o número de tua conta aí nos EUA para que eu possa te enviar a soma correspondente. Quanto ao livro, eu segui tuas instruções.

'Tell me the total price and send me the number of your account there in the USA so that I can send you the corresponding amount. As for the book, I followed your instructions.'

(Paredes Silva 1993:43-44)

The difficulty of distinguishing in the data these six degrees of connectedness has led most researchers to construct this variable in a simpler, clear-cut way. Otheguy, Zentella and Livert (2007) and de Prada (2009) reduce the six degrees to three: *same reference and same TMA in preceding and target clause*, *same referent and different TMA*, and *different referent*, and they have obtained very similar results to the ones arrived at by Paredes Silva: more connected discourse favors null subjects, and less connected discourse favors overt subject insertion.

In sum, continuity has been interpreted in previous variationist work as a broad notion that covers subject form, reference and, and one version or another of the continuity effect has been found in most studies on Spanish subject expression.

Not all studies, however, present such uniform results with respect to discourse factor groups not associated with continuity. A case in point is *Clause Type*. Variants for this factor group are *main* and *embedded* clauses. This variable is not examined in a number of investigations; among the studies that examined it, Morales (1997), Otheguy, Zentella and Livert (2007) and Silva-Corvalán (1994) conclude that embedded clauses favor null subjects, while neither overt nor null

subjects are favored in main clauses, which researchers take as an indication that factor groups other than *Clause Type* regulate subject use in main clauses. As far as embedded clauses are concerned, note that issues related to co-referentiality interact with null or overt subject insertion, particularly with respect to third person subjects. It could be thought, then, that co-referentiality may at least in part account for the fact that embedded clauses tend to feature null subjects.

Finally, some factor groups that regulate subject expression in Spanish sit at the edge between morphosyntax and discourse. This is the case of *Person*. This variable teases apart first, second, and third persons in singular and plural. While this distinction is encoded in the morphosyntax, *Person* can also be considered a discourse-related variable since first and second person, as opposed to third person, have differing discourse properties: first and second persons are always given information (Chafe 1994) and are essentially deictic, while the third person does not necessarily constitute given information in the discourse, and it is potentially anaphoric. De Prada (2009) points out that if *Person* were to be relevant for subject expression because of its discursive properties we would expect for first and second persons to feature less overt subjects than the third person; however, she notes, the opposite trend is found in previous work. Generally, first person subjects have been found to exhibit higher rates of overt subjects in Non-Caribbean and language contact varieties (Enríquez 1984, Travis 2005, de Prada 2009), while the highest rates of expressed pronominals arise in the second person singular in Caribbean dialects (Cameron 1992, 1993, 1996, Otheguy, Zentella and Livert 2007).

In any case, the fact that specific constraints on subject use have discourse-related as well as morphosyntactic properties argues in favor of analyzing subject expression with a variationist methodology, which allows us to consider the role of these two types of factors simultaneously in a given data set.

Having discussed discourse-related constraints on subject expression, the next section centers on the morphosyntactic properties relevant for overt and null subject insertion.

4.1.2. Morphosyntactic variables.

The three morphosyntactic factor groups most studied in variationist investigations are *Person* (of the subject), *Verb Ambiguity* and *Tense*. The latter examines whether different verb tenses favor or disfavor null or overt subjects; *Person* distinguishes between first, second and third person in the singular and plural, and *Verb Ambiguity* distinguishes between ambiguous and unambiguous verb forms.

Unambiguous forms are those for which the verb inflection is exclusive to one person, while ambiguous forms are those that share inflectional endings for more of one person of the paradigm. An example of the former would be the present tense ending *–mos*, exclusive to the first person plural across conjugations (*bail-a-mos*, *com-e-mos*, *dorm-i-mos*, 'we dance', 'we eat' 'we sleep', respectively); ambiguous forms are for instance the *–ba* and *–ía* imperfect endings, as they are shared by the first and third persons (*yo/el/ella/usted bail-a-ba, comía, dormía*, 'I/he/she/you (formal) danced, ate, slept'). While there is consensus in the literature that *Person* has a relevant role in subject expression phenomena, contradictory results for *Verb Ambiguity* have been obtained in previous work. Since *Person*, *Verb Ambiguity* and *Tense* are closely tied to one another, I discuss them jointly in what follows.

As far as *Person* is concerned, in addition to the main difference between Non-Caribbean and Caribbean varieties regarding first and second person pointed out in the previous section, studies generally converge in finding that singular contexts feature higher rates of overt subjects than plural contexts (Enríquez 1984, Bentivoglio 1987, Cameron 1992, Silva-Corvalán 1994, Travis 2005) in all Spanish dialects studied up to date[26].

The singular-plural asymmetry has been suggested (Bentivoglio 1987) to stem from the fact that Spanish plural verb forms feature 'heavier' morphology than singular ones (cf. 'com-*o*', first person singular, versus 'com-e-*mos*', first person plural). Following this line of reasoning, i.e., that more overt subjects should be featured where morphological distinctions are not as 'heavy',

[26] See the discussion on Cameron's (1992) work on Puerto Rican Spanish in 4.2.2.2. below for a different explanation for the singular/plural asymmetry.

Hochberg (1986) proposed the *Functional Compensation Hypothesis* to explain the differences in pronominal expression among Caribbean and Non-Caribbean varieties.

In this hypothesis the increased use of overt subject pronouns in Caribbean dialects is explained as a compensation strategy for the loss of inflectional distinctions: specifically, as –s is the second person inflectional morpheme in Spanish (cf. *bail-a-s*, *com-e-s*, *duerm-e-s*, 'you dance', 'you eat', 'you sleep') and final –s is deleted in Caribbean dialects, not only first and third persons, but first, second and third persons would share inflectional endings in ambiguous verb forms. Furthermore, verb forms that are non-ambiguous in the dialects that do not delete final –s share their second and third person inflectional endings in Caribbean Spanish (cf. tú *com-e-s*, *élla com-e*, in a Non-Caribbean dialect, versus *tú come_*, *ella come* in a Caribbean dialect).

However, the *Functional Compensation Hypothesis* has been questioned in the literature: on the one hand, as noted by Lipski (1994) and quantitatively contrasted in Ranson (1991), there are s-deleting dialects in Southern Peninsular Spanish, and these dialects do not show particularly high rates of overt subject insertion. On the other, Cameron (1992, 1993, 1996) shows that the high rates of second person pronouns in Caribbean varieties are to be attributed to a reanalysis of specific and non-specific *tú* in these dialects, as thoroughly discussed in 4.2.2. below. Outside of the Caribbean, while some studies (Silva-Corvalán 1994, Travis 2005) find that morphological ambiguity in verb form is a significant factor group in subject expression, it is concluded in several other investigations that discourse-related variables override *Verb Ambiguity*; in other words, contextual ambiguity seems to play a larger role than morphological ambiguity (Ranson 1991, Cameron 1996, Morales 1997).

Also related to *Person* and *Verb Ambiguity* is *Tense*, since as pointed out above morphological ambiguity in inflectional endings only arises in specific tenses in Spanish, namely in the imperfect and past perfective indicative, in the present perfect and past perfect subjunctive, and in the conditional, in dialects that do not delete final –s. As was the case with *Verb Ambiguity*, some investigations have found that the discursive role of specific verb tenses overrides the significance

of their potential morphological ambiguity. This is the conclusion arrived at by Silva-Corvalán (1997, 2001): she notes that verb tenses that happen to be morphologically ambiguous, such as the imperfect or the conditional, are also the tenses that mark background events in speech and are non-factual and non-assertive; factual-assertive narrative verb tenses which mark foregrounded events, such as the preterite, happen to be non-morphologically ambiguous in Spanish. In Silva-Corvalán's work, the effect of verb tenses classified according to their discursive function was found to be stronger than the effect of morphological ambiguity. Specifically, while background tenses favored overt subjects and foreground tenses disfavored them, the present tense neither favored nor disfavored them. Similar results were obtained by Bayley and Pease-Álvarez (1997) and Travis (2007).

Having discussed the role of discourse-related and morphosyntactic constraints in Spanish subject expression, I now turn to discuss the relevance of lexico-semantic variables.

4.1.3. Lexico- semantic variables.

Even though in general terms lexico-semantic variables have been found to be less robust than discourse-related and morphosyntactic variables, two factor groups related to the lexical properties of subjects and verbs display significant effects in some studies. This factor groups are *Semantic Verb Type* and *Animacy*.

Animacy classifies referents as animates or inanimates. Since there is no overt pronominal form for inanimates in Spanish, the investigations that do not include full DP subjects do not consider this variable. However, inanimate referents can bear either full DP or null subjects. For this reason studies like de Prada (2009) include this variable, with significant *Animacy* findings.

Semantic Verb Type classifies verbs according to their meaning. Factors included in this group across studies are *psychological*, *speech act*, *motion*, and *copulative* verbs. This constraint has been found to be significant by Bentivoglio (1987), Silva-Corvalán (1994) Travis (2007), and de Prada (2009), among others, who find that verbs used to express opinions typically favor overt subject insertion. Note in this respect that these verbs feature the first person singular frequently

(Torres Cacoullos and Walker 2008), and that first person singular subjects favor overt subjects, as discussed in §4.1.2. Travis (2007) further concludes that copulative verbs favor overt subjects as well.

The syntactic properties of verbs derived from their lexical structure have been only rarely investigated in previous quantitative work. De Prada (2009) is an exception to this trend, as she investigates a *Syntactic Verb Type* variable, which distinguishes among *transitive*, *unergative*, *unaccusative* and *copulative verbs*. Her results show that this factor group is not ranked as significant in the statistical analysis of her data set.

In sum, what emerges from the investigation of different types of constraints is that discourse-related factors strongly constrain subject expression in Spanish. As far as morphosyntactic variables are concerned, *Person* has arisen as a very relevant factor group, particularly for explaining the differences between Caribbean and Non-Caribbean dialects. However, there is disagreement in the literature with respect to the role of *Verb Ambiguity* and *Tense*. Language-external variables, with the exception of *Genre*, are typically not significant, except in contexts of language contact. Nonetheless, external variables have been found significant in a few studies that investigate Caribbean Spanish (Lizardi 1993, Ávila-Jiménez 1996). Table 4.1. offers a summary of the variables and associated constraints that have been discussed thus far.

The discussion of the relevant factor groups in Spanish subject expression has also shown the specific advantages of variationist analysis for dealing with constraints that do not surface as categorical rules on subject insertion, such as the ones triggered by discourse-related properties. Furthermore, the fact that specific variables, such as *Person*, lie at the intersection between morphosyntax and discourse also speaks to the utility of a methodology in which all factors operating on a given linguistic phenomenon can be assessed at the same time.

The investigation of the relative role of morphosyntactic and discourse-related factors in subject expression is particularly relevant for Caribbean dialects, as evidenced by the substantial amount of available research on this issue (Cameron 1992, 1993, 1996, Paredes-Silva 1993, Ávila-

Jiménez 1996, Orozco and Guy 2008, among others). In short, the study of varieties that display loss of agreement morphology presents a suitable venue for exploring the role of inflectional morphology in subject licensing; conversely, the study of the shifts, if any, that discourse-related properties have undergone in Caribbean varieties might also shed light on the direction of grammatical restructuring in these dialects.

Table 4.1. Spanish subject expression: Factor groups and constraints.	
Dependent variable	***Constraints***
SUBJECT FORM	*Null* *Pronominal subject* *DP subject*
Independent variables	
DISCOURSE-RELATED VARIABLES	
CO-REFERENCE	*Co-referent* *Non-co-referent*
TOPIC CONTINUATION	*Topic* *Topic shift* *Emphasis*
DISTANCE FROM PREVIOUS MENTION	*ø* *1* *2* *3* *4* *more than 4*
TMA CONTINUITY	*Same TMA* *different TMA*
PRIMING	*previous subject is expressed* *previous subject is unexpressed*
SPEECH CONNECTEDNESS	*same referent, same TMA* *same referent, different TMA* *different referent*
CLAUSE TYPE	*main* *embedded*
MORPHO-SYNTACTIC VARIABLES	
PERSON	*first, second, third person singular* *first, second, third person plural*
VERB AMBIGUITY	*Ambiguous verb form* *Unambiguous verb form*
TENSE	*Present* *Imperfect* *Future* *…*
LEXICO-SEMANTIC VARIABLES	
SEMANTIC VERB TYPE	*Psychological* *Speech act* *motion* *copulative*
ANIMACY	*Animate* *Inanimate*
SYNTACTIC VERB TYPE	*transitive* *unergative* *unaccusative* *copulative*

4.2. Subject expression in Caribbean Spanish.

Early descriptions of Caribbean dialects (Henríquez Ureña 1940, Navarro Tomás 1948) already registered high rates of overt subject use. In contemporary variationist work on these varieties the general pattern that has been described is as follows: while percentages of overt subject insertion are much higher in Caribbean than in Non-Caribbean Spanish, they do not translate into significantly different factor group rankings for the two groups of dialects. In fact, subject expression seems to be regulated by the same factor groups in all Spanish dialects. Nonetheless, relevant differences have been found between Mainland and Caribbean Spanish with respect to constraint ranking. These differences, and what they entail for an assessment of the extent of grammatical restructuring in Caribbean subject expression, are discussed in the next paragraphs.

4.2.1. Subject expression: Overall rates of expression and factor group rankings.

Table 4.2, adapted from Silva-Corvalán (2001) and de Prada (2009), shows a continuum in rates of subject expression from European Peninsular varieties of Spain (Valladolid and Madrid), with the lowest rate, to the Spanish of San Juan, in Puerto Rico and Cali, in coastal Colombia, with the highest rates. Generally, Latin American Spanish seems to favor more overt subjects than Spanish Peninsular varieties, and within Latin America, Caribbean dialects significantly depart from Mainland varieties.

However, as Travis (2007) and de Prada (2009) note, comparing percentages across Spanish dialects does not prove very fruitful. Not all studies include the same variants within the dependent variable subject form, as they aim to answer different questions: while most studies only take into account overt pronominal subjects, a few of them also include full DP subjects in the envelope of variation (Silva-Corvalán 1982, de Prada 2009). Furthermore, within the pronominal paradigm, Hochberg (1986) takes into account only singular subjects; Bentivoglio (1987) and Travis (2007) examine only first person subjects; on the contrary, a number of studies investigate the whole paradigm (Otheguy, Zentella and Livert 2007, Cameron 1992, Orozco and Guy 2008, de Prada 2009). As *Person* of the paradigm has strong effects on subject use, it is very likely that the

asymmetries in rates across studies stem from considering only specific grammatical persons. By way of an example, both Travis (2007) and Orozco and Guy (2008) obtain data in Caribbean Colombian communities; however, subject rates obtained range from 35% in the former to 48% in the latter.

Table 4.2 Dialectal variation in overall rates of subject expression.

STUDY	PLACE	% OF OVERT SUBJECTS
de Prada (2009)	Valladolid, Spain	12%
Enríquez (1984)	Madrid, Spain	21%
Cameron (1992)	Madrid, Spain	21%
Otheguy et al. (2007)	Mainland (Mexico, Colombia and Ecuador) Newcomers in NYC	24%
Miró Vera & Pineda (1982)	Sevilla, Spain	27%
Silva-Corvalán (1982)	Los Angeles	35%
Orozco & Guy (2008)	Barranquilla, Colombia	35%
Barrenechea & Alonso (1977)	Buenos Aires, Argentina	36%
Otheguy et al. (2007)	Caribbean Newcomers in NYC	36%
Hochberg (1986)	Puerto Ricans in Boston	37%
Cifuentes (1980)	Santiago, Chile	38%
Bentivoglio (1987)	Caracas, Venezuela	40%
Morales (1982)	Puerto Rico	40%
Travis (2007)	Cali, Colombia	48%
Cameron (1996)	San Juan, Puerto Rico	60%

More interesting and viable comparisons emerge from exploring differences in patterns of subject use. Comparisons among factor groups, as compared to overall rates, offer quite different results across varieties: generally the same factor groups constrain subject use across communities. Otheguy, Zentella and Livert (2007), in their comparative study of two generations of Latin Americans living in New York City regard this as an unexpected result: "because Caribbean and

Mainland newcomers differ markedly in pronoun occurrence rates, they might be expected to differ markedly as well with regard to variable hierarchies, but the results do not confirm such expectation" (Otheguy, Zentella and Livert 2007:789). In general, however, as pointed out in Cameron (1996, 2004) under the variationist view of language variation the patterning in the factor groups across dialects is not a surprising result, and it confirms that Spanish dialects, as far as general constraints on subject licensing are concerned, share an underlying grammar. In fact, this is the main asset of variationist methodology, namely that it can "find basic similarities of grammatical systems and constraints on variation that may be masked by differences in frequency" (Cameron 1996:327). The fact that discourse-related and morphosyntactic factors regulate subject use across Spanish dialects does not entail dismissing, however, the possibility of grammatical restructuring in Caribbean dialects, or assuming that this restructuring cannot be detected by a variationist analysis. As will be discussed below, cross-dialectal variation is best accounted for by comparing differences in constraint rankings.

The parallels in factor group ranking across varieties are illustrated by the analysis of subject expression in Otheguy, Zentella and Livert (2007). The authors examine the role of language contact with English and dialect leveling across the six largest communities of Latin Americans living in New York City. The sample includes communities that have populated New York for decades (Puerto Ricans and Mexicans) as well as communities that have had a presence in New York for shorter periods of time, like Dominicans. The speech of Ecuadorians, Colombians and Cubans is also examined. The study tests the role of 10 linguistic and 35 sociodemographic variables in two generations of speakers, recent arrivals in New York (Newcomers) and born and raised in New York speakers (NYBR) respectively.

Statistical analysis shows that both contact with English and dialect contact among communities have an effect in shaping the subject expression grammar of NYC Spanish speakers. Of most interest for the study carried out in this dissertation are the results on dialect leveling, and more specifically the results for newcomers from different regions. The speech of newcomers in

Otheguy et al. is taken to reflect the dialectal differences in their countries of origin. Table 4.3 provides the overall rates of overt subject use for the six communities:

Table 4.3. Newcomers' pronoun rate by national origin (Otheguy et al. 2007:785)

	N speakers	*% overt pronouns*	*N verbs*
DOMINICAN REPUBLIC	6	41	2,217
PUERTO RICO	6	35	3,805
CUBA	7	33	2,278
ECUADOR	8	27	3,735
COLOMBIA	6	24	1,926
MEXICO	6	19	2,569

$p<0.01$

Newcomers: Age of arrival >16 and YrsNYC<6

Table 4.3 reveals a contrast in rates of subject expression between speakers from Mainland varieties (Ecuador, Mexico and Colombia) and those from the Caribbean (Dominican Republic, Puerto Rico and Cuba).[27] For this reason the authors reduce their ensuing statistical analysis to a contrast between Mainland and Caribbean speakers.

Table 4.4 shows the results of Otheguy et al.'s statistical analysis of 10 linguistic variables; apart from the familiar variables *Person*, *Discourse Connectedness*, *Tense*, *Clause Type, Semantic Verb Type* (labeled *Lexical Verb Type* in this study) and *Genre*, they also tested whether reflexive and non-reflexive verbs behave differently (*Reflexive*). In addition, *Set phrases*[28], usually left out of the envelope of variation in most studies, were also considered. Finally, the variable *Picture* refers to the section of the interview in which the target subject appears; the speakers in this study, in addition to participating in a standard sociolinguistic interview, filled a picture-based questionnaire about lexical usage; these data, the authors inform, were gathered for an unrelated project but nonetheless were included in the statistical analysis.

Variable hierarchies for Mainland and Caribbean Newcomers are quite similar: *Person, Connect*, *Tense* and *Clause Type* are ranked among the top five variables in the hierarchy for both

[27] The Colombian speech represented in this corpus pertains to Mainland Colombian dialects. No Coastal Colombia speakers are part of the sample.

[28] Set Phrases such as *(tú) sabes*, 'you know' or *(tú) entiendes*, 'do you understand?' are excluded from the envelope of variation in most studies because they feature (almost) categorically overt subjects in Caribbean varieties, and null pronouns in Peninsular Spanish.

groups of speakers. With regard to the lower ranked variables, *Reflexive* and *Set Phrase* are significant in the Mainland but not in the Caribbean, while *Picture* is significant in the Caribbean but not in the Mainland.

It should be noted that the authors carry out their analysis with the *Statistical Package for the Social Sciences* (SPSS), rather than with a version of the *Variable Rule Program* (D.Sankoff 1975, D.Sankoff, Tagliamonte and Smith 2005) usually employed for multivariate analyses in variationist work. In Table 4.4 the relative strength of the independent variable is assessed by the Wald statistic associated with each of the variables, rather than with the factor weights provided by the Variable Rule Program, and the significance of the calculation by one asterisk or two:

Table 4.4. Variable hierarchies in overt pronoun expression: Newcomers (Otheguy, Zentella and Livert 2007:789)

CARIBBEAN NEWCOMERS			MAINLAND NEWCOMERS		
Rank	*Variable*	*Wald*	*Rank*	*Variable*	*Wald*
1st	PERSON	495.218**	1st	PERSON	408.05**
2nd	CONNECT	293.86**	2nd	CONNECT	194.11**
3rd	TENSE	239.48**	3rd	SET PHRASE	71.17**
4th	CLAUSE	87.33**	4th	TENSE	55.10**
5th	LEXICAL	22.20**	5th	CLAUSE	45.49**
6th	GENRE	12.30*	6th	LEXICAL	45.14**
7th	PICTURE	4.86*	7th	REFLEXIVE	31.89**
8th	SET PHRASE	2.23	8th	GENRE	13.45**
9th	REFLEXIVE	1.98	9th	SPECIFIC	1.50
10th	SPECIFIC	0.96	10th	PICTURE	1.35
N verbs=8,583			N verbs=8,155		
N speakers=19			N speakers=20		

*$p < 0.05$; **$p < 0.01$

Somewhat similar results can be found in Travis (2007), who compares New Mexico Spanish with Coastal Colombian data. Travis includes the following variables in her analysis: *Semantic Verb Type* (which she labels *Verb Class*), *Priming*, *Distance from previous mention*, *Verb Ambiguity* (labeled *(Un)ambiguous TMA*), *Clause Type*, *TMA Continuity* and *Position in turn*. This last factor group is included to probe for the differences in subject expression in conversational versus narrative discourse. As evidenced by her results, illustrated in Table 4.5, Caribbean speakers feature higher rates of overt subjects than New Mexican speakers. As far as hierarchies of

significant factor groups are concerned, since Travis only considers first person subjects, no comparison is possible with the variable *Person* in Otheguy, Zentella and Livert (2007). The factor group ranked first in Table 4.5 is *Verb Class*. This result probably stems from including only first person subjects in the analysis (cf.§4.1.2.). Immediately following are two discourse-related factor groups, *Distance from previous mention* and *Priming*, and next in the hierarchy is *TMA ambiguity*. *Clause Type*, *TMA continuity* and *Position in turn* were not selected as significant.

Table 4.5. Two independent variable rule analyses of the contribution of factors selected as significant to the probability of expressed subjects in NM and Colombian data. (Travis 2007:115).

	NEW MEXICO			COASTAL COLOMBIA		
Total N	853			878		
% expressed subjects	33%			48%		
Corrected mean	.31			.48		
	Weight	*%*	*% of data*	*Weight*	*%*	*% of data*
VERB CLASS						
PSYCHOLOGICAL	.70	55	18	.68	67	20
COPULA	.55	39	7	.63	60	6
SPEECH	.53	35	20	.53	51	16
OTHER	.43	25	41	.42	39	47
MOTION	.35	20	12	.36	31	9
Range	35			32		
DISTANCE						
5+ CLAUSES	.63	46	23	.58	55	38
2-4 CLAUSES	.58	39	16	.51	49	17
1 CLAUSE	.53	34	16	.45	44	13
Ø (SUBJECT CONTINUITY)	.39	23	44	.42	38	30
Range	24			16		
PREVIOUS REALIZATION						
EXPRESSED SUBJECT	.67	50	35	.57	55	49
UNEXPRESSED SUBJECT	.41	23	64	.43	40	50
Range	26			14		
TMA						
AMBIGUOUS TMA	.62	37	22	.62	60	10
UNAMBIGUOUS TMA	.47	32	77	.48	46	89
Range	15			14		

NMCOS: p=.003; log likelihood= -476.040; chi-square/cell=1.1082. Factor groups not selected as significant: *Clause Type, TMA Continuity,* and *Position in turn*.
Colombia: p=0.13; log likelihood=-563.736; chi-square/cell=1.2097. Factor groups not selected as significant: *Clause Type, TMA Continuity,* and *Position in turn.*

Travis, like Otheguy et al. (2007), finds similar factor hierarchies across dialects despite differences in rates. Both studies rank discourse-related factors at the top of the hierarchy, *Connect* in Otheguy et al. and *Distance from previous mention* and *Priming* in Travis, and both find *Tense* significant[29].

Note that we cannot compare constraint rankings of significant factor groups across these studies, as Otheguy et al. (2007) do not provide information on the ranking of specific constraints in their overall results. Nonetheless, it is interesting to note for the present study that constraint rankings within the factor group *Person* can differentiate Spanish dialects, and thus allow us to investigate the extent of grammatical restructuring in subject expression in Caribbean varieties.

4.2.2. On the site of cross-dialectal differences in subject expression: *Person* constraint rankings.

Studies that examine subject use in all persons of the paradigm have consistently shown that not only overall rates of expressed subjects vary across dialects. Rates also differ significantly among specific pronominals. For instance, Morales (1997) compares the overt subject rates she found in Puerto Rico with the ones attested in Madrid, Spain and Buenos Aires, Argentina[30]. The results of the comparison are provided in Table 4.6:

[29] Travis (2007), in addition to *percentages* and *factor weights*, provides the following information:

(i) The *corrected mean* provides a global measure of the rate of rule application, and therefore constitutes an overall indication of the strength of a given rule.

(ii) The *range* for each factor informs on its relative strength: the factor group with the largest range will be the most significant, and the factor group with the smallest range will be the least significant.

(iii) *Log likelihood* indicates which factor group increases the most the probability of occurrence (i.e. likelihood) of application of a rule. The closest this value is to zero, more likely it is that this factor group constrains rule application.

(iv) *Chi square per cell* gives information on the overall significance of the distribution of variants.

All these measures, as well as the factor groups not selected as significant, are returned from the application of the Variable Rule Program to a given data set. When this information is not provided in the tables that follow in the remainder of this chapter, it is because the authors have not included it in their studies.

[30] The Madrid data in Morales (1997) are taken from the *Habla Culta* corpus (Esgueva and Cantarero, 1981). Buenos Aires data are borrowed from Barrenechea and Alonso (1977).

Table 4.6. Overt subject rates for specific pronouns: San Juan, Puerto Rico, Madrid, Spain, and Buenos Aires, Argentina (Morales 1997:155).

	SAN JUAN	MADRID	BUENOS AIRES
YO, TÚ	421/777 (54%)	195/708 (28%)	218/721(30%)
ÉL/ELLA, ELLOS/ELLAS	180/517 (35%)	33/419 (8%)	70/397 (18%)
NON-SPECIFIC *UNO/TÚ*	132/191 (69%)	19/85 (22%)	58/121 (48%)

For all persons, Puerto Rico shows the highest rates of overt expression. As far as third person pronouns are concerned, singular and plural percentages differ in all three dialects. Since singular and plural third person pronouns are conflated in Morales' data, it is not possible to assess whether high percentages correspond to singular or plural pronouns. Similarly, first person singular *yo* and second person singular [+specific] *tú* are treated jointly, and therefore differences in first person among dialects, were they to arise, are not visible. The strongest differences, however, arise in pronouns with a non-specific meaning, *uno* and *tú*: while rates of overt expression in Puerto Rico and Buenos Aires are very close to 70% and 50% respectively, they only reach 22% in Madrid Spanish.

These results suggest that the major constraint on rates is the distinction specific/non-specific, rather than the differences in inflectional morphology across dialects. Interestingly, Morales (1997) also finds that probability weights for the insertion of overt non-specific pronouns also differ significantly in the three dialects she compares: while Puerto Rican Spanish favors expressed subjects with *tú* and *uno* at a .68 weight, Madrid speakers strongly disfavor them at .19; in Buenos Aires, overt subjects are nor favored neither disfavored, with a weight of .45.

Cameron (1992) carries out an investigation of the role of specificity in accounting for differences among Caribbean and Non-Caribbean dialects. He studies Puerto Rican subject expression in the metropolitan area of San Juan, and compares the results obtained with data gathered in Madrid, Spain. His results are provided in Table 4.7:

Table 4.7. Rates of Subject Pronoun Expression in San Juan, Puerto Rico and Madrid, Spain (Cameron 1992: 233)		
SINGULAR		
FIRST PERSON: YO		
	N overt subjects	*%*
San Juan	949	50
Madrid	1010	31
SECOND PERSON: TÚ [+SPECIFIC]		
	N overt subjects	*%*
San Juan	145	48
Madrid	58	40
SECOND PERSON: TÚ [-SPECIFIC]		
	N overt subjects	*%*
San Juan	188	69%
Madrid	150	19%
THIRD PERSON: ÉL/ELLA		
	N overt subjects	*%*
San Juan	443	39
Madrid	286	8%
THIRD PERSON: UNO (-SPECIFIC)		
	N overt subjects	*%*
San Juan	39	85
Madrid	38	8
PLURAL		
FIRST PERSON: NOSOTROS/NOSOTRAS		
	N overt subjects	*%*
San Juan	220	15
Madrid	427	6
THIRD PERSON: ELLOS/ELLAS		
	N overt subjects	*%*
San Juan	138	25
Madrid	122	8

Table 4.7. shows that when first and second person singular are distinguished, the strongest differences in overt subject rates arise in the second person. Cameron's results also demonstrate that distinguishing singular and plural occurrences is key for an accurate description of Caribbean and Non-Caribbean patterns: while third person singular pronouns are expressed overtly at a percentage of 39% and 8% in San Juan and Madrid respectively, third person plurals are overt 25% of the time in Puerto Rico and 8% of the time in Madrid. Finally, Morales and Cameron arrive at similar findings regarding [±specificity]: when considering non-specific *uno* and *tú* separately, *uno* features an 85% rate of overt use, in sharp contrast with Madrid's 8%; differences in [+specific] *tú* are not as

sharp (48% in San Juan and 40% in Madrid, with a not significant chi-square value of the difference of 1.240)), whereas [-specific] *tú* behaves very differently the dialects under comparison, with a 69% rate in Puerto Rico and a 19% rate in Madrid, and a significant ($p < .001$) chi-square value of 83.105.

Differences in [±specificity] reflect a process of grammatical restructuring in Caribbean dialects. Recall from the discussion on Brazilian Portuguese in Chapter Three that in this language a number of strategies for avoiding null subject use were detected in the quantitative literature (Kato and Taralho 1983, Duarte 1992, 1993). For instance, structures used for the expression of genericity or impersonality that featured a null subject in NSL European Portuguese were found to be substituted in Brazilian Portuguese for overt subject constructions.

The results in Cameron and Morales point in the same direction: in NSL Spanish, the options for the expression of impersonality are either a structure with a generic null subject *tú*, or a structure with the clitic *se* (cf. for instance *Tienes muchas oportunidades de trabajo en Norteamérica*, 'You (generic) have/One has a lot of work opportunities in North America' or *Se trabaja mucho aquí*, 'You (generic) work/One works a lot here'). Puerto Rican speakers seem to avoid these structures by resourcing to the use of non-specific, overt *uno* and *tú*.

Puerto Ricans have been shown to display similar patterns of overt subject use in other studies. As shown in Table 4.8, percentages are very similar for Puerto Ricans in the island (Ávila-Jiménez 1996) and in the diaspora (Flores-Ferrán 2002, who investigates Puerto Rican speech in New York City):

Table 4.8. Overt pronominal subjects for person and number: San Juan and NYC (Flores-Ferrán 2004:59).

	NYC (Flores-Ferán2002)	San Juan (Cameron 1992)	San Juan (Ávila-Jiménez, 1996)
	%	%	%
1ST PERSON (*yo*)	52	50	53
2ND PERSON SG SPECIFIC	53	48	59
2NDPERSON SG NON-SPECIFIC	60	69	63
3RD PERSON (*él/ella*)	48	39	38
1ST PERSON PLURAL	17	15	16
3RD PERSON PLURAL	22	25	24

Differences between Puerto Rico and Non-Caribbean dialects are also reflected in constraint rankings when variable rule analysis is applied, as shown in Table 4.9:

Table 4.9. Constraint rankings in Person: San Juan versus Madrid (Cameron 1992:239)

SINGULAR		
	San Juan	Madrid
FIRST PERSON: *yo*	.59	.69
SECOND PERSON: *tú* [+SPECIFIC]	.51	.72
SECOND PERSON: *tú* [-SPECIFIC]	.72	.50
THIRD PERSON: *uno* [-SPECIFIC]	.86	.72
THIRD PERSON: *él/ella*	.47	.32
PLURAL		
FIRST PERSON: *nosotros/nosotras*	.15	.22
THIRD PERSON: *ellos/ellas*	.26	.31

Note that factor weights for [+specific] and [-specific] *tú* are reversed in San Juan and Madrid: in the latter overt subjects with [+specific] *tú* are favored (.69), while neither favored neither disfavored (.50)for [-specific] occurrences. The opposite trend is found in San Juan, where [-specific] *tú* is favored at .72, and overt [+specific] *tú* is slighthly favored at .51.

Otheguy et al. (2007), in their comparison of Caribbean and Mainland communities in NYC, arrive at similar results: while second person non-specific ranks at the top of the variable hierarchy for Caribbeans, this trend is reversed in the speech of Mainland speakers[31]:

[31] In this table, Exp(B) represents the change in the log odds in the dependent variable (*null pronoun* versus *overt pronominal* in Otheguy's and colleagues work) triggered by the constraints in the independent variables, in this case, *Person*; a constraint with an Exp(B) value above 1.0 increases the odds of appearance of an overt pronoun; a constraint with a value of Exp(B) below1.0 reduces the odds of appearance of an overt pronoun; in other words, the change point between increasing or reducing probability of occurrence is not 5.0, as in the Variable Rule Program, but rather 1.0.

Table 4.10. Constraint ranking within the Person variable (Otheguy, Zentella and Livert 2007:790).

CARIBBEAN NEWCOMERS			MAINLAND NEWCOMERS		
	N verbs	*Exp(B)*		*N verbs*	*Exp(B)*
2NDPSG +SPECIFIC	716	3.22**	3RDSG	1,1668	3.21**
2NDPSG -SPECIFIC	414	3.13**	1PSG	3,907	1.98**
3SG	1,395	1.09	2NDPSG SPEC	279	1.19
1SG	4,458	1.04	2NDPSG NONSPEC	485	0.61**
3PL	1,045	0.30**	3PL	1,117	0.58**
1PL	555	0.29**	1PL	699	0.37**
N verbs=8,583 N speakers=19			N verbs=8,155 N speakers=20		

Exp(B) > 1.0 indicates that factor favors overts
Exp(B) < 1.0 indicates that factor favors nulls
Newcomers: Age of arrival > 16 and YrsNYC < 6
*$p < 0.05$
**$p < 0.01$

In Cameron (1992) and Otheguy et al. (2007) the least favoring contexts within *Person* are first and third person plural. First person singular differs markedly between Caribbean and Mainland speakers in Otheguy et al.: it is ranked second for Mainland speakers, strongly favoring overt subject insertion, while it is not significant for Caribbeans. In this respect Mainland speakers in New York pattern with other (Non-Caribbean) Spanish varieties studied up to date (cf.§4.1.2 and Cameron's results in Table 4.8, where first person singular favors expressed subjects in Madrid at .69), with the only difference that for New York speakers third person overrides first person in the singular. In fact, third person is the most favoring context for Mainlanders, while for Caribbeans, this constraint ranks third. Cameron (1992) and Otheguy et al. (2007) differ in their findings for second person singular, as in the latter non-specific tú is ranked second in the hierarchy, while specific tú occurrences have the strongest effect on overt pronominal insertion. Both studies, however, converge in finding that non-specificity favors null subject use in Non-Caribbean dialects, as indicated by a **0.61 weight in Table 4.10.

The results in Otheguy et al. (2007) lead the authors to conclude that Mainland and Caribbean Newcomers constitute two different speech communities when they arrive in New York:

while as speakers of the same language they share subject expression variability and the ranking of variables that favor either null or overt subject insertion, as different speech communities they differ markedly in the constraints on subject expression.

Cross-dialectal differences related to [±specificity] have been also found to constrain subject expression for third person plurals. Nonetheless, differences among communities are not as marked as the ones found for second person singular pronouns:

Essentially, third person plural non-specific pronouns are categorically null in Non-Caribbean Spanish varieties. Recall from the discussion in Chapter Two that for instance overt pronouns cannot bear an arbitrary reading in NSL Spanish. Recall also that while both null second person singulars and null third person plurals in NSL may be non-specific, there is a semantic difference between the two. Null second person singular pronouns denote people in general, including speaker and addressee and translate in English 'one', i.e., they bear a generic reading as in (11a). Null third person plural pronouns bear an arbitrary reading, and as such denote people in general in some domain, excluding speaker and addressee, and translate into English 'they':

(11) a. Tienes muchas oportunidades de trabajo en Norteamérica.
You (generic) have a lot of job opportunities in North America
(=There are a lot of job opportunities in North America).

b. (Se) hablan muchas lenguas en la India.
'They speak a lot of different languages in India.'

Variationist work, however, has shown that non-specific third person plurals are not categorically null in all Spanish varieties. These pronouns are overt (albeit at low rates) in contexts of language contact (Lapidus and Otheguy 2005) and in Caribbean areas (Cameron 1992, Ávila-Jiménez 1996). Lapidus and Otheguy (2005) attest overt third person pronouns that lack an explicit antecedent in the discourse and whose identity is unknown to speaker and addressee in the speech of Latin American Newcomers in New York. They find three degrees of (non-) specificity in their corpus:

(i) Non-referential. These pronouns do not designate any referent. They are like the bleached pronoun 'they' in English (cf. (11b)).

(ii) Semireferential. The set of possible referents is restricted, for example by a locative in the context.

(iii) Corporate. These pronouns refer to some unspecified but socially significant group of people such as governments, bosses, criminals, doctors, school officials, and so on and so forth.

Examples illustrating the possibilities in (i)-(iii) are provided in (12), (13) and (14) respectively:

(12) S: Es una ciudad que…que luce más como un pueblo y se llama xxx y nadie en la vida ha oído de esa ciudad, porque está perdida por allá en xxx.

A: ¿Dónde queda?

S: En la…eh…queda en la parte norte, en el… **ellos** la llaman, la xxx. Está en la bahía de San Francisco, alrededor de… está como a una hora de San Francisco.

(194C).

S: 'It's a city that…that looks like more like a town and it's called xxx and nobody in the world has heard of this city because it's lost out there in xxx.

A: Where is it?

S: In the… it's in the northern part, in the…**they** call it, the xxx. It's in the San Francisco bay area around… it's about an hour from San Francisco.'

(13) Si yo voy a Santo Domingo o Venezuela, o Ecuador, cualquiera de estos países a buscar trabajo, solamente por la simple razón de que estudié en Nueva York y sé inglés, **ellos** son capaz de quitarle el trabajo a un empleado de ellos para dármelo a mí simplemente porque yo soy un americano.

'If I go to Santo Domingo or Venezuela, or Ecuador, or any of those countries to look for work, only for the simple reason that I studied in New York or I know English, **they** are capable of taking away a job from an employee of theirs to give it to me simple because I'm an American.'

(14) Mandaron a hacer uno de esos *labors* a máquina de escribir, pero resulta que da la casualidad que me encontré con el presidente de la compañía en la oficina y él me preguntó que cómo me llamaba, cuántos años tenía, se sorprendió que yo le contesté y le hablé el ingles de lo más bien, y me dijo que qué yo hacía trabajando ahí teniendo yo un idioma, que hablaba muy bien el idioma. Después como al mes me llamó y me dijo que **ellos** se iban a mudar y que necesitaba un…una persona en la oficina y me dijeron que iban a tener el puesto pendiente, antes de mudarse la compañía para el piso en que estamos, **ellos** me llamaron de la oficina y me dijeron que si quería el trabajo.

'(They) sent for someone to do one of those typewriting jobs, but it turns out that I coincidentally met the president of the company in the office and he asked me my name, how old I was, he was surprised that I answered him and that I spoke English so well and he said to me, what I was doing working there having a language…that I spoke the language so well. After a month, he called me and told me that **they** were going to move and that they needed a… a person in the office and they told me that (they) were going to have the job pending, when…before the company moved to the floor that we're on, **they** called me to the office and (they) asked me that if I wanted the job.'

In Lapidus and Otheguy's study (2005) (N=8496 tokens, 141 informants), non-specific third person plural pronouns account for 33% of all third person plural occurrences (N=2834). Of these, 2727 are null, and only 107 are overt, too few for full-fledged variationist analysis. However, when isolating percentages for the types of non-specific pronouns outlined in (i)-(iii), the authors find that

as non-specificity decreases, the use of overt non-specific *ellos* increases, as illustrated in Table 4.11:

Table 4.11. Frequencies of types of overt non-specific (=arbitrary) *ellos* (Lapidus and Otheguy 2005:170)

	N	%
NON-REFERENTIAL	15	14
SEMI-REFERENTIAL	36	34
CORPORATE	56	52
TOTAL	107	

Lapidus and Otheguy (2005) compare in their study recently arrived immigrants in New York with New York born and raised (NYBR) speakers. Overt non-specific *ellos* instances appear in both groups of speakers, although the authors note that NYBR speakers feature higher numbers of overt non-specific occurrences, which they attribute to language contact with English. Nevertheless, given that these pronouns are also available to Newcomers, whose language is taken to represent their dialects of origin, it might be worthwhile to explore whether these pronouns are also attested in the speech of Spanish speakers that have not left their countries of birth. In fact, in his study of San Juan Spanish, Cameron (1992) found instances of overt non-specific *ellos*; however, judging that the sample was too small, he did not distinguish specific and non-specific uses of third person plurals in his analysis. On the other hand, Ávila-Jiménez did isolate both types of third person plural pronouns in her study, and she found, along the lines of Lapidus and Otheguy (2005), that non-specific *ellos* can be featured overtly in the speech of Puerto Ricans: while [+specific] third person plurals are null 81% of the time, [-specific] *ellos* is null at a rate of 93%. Even if rates of overt non-specific *ellos* are very small, their mere occurrence in Caribbean dialects is worth noting, as they are categorically null in Mainland varieties.

In sum, previous studies suggest that the shifts in the form of [±specific] pronouns are key to explaining patterns of grammatical restructuring in Caribbean Spanish. As will be discussed in the next section, [±specificity] seems to be more relevant for this explanation than morphological ambiguity, which was assumed to be at the source of the high rates of overt subject use in early studies on Caribbean dialects. Furthermore, as the ensuing discussion will make clear,

morphological ambiguity becomes relevant for explaining patterns of subject expression only when the relative roles of morphological and contextual ambiguity (i.e. *Co-reference* or lack thereof) are assessed jointly.

4.2.2.1. *Person* constraint rankings and *Verb Ambiguity*.

The working hypothesis in early studies on Caribbean Spanish (Jiménez Sabater 1975, Terrell 1976, Hammond 1979) was that overt subjects were inserted to resolve morphological ambiguity. This sort of ambiguity, reflected in the factor group *Verb Ambiguity*, is the result of morpho-phonological processes, particularly final –s deletion. Hochberg (1986) formalized this view in the *Functional Compensation Hypothesis*, working with data from Puerto Rican women living in Boston. Hochberg hypothesized that more overt subjects were to be found precisely where morphological ambiguity arises, i.e. in second person singular subjects.

In order to test this hypothesis, Cameron (1992) examines closely rates of final –s deletion in Puerto Rican second person pronouns. Interestingly, as illustrated in Table 4.12, he finds that rates of –s deletion are different in specific and non-specific *tú* instances; even though the chi-square difference for both types of pronouns is not significant (chi=1.789), he suggests that if the loss of inflectional distinctions alone was responsible for overt subject insertion, no difference in rates between [+specific] and [-specific] pronouns should arise:

Table 4.12. Rates of /s/ deletion in Puerto Rican Spanish: Specific versus nonspecific *tú* (Cameron 1992:292).

Reference Type		*Overt*	*Null*	*Total*
[+SPECIFIC]	N	131	212	343
	%	38	62	
[-SPECIFIC]	N	286	386	672
	%	43	57	
TOTAL	N	417	598	1,015
	%	41	59	

Two further basic findings emerge from Cameron's results: first, they replicate the findings in Poplack (1979) and López Morales (1983), who detected variation in -s deletion patterns across stylistic and social dimensions in Puerto Rican Spanish. Second, patterns of –s deletion do not

correlate with results for *Verb Ambiguity* across these social groups. If the *Compensation Hypothesis* were on the right track, the groups of speakers that favor –s deletion should also favor overt subject insertion. As the results in Table 4.13-4.15 show, this is not the case.

It can be observed in Table 4.13 that males favor –s deletion more than females:

Table 4.13. Overall rates of /s/ deletion: males versus females (Cameron 1992:298)

Gender	*/s/ %*	*/s/ N*	*ø %*	*ø N*
MALE	26	1183	74	3292
FEMALE	41	1996	59	2888
TOTAL	34	3179	66	6180
Chi-square= 216.868 (p < .001)				

Furthermore, when *Gender* is crosstabulated with *Occupational Class*, we can observe that this latter factor group also regulates –s deletion, and that differences between males and females are maintained in the *Professional, Technical/Sales* and *Skilled Labor* occupational classes, while blurred in the *Unskilled Labor* class:

Table 4.14. Overall rates of /s/ deletion and chi-square values: occupational class crosstabulated with males versus females (Cameron 1992:302).

Occupational class		*Male*		*Female*		*Total*		*Chi-square of male vs female*
		N	*%*	*N*	*%*	*N*	*%*	
PROFESSIONAL	/s/	241	38	817	54	1058	50	
	ø	387	62	683	46	1070	50	Chi= 45.847
	Total	628		1500		2128		p < .001
SALES	/s/	502	43	232		734	46	
	ø	670	57	207	47	877	54	Chi= 12.914
	Total	1172		439		1611		p < .001
CLERICAL	No males represented							
SKILLED LABOR	/s/	109	11	104	26	213	15	
	ø	912	89	303	74	1215	85	Chi= 56.246
	Total	1021		407		1428		p < .001
UNSKILLED LABOR	/s/	102	25	193	24	295	24	
	ø	311	75	610	76	921	76	Chi= .065
	Total	413		803		1216		not significant
TOTAL	/s/	954	29	1596	45	2550	38	
	ø	2280	71	1964	55	4244	62	Chi= 169.910
	Total	3234		3560		6794		p < .001

The results in Tables 4.13 and 4.14 also suggest that –s deletion and subject expression instantiate different types of phenomena. Cameron (1992) points out with López Morales (1983)

that the speakers who represent lower sociocultural classes in his study show linguistic insecurity with respect to –s deletion and consequently display hypercorrection. As will be discussed in §4.2.3, shifts in subject expression among Caribbean speakers occur below the level of awareness, and therefore no stigma is associated with higher of lower rates of subject use.

Furthermore, it is not the case that speakers who favor /s/ deletion the most show higher overt subject rates, or differences in the factor weights assigned to each of the constraints that have a role in subject expression: Cameron reports that no evidence of functional compensation is found in either male or female speakers for non-specific second person singular pronouns. There are not any significant differences either between these two groups as far as the frequencies of overt specific *tú* expression. Differences do arise, on the other hand, with respect to specific *tú*. In this case, however, what Cameron finds is that even if male speakers are the ones who delete –s in higher numbers, when *Gender* and *(Un)expressed Inflection*, i.e., -s deletion, are crosstabulated, a pattern of functional compensation is found not for males but rather for female speakers. In fact, males display a counter-functional pattern, as they produce more overt subjects where –s is not deleted and therefore inflection is present:

Table 4.15. Pronominal expression crosstabulated with +/- inflection: specific *tú* in male versus female speakers (Cameron 1992:315).

MALES			
	tú	*null*	*total*
/s/	70% (24)	30% (10)	34
ø	52% (58)	48% (54)	112
Chi-square = 3.745 not significant at .05			
FEMALES			
	tú	*null*	*total*
/s/	45% (43)	55% (53)	96
ø	63% (63)	37% (37)	100
Chi-square = 6.539 significant at .05			

In short, results point in the direction of rejecting the *Functional Compensation Hypothesis* as stated in Hochberg (1986): since–s deletion varies across social groups, we would expect to find functional compensation in those social groups (males and lower occupational classes) which

showed the higher rates of deleted –s. This is not the case; moreover, contrary to Hochberg's hypothesis, female speakers show a pattern of functional compensation while males do not.

Cameron (1992) completes his study of functional compensation in Puerto Rican Spanish by comparing the effect of *Verb Ambiguity* on subject expression in that dialect and in a non-s-deleting dialect in Madrid. Following the line of reasoning of the *Compensation Hypothesis*, those verb tenses in which ambiguity is possible should feature a higher rate of expressed subjects. Cameron classifies verb tenses thus: unambiguous verb tenses are the present indicative of the verb *ser*, 'to be', a highly irregular verb, and the preterite of all verbs, in which first, second and third person singular bear exclusive morphology in all Spanish dialects; ambiguous verb forms are divided in two categories: on the one hand, two-way ambiguous tenses, such as the present indicative, in which when –s is deleted second and third person singular become homophonous (cf. *tú com-e-~~s~~, él/ella com-e*); on the other, three-way ambiguous tenses, such as the imperfect or the conditional, for which first, second, and third person are homophonous in –s deleting dialects (cf. *yo com-ía, tú com-ía~~s~~, él/ella com-í-a*). Results for *Verb Ambiguity*, which Cameron labels *Verb Class*, are presented in Table 4.16.

Table 4.16. Rates of pronominal expression associated with *Verb Class* in San Juan and Madrid (Cameron 1996:317).

		overt %	*overt N*	*null %*	*null N*	*total*
SAN JUAN						
UNAMBIGUOUS VERB FORMS	*Present of ser*	62	29	38	18	47
	Preterite	42	184	58	257	441
TWO-WAY AMBIGUITY		50	330	50	325	655
THREE-WAY AMBIGUITY		55	340	45	281	621
MADRID						
UNAMBIGUOUS VERB FORMS	*Present of ser*	31	14	69	31	45
	Preterite	22	51	78	186	237
TWO-WAY AMBIGUITY		26	263	74	735	938
THREE-WAY AMBIGUITY		28	65	72	167	232

Chi-square values for the differences between verb classes in San Juan:
Preterite vs. Two-way ambiguity: chi square 7.934 $p < 0.1$.
Two-way vs Three-way ambiguity: chi square 2.440 not significant.
Preterite vs Three-way ambiguity: chi square 17.508 $p < .001$

The unambiguous form of *ser* in the present tense contradicts the working hypothesis: in spite of being a maximally unambiguous tense, overt subjects appear at a rate of 62% in San Juan and 31% in Madrid. The fact that *ser* is a stative verb might explain this trend, as stative verbs typically favor overt subjects. In San Juan, two-way ambiguous verbs display the exact same percentages of null and overt subjects, while three-way ambiguous forms feature higher numbers of overts. The chi-square values for the differences among unambiguous and ambiguous forms return a significant difference between preterits and three-way ambiguous verbs, and a non-significant difference between two-way and three-way ambiguous verbs.

We might conclude from the results in Table 4.16 that *Verb Ambiguity* plays a relevant role in subject expression in both dialects, since if we leave out of the comparison the verb *ser*, overt subject use increases as verb forms increase in ambiguity (41% for preterits, 50% for two-way ambiguous forms and 55% for three-way ambiguous forms in San Juan, and 22% for preterits and 26%, 28% for ambiguous forms). However, it does not seem to be the case that *Verb Ambiguity* has a stronger effect in a s-deleting dialect like the one spoken in San Juan.

Furthermore, as numerous researchers have pointed out (Cameron 1992, Flores-Ferrán 2002, Travis 2005, 2007, among many others), the relative role of morphological ambiguity cannot be assessed accurately without taking into account the other type of ambiguity that may arise with respect to Spanish subjects, i.e., contextual ambiguity, as morphological ambiguity is usually resolved in speech with contextual information or pragmatic knowledge.

4.2.2.2. Morphological ambiguity and contextual ambiguity: On the role of *Co-reference* in Caribbean Spanish subject expression.

Co-reference has been identified as the single variable that uncontroversially has an effect on subject expression in all Caribbean and Non-Caribbean Spanish dialects. This result has been regarded as puzzling by some researchers (Cameron 1992, 1996). In light of the high overt subject rates displayed by Caribbean dialects, one might expect that the strength of *Co-reference* be blurred in these varieties. Rather, what has arisen from variationist analyses is that in fact Caribbean

dialects feature more overt subjects in context of co-reference than their Non-Caribbean counterparts, but this does not preclude *Co-reference* ranking among the strongest variables that regulate subject use. By way of an example, Table 4.17 offers the results for this variable in Puerto Rican Spanish:

Table 4.17. Rates of pronoun expression for Co-reference (Cameron 1992:177)

Preceding subject is >	*Co-referent*		*Non-co-referent*	
	%	*N*	*%*	*N*
OVERT SUBJECTS	31%	316	57%	630
NULL SUBJECTS	69%	689	43%	475
TOTAL	1005		1105	
Chi-square: 139.141 $p < .001$				

Given that overt subject rates have been found to differ significantly among singular and plural pronouns, it is worth examining the relative role of *Co-reference* for each type of pronoun. Table 4.18 offers percentages of overt subject use for singular and plural subjects for the two dialects investigated by Cameron, Madrid, Spain, and San Juan Puerto Rico:

Table 4.18. Rates of subject expression by plurality (Cameron 1992:241).

	San Juan		***Madrid***	
	Overt	*Total*	*Overt*	*Total*
SINGULAR	50%	1,764	26%	1,512
PLURAL	19%	358	7%	549

The data in Table 4.18 show that differences in rates occur in both dialects with respect to subject plurality, although rates are much higher in both contexts in Puerto Rican Spanish. When *Co-reference* is crosstabulated with subject plurality it can be observed that both dialects approximately double overt subject rates when the preceding and target subjects are non-co-referent:

Table 4.19. *Co-reference* crosstabulated with singular and plural subjects (Cameron 1992: 178).

Preceding subject is >	*Co-referent*		*Non-co-referent*	
NUMBER	*Overt %*	*Overt N*	*Overt %*	*Overt N*
SINGULAR	35%	876	66%	879
PLURAL	10%	129	24%	228
Chi-square singular subjects: 168.141 Chi-square plural subjects: 10.121 $p < .001$				

Even though the increase in rates when subjects are not co-referent is very similar in singular and plural subjects, Cameron (1992) points out that there is a very relevant qualitative difference between singular and plural pronouns: while a singular subject is always clearly co-referent or non-co-referent with a preceding subject, a plural subject may not be co-referent with a preceding subject and yet share its reference with any combination of subjects and objects in the preceding discourse.

In order to test whether this difference between singular and plural subjects has any effect on the workings of *Co-reference*, Cameron designs it as a scalar variable in which plural subjects are analyzed as sets. Constraints for plural subjects reflect whether elements of a set have been explicitly mentioned in the preceding discourse, and at what distance (within 5, 6 to 10, or more than 10 clauses); constraints measuring whether the subject is inferable from the preceding discourse, even if not explicitly mentioned, are also included within this factor group. A steady increase of plural pronominal subjects is obtained for each of these constraints, as illustrated in Table 4.20. The comparison between San Juan and Madrid in this table evidences that similar results are obtained in varieties that display different degrees of morphological ambiguity, suggesting thus that co-reference issues override morpho-syntactic properties in plural subject expression:

Table 4.20. Set to elements saliency: San Juan versus Madrid (Cameron 1992:249)			
San Juan, P.R.			
DISTANCE BACK OF SET-ELEMENTS	*% of overt subjects*	*Total N of subjects and %*	
WITHIN 5 CLAUSES	14%	317	(89%)
WITHIN 6 TO 10	44%	18	(5%)
BEYOND 10	67%	12	(3%)
NEW	67%	9	(3%)
TOTAL		356	
Madrid, Spain			
DISTANCE BACK OF SET-ELEMENTS	*% of overt subjects*	*Total N of subjects and %*	
WITHIN 5 CLAUSES	5%	507	(93%)
WITHIN 6 TO 10	17%	18	(3%)
BEYOND 10	35%	17	(3%)
NEW	67%	6	(1%)
TOTAL		548	

Variable Rule programd factor weights further evidence that plural subjects pattern similarly in both dialects:

Table 4.21. Variable rule program weights for set-to-elements saliency: San Juan versus Madrid (Cameron 1992:250).		
DISTANCE BACK	**San Juan**	**Madrid**
WITHIN 5 CLAUSES	.45	.46
WITHIN 6 TO 10 CLAUSES	.78	.72
BEYOND 10	.90	.88
NEW	.91	.97

These results cast doubts on the hypothesis sustained in previous work (Bentivoglio 1980) that the low rate of expressed plural subject pronouns results from the fact that plural morphology is 'heavier' and does not display ambiguities in Spanish. While this may in fact be a contributing factor, Cameron's results strongly suggest that the mention of members in the set represented by the plural subject in the immediately preceding discourse is the most favouring constraint on null plural subjects.

Cameron also notes that those contexts favoring an expressed subject in speech occur infrequently: in his study, contexts in which set members were either not mentioned at all in the preceding discourse or mentioned at a distance of more than 10 clauses represented only 6% of all

tokens. Consequently, independently of morphological ambiguity or lack thereof, because the majority of plural subjects happen to occur in contexts that favor null subjects, these pronouns are frequently null overall.

As far as singular subjects are concerned, San Juan and Madrid also behave very closely with respect to the strength of *Co-reference*:

Table 4.22. *Co-reference* in singular subjects only: San Juan versus Madrid. (Cameron 1992: 244).

	San Juan		*Madrid*	
	Overt	*Total*	*Overt*	*Total*
NON-CO-REFERENT	66%	879	38%	740
CO-REFERENT	35%	876	14%	769

The parallels become even clearer when rates are translated into Variable Rule Program probabilities. San Juan and Madrid are practically identical:

Table 4.23. Variable Rule Program weights for *Co-reference* (singular subjects only) San Juan versus Madrid (Cameron 1996: 244).

	San Juan	*Madrid*
NON-CO-REFERENT	.65	.66
CO-REFERENT	.35	.35

The numbers in Table 4.23 indicate that the two dialects share their grammar of subject expression as far as *Co-reference* is concerned. Very similar results were arrived at by Silva-Corvalán (1989), who compared probabilities of occurrence of null and overt first person subjects as related to *Co-reference* in her data from Mexican-American Spanish with Bentivoglio's (1987) results for the Spanish of Caracas, Venezuela. Again, identical factor group weights were returned after Variable Rule Program application: both in Los Angeles and in Caracas probabilities of .66 and .34 were found for non-co-referent and co-referent subjects respectively.

The next relevant question is whether contextual ambiguity overrides morphological ambiguity or the other way around. The results presented in Table 4.16 above suggested that *Verb Ambiguity* does not arise as a more significant variable in s-deleting than in non-s-deleting dialects. When *Verb Ambiguity* and *Co-reference* are crosstabulated in both types of dialects, findings are as

follows: on the one hand, co-reference seems to nullify the contribution of verbal morphology; on the other, non-co-reference causes the contrasts in degrees of ambiguity among different verb classes to emerge:

Table 4.24. *Verb class* crosstabulated with *(Non-)Co-reference*: San Juan versus Madrid (Cameron 1992: 268).

***Verb class* crosstabulated with *Co-reference* in San Juan and Madrid**				
	San Juan		*Madrid*	
	Overt subjects %	*Total*	*Overt subjects %*	*Total*
PRETERITE	32%	238	13%	143
TWO-WAY AMBIGUITY	32%	280	14%	477
THREE-WAY AMBIGUITY	38%	333	13%	120
***Verb class* crosstabulated with *Non-coreference* in San Juan and Madrid**				
	San Juan		*Madrid*	
	Overt subjects %	*Total*	*Overt subjects %*	*Total*
PRETERITE	55%	200	34%	93
TWO-WAY AMBIGUITY	64%	371	37%	519
THREE-WAY AMBIGUITY	73%	286	44%	112

Parallels are maintained when rates are translated in Variable Rule Program probabilities; in fact, virtually no differences are detected between Madrid and San Juan:

Table 4.25. Variable rule program weights associated with *Verb Ambiguity* and *Co-reference* in San Juan and Madrid (Singular subjects) (Cameron 1992:270).

		San Juan	*Madrid*
PRETERITE			
	Non-co-referent	.57	.61
	Co-referent	.33	.32
TWO-WAY AMBIGUITY			
	Non-co-referent	.61	.63
	Co-referent	.30	.34
THREE-WAY AMBIGUITY			
	Non-co-referent	.74	.76
	Co-referent	.39	.39
PRESENT TENSE OF 'SER'			
	Non-co-referent	.85	.83
	Co-referent	.45	.48

As the data in Table 4.25 make clear, not only there are no differences between San Juan and Madrid with respect to the relative effects of *Co-reference* and *Verb Ambiguity*; this similarity also ranges over all eight categories that result from the intersection of these two variables.

According to Cameron (1996), these results also contribute to the conclusion that San Juan and Madrid share an underlying grammar of subject expression. Despite significant differences in frequency of overt subjects, in both dialects the strength of the factor groups involved is very similar. However, differences in constraint rankings within *Person*, for instance, suggest that Caribbean dialects are undergoing grammatical restructuring. Cameron explains this apparent paradox by resorting to Kroch's (1989) *Constant Rate Hypothesis*. Kroch argues that in language processes, "the strength of these effects [i.e. constraints on variation related to discourse and processing] remains constant as the change proceeds (1989:238)". As a result of the *Constant Rate* property, the sources of language change cannot always be detected in contextual effects, i.e., in constraints on variation. Kroch further proposes that causes for language change are to be searched "at more abstract levels of structure" (1989:239).

Cameron (1996), building on Morales (1989), suggests that these structural shifts are to be found in Caribbean dialects in word order patterns. Morales, in light of the results she obtained from comparing Puerto Rican, European and Argentinean Spanish, proposed that focusing on the shift towards a fixed SVO order in Caribbean dialects rather than on overt subject rates might contribute significantly to our understanding of grammatical restructuring in subject expression patterns.

In fact, from a structural perspective, as discussed in Chapter Two, what emerges as a significant grammatical shift is precisely that preverbal subjects are widespread in Caribbean dialects, even in constructions that categorically call for a null or postverbal subject in Non-Caribbean Spanish, such as *wh*-interrogatives or personal infinitives. If a connection could be found between the shift to a fixed SVO order and high rates of overt subject insertion, this would lend support to Morales' proposal and the findings in theoretical work that structural shifts are at the source of Caribbean restructuring of subject grammar.

Cameron (1996) points out that such a connection has been already documented in the theoretical literature. He discusses Kroch's (1989) work on Middle French and Platzack's (1987) investigation of Swedish Stylistic Fronting. Kroch reports that in Middle French (1400-1700)

subject-verb inversion and null subject use disappeared at the same time. Similarly, Platzack found a parallel decline in the frequency of null subjects in finite clauses and Stylistic Fronting in subordinate clauses in the Swedish texts written between 1530 and 1730 he examined. This led him to propose that these two phenomena are "reflexes of a single parameter in the grammar." (Platzack 1987:377).

Cameron also stresses that in order to determine whether SVO word order and frequency of overt subject insertion are related to one another (either causally or parametrically) in Caribbean dialects, the inquiry could take two alternative paths: from a quantitative perspective, researchers could seek an implicational relationship between rates of overt subject use and rates of preverbal subject insertion. From a structural point of view, researchers could investigate whether widespread Caribbean overt subjects appear in SVO orders, and a principled explanation of this relationship should be provided.

As far as the structural view on Caribbean facts is concerned, recall from the discussion in Chapter Two (cf.§2.3.4.1.) that overt Caribbean pronouns are indeed mainly inserted in a preverbal position. Furthermore, the position of Caribbean pronominals has been derived from their feature specification, in which they relevantly differ from Non-Caribbean Spanish (Toribio 2000, 2001, Ordóñez and Olarrea 2006, Martínez Sanz 2007, Camacho 2008). Essentially, Caribbean pronouns have been described as 'weak' elements, as opposed to NSL Spanish 'strong' nominative pronouns. As such, they lack the anaphoric meaning of NSL Spanish pronouns and have differing syntactic and prosodic properties. A further consequence of these differences is that Non-Caribbean and Caribbean Spanish pronouns are merged in distinct positions in the structure. Specifically, while Non-Caribbean pronominals are merged higher in the structure, where they can be separated from the verb, Caribbean weak pronouns are merged in lower positions to keep, among other properties, subject-verb adjacency and SVO word order.

Going back to variationist investigations of Spanish subject grammar, a quantitatively motivated connection between SVO word order and high rates of overt subject use has also been

found in previous work. Table 4.26, borrowed from Cameron (1996), features a comparison between word orders in a Caribbean dialect (Puerto Rican Spanish as described by Morales 1982) and the Spanish spoken by Mexican-Americans living in Los Angeles. If frequencies of overt subject use are examined, the familiar pattern in which the Caribbean dialect features high rates of overt subject use is obtained: Morales (1982) reports a 53% rate of overt occurrences, and Silva-Corvalán a 39% rate. If there was an implicational relationship between rates of overt use and rates of preverbal subjects, Puerto Rican speakers should be expected to display significantly higher numbers of SVO orders as compared with Mexican-Americans. This is precisely what the comparison between these varieties show. While Puerto Ricans produce overt subjects in a preverbal position at an 81% rate, Mexican-Americans do so only 65% of the time[32].

Table 4.26. Comparison of Puerto Rican and Mexican-American Spanish rates of subject-verb inversion (Cameron 1996:322).

	Puerto Rican Spanish (Morales 1982)	***Mexican-American Spanish*** (Silva-Corvalán 1977)
PRE-VP SUBJECTS	81% (N=1,231)	65% (N=326)
POST-VP SUBJECTS	19% (N=298)	35% (N=175)
Chi-square value of the difference 54.039, $p < .001$		

In light of the converging results arrived at in formal and variationist work, it is argued throughout this dissertation that these available paths to disentangling Caribbean restructuring of subject grammars are not necessarily opposed alternatives, as presented in Cameron (1996), but rather they may complement each other. In other words, it is submitted here that in order to understand the Caribbean shifts what is ultimately needed is an explanation that synthesizes quantitative analysis of attested patterns of use with a syntactic investigation of the speaker's mental grammar from which such patterns are derived.

I should add in this respect that while a explicitly cohesive approach to Caribbean syntactic variation has not been featured in the relevant literature on Spanish subject expression, a few quantitative studies on specific aspects of Spanish subject grammar have drawn theoretical

[32] Note that in both dialects rates of preverbal subjects are very high. This is an expected result, since SVO is the default word order in Spanish.

implications from variationist analysis. Examples include Lizardi's (1993) work on *wh*-interrogatives in Puerto Rican Spanish and Morales' (1997) investigation of word order patterns in Puerto Rican, European and Argentinean Spanish, discussed in the following section.

4.2.2.3. Structural constraints on Spanish subject expression.

While subjects may arise in a preverbal position in several types of structures in Caribbean dialects, it has been proposed in theoretical studies that differences in patterns are attested in specific constructions. For instance, pronominal subjects are overt in a preverbal position in interrogative structures in very high numbers (Ordóñez and Olarrea 2006), whereas it is not clear whether overt preverbal subjects are the norm in non-interrogative constructions. Furthermore, as discussed in Chapter Two, there is disagreement in the relevant literature with respect to the differing behavior of pronominals and full DP subjects. Variationist analyses of the structural constraints that have a role in subject expression have contributed to disentangling these issues.

4.2.2.3.1. Word order patterns in interrogative constructions: pronouns versus DP subjects.

Lizardi (1993) examines the inversion patterns displayed by pronominal and DP subjects in Puerto Rican Spanish. Her study takes as a point of departure the observation made in theoretical work (Torrego 1998) that in Non-Caribbean Spanish subject-verb inversion is obligatory in questions that include an argumental *wh*-word, such as *qué*, 'what', whereas it is optional with non-argumental *wh*-elements like *cuándo*, 'when'. Caribbean Spanish, on the contrary, has been typically described as not featuring subject-verb inversion in *wh*-questions. (cf.§2.3.4.4.). The aim of Lizardi's investigation is to assess to which extent preverbal subjects are possible in Puerto Rican Spanish, and whether the [±argumental] distinction is relevant in this dialect.

The results in Table 4.27 show that in fact the choice of subject position in Puerto Rican Spanish is constrained by the [±argumental] distinction:

Table 4.27. The effect of argumentality on overt subjects (Lizardi 1993:49)						
TYPE OF WH-PHRASE	*Preverbal*		*Postverbal*		*Total*	
	%	*N*	*%*	*N*	*%*	*N*
+ ARGUMENTAL	64	380	36	217	51	597
- ARGUMENTAL	80	462	20	116	49	578
$p > 0.001$						

Two conclusions can be drawn from the data in Table 4.27:

First, Puerto Rican speakers favor preverbal subjects with argumental (64%) and non-argumental (80%) *wh*-words, but more strongly with the latter. This is an expected result, since in other Spanish dialects absence of subject-verb inversion is allowed with non-argumental *wh*-words. Second, the 64% rate of explicit subjects among argumental *wh*-words is striking when compared to Non-Caribbean Spanish, in which these type of *wh*-words categorically undergo subject-verb inversion.

Lizardi (1993) also concludes that whether the subject is a DP or a pronominal also constrains word order possibilities. Her results are provided in Table 4.28[33]:

Table 4.28. The interaction of *Argumentality* and *Subject Type* (Lizardi 1993:52).				
Type of subject	*Type of wh-phrase*		*PreV-Null*	*PostV*
PERSONAL PRONOUNS	+ ARGUMENTAL	*%*	88	12
		N	724	99
	- ARGUMENTAL	*%*	95	5
		N	762	38
	TOTAL	*%*	92	8
		N	1496	137
NOMINALS AND DEMONSTRATIVES	+ ARGUMENTAL	*%*	9	91
		N	12	118
	- ARGUMENTAL	*%*	51	49
		N	81	78
	TOTAL	*%*	32	68
		N	93	196
$p < 0.001$				

[33] Lizardi (1993) conflates preverbal and null pronouns in her statistical analysis. She justifies this procedure in the finding that no relevant differences arise among null and preverbal pronouns with respect to the workings of the [±argumental] distinction.

The [±argumental] distinction interacts strongly with *Subject Type*, defined as [±pronominal]. The data in 4.27 and 4.28 lead Lizardi to propose the following hierarchy of likelihood for preverbal subject occurring, ordered from most to least favoring context:

(15) Hierarchy of preverbal occurrence:

a. -arg pronouns *¿Adónde **tú** vas?*

'Where are you going?'

b. +arg pronouns ¿Cómo **yo** me sentiría?

'How would I feel?'

c. -arg non-pronouns ¿Cómo **los padres** pueden ayudar a sus hijos?

'How can parents help their children?'

d. +arg non-pronouns ¿Qué **ese método** les está haciendo?

'What is that method doing for them?'

This hierarchy is reflected in Lizardi's data set as follows: in Table 4.28, among non-argumentals, preverbal and null subjects are almost categorical for personal pronouns (95%), while nominals display a balanced distribution, with 51% of preverbal and null subjects and 48% of postverbal occurrences. As far as argumental *wh*-words are concerned, the reverse pattern emerges: postverbal subjects are almost categorical for nominals, at a rate of 81%, but marginal among personal pronouns. Lizardi also stresses that the influence of *Argumentality* varies across pronouns: while second person singular *tú* displays a very similar rate of preverbal occurrence with argumental and non-argumental *wh*-words, 96% and 98% respectively, third person pronouns are only featured preverbally 64% of the time with argumental forms.

Lizardi finds a principled explanation for this asymmetry in the comparison of her results to those obtained in Navarro Tomás' (1948) early description of Puerto Rican Spanish, carried out by reference to data collected though a series of questionnaires in the late 1920's. Navarro Tomás pointed out that preverbal *tú* was already widespread and not regulated by argumentality constraints at the beginning of the XXth century. On the contrary, the [±argumental] distinction strongly

constrained the position of third person subjects, which in addition were consistently rejected by his informants when featured in preverbal position and plural form. According to Lizardi, her results essentially mirror those arrived at by Navarro-Tomás, with the difference that hers represent a much more advanced stage of progress.

A breakdown of results for different pronouns of the paradigm is offered in Table 4.29:

Table 4.29. Distribution of preverbal, postverbal and null variants among individual pronouns (Lizardi 1993:65)

PRONOUN		*Preverbal*	*Postverbal*	*Null*	*Total*
YO	*N*	120	17	126	263
	%	46	6	48	16
TÚ	*N*	429	14	265	708
	%	61	2	37	44
ÉL/ELLA, ELLOS/ELLAS	*N*	85	38	170	293
	%	29	13	58	18
UNO	*N*	18	4	-	22
	%	82	18	-	1
NOSOTROS	*N*	5	3	87	95
	%	5	3	92	6
USTEDES	*N*	92	61	89	242
	%	38	25	37	15
TOTAL	*N*	749	137	737	1623
	%	46	8	45	
$p<0.001$					

Note that first and second singular pronouns *yo* and *tú* account for 60% of all tokens in the corpus, and that they have a skewed postverbal to preverbal ratio. Changes in overall rates can be observed if *yo*, *tú* and *uno*, the other pronoun that has a very high preverbal rate, are isolated from the rest of the pronouns. In Table 4.30, 'pronouns (b)' is a reduced sample that conflates *yo*, *tú* and *uno*, and pronouns (a) includes all pronominal forms:

Table 4.30. Distribution of preverbal, postverbal and null variants: Before and after the exclusion of *yo*, *tú* and *uno* (Lizardi 1993:66).

SAMPLE		*Preverbal*	*Postverbal*	*Null*	*Total*
PRONOUNS (A)	*N*	749	137	737	1623
	%	46	8	45	
PRONOUNS (B)	*N*	182	102	346	630
	%	29	16	55	
p<0.001					

In Table 4.30, overall preverbal occurrences have decreased from 46% to 29%, and the frequency of postverbal subjects has doubled. Null subject use has also increased by 10%, from 45% to 55%. Given these results, the logical expectation is for argumentality constraints to be weaker in pronouns (a), which show a higher degree of grammaticalization of the preverbal position, and stronger in pronouns (b). This expectation is borne out, as illustrated in Table 4.31:

Table 4.31. The effect of argumentality on pronouns (b) (Lizardi 1993:67)

[±ARGUMENTAL]		*Preverbal-Null*	*Postverbal*	*Total*
[+ARGUMENTAL]	*N*	255	75	330
	%	77	23	52
[-ARGUMENTAL]	*N*	273	27	300
	%	91	9	48
	N	528	102	630
	%	84	16	
p<0.001				

In addition to *wh*-questions, Lizardi examines clefted questions, which are also very productive in Caribbean dialects (cf.§1.4). In her data set, clefted questions appear with both pronominal and DP subjects, as in (16a-b):

(16) a. ¿Y qué es lo que *él* quiere?

'What is it that he wants?'

b. ¿Qué es lo que *el abuelo* nos está mandando?

'What is it that grandpa is sending us?'

In Lizardi's corpus, which comprises 1912 tokens, there are only 130 clefted questions. They display a slightly higher rate of preverbal subjects as compared with non-clefted interrogatives

(53% for clefted questions versus 44% for non-clefted interrogatives. Both types of questions present very similar figures for the postverbal (13% and 17% respectively) and the null (34% versus 39%) variants. Finally, most clefted questions feature pronominal subjects (81%) and argumental *wh*-forms (62%). Lizardi reports that for the speakers' choice of subject position differences in *Subject Type* and *Argumentality* are not returned as statistically significant.

To sum up, the investigation of structural properties such as *Subject Type* and *Argumentality* contributes to the accurate description of subject expression in Caribbean interrogatives. The effects of *Argumentality* vary among nominals and pronouns, and among the pronouns in the paradigm. In clefted questions, where subject-verb inversion is not categorical in Spanish, *Argumentality* is not significant.

Another factor group that probes into structural properties, *Clause Type*, has also been found to be significant in both interrogative and non-interrogative Caribbean constructions. Relevant results available in previous work are discussed in the following paragraphs.

4.2.2.3.2. Clause Type.

The factor group *Clause Type* has been examined in only a subset of variationist studies that deal with subject expression. Studies generally find that main clauses neither favor nor disfavor overt subject use, whereas embedded and conjoined clauses favor null subjects (Enríquez 1984, Cameron 1992, de Prada 2009).

However, more interesting results obtain when embedded clauses have been broken down into subtypes and, importantly, subject position, as in Morales (1997). This analysis distinguishes *Relative Clauses*, *Subordinate Clauses*, *Clauses with preposed NP or PP objects* and *Other*, and compares their effects in the dialects of Madrid, San Juan and Buenos Aires. As Table 4.32 illustrates, relative clauses favor overt subject use the most. Interestingly, clauses with a preposed NP or PP object disfavor overt preverbal subjects.

Table 4.32. *Clause Type* crosstabulated with subject position in San Juan, Madrid and Buenos Aires (Morales 1997:156-158)

		Preverbal N	*Preverbal %*	*Preverbal weight*
RELATIVE CLAUSES				
Pronoun	1ST AN 2ND PERSON (*yo & tú*)	94/207	46	.58
	3RDPERSON (*él/ella/ellos/ellas*)	38/105	36	.64
SUBORDINATE CLAUSES				
Pronoun	1ST AN 2ND PERSON (*yo & tú*)	111/287	39	.54
	3RDPERSON (*él/ella/ellos/ellas*)	54/215	25	.54
CLAUSES WITH A PREPOSED NP/PP OBJECT				
Pronoun	1ST AN 2ND PERSON (*yo & tú*)	42/171	25	.36
	3RDPERSON (*él/ella/ellos/ellas*)	7/90	8	.25
OTHER				
Pronoun	1ST AN 2ND PERSON (*yo & tú*)	587/1541	38	.50
	3RDPERSON (*él/ella/ellos/ellas*)	54/215	25	.50
TOTAL		3,540		

When *Relative Clauses* are crosstabulated with *Dialect*, it becomes clear that San Juan is the community responsible for the high rates of preverbal subjects and the factor weights favoring them in speech:

Table 4.33. Preverbal subjects in Relative Clauses: San Juan, Madrid and Buenos Aires (Morales 1997:161)

	San Juan		**Madrid**		**BBAA**	
	N	*%*	*N*	*%*	*N*	*%*
PRONOUN						
1ST AND 2ND PERSON (*yo, tú*)	56/85	66	16/60	27	23/62	37
3RD PERSON (*él/ella, ellos/ellas*)	28/46	61	5/26	19	5/32	16

As to why relative clauses[34] should favor overt subjects, Morales suggests that this result might derive, at least in part, from the properties of relativization. Specifically, she notes that relative clauses are frequently associated with the Topicalization of elements; in fact, the antecedent of the relative clause is always topicalized. When a DP subject or a subject pronoun is inserted

[34] It should be noted that in this discussion I am always referring to object relative clauses, as subject relative clauses categorically bear a null subject in Spanish (cf. *Las cosas [que trajo Mario ayer]*, 'The things Mario brought over yesterday' versus *La chica [que (*ella) vino ayer]*, 'The girl that came over yesterday'). Nonetheless, subject relative clauses with overt pronominal subjects have been documented in Dominican Spanish (Camacho 2008) and Brazilian Portuguese, and they were marginally attested in the corpus gathered for this dissertation. These constructions, and what their occurrence entails for a theory of Caribbean grammatical restructuring, will be discussed in Chapter Six.

overtly in the relative clause it also conveys the meaning of a topicalized, contrastive element, as illustrated in the contrast between (17a) and (17b):

(17) a. Y posiblemente esas formas que ø adoptan, no son reveladoras de un
(18) fondo muy positivo.

b. Y posiblemente esas formas que *ellos* adoptan, no son reveladoras de un fondo muy positivo.

'And probably the ways they (=youngsters) adopt do not reveal a very positive background.'

(M MI, Morales 1997:159)

Example (17a) corresponds to Madrid Spanish. However, when examining the Puerto Rican corpus, overt subjects in relative clauses do not always convey a contrastive reading, which suggests that, as attested elsewhere in the Puerto Rican grammar, overt subjects might have lost their discursive properties to some extent:

(19) Entonces cuando los has leído, o cuando has leído las veinte páginas, o las cuarenta, o las ochenta que te asignaron, que tú crees que vienes bien preparado, el profesor te pregunta ¿qué se decide en el caso tal? y tú inmediatamente coges el *issue* importante y tú dices, aquí estoy yo, pues mire, ahí se dice tal y tal cosa. Dice sí, sí, pero cuénteme los muñequitos, qué dijo Fulano, que dijo Zutano. En otras palabras, son veinte casos que *tú* te los tienes que aprender bien.

'And then, when you have read them, or when you have read the twenty pages, o the forty pages, or the eighty pages you have been assigned, when you think you are well prepared, the professor asks you 'what is it said in this or that case? And you immediately catch what is the relevant issue and you say 'here I am, look, in this case this and that was said'. And he says, yes, yes, but act it out for me, what did this person say, what did the other person say. In other words, those are twenty cases that you need to learn through and through.' (SJ MIII, Morales 1997:160).

In the discourse string in (18) even though there is no shift in reference (i.e. no contrast and therefore no need for Topicalization) the speaker inserts an overt subject in the relative clause. However, note that in the example at hand the subject is a non-specific second person singular pronoun, which in Puerto Rican Spanish has been shown to be frequently overt in all clause types.

Lizardi (1993), in her quantitative study of *wh*-interrogatives, finds significant differences between matrix and embedded questions. Interestingly, these differences are significant in the sub-sample of *wh*-questions that feature nominals and demonstratives ($p<0.01$). On the contrary, *Clause Type* is not significant in *wh*-questions with pronominal elements ($p<0.98$). Table 4.34 provides results for nominals and demonstratives:

Table 4.34. Differences in *Clause Type*: Nominals and demonstratives (Lizardi 1993:69)

TYPE OF CLAUSE		*PreV*	*PostV*	*Total*
MATRIX	*%*	27	73	67
	N	52	142	194
EMBEDDED	*%*	43	57	33
	N	41	54	95
TOTAL	*%*	32	68	
	N	93	196	289

$p<0.01$

As Table 4.34 shows, while postverbal subjects are generally favored in interrogative structures with preverbal full DPs, preverbal subjects are more frequent in embedded than in main clauses (43% versus 27%). According to Lizardi, these numbers are not surprising in light of the results obtained by Navarro-Tomás (1948) half a century before her: when Navarro-Tomás first reported the widespread occurrence of preverbal subjects in Puerto Rican Spanish, he noted that embedded contexts allowed for a wider variety of subject elements than questions in main clauses. Specifically, he found that while both first and second person singular pronominal subjects could arise preverbally, first person preverbal subjects appeared only in embedded environments. In the more advanced stage of change represented by Lizardi's corpus, all pronouns can arise preverbally; DP subjects, on the other hand, can also be featured preverbally in matrix and embedded clauses,

but rates of preverbal use are higher in embedded contexts. According to Lizardi, this asymmetry can be explained as a result of the more conservative behavior of full DPs as compared to pronouns. She adds that the fact that preverbal subjects are expanded in higher numbers in embedded clauses than in matrix ones appears to confirm the claim that grammaticalization operates earlier in embedded environments (Trauggot and Heine 1991).

From the examination of Lizardi's (1993) and Morales' (1997) work it can be concluded that the relative significance of the variables that regulate subject use shifts when word order patterns, rather than rates of subject use, are investigated. So far, these shifts in relevance have only been discussed for language internal factors. Interestingly, differences in strength also arise with respect to language external factors when the focus is set on structural properties.

4.2.3. Language external fac tors and grammatical restructuring in Caribbean dialects.

Variationist studies have tested the relevance of a number of language external, sociodemographic variables for Spanish subject expression. While such variables are typically non-significant in Mainland varieties, a few studies on Caribbean dialects have found *Gender*, *Age* and *Occupation* or *Education* significant. Cases in point are Lizardi (1993) and Ávila-Jiménez (1996), who investigate differences in subject expression across generations in Puerto Rican Spanish. Probing for the role of language external variables becomes particularly relevant in Caribbean Spanish, as these dialects have been suggested to be in the process of restructuring their grammar of subject expression. Specifically, finding different patterns of null and overt subject use in younger and older speakers would lend support to the hypothesis of ongoing language change in the Caribbean.

Change in progress may be examined in a variationist study either in 'real time', by examining the speech of the same group of individuals in two different periods of time, or in 'apparent time', by investigating the synchronic distribution of a linguistic variable among different age groups in a community. If an apparent time study reflects a correlation between *Age* and the dependent variable, one might conclude that there is change in progress in the community under

study. However, as noted by Labov (1991, 1994), matters are more complex when dealing with change in syntactic systems, where the consequences of change are not always readily observed, as they can be masked by an 'age grading' effect, defined by Labov (1994:124) as "a regular change of linguistic behavior with age that repeats in each generation." According to this author, age grading effects are a natural feature of syntactic patterns because of the 'permeability' of syntactic structures. As opposed to phonological or phonetic systems, which are fixed in infancy or adolescence and resist change, syntactic structures can incorporate change in a lifespan.

A variationist analysis can also provide information related to the social groups involved in language change. Two types of language change have been defined in this respect (Labov 1991, 1994): *change from above* and *change from below*. Changes from above might be introduced by dominant social classes and usually carry high levels of social awareness. They also show higher rates or probabilities of occurrence in formal styles, are more likely to stereotype those forms that are being substituted by the new, prestigious form, and are often subject to hypercorrection. Changes from below, on the contrary, might be introduced by any social group and usually occur from beginning to completion below the level of social awareness. Labov states that change from below is "the basic form of linguistic change that operates within the system".

In the description of the phenomena related with subject expression in Caribbean Spanish carried out in this chapter, language external factors in Caribbean Spanish have only become relevant when discussing the power of Hochberg's (1986) *Functional Compensation Hypothesis* for explaining rates of overt subject use. Recall in this respect that Cameron (1992) found that while –s deletion had a variable distribution across social groups, with male speakers displaying higher rates of deletion, this fact did not make any difference in the cumulative rate with which the second person subject pronoun *tú* is expressed. Contrary to the logical implications of the *Functional Compensation Hypothesis*, female speakers showed a pattern of functional compensation in overt subject insertion while male speakers did not, even though female speakers show a significantly lower rate of /s/ deletion ($p<0.001$). Moreover, when both specific and non-specific *tú* were

combined in the analysis of functional compensation in male speakers, a counter-functional pattern emerged, as males more frequently expressed the subject pronoun when inflection was present, not absent.

The strength of language external factors changes, however, when patterns of preverbal and postverbal subject use, rather than patterns of null and overt subject insertion, are examined, in Lizardi's (1993) study of *wh*-structures. Specifically, the rate of preverbal subjects varies across *Gender* and *Age* differences. These results, in turn, have triggered a debate on the nature of grammatical shifts in subject expression in Caribbean dialects.

4.2.3.1. *Gender* and *Age*.

As far as *Gender* is concerned, women exhibit higher rates of preverbal subjects than men in *wh*-questions:

Table 4.35. Overall gender differentiation among Puerto Rican speakers (Lizardi 1993:78)

Gender		*PreV-Null*	*PostV*	*Total*
WOMEN	*%*	87	13	56
	N	937	141	1078
MEN	*%*	77	23	44
	N	642	192	834
TOTAL	*%*	83	17	
	N	1579	333	1912
$p<0.001$				

Further statistical analysis shows that *Gender* differentiation is maintained uniformly across subject types, as evidenced in the divergent percentages of preverbal subject use among men and women:

Table 4.36. Overall gender differentiation among Puerto Rican speakers, preverbal occurrences: Pronouns versus nominals (Lizardi 1993:79)

	WOMEN	**MEN**
PRONOUNS (A)	95%	88%
PRONOUNS (B)	88%	80%
NOMINALS	37%	25%

According to Lizardi, the percentages in Table 4.36 suggest that women might be one step ahead of men in the grammaticalization of preverbal subjects. In this respect, she stresses that men's

percentage of preverbal subjects for the less conservative groups of pronouns, pronouns (a) is identical to the percentage displayed by women for the most conservative pronouns (b).

Regarding *Age*, Lizardi (1993) initially classified speakers into seven different age groups. Initial tests for statistical significance, however, showed that two natural age groups emerged among Puerto Rican speakers, below 50 years old and above 50 years old. In other words, 50 years old arises as the cut-off point for the seemingly ongoing Puerto Rican grammaticalization process. When these two age groups are compared, percentages reflect that proportions are more skewed proportions among younger speakers than among speakers over 50 years old:

Table 4.37. Overall age differentiation among Puerto Rican speakers (Lizardi 1993:80)

AGE		*PreV-Null*	*PostV*	*Total*
UNDER 50	*%*	90	10	
	N	1033	117	1150
OVER 50	*%*	72	28	
	N	546	216	762
TOTAL	*%*	83	17	
	N	1579	333	1912
$p<0.001$				

The results of the analysis across subject types indicate that younger speakers exhibit higher frequency of the preverbal-null variant in all three types, pronouns (a), pronouns (b) and nominals. They also display consistent rates for pronouns (a) and pronouns (b) (96% and 92% respectively). Lizardi argues that these percentages point to syntactic leveling in pronominal use among young speakers. On the contrary, speakers above 50 years old display a sharper decrease in percentages (from 84% in pronouns (a) to 75% in pronouns (b)), which reveals in Lizardi's view different stages of grammaticalization of the preverbal position for each type of pronouns.

Table 4.38. Overall age differentiation among Puerto Rican speakers: Percentages for pronouns and nominals (Lizardi 1993:81)

	AGE GROUP 1 (-50)	AGE GROUP 2 (50+)
PRONOUNS (A)	96%	84%
PRONOUNS (B)	92%	75%
NOMINALS	40%	28%

Age differences have also been found significant in Ávila-Jiménez (1996). In her study, the aim is to give an accurate description of null and overt subject insertion patterns in Puerto Rican Spanish, and therefore she does not explore word order differences. Before I move on to discuss Age results in Ávila-Jiménez's work, overall percentages for overt and null subject distribution in her data set are provided in Table 4.39:

Table 4.39 Distribution of Personal Pronouns in Puerto Rican Spanish (Ávila-Jiménez 1996:97)

	Null		***Overt***	
PERSONAL PRONOUN	*%*	*N*	*%*	*N*
YO	49	1045	51	1077
TÚ [+ SPECIFIC]	41	52	59	76
TÚ [– SPECIFIC]	37	31	63	53
ÉL	66	443	34	227
ELLA	62	326	38	203
USTED	50	17	50	17
UNO	8	7	92	80
NOSOTROS	85	477	15	104
USTEDES	58	7	42	5
ELLOS [+ SPECIFIC]	76	75	24	24
ELLOS [– SPECIFIC]	95	386	5	22
ELLAS	58	14	42	10
TOTAL	60	2898	40	1902
$p<0.001$				

Percentages of null and overt subject in Ávila-Jiménez's study are very close to those provided in other studies of Puerto Rican Spanish (Cameron 1992, 1996, Flores-Ferrán 2002). Her study departs from previous work in that she does find *Age* as a significant variable for explaining Puerto Rican patterns. Similarly to Lizardi (1993), a cut-off point of 50 years old emerged in her data set to differentiate younger and older speakers. Since in her corpus speakers between 10 and 19 years old behave quite differently from other age groups, Ávila-Jiménez analyzes them separately. Pronominal distribution by age group in her data set is offered in Table 4.40:

Table 4.40 Pronominal distribution by age group in Puerto Rico (Ávila-Jiménez 1996:98)				
	Null		***Overt***	
AGE	*%*	*N*	*%*	*N*
GROUP A (5-49)	57	1558	43	1172
GROUP B (50+)	63	879	37	506
GROUP C (10-19)	67	461	33	224
$p<0.001$				

Although differences in percentages seem small, they are statistically significant for *Group A* (5-49) and *Group B* (50+). *Group C* results were not significant, but Ávila-Jiménez still distinguishes them as a group because these speakers showed the highest rates of use of plural pronominals (20% of the total plural occurrences in the data set), as a result of participating in group interviews; since plurals tend to bear null pronouns, the speech of *Group C* displays the highest rates of overt plural subjects.

Ávila-Jiménez compares her results with the ones obtained by Morales (1986) with reference to a corpus of urban Puerto Rican speech gathered in 1982-1983:

Table 4.41. Pronominal distribution by age group (Morales 1986)				
	Null		***Overt***	
AGE GROUP	*%*	*N*	*%*	*N*
16-50	58	3829	42	2796
50+	72	1595	28	606
$p<0.001$				

The comparison between the age groups with speakers under 50 years old in Morales' and Ávila-Jiménez's studies, the latter reports, is not statistically significant. In both studies, however, the differences between speakers under and above 50 are significant ($p<0.001$). Taking as a point of departure Labov's (1994) claim that there should be at least half a generation difference between studies in order to analyze linguistic change, Ávila-Jiménez suggests that the comparison between the two corpora at hand, gathered in the early eighties and early nineties respectively, suggests that

older speakers are showing less use of null pronominals. Therefore, they are showing a tendency towards the adoption of the new, overt subject trend.

In sum, interesting patterns that suggest ongoing language change have been found in Puerto Rican Spanish. In light of these facts, Lizardi (1993) explores jointly *Age* distribution and *Gender* differentiation. Her results, provided in Tables 4.42-4.44, show the following trends:

First, salient *Age* differences arise only within male speakers:

Table 4.42. Uniform grouping of *Gender* according to the cut-off point of 50 years of age (Lizardi 1993:83)

GENDER AND AGE		*Preverbal-Null*	*Postverbal*	*Total*
WOMEN UNDER 50	*N*	678	82	760
	%	89	11	
MEN UNDER 50	*N*	355	35	390
	%	91	9	
WOMEN OVER 50	*N*	259	59	318
	%	81	19	
MEN OVER 50	*N*	287	157	444
	%	65	35	
TOTAL	*N*	1579	333	1912
	%	83	17	

Second, while the cut-off point of 50 years is valid for men ($p<0.001$), more statistically significant cut-off points are found for women:

Table 4.43. Apparent time redistribution for women (Lizardi 1993:84)

AGE GROUP		*Preverbal-Null*	*Postverbal*	*Total*
3-18	*%*	94	6	
	N	207	14	221
19-59	*%*	86	14	
	N	627	98	725
60+	*%*	79	21	
	N	88	24	112
TOTAL	*%*	87	13	
	N	922	136	1058
$p<0.001$				

Third, contrasts between male and female speakers are almost non-existent among young speakers, which in Lizardi's view indicates the advanced movement of preverbal subject towards categoricity in this age group:

Table 4.44. Fine age distribution by *Gender* (Lizardi 1993:85)		
AGE GROUP	*Preverbal-Null subject use*	
3-18	MEN	91%
	WOMEN	92%
19-35	MEN	91%
	WOMEN	89%
36-49	MEN	88%
	WOMEN	86%
50+	MEN	65%
	WOMEN	81%

Lizardi suggests that if the distributions in Table 4.44 are representative of real time, then the youngest and oldest age groups in the Table should roughly correspond to the most advanced stage and to an earlier stage respectively.

Other language external variables such as *Occupation* or *Education*, are not returned as significant in Lizardi's work. *Occupation*, however, was found to constrain subject expression in Puerto Rican Spanish by Ávila-Jiménez (1996).

Regarding *Education*, relevant differences were found when the initial seven-group classification of speakers was reduced to a two-group opposition. As depicted in Table 4.45, more educated speakers display higher rates of overt subject use:

Table 4.45 Pronominal distribution based on educational level (Ávila-Jiménez 1996:113)				
EDUCATION	*Null*		*Overt*	
	%	*N*	*%*	*N*
LEVEL 1 (NO SCHOOL TO HIGH SCHOOL)	63	1486	37	869
LEVEL 2 (SOME COLLEGE-PROFESSIONAL SCHOOL)	58	1412	42	1033
$p<0.001$				

When *Age* and *Education* are crosstabulated, the percentages obtained resemble very closely the ones in 4.45 above, i.e., rates obtained by submitting only *Age* to statistical analysis; Ávila-Jiménez interprets these results as evidencing that *Age* and *Education* have a similar strength in constraining subject use:

Table 4.46. Pronominal distribution based on educational level and age (Lizardi 1996:115)												
AGE	AGE GROUP A (5-49)				AGE GROUP B (50+)				AGE GROUP C (10-19)			
	Null		***Overt***		***Null***		***Overt***		***Null***		***Overt***	
EDUCATIONAL LEVEL	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*
LEVEL 1	58	417	42	301	65	673	35	369	67	396	33	199
LEVEL 2	57	1141	43	871	60	206	40	137	72	65	28	25

A similar effect obtains with respect to *Occupation*. When this variable is considered on its own, speakers in *Occupational Level A* (*Professional*) display lower rates of overt subject use than speakers in *Occupational Level B* (*Managerical, Clerical, Skilled and Unskilled workers*):

Table 4.47. Pronominal distribution by occupation (Ávila-Jiménez 1996:117)				
	Null		***Overt***	
OCCUPATIONAL LEVEL	%	N	%	N
LEVEL A (PROFESSIONAL)	57	1189	43	908
LEVEL B (MANAGERIAL, CLERICAL, SKILLED, UNSKILLED)	63	1709	37	994
$p<0.001$				

However, Ávila-Jiménez reports that when *Age* and *Occupation* are crosstabulated, the following pattern emerges, again percentages resemble closely the ones in Table 4.40: results for *Age group A* and *Occupational level A* are the same (57%) and results for *Age groups B* and *C* and *Occupational Level B* are also very similar (63% and 67%). Since no parallel results in the previous literature have been found for *Education* and *Occupational Levle*, Ávila-Jiménez does not elaborate further on an interpretation of these trends beyond noting that like *Education*, *Occupational Level* seems to have a similar strength to Age in regulating subject use.

To sum up, while both Lizardi (1993) and Ávila-Jiménez (1996) found *Age* significant, they differ with respect with other sociodemographic variables: the crosstabulation of *Age* and *Gender* offered significant results in Lizardi's study, while Ávila-Jiménez found that *Education* and *Occupation* patterned with *Age* in significance. The next question that arises is what consequences these patterns may have for a characterization of the seemingly ongoing change in Caribbean subject expression.

4.2.3.2. Interpreting distributions in apparent time: on the nature of grammatical restructuring.

Lizardi (1993) and Ávila-Jiménez (1996) discuss whether the distribution of subjects in apparent time in their Puerto Rican corpora reflects ongoing language change or rather represents a pattern of age grading, and both arrive at the conclusion that Caribbean grammatical restructuring of subject grammar represents a case of change from below:

On the one hand, it definitely seems to take place under the level of awareness, as speakers are reported to not associate any stereotype or stigma with widespread preverbal, overt subject use. On the other, Lizardi (1993) points out that changes from below have been described as typically led by women (Labov 1994), and women are precisely the ones displaying more advanced word order patterns in her study. Clefted questions, which might be regarded as an innovative strategy in interrogative constructions, are also concentrated largely in the speech women (70%), and the other group identified as innovators, speakers under 50 years old (67%)[35].

Furthermore, the comparison that Lizardi (1993) carries out between her data set and the corpus gathered by Navarro-Tomás (1948) depicts the grammaticalization of overt preverbal subject in Puerto Rican Spanish as an internally-motivated change, which has advanced along the pronominal paradigm and is extending to nominals in contemporary speech: in the late 1920s, as per Navarro-Tomás' description, SV word order was only available with pronominal elements, and first person singular and plural pronouns were only judged grammatical by speakers in embedded contexts. According to Lizardi, in the early 90s all pronouns of the paradigm may arise preverbally in matrix and embedded contexts, and nominals might also precede the verb, although in smaller numbers.

[35] Note that clefting is a much more productive strategy in Caribbean than in Non-Caribbean Spanish. Even if this is the case, clefted questions only account for 6% of the overall number of tokens in Lizardi's corpus, probably because they are less frequent in speech that non-clefted *wh*-structures. She points out that this lesser frequency might be associated with these constructions bearing a specific discursive meaning. The role of clefting in Caribbean varieties with respect to Focus phenomena will be approached in Chapter Seven.

The extent of grammatical restructuring is not easy to assess, however, given that, as stated by Labov (1994) adult systems seem to be permeable to syntactic innovations. Labov suggests that the fact that flat age distributions characterize syntactic change might explain the lack of reports on syntactic shifts in the literature. However, he also points out that finding age grading in ongoing syntactic changes might not be necessarily interpreted as a contradictory result: "[finding both processes together] suggests that that we may have been setting up a misleading opposition between age grading and generational change (…). It is possible that age grading is involved in the mechanism of real-time change for certain types of linguistic change." (Labov 1994:97).

In other words, change in progress might be evident, nonetheless subordinated to age grading. Labov (1994) illustrates his point with a re-study of Cedergren's (1973, 1983) data of *ch*-lenition in Panama City. Cedergren gathered data on this phenomenon in 1969 and again in 1982. Working with these two data sets, Labov (1994) takes the values obtained for *ch*-lenition in each generation for the 1969 data and plots them in a graph; he then compares the expected values for those same age groups in 1982[36]. He observed that the results did not pattern with those actually reported in Cedergren (1983). Specifically, Cedergren's (1983) results were closer than expected to those obtained in 1969 for the same generations. In principle, this asymmetry between expected and actual results points to age grading. However, when results are looked at closely, it may be observed that except for the two lowest age groups, the 1983 values were consistently 10-15% higher than those obtained in 1969. In conclusion, even if age grading was the dominant configuration, there was a steady increase in lenition, which in turn reflected ongoing change in real time.

In light of these results, Labov (1994) proposes carrying out real time studies in combination with apparent time investigations. He adds that another factor to take into account when assessing grammatical restructuring is the age of the linguistic change at hand in a speech community, since while newer changes tend to show sharp differentiations, changes nearing

[36] Expected values stand here stand for the predicted values, were ongoing change actually happening.

completion typically show moderate age vectors. In this respect, Lizardi (1993) points out that the high numbers she obtained for all pronouns in her study of Puerto Rican interrogatives suggest that the shift to preverbality in this dialect does not represent a new, vigorous change, but rather a late-stage change, which would explain that *Age* results do not display particularly sharp differences.

Both Lizardi (1993) and Ávila-Jiménez (1996) agree in further arguing that their results support the notion of syntactic permeability in adult systems: quantitative evidence of a cut-off point at 50 years old in age groups suggests that older male speakers have retained their syntactic systems despite the advance of the shift towards preverbality and overt subjects; however, the fact that finer significant generational differences arise among women argues in favor of the permeability hypothesis. In addition, Lizardi notes, qualitative evidence also lends support to this hypothesis. In her corpus, speakers older than 50 years old produce utterances that illustrate the two most advanced stages of grammatical restructuring, namely, non-inverted nominals and DPs with argumental *wh*-words:

(20) a. ¿Tú sabes qué *Torres* decía? [70, female]

b. ¿Qué diantres *esa mujer* va a hacer al pueblo todos los días? [56, female]

c. ¿Qué *ese hombre* le ha quitado a eso? [58, male]

It can be concluded from the above discussion that the evidence in Lizardi (1993) and Ávila-Jiménez (1996) points to a relatively stable situation in Puerto Rican Spanish: comparisons with previous studies on this variety (Morales 1986, cf. Table 4.41) showed that there have not been significant changes in the last twenty years in the overall rates of expressed pronominals among the youngest generations of Puerto Rican speakers. As far as speakers over 50 are concerned, a gradual rather than sharp increase of overt subject rates is attested, which also suggests a certain degree of stability in the language change process.

4.3. Summary. Bridging paradigms: where variationist methodology and theoretical concerns meet.

In this chapter I have reviewed variationist approaches to Spanish subject expression. From the above discussion, the following conclusions might be extracted:

(i) Variationist and theoretical work, in spite of differences in research agendas, converge to some extent both in the aims of their investigations and in the challenges they have faced over the years:

In both paradigms the concept of 'rich' inflectional morphology has resisted an accurate description. As discussed in Chapter Two, diachronic facts such as the erosion of inflectional endings in Old French and subsequent loss of NSL properties (Roberts 1993) have been interpreted as evidence for the close relationship between morphology and syntax. However, as typologically different languages were investigated, it became clear that inflectional morphology or lack thereof was definitely not the sole factor involved in subject licensing (Holmberg 2005, 2009, among others). In fact, mechanisms other than agreement are available even in Romance languages, as the investigation of dialectal differences such as the ones displayed in Non-Caribbean and Caribbean Spanish has suggested (Toribio 2000, 2001, Camacho 2008). Simply stated, when these findings are translated into the investigation of synchronic syntactic variation, it cannot be predicted that future Caribbean dialects will develop like French has in the past, and that pronouns will ultimately serve as markers of person that will substitute for person inflectional endings.

Variationist investigations of the synchronic profile of Caribbean dialects confirm this: for instance, the comparison of present-day Puerto Rican and Madrid Spanish carried out in Cameron (1992, 1993, 1996) makes clear that inflectional morphology is not the only factor involved in subject expression in neither of these dialects. While factors such as *Verb Ambiguity* might have a role in patterns of subject use, discourse-related constraints arise as very significant for speakers' choices. Moreover, the relative role of morpho-syntactic constraints such as morphological

ambiguity cannot be accurately described without also taking into account discourse-related, contextual ambiguity. The simultaneous investigation of the multiple factors involved in subject expression is in fact one of the main contributions that variationist work has made to our understanding of syntactic variation in Spanish. This possibility of joint analysis of numerous factors naturally leads to the second relevant conclusion that stems from the discussion carried out so far:

(ii) In probing for the number of factors that are relevant for null and overt subject insertion in Spanish, variationist studies have shown where in the grammar we should look for the traces of grammatical restructuring in Caribbean Spanish, and perhaps most importantly, where *not* to look for them.

In the discussion on the role of discourse-related constraints in Caribbean Spanish in 4.2.1. it was explained that contrary to what was expected by some researchers (Otheguy, Zentella and Livert 2007) from the high rates of overt subjects found in these dialects, discourse-related constraints strongly constrain subject expression in the Caribbean. Cameron (1992) showed that factor weights for the variable *Co-reference* were virtually identical in European and Puerto Rican Spanish. These results might be taken to suggest that given that all Spanish varieties share significant factor groups in subject use, they also share underlying grammars and therefore grammatical restructuring is not attested in Caribbean dialects. Further research into these issues has shown that this is not actually the case: on the one hand, constraint rankings within factor groups reflect significant differences between Non-Caribbean and Caribbean Spanish, which suggests that the signs of grammatical restructuring are to be looked for in these kind of measures. On the other, when not only null and overt subject insertion but also structural properties such as word order patterns are explored, specific patterns unique to Caribbean Spanish are detected (Lizardi 1993, Morales 1997). For instance, significant differences are found with respect to the behavior of pronominal and DP elements.

(iii) The fact that structural properties arise as particularly relevant for disentangling grammatical restructuring in subject expression argues, in my view, in favor of applying a cohesive approach to syntactic variation to arrive at an accurate description of Caribbean subject grammars. The available theoretical investigations can inform quantitative research on the structural variables that might potentially regulate these grammars and that have been overlooked in previous variationist work. In turn, taking a corpus (to which variationist methodology is applied) as a point of departure to draw theoretical conclusions will allow syntactic investigations to build accurate theories on the possibilities of subject expression afforded by Caribbean grammars. So far, variationist investigations have already shown that subject expression in Spanish is indeed a variable phenomenon. Specifically, some of the rules associated with subject insertion that were assumed to be categorical in Spanish have been shown to allow for a certain amount of 'optionality'. It is the stand taken in the present study that accurately defining these potential sites of variability is a necessary step to build syntactic theories that can explain the limits afforded by formal grammars for cross-linguistic and cross-dialectal variation.

Chapter Five. Methodology. Probing into a restructured subject expression grammar.

5.1. Introduction. Restructured grammars and the intersection between language variation and language change.

Three broad inquiries articulate this study:

(i) How can we describe the Dominican subject expression system in its synchronic state?

It is generally agreed in the theoretical and variationist literature that Caribbean varieties of Spanish are undergoing grammatical restructuring with respect to their subject expression grammars. In theoretical work, traces of grammatical restructuring have been found in the existence of Caribbean-specific constructions, such as *wh*-interrogatives with no subject-verb inversion, personal infinitives with preverbal nominative subjects, or widespread clefted structures (cf. 2.3.4). In variationist investigations, restructuring is found mainly in the different constraint rankings within factor groups like *Person* (cf.4.3.2).

Despite differences in research agendas and methodological approaches, both theoretical and variationist studies converge in locating differences between Caribbean and Non-Caribbean dialects in morpho-syntactic properties: while in theoretical investigations Caribbean patterns of subject licensing have been attributed to the distinct nature of pronominal subjects in terms of feature assembly (Ordóñez and Olarrea 2006, Camacho 2008), in variationist work differences in the meaning conveyed by overt and null subjects, such as the [±specific] distinction, have been deemed crucial for explaining differences in subject use.

Furthermore, theoretical and variationist investigations have also found similar challenges in explaining how far embedded is grammatical restructuring in Caribbean dialects. In this respect, differences in the licensing of pronominal and full DP subjects seem to be very relevant for

disentangling this sort of questions: while the theoretical explanations that focus only on pronominal subjects can accurately account for overt subject pronoun structures (*The Weak Pronoun Hypothesis*, Ordóñez and Olarrea 2006), it has been repeatedly shown that full DPs are also available to speakers in Caribbean-specific constructions (Lizardi 1993, Toribio 2000). Similarly, while most variationist work centers only on pronominal subjects, in the few investigations that take into account both pronominal and DP subjects (Silva-Corvalán 1982, 1984, Lizardi 1993) it has been found that grammatical restructuring is far more advanced in pronominal subjects, while overt, preverbal DP subjects are making their way into the grammar.

In light of theoretical and variationist findings, it is proposed in this dissertation that the explanation of the DS subject system must include an assessment of its morpho-syntactic properties, as well as and accurate description of pronominal versus full DP subject expression. It is further argued that syntactic facts such as the attested word order possibilities for both types of subjects in different Caribbean specific constructions are central for defining the nature and extent of grammatical restructuring.

(ii) Does the Dominican subject expression system reflect a situation of ongoing language change? Is language change reflected in competing grammars in the minds of the speakers?

Previous syntactic work (Toribio 2000, 2001, Martínez Sanz 2007) has attributed the properties of the DS subject grammar to competing grammars in the minds of the speakers. (Kroch 1989, 1994). Kroch's *Competing Grammars Hypothesis* rests on a Government and Binding view of language variation and change in which parameters are conceived as binary, large-scale devices that entail a fixed cluster of properties. In this view the availability in a language of NSL-like null subject constructions along with vernacular overt subject structures can only be taken as reflecting two distinct parameter settings in the speakers' minds. The investigations carried out in the last decade have shown, however, that parameters are best defined as smaller-scale devices with local consequences in the speakers' grammar (Kayne 2005, Holmberg 2009, among others). This shift in

view has allowed for the description of partial, or non-standard null and overt profiles that do not reflect (large-scale) parameter clustering.

Taking this novel view as a point of departure, it is hypothesized in the present study that an explanation of the DS subject grammar does not necessarily require appealing to competing grammars in the speakers' minds. In other words, it is argued that the fact that construction X and construction Y co-exist in a given system, a common scenario in ongoing language change, does not entail the existence of two distinct grammars; rather, what is required is a theory that can account for such an attested (single) grammar.

(iii) How can formal grammars account for cross-dialectal variation?

Investigations on non-standard varieties (Henry 1995, 1997, Toribio 2000, 2001) have shown that cross-dialectal differences may mirror cross-linguistic variation. While this fact could pose a problem for early views on parametrization (Chomsky 1981, 1986), it is resolved in the Minimalist Program (Chomsky 1995, 2001, 2004), in which the primitives of language variation are located in the features that make up Lexical items.

A pervasive problem in generative models is represented by so-called 'optionality'. In a model that is guided by economy principles and that poses strong, categorical restrictions on linguistic structures, the availability of more than one possible derivation to convey the same meaning is not straightforwardly accounted for. However, this sort of semantically-vacuous 'optionality' has been found to be a core property of both developmental and end-state grammars (Liceras 1998, 2007, Sorace 2000, 2005, Fukui 1993, Koopman and van der Wurff 2000). To overcome this challenge, recent minimalist theories propose that there is space for a certain amount of optionality within formal grammars (Biberauer and Richards 2005). In these theories, optionality arises from the different possibilities allowed in a system for the valuation of a given set of features (cf.§1.1).

In the variationist paradigm, so-called 'optionality' does not represent a challenge. In fact, this framework is designed to account for it in a principled way. 'Optionality' is defined as highly

constrained variation, and it is conceived as an inherent property of language, whether ongoing change is taking place at a specific point in time or not; i.e., while language change always implies variation, variation is not to be interpreted necessarily as evidence for language change.

It is argued in this dissertation that the specific type of optionality displayed in DS with respect to null and overt subject insertion can be accounted for within the space provided by formal grammars. It will be proposed that different mechanisms are available in Spanish for Case and EPP-licensing. Consequently, available word order patterns will be derived from the specific ways in which features related to subject licensing are valued.

The methodology for this study, outlined in the following sections, was designed with these three broad research questions in mind: On the one hand, data were collected using standard variationist methodology, as it allows us to "catch a glimpse of linguistic structure" (Meechan and Foley 1995:82). On the other, both theoretical and variationist contributions to our understanding of subject licensing were taken into account when defining the variables that are quantitatively examined in synchronic DS. These variables operationalize the hypotheses on the Dominican subject grammar tested in this study. Theoretical implications will be drawn from quantitative results, as well as from the qualitative evidence on grammatical restructuring that is found in the corpus at hand.

5.2. Research questions and hypotheses.

A number of research questions stem from the broad inquiries depicted in §5.1. Among them are questions probing into the different language-internal variables, morpho-syntactic and otherwise, that regulate subject expression in DS. In turn, investigating the role of language-external questions is also relevant to assess ongoing language change. Questions related to the language-internal and language-external variables involved in subject licensing also provide information on the specific areas of the DS system where grammatical restructuring is to be looked for. Finally, specific questions designed to draw theoretical implications from the quantitative results obtained in this study are also formulated.

5.2.1. Internal factors.

Subject expression has been shown to constitute a variable phenomenon in Spanish. In spite of differences in overall rates of expressed subjects between Caribbean and Non-Caribbean varieties, a core set of variables or factor groups has been shown to regulate subject expression in all dialects: discourse-related variables such as *Co-reference*, *Distance from previous mention*, or *Discourse Connectedness* strongly constrain null and overt subject insertion (Bentivoglio 1987, Silva-Corvalán 1982, 1994, Travis 2005, 2007, among others). Morpho-syntactic variables, such as *Person*, or to a lesser extent, *Verb Ambiguity*, have also been argued to be relevant across dialects (Cameron 1992, Otheguy, Zentella and Livert 2007). Although less robust, lexico-semantic variables such as *Semantic verb type* have also been found significant in subject expression patterns. (Bentivoglio 1987, de Prada 2009).

In light of the findings in previous variationist work and the attested differences between the Dominican and other Spanish dialects in theoretical investigations, relevant research questions in approaching the Dominican pattern of subject use are as follows:

(1) Overt and null subjects in Dominican Spanish: Rates.

Research question 1:

What are the rates of overt subject expression in DS?

Hypothesis:

Overall rates are expected to be higher in DS than in other Caribbean dialects. On the one hand, DS has been hypothesized to be "at the forefront of linguistic innovation" (Toribio 2000:343) with respect to its Caribbean counterparts. On the other, if ongoing language change is confirmed, rates should have increased in the 20-year span that exists between the data set for this study, gathered in 2008, and the comparable data sets from other Caribbean varieties, which were collected in the early 1990s (Cameron 1992, Lizardi 1993, Ávila-Jiménez 1996).

(2) Language-internal variables in Dominican subject expression: significant factor groups and factor group ranking.

Research question 2:

What are the significant language-internal variables in DS? Do they differ from the variables found significant in other studies?

Hypothesis:

Significant factor groups in DS are not expected to depart from the core set of variables found significant in other dialects, as high rates of overt subject use have not been shown to translate into differences in factor groups (Tagliamonte 2006). Therefore, discourse-related variables such as *Co-reference* or *Distance from previous mention* should be ranked at the top of the factor group hierarchy in DS as well. Moreover, the morpho-syntactic variable *Person* is expected to strongly constrain subject expression in DS, as it has been found in previous work that differences in the behavior of specific pronominal elements tear apart Caribbean from Non-Caribbean dialects (Otheguy, Zentella and Livert, 2007).

Research question 3:

What is the relative role of discourse-related variables and morpho-syntactic variables?

Hypothesis:

Both types of variables are expected to strongly constrain subject expression in DS. However, *Person* is expected to override discourse-related variables in DS, similarly to what has been found for other Caribbean dialects (Otheguy, Zentella and Livert 2007, Orozco and Guy 2008). It is argued in this dissertation that the preeminent role of *Person* in Caribbean Spanish stems from the substantial shifts in the semantic and syntactic properties of certain pronouns.

Research question 4:

What is the role of morphological ambiguity? What is the relative role of morphological versus contextual ambiguity?

Hypothesis:

Morphological ambiguity is not expected to override contextual ambiguity in the corpus under study. The former has been measured in previous variationist work with the factor group *Verb Ambiguity* (Cameron 1996, Travis 2005, 2007), which distinguishes morphologically ambiguous and unambiguous forms. Since morphological ambiguity only arises in specific verb tenses, the variable *Tense* also tests for the role of inflectional morphology in subject licensing. While in early studies on Caribbean dialects it was proposed that the loss of inflectional distinctions was at the source of shifts in overt subject rates (*The Functional Compensation Hypothesis*, Hochberg 1986) subsequent investigations have shown that: first, the relative weight of *Verb Ambiguity* does not significantly depart in Caribbean and Non-Caribbean dialects (Cameron 1996); and second, that morphological ambiguity cannot be assessed accurately without taking into account contextual ambiguity (Cameron 1992, 1993, 1996, Paredes Silva 1993).

(3) Language-internal variables in Dominican subject expression: Constraint rankings within *Person*.

Research question 5:

a. What are the constraint rankings within the factor group *Person*?

b. Person and the discursive status of the different persons in the paradigm: Are there differences between first and second persons versus the third?

Hypothesis:

a. Dialectal variation is not necessarily detected on the specific factor groups that are involved in subject expression, but rather in the relative ordering of

these factor groups across varieties, and even more so, in the constraint rankings within factor groups (Tagliamonte 2006). In fact, constraint rankings within *Person* diverge significantly in Caribbean and Non-Caribbean varieties: plural pronominal subjects have been shown to disfavor overt occurrences in both Mainland and Caribbean varieties (Cameron 1992, Travis 2005, 2007 de Prada 2009). As far as singular pronominal subjects are concerned, while in Mainland dialects the first person (Silva-Corvalán 1982,1994, de Prada 2009) favors overt subjects the most, in Caribbean dialects these expressed subjects are most favored with second person singular specific and non-specific subjects. In light of these results, second person singular specific-and non-specific is expected to favor overt subjects the most in this study. Plural subjects are expected to conform to the patterns found in other Spanish dialects in not favoring overt subject insertion (Cameron 1992, Otheguy, Zentella and Livert 2007.

b. Further differences are hypothesized to arise in the singular paradigm between first and second persons versus the third person singular. First and second persons are deictic, whereas the third person is potentially anaphoric. This fact entails differences in reference tracking. For this reason, third person singular pronouns are expected to favor overt subjects less than first and second pronouns.

In other words, while the hypothesis in (3a) that second person singular will favor overt subject use the most in DS is based on the shifts in the referential meaning of pronouns attested in Caribbean Spanish, the hypothesis in (3b) that first and second person will pattern differently from the third stems from the differing discursive properties of pronominals. Further support for this claim is found in the fact that in other restructured varieties, such as

Brazilian Portuguese (cf.§3.3), third person singular has been shown to allow for zero subjects in contexts where first and second expressed pronouns are no longer available (Modesto 2000, 2009).

(4) Language-internal variables in DS: structural properties.

Generally, previous variationist work has focused on defining patterns of null and overt subject use, and similar variables have been found to constrain this phenomenon across dialects (cf.(2) above). The examination of structural properties related for instance with word order patterns displayed with pronominal and DP subjects has been approached mainly in theoretical work. However, when structural properties have been explored quantitatively (Lizardi 1993, Morales 1997), results have shown that relevant differences arise between Caribbean and Non-Caribbean varieties. Among the variables previously examined that operationalize hypotheses on word order differences across varieties are *Subject Position* and *Clause Type*. Morales (1997) and Otheguy, Zentella and Livert (2007) report that pronominal subjects are almost categorically preverbal in Caribbean Spanish. Regarding *Clause Type*, for Non-Caribbean Spanish the basic conclusion reached in previous work is that while main clauses neither favor nor disfavor overt subject insertion, embedded clauses favor null subjects (cf. 4.3.2). Morales (1997) found, however, that Puerto Rican Spanish favors overt subject insertion in relative clauses. Finally, in her study of subject position in *wh*-interrogatives, Lizardi (1993) found very relevant differences among pronominal and DP subjects, and among different pronouns in the paradigm.

The relevance of variables that operationalize claims on the differing syntactic properties of Caribbean and Non-Caribbean Spanish is explored in this study. Specifically, the hypothesis that these variables will arise as significant for subject insertion in DS is put forward. *Subject Position* and *Clause Type* will be included among the set of variables tested in the data set. In addition, I will explore the potential differences that arise in interrogative versus non-interrogative constructions. No explicit comparison of the extent of grammatical restructuring in interrogative and non-interrogative constructions has been carried out in previous work; however, the results in Lizardi

(1993) suggest that grammatical restructuring is in a more advanced stage in interrogatives. I will also investigate whether different (syntactic) verb types favor overt preverbal subjects. Since in Non-Caribbean Spanish specific syntactic verb types, such as unaccusatives, tend to bear postverbal subjects (cf.2.3.4 and Burzio 1986), it is sustained here that it may be worth exploring whether these differences are attested in a restructured variety like DS.

Research question 6:

Are variables related to structural properties such as *Subject Position*, *Clause Type*, *Interrogative* or *Syntactic Verb Type* significant in DS?

Hypothesis:

Variables that operationalize syntactic properties are expected to arise as significant in the corpus examined. Moreover, the hypothesis put forward in this dissertation is that grammatical restructuring will not be as visible in null and overt patterns, as zero and expressed subject insertion is guided in DS by the same set of factors that are relevant across dialects. On the contrary, shifts in structural properties will inform us on the effects of grammatical restructuring.

Research question 7:

What differences emerge between pronominal and DP subjects with respect to *Subject Position*?

Hypothesis:

Significant differences are expected to arise between pronominal and DP subjects. Specifically, overt preverbal pronominal subjects are expected to be strongly favored across constructions, whereas substantial variation is predicted with respect to DP subjects.

Research question 8:

What differences emerge among verb types with respect to *Subject Position*?

Hypothesis:

Overt subjects have been argued to be widespread in Dominican Spanish (Toribio 2000, Camacho 2008). In the data set at hand, transitive and unergative verbs are predicted to strongly favor overt subjects, and unaccusative verbs are not expected to favor postverbal subjects.

Research question 9:

a. What differences emerge between main and embedded clauses?

b. What are the patterns of subject use in finite and non-finite structures?

Hypothesis:

a. I predict for relevant differences to emerge between main and embedded clauses. Specifically, object relative clauses are expected to favor overt preverbal subjects, in line with the results in Morales (1997). The source of this tendency will be argued to be in the development of (vernacular) clefting structures in DS, which involve preverbal subject insertion (cf 1.3.1).

b. The availability of pronominal and DP preverbal subjects in non-finite structures has been pointed out in the previous literature (cf. Sitaridou (2002), Suñer (1982), Lipski (1991) and §2.3.4.5.2). According to these descriptions, they are mainly featured with the preposition *para*, 'for'. These constructions are predicted to arise in the corpus gathered for this study, although they are not expected to be widespread in the speech of Dominicans, as they can alternate with constructions in the subjunctive (cf. *Ven acá* [*para yo verte*] versus *Ven acá* [*para que yo te vea*], 'Come here for me to see you').

Research question 10:

a. What differences emerge between interrogative and non-interrogative constructions?

b. Are preverbal pronominal subjects categorical in any given construction?

Hypothesis:

a. Interrogative constructions are predicted to strongly favor overt preverbal subjects as compared with their non-interrogative counterparts. This hypothesis stems from the findings in previous work with respect to the long-standing availability of non-inverted *wh*-interrogatives in Caribbean Spanish: Navarro-Tomás (1948) documented them in Puerto Rico in the late 1920s, and de Granda (1987) reported their existence in DS in the 19th century already.

b. Pronominal subjects are expected to be almost categorical in *wh*-interrogatives, in line with the results in Lizardi (1993). Furthermore, differences are predicted to emerge among different pronouns of the paradigm: first and second person subjects, which also feature the highest rates of overt use in Caribbean Spanish (Cameron 1992, 1996), are expected to favor preverbal subject insertion the most.

5.2.2. External factors.

Research questions that probe into the language-external variables that may be potentially significant in DS were designed to test the hypothesis that ongoing language change is attested in this variety. In this respect, the variable *Age* is particularly relevant. As detailed in section 5.3, speech samples pertaining to speakers between 18 and 81 years of age were collected for this study with the aim of comparing their subject expression grammars. Other sociodemographic variables

usually examined in variationist work, such as *Gender*, *Social Class* or *Education*, were also tested in this investigation. Furthermore, urban and rural speech is compared.

(5) Language-external variables in DS subject expression: *Age*.

Research question 11:

Are there apparent time differences? Do these differences reflect real time ongoing language change?

Hypothesis:

Significant differences across age groups in rates and variables constraining subject use are not expected to arise in the data set examined, whereas apparent time differences are predicted to emerge in the corpus with respect to the strength of structural variables such as *Subject Position*. This hypothesis stems from the results reported in previous work: while clear *Age* differences have been found in the study of other restructured varieties such as Brazilian Portuguese (Duarte 2000, cf.3.3), this variable is typically returned as a non-significant for subject expression investigations that examine the distribution of null and overt subjects (Bentivoglio 1987, Cameron 1992) in Caribbean dialects. An exception to this trend is Ávila-Jiménez (1996), who found age differences significant in her corpus of Puerto Rican Spanish. A cut-off point of 50 years old was found in her data, with speakers below 50 showing significantly higher rates of overt subject use. *Age* differences have also emerged in the studies that examined structural word order properties rather than null and overt subject insertion. A case in point is Lizardi (1993). In this investigation, 50 years old seems to be the cut-off point that distinguishes age groups as well, and significant differences emerge between younger and older speakers with respect to the preverbal positioning of both pronominal and DP subjects.

(6) Language-external variables in DS subject expression: *Gender*.

Research question 12:

Are there differences between men and women? How can these differences be interpreted for the characterization of ongoing change?

Hypothesis:

Similarly to what is the case with *Age*, *Gender* has not been found as a significant variable for the distribution of null and overt subjects, and it has only been found relevant for the description of structural phenomena such as word order patterns of pronominal and DP subjects. Lizardi (1993) reports that women, along with younger speakers, show the most advanced trends in preverbal subject insertion: they almost categorically favor preverbal insertion of pronominal subjects and they display significantly higher rates of preverbal DP subject insertion. This leads her to suggest that Caribbean ongoing change in subject expression constitutes an instance of 'change from below' (cf. Labov 1991, 1994 and §4.2.3). In the present study, *Gender* differences are only expected to arise in the cross-tabulation of *Subject Position* with pronominal and DP subjects.

(7) Language-external variables in DS subject expression: *Social class* and *Education*.

Research question 13:

Are there differences among social classes and/or educational levels?

Hypothesis:

No significant differences at the level of factor groups and constraint rankings are predicted in this study's data set with respect to *Social Class* and *Education*, in spite of potential differences in overt subject use overall rates. Lizardi (1993) and Ávila-Jiménez (1996), find the variable *Occupation* significant: as discussed in §4.2.3, in the latter the speaker group that included professional and skilled workers featured

higher rates of overt subject expression than the speaker group conformed by unskilled workers; she reports very similar results with respect to *Education*, with more educated speakers producing displaying the highest percentages of expressed subjects. Similar results were returned from Lizardi's analysis of interrogative constructions. Generally, however, *Social Class*, *Education* and related variables such as *Occupation* were not significant in previous analyses.

(8) Language-external variables in DS subject expression: *Urban* and *Rural*.

Research question 14:

a. Are there differences between Urban and Rural speakers?

b. Do Urban and Rural DS conform two distinct speech communities as far as subject expression is concerned?

Hypothesis:

The differences between urban and rural Caribbean speech have only very rarely explored in previous work. To the best of my knowledge, only Ávila-Jiménez (1996) gathered a subset of her data in rural areas of Puerto Rico. She did not find significant differences between urban San Juan and rural speakers. In other contemporary investigations, Caribbean subject expression data sets have been gathered in urban settings (Bentivoglio 1987, Cameron 1992, Flores-Ferrán 2002, 2004, Guy and Orozco, 2008). On the other hand, early descriptions of Caribbean varieties, carried out within traditional dialectology frameworks (Navarro-Tomás 1949, Henríquez Ureña 1940), typically report the speech of rural areas. In recent work, early rural descriptions and contemporary urban corpora are compared in order to address progress in language change. While relevant differences between urban and rural speech are attested in morpho-phonological phenomena, for instance (Bullock and Toribio 2008), no qualitative differences have been reported with respect to patterns of null and overt subject insertion. Taking into account this partial, indirect evidence,

the null hypothesis in this study is that urban and rural speakers share an underlying grammar of subject expression.

5.2.3. Grammatical restructuring and formal grammars.

A number of research questions outlined in this section explore the theoretical implications that can be drawn from the quantitative investigation carried out in the present study. In addition, the qualitative evidence on the availability of specific constructions in DS as reflected in the corpus under study will also be taken into account to explain variable subject expresssion. The theoretical concerns that can be explored, at least to some extent, in this dissertation's data set range from the assessment of methodological approaches to the study of subject expression, to specific issues related with Spanish subject syntax, such as the position of subjects in clausal structure or the feature composition of null and overt pronouns. As outlined in Chapter One (cf.§1.3), it is sustained in this dissertation that whether or not synchronic patterns of null and overt subject use in DS derive from ongoing language change, as these patterns reflect an attested, possible grammar, they should be accounted for within the limits offered by formal theories on language variation.

(9) On the theoretical implications of quantitative evidence.

Research question 15:

Are parametric connections expressed quantitatively?

Hypothesis:

In recent work (Adger and Smith 2003, 2005, Cornips and Corrigan 2005) the analysis of quantitative data is deemed relevant not only as far as it can lend support to structural analysis; rather, the potential of bridging quantitative and formal paradigms is conceived as an instrumental tool for the explanation of syntactic variation. Following this line of research, the hypothesis put forward in the present study is that quantitative analysis may reveal unattested correlations in the DS subject expression grammar, which in turn may contribute to an accurate description

of the specific instantiation of the Null Subject Parameter represented by this variety.

(10) On the role of qualitative evidence.

Research question 16:

What does it entail that specific constructions are either not attested or marginal in other dialects while possible in DS?

Hypothesis:

It is argued in this dissertation that the examination of specific structures, even if they do not arise in high numbers in the corpus, is also relevant for the explanation of the Dominican subject grammar. Recall for instance from the discussion of interrogative structures in §4.2.2.3.1 that clefted questions, even if they are widespread in Caribbean dialects when compared to their Non-Caribbean counterparts, do not arise in spontaneous speech as often as *wh*-interrogatives (in Lizardi's (1993) data set, which comprised 1912 tokens, only 130 clefted questions were reported). However, their existence is relevant for the explanation of Caribbean subject grammars, as they are vernacular constructions that feature the overt preverbal subjects, i.e, the word order pattern associated with grammatical restructuring in these varieties.

(11) On the nature of dialectal variation: The formal definition of the Dominican subject grammar.

Research question 17:

a. What is the specific role of rich agreement in subject licensing in DS? What mechanisms are available for Nominative Case assignment?

b. What is the nature in terms of semantic properties and feature assembly of Dominican overt subjects? Are they distinct from prototypical, NSL Spanish overt subjects?

c. Does widespread use of overt subject affect the discourse-pragmatic functions already conveyed through overt pronominalization in other Spanish dialects? If this is the case, are there innovative strategies in DS for the expression of discourse-related phenomena such as Focus?

Hypothesis:

a. As discussed in § 2.2.2. extensive research has shown that mechanisms other than 'rich' agreement are available cross-linguistically for the licensing of null subjects. Taking this finding as a point of departure, it will be argued, with Toribio (2000) and Sheehan (2006), that Case might be assigned configurationally in DS. Furthermore, it will be proposed, following Sheehan (2006), that subjects in Caribbean and Non-Caribbean Spanish belong in the TP domain, which entails that EPP features are licensed via subject movement to Spec, TP.

b. Following Camacho (2008), it will be proposed that Dominican overt pronominal subjects differ from their Non-Caribbean counterparts with respect to their feature composition. Consequently, the quantitative and qualitative evidence offered in this study lends support to theories that propose that different subject positions are available within TP in the Spanish clausal structure for the landing of subjects (Suñer 2003, Cardinaletti 2004, Camacho 2008). DS subjects are merged lower in the structure, which would explain their limited distribution, as well as the specific properties of Caribbean structures that require subject-verb adjacency.

c. Variationist studies have repeatedly shown that very similar discourse-related variables constrain overt and null subject insertion in both Non-Caribbean and Caribbean dialects. In theoretical work, however, a number of

> Caribbean-specific constructions related with the expression of discourse-related phenomena have been attested. Cases in point are clefted structures, which can convey Focus in Spanish and are widespread in Caribbean dialects. Carrying out a thorough investigation of the Focus structures available to Dominican speakers is not among the main goals of this dissertation. However, from the observation of the clefted structures featured in the corpus at hand it will be suggested that the availability of vernacular Focus constructions in which the preferred SV word order is maintained is in fact related with the process of restructuring of the DS subject grammar.

In order to address the research questions outlined in (1)-(11), spontaneous speech samples from urban and rural Dominicans were collected. Details related to the speakers' background are offered in the next section.

5.3. The speaker sample.

In gathering the data for this study, the goal was to gather a corpus of Dominican Spanish in which men and women of different ages and social profiles were represented. Data were collected in two different locations: the capital city of Santo Domingo and a rural community a few kilometers away from the Haitian border in the province of Dajabón. Over 21 hours of spontaneous speech were recorded in a period of approximately two months (April-May 2008). A total of 50 monolingual Dominican speakers participated in 38 sociolinguistic interviews, from which 34 speakers and 25 interviews were selected for analysis. The final selection of speakers was based on two criteria: the number of available tokens per speaker, and the speakers' background. Each of the speakers selected contributes approximately 200 tokens to the data set, and to the extent that the available data made it possible, homogeneous groupings, in terms of socio-demographic profiles, were constructed. The final corpus comprised 6,800 tokens. After eliminating from the data set the tokens that appeared in contexts in which either null or overt subjects were inserted categorically (cf.§5.4), 6,005 tokens were submitted to statistical analysis.

Table 5.1. details the distribution of speakers according to their socio-demographic profiles[37]:

Table 5.1. The present study: Speakers. (N=34; 6006 tokens)					
Santo Domingo (N=21; 3801 tokens)			**El Cibao** (N=13; 2205 tokens)		
AGE			**AGE**		
18-50 9	50+ 12		18-50 7	50+ 6	
GENDER			**GENDER**		
MEN 12	WOMEN 9		MEN 6	WOMEN 7	
SOCIAL CLASS			**SOCIAL CLASS**		
HIGH 12	LOW 9		HIGH ------	LOW 13	
EDUCATION[38]			**EDUCATION**		
ILLITERATE -------	SOME SCHOOL 10	MORE SCHOOL 11	ILLITERATE 4	SOME SCHOOL 9	MORE SCHOOL -------

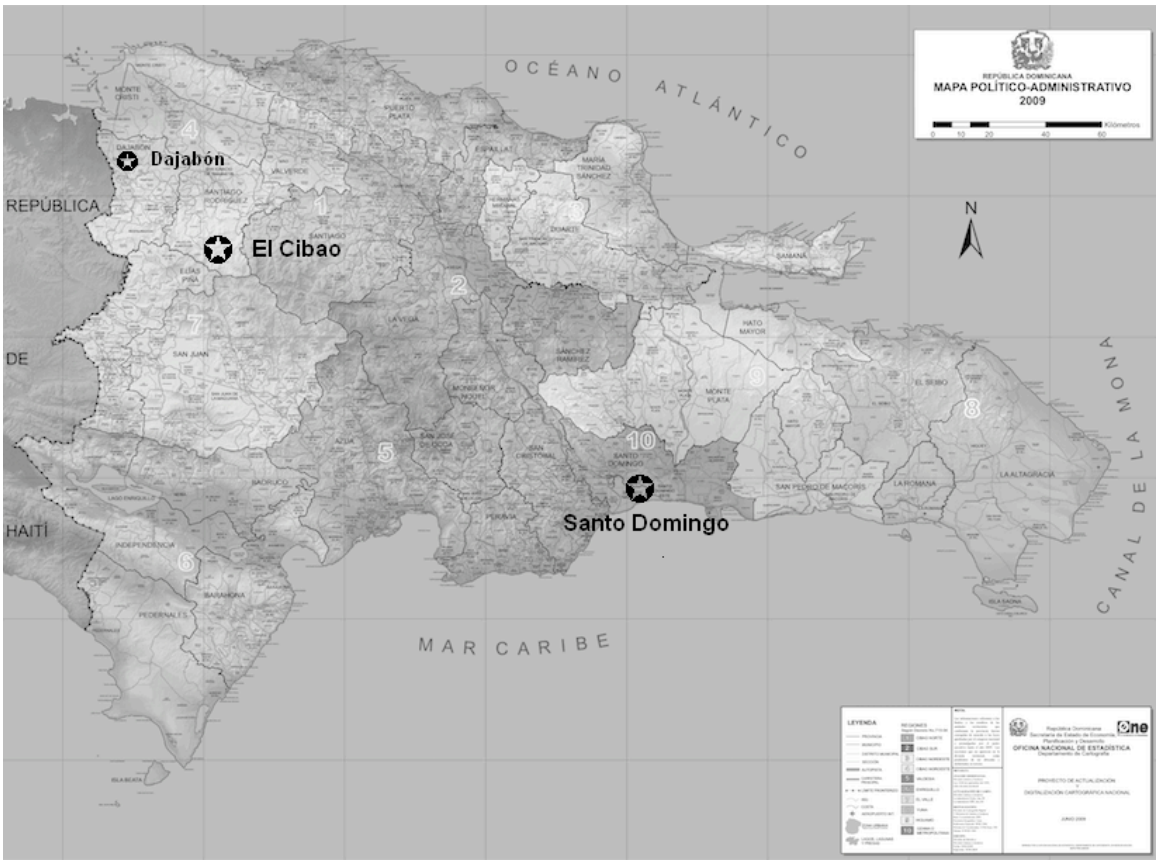
Out of the 34 speakers interviewed for this study, 21 were born and raised in the metropolitan area of Santo Domingo and 13 in the rural area in the province of Dajabón. A higher number of speakers were selected for analysis in Santo Domingo because among them there are significant social class differences, and a higher number of tokens were needed in order to carry out comparative statistical analyses of high/middle class and working class speakers. Rural participants, on the contrary, conform a more homogeneous group as far as social class is concerned. Further asymmetries emerge between urban and rural speakers with respect to education: no illiterate urban speakers were found when collecting the data in Santo Domingo, while access to illiterate participants was granted in Dajabón. Conversely, no university-educated students were found in the rural setting, whereas speakers with a university background were interviewed in Santo Domingo. Bearing in mind the inherent asymmetries in the speech sample collected for this study, the number of speakers in each of the cells in Table 5.1. was chosen in such a way that comparisons across groups were made possible in the quantitative analysis: the number of tokens in the urban and rural

[37] A table that provides these data for each individual speaker is provided in Appendix A.
[38] In this table, 'some school' includes speakers that have a middle/high school background, while speakers with a professional school or university background were ascribed to the 'more education' group.

cells respectively makes it possible to carry out statistical analysis of these two groups separately. Moreover, where comparisons across rural and urban speakers are relevant and numbers of speakers and tokens differ largely, subsets of speakers in the cells with higher numbers can be selected in order to obtain a comparable number of tokens to submit to statistical analysis. By way of an example, if constraints on subject expression were to be compared in the urban and rural populations with respect to the *Age* group 50+ (N=12 in Santo Domingo and N=6 in Dajabón), a subset of 6 Santo Domingo speakers can be compared with the full set of Dajabón speakers.

Data were gathered in an urban and a rural setting with the aim of exploring whether there are differences in subject expression in urban and rural speakers. As pointed out in §5.2.2., with the exception of Ávila-Jiménez (1996), previous investigations on Caribbean dialects have always been based on urban speech. In this study, rural speech is also analyzed in order to provide an accurate description of Dominican patterns of subject use. In Map 5.1, the two provinces were fieldwork for this dissertation was carried out are highlighted in bold:

Map 5.1. The Dominican Republic[39].



The province of Dajabón is located in the northwest of El Cibao, a region that extends West to East from Manzanillo Bay to Samaná Bay and is bound North to South by the Cordillera Septentrional and the Cordillera Central. Agricultural production concentrates in the fertile valleys in Central and Eastern Cibao provinces. Data for this study were collected in the western region, usually referred to as *La Frontera* because of its proximity to the Haitian border. In fact, Haitian presence is visible all along this permeable border. As pointed out by Bullock and Toribio (2008), this area is semi-arid; as a result, most of its inhabitants live in extreme poverty.

In Map 5.2 the rural areas where data for this study were collected are highlighted in bold. All speakers were interviewed in the small settings or 'campos' of El Pino and Los Compos.

[39] Available at: http://www.one.gob.do/index.php?module=articles&func=view&catid=126.

Map 5.2. Rural areas in Dajabón[40].

Santo Domingo is located in the Southeast coast of Hispaniola. The city, with a population of approximately 3,800,000, displays the characteristics of a typical Latin American urban setting, in that it attracts immigration from the rural areas throughout the country and social and economic differences are enormous. Wealth is mainly concentrated in the extended downtown core of Santo Domingo, which is surrounded by a number of impoverished areas that host the working class population. In Map 5.3. the areas where data for this study were collected are highlighted. Interviews with high/middle class speakers were carried out in the neighborhoods of Naco-Ensanche Piantini, La Esperilla, Evaristo Morales and Zona Universitaria. Working class speakers were reached in Villa Consuelo, Sabana Perdida, Gualey, Cristo Rey and Herrera:

[40] Available at http://www.dominicana.cc/dop_maps.html

Map 5.3. Santo Domingo[41].

5.4. Data collection: fieldwork protocol and materials.

Spontaneous speech samples were collected by the author through a series of interviews with the 50 speakers that participated in this study. Speakers were recruited using snowball sampling[42], and were interviewed at their homes or workplaces. Interviews were recorded with a Marantz PMD660 digital recorder, and they vary in length from 25 minutes to an hour and a half. In a number of cases speakers were encouraged to participate in group interviews, with the aim of creating a comfortable environment for the speakers in which vernacular language could be recorded.

All interviews had the structure of a standard sociolinguistic interview. As such, the interview protocol was designed to encourage informal conversation on a range of topics. Interviews consisted in a hierarchically structured set of questions that conformed conversational modules. They usually began with questions related to the community or neighborhood, and

[41] Available at: http://maps.google.es/

[42] An initial set of participants was recruited by the researcher through a social network of Domican speakers she had access to. That existing network of speakers recruited potential participants for the study among their acquaintances.

progressed into conversational modules that allowed for eliciting narratives of personal experience, as this type of discourse has been shown to typically trigger vernacular language use (Tagliamonte 2006). Bearing in mind the characteristics of the target culture, eight conversational modules were designed for this study[43]:

(12) a. Demographics.

b. Neighborhood/Community and Social Practices.

c. Parents and family.

d. Traditions in the Family/Community and Folk Remedies.

e. Travel.

f. Spirituality/Uncommon Experiences.

g. Language/Language and Education.

Within each conversational module, questions were hierarchically organized as well, beginning with exploratory queries, and moving on into more detailed questions if the speaker was willing to talk about a given topic. In designing interview questions, the aim was twofold: on the one hand, to bring up topics of conversation in which the speakers were interested, in order to elicit production; on the other, to obtain relevant socio-demographic data to carry out the subsequent analysis. When possible, relevant information on language attitudes and the perception of Dominican speech by the speakers was gathered as well. This information proved very useful for the observation of the speakers' perception of Dominican speech, and for the exploration of differences in perception versus actual use of specific structures as attested in the corpus. Generally, as observed in previous work (Ávila-Jiménez 1996), the specific Caribbean pattern of null and overt subject use goes unnoticed to the speakers, with the exception of overt expletive constructions, which stigmatize speakers. The participants' awareness of geographical or social differences in Dominican speech, aside from overt expletive constructions, was limited to morpho-phonological patterns, as illustrated in (13)-(14):

[43] A full listing of the guideline questions designed for the interviews is provided in Appendix B.

(13) M: Como la gente habla [en Barahona] depende también del nivel social, y como la mayoría de gente de allí es pobre, o gente de campo, entonces hablan super raro, y hablan mucho con la r, muchas malas palabras y eso, porque es que yo no hablo…yo hablo normal.

O: No, ella habla medio del campo a veces.

M: 'How people speak [in Barahona] depends also on their social level, and because most people there are poor, or *campo* people, they speak weird, and speak a lot with the r, a lot of bad words and stuff like that, because I don't speak…I speak normally.

O: No, she talks like she comes from the *campo* sometimes'.

(Marigabi, 19, UH; Omar, 19, UH)

(14) I: ¿Y hablaba con *ello* también [la muchacha que te cuidaba de chiquita]?

N: Todo el tiempo yo misma digo *ello*, yo digo, cuando estoy conversando con mi mamá, digo "ello hay algo en el ambiente, que no me sé explicar".

I: Mira, ¿y eso tú sólo lo has escuchado en El Cibao, o se escucha en la capital también?

N: ¡Eso no es verdad! ¡Aquí se habla bien malo también!

I: Did she also speak with *ello* [the woman that took care of you when you were little]?

N: Yes, and myself I say *ello*, I say, when I'm chatting with my mom, I say "there (=*ello*) is something in the air that I don't know how to explain".

I: Look, and that you've heard only in the Cibao, or here in Santo Domingo you can hear that as well?

N: That's not true! Here [in Santo Domingo] people speak very badly as well!

(Tatica, 51, UH)

All interviews selected for analysis were transcribed and coded following the procedure outlined in the next section.

5.5. The envelope of variation. Transcribing and coding procedure.

In variationist research, the 'envelope of variation' is defined as the set of contexts that allow for variability in the application of the examined dependent variable. Contexts in which the application of the dependent variable is categorical are left out of the analysis. Therefore, in a study of subject expression, the envelope of variation comprises all the sentences in the data set that feature a null subject in a lexical or syntactic environment where overts are also likely to occur. Conversely, all clauses in which an overt subject is used in a lexical or syntactic environment where nulls may occur are also part of the envelope of variation.

Defining contexts of potential variation in the study of the null and overt subject grammar of Spanish has proved to be a challenging task. According to Cameron (1992), two main sets of problems have been faced in defining the envelope of variation: first, there is no agreement in the literature as to the factors that should be investigated, which has resulted in differences in the nature and number of variables investigated by researchers; second, consensus has not been reached either in the definition of syntactic and pragmatic contexts in which variation is not attested.

A relevant example is provided by contrastive focus environments. In theoretical studies these contexts are claimed to categorically trigger overt subject insertion, as in (15a). A number of variationist investigations incorporate this assumption and exclude these contexts (Otheguy, Zentella and Livert 2007). However, Amaral and Schwenter (2005) stress that in spontaneous speech null subjects are also found in contrastive contexts, as in (15b):

(15) a. Nosotros queríamos que asistiera pero ella no quiso ir.

b. Queríamos que asistiera pero no quiso ir.

'We wanted for her to go but she didn't want to.'

Amaral and Schwenter (2005) further argue that these contexts should be included in subject expression studies because in transcribing and coding naturalistic speech data it is very difficult to

distinguish with a reliable degree of consistency whether a specific context is actually an instance of contrastive focus in which overt subject insertion is mandatory. Following Amaral and Schwenter (2005), contrastive focus environments were included in the envelope of variation for this study.

More generally, in delimiting the envelope of variation for the quantitative analysis, my goal was on the one hand, to include in the study the core set of variables that are generally agreed to regulate subject expression in Spanish, and on the other, to also examine a number of factors that have been found relevant in theoretical investigations but usually overlooked in variationist work.

As illustrated in examples (16) and (17), even if overt subjects are widespread in DS, as in other Spanish dialects, null and overt subject insertion is a variable phenomenon:

(16) *Yo* tengo mi hermana que vive allá, ella estaba aquí, *ø* se fue hoy, si *ø* me encuentra un trabajito, *yo* me voy, *yo* no quiero volver a la zona franca porque *yo* no sé coser en máquina, entonces *yo* trabajaba en empaque, dobladora, entonces *yo* le tengo miedo a la máquina, entonces si *uno* le tiene miedo, *uno* puede cortarse fácilmente, pero *yo* quiero trabajar para *ø* ayudar a mi familia, este tamaño me lo dieron *ellos* sacrificándose su vida.

'I have my sister who lives there, she was here, left today, if she finds me a little job I will leave, I don't want to go back to the factory because I don't know how to sew with the machine, so I worked in the packing department, I was folding, and I am scared of the machines, and if you are scared, you get hurt easily, but I want to work for my family, the size I have today they gave it to me working hard themselves.'

(Ramona 25, RW).

(17) Lo que pasa que *yo* empecé estudiando administración de empresas, *yo* tenía dos semestres y un verano. *ø* Dije *ø* me voy a cambiar de universidad y de carrera, y ese fue un año a pique, porque cuando *ø* entré a la otra universidad, no me convalidaban los créditos, porque *ø* tenía que tener más créditos para que *ø* me convalidaran,

entonces *ø* tuve que arrancar de cero otra vez, todas las materias que *yo* cogí en la primera universidad, *ø* las tuve que coger en la segunda.

'What happens is that I started studying business management, I had two semesters and a summer. I said I'm going to change schools and programs, and that was a lost year, because I needed more credits for them to transfer me, and then I had to start over from scratch, all the courses I had taken in the first university I had to take them again in the second one.'

(José Gabriel, 34, UH)

In (16) and (17), overt pronominal subjects alternate with null pronouns for instance in contexts of Topic continuation, following the general Spanish pattern (*ella estaba aquí, ø se fue hoy*, in (16), and *todas las materias que yo tuve que coger en la primera universidad, ø las tuve que coger en la segunda*, in (17)). However, overt pronominal subjects may also be inserted in these same contexts (*yo me voy, yo no quiero volver a la zona franca porque yo no sé coser en máquina*, in (16), and, in (17), *yo empecé estudiando administración de empresas, yo tenía dos semestres y un verano*).

Overt pronominal subjects alternate not only with zero subjects, but also with DPs:

(18) Y *Máximo Gómez*, el famoso libertador de Cuba, que es dominicano, *él* se fue con los españoles, porque *él* era santanista.

'And Máximo Gómez, the famous liberator of Cuba, who was Dominican by the way, he took the Spanish side, because he was with Santana.'

For this reason, while the general trend in variationist studies is to only examine pronominal subjects, in this study DP subjects are also included in the envelope of variation, following the procedure in Silva-Corvalán (1982, 1994) and de Prada (2009). Examining DP subjects is particularly relevant in a study on the Dominican dialect, since relevant differences in patterns have been found for pronominal and lexical subjects in previous work (Lizardi 1993, Ordóñez and Olarrea 2006). Therefore, the dependent variable for this study has three variants:

(19) Dependent variable:

a. Null subject

b. Overt pronoun

c. Overt lexical subject.

In delimiting the variable contexts in which independent factor groups may have an impact in subject form choice, a number of environments were excluded from quantitative analysis. Cases in point are the contexts that are categorically null, such as impersonal 'se' constructions:

(20) Había una mata de naranja dulce, que esa era la mata preferida de mi abuelo, porque entonces las naranjas eran dulces, dulces; *pero no se podían coger*, porque esa era para hacerle jugo al recién nacido de turno.

'There was a sweet orange tree, which was my grandpa's favorite tree, because the oranges were sweet, sweet; but those oranges you could not pick, because that tree was for making juice for the newborn'.

(Néstor, 56, UH)

Conversely, contexts that categorically bear an overt subject in speech, such as the verbs *saber*, 'to know', or *entender*, 'to understand' in set phrases, are also excluded[44]:

(21) Por ejemplo yo viviría pendiente de mis hijos, si se suben en el cogollito de una mata, ¿*tú entiendes*?

'For example I would live always watching my kids, if they climb to the top of a tree or whatever, you understand?'

(Danilo, 52, UH)

The contexts illustrated in (20)-(21) are usually excluded in previous work as well. Nonetheless, specific contexts that are excluded from investigations of Non-Caribbean varieties have been proved to be a part of the envelope of variation in Caribbean Spanish, and therefore are

[44] Set phrases are included in the envelope of variation and coded separately in other studies (Otheguy, Zentella and Livert 2007).

considered in this study. This is the case of second person singular non-specific pronouns, which were found to favor overt subject insertion in Cameron (1992, 1996):

(22) Creo que el que fue a la universidad se nota que fue a la universidad y el que no fue se nota que no fue. Algo *tú aprendes*, no sé cuál es la palabra, algo que *tú aprendes* que es por *tú haber ido* a la universidad, y si *tú no vas*, a lo mejor *tú no lo aprendes*.

'I think that the person that went to university, you can tell that he went, and who didn't went you can tell he didn't went. There is something you learn, I don't know what the word is, something you learn because you went to university, and if you don't go, maybe you won't learn it.'

(José Gabriel, 34, UH)

Third person plural subjects can also convey a non-specific, generic meaning in Spanish. As they are categorically null in most dialects, they have been excluded in previous work. However, based on the fact that overt third person plural non-specific pronominal subjects have been documented in other Caribbean varieties (Ávila-Jiménez 1996, Otheguy and Lapidus 2005) and they are attested in the data set for this study, they are included in the quantitative analysis. Example (23) illustrates the variability attested in the corpus:

(23) a. S: Yo pertenezco al PRD, soy secretario, y estamos trabajando para llevar a Miguel Vargas a la presidencia.

I: Sí, yo vi en el periódico que va ganando.

S: Es lo que esperamos. Siempre *ellos* dicen…*ellos* publican encuestas. La encuesta que yo siempre he creído es el tú a tú.

S: 'I belong to the PRD, I am secretary, and we are working towards making Miguel Vargas the president.

I: Yes, I saw in the newspaper that he is winning.

S: That's what we're hoping for. They always say…they publish surveys. The survey I have always believed in is word of mouth'.

(Melvyn, 31, UW)

b. Y hace como diez o quince años que renovaron la educación, y ahí fue que la pusieron mala de verdad.

'And ten or fifteen years ago they renovated the education system, and it was at that time that they made it really bad.'

(Tatica, 51, UH)

There are other contexts that systematically bear a null subject in other dialects and may display an expressed pronoun in DS. Cases in point are impersonal constructions with the expletive *ello* and subject relative clauses with overt pronouns (cf. §1.3). These constructions are attested in the data set gathered for this study, as illustrated in (24) and (25). However, the small number of these tokens did not allow including them in the statistical analysis.

(24) Antes yo lo hacía, cuando era muchacha, ya no; ahora ya *ello* hay otros jóvenes que lo hacen.

'Long time ago I used to do it, when I was a young girl, not now; now there are other young people who do it.'

(Ramona, 25, RW)

(25) Entonces la mala es una muchacha que *ella* estuvo en República Dominicana.

'And the bad one [in the soap opera] was a girl that she was visiting the Dominican Republic…'

(Ambiori, 28, RW).

Finally, non-finite clauses are excluded from the envelope of variation in the previous literature. However, they are considered in the present investigation as non-finite clauses may carry

an expressed subject in DS (cf. §1.3. and the discourse string *algo que tú aprendes que es por tú haber ido a la universidad* in (22) above).

All tokens selected for statistical analysis were transcribed and coded in Excel for the 18 independent variables examined in this study, and subsequently submitted to statistical analysis with Goldvarb X. As far as language-external variables are concerned, all speakers were assigned a pseudonym and a number, and the rest of language external factor groups were determined by their socio-demographic profiles:

(26) Coding of language external variables

a. Speaker

b. Speakers's age

c. Speaker's gender

d. Speaker's geographical origin

e. Speaker's social class

f. Speaker's education

As outlined above, language-internal factor groups selected for this study are informed by previous results in the variationist and theoretical literature on subject expression in Spanish. A total of 12 factor groups with their associated constraints were designed. These factor groups operationalize the hypotheses put forward in this study with respect to the discourse-related, morpho-syntactic, lexico-semantic and structural constraints that regulate subject expression.

Taking as a point of departure that within discourse-related constraints the basic finding in the variationist literature is that while continuity (in reference, TMA) favors null subjects and non-continuity disfavors them (cf.§4.1), the following discourse-related factor groups were considered:

(27) Discourse-related variables: Continuity.

a. Discourse function (Topic).

b. Co-reference.

c. Distance from previous mention.

d. TMA continuity.

As in previous work, *Discourse function* classifies subjects according to whether they have been mentioned or not in the previous discourse. *Co-reference* measures whether preceding and target subject share the same referent. *Distance from previous mention* classifies subjects in New, mentioned within ø, 1, 2, 3, 4 or more than 4 clauses. *TMA continuity* distinguishes whether the tense, aspect and mood of the verb accompanying the target subject are the same in the preceding and target clauses.

Morpho-syntactic and lexico-semantic factor groups examined in this study are as follows:

(28) Morphosyntactic and lexico-semantic variables:

a. Person of the subject.

b. Verb.

c. TMA.

d. Animacy.

Person of the subject was designed to distinguish different persons of the paradigm, as well as specific and non-specific occurrences of second person singular and third person plural subjects. The pronominal element *uno* was also coded separately, as it has been pointed out in the previous literature that its overt, preverbal use is extended in DS (cf.1.3).

As far as *Verb* is concerned, recall that *Semantic Verb Type* has been found significant in previous work (Bentivoglio 1987, Silva-Corvalán 1994). *Syntactic verb type*, in the few investigations that have examined it, has not arisen as a particularly significant variable. Most variationist studies deal only with null and overt subject insertion, and do not consider word order

patterns. However, in theoretical investigations the lexico-syntactic nature of predicates is regarded as relevant for specific word order patterns. In this study, in order to test for the significance of semantic and syntactic properties of verbs, predicates were coded for independent lexical items: most frequent verbs were assigned an independent code; the same procedure was followed with reflexive uses, found significant in previous literature (Otheguy, Zentella and Livert 2007), and psychological predicates. The latter were coded separately since they have been shown to have distinct properties in Spanish (Belletti and Rizzi 1988). This coding system allows us to group verbs according to semantic or syntactic properties when they are submitted to statistical analysis with Goldvarb X.

Predicates were also coded according to their TMA. This coding decision stems from two basic findings in previous variationist literature: morphological ambiguity with respect to person only arises with specific verb tenses in Spanish (cf.§4.2.3), so I coded for this factor group, and differences in the discourse properties of verb tenses have also been found relevant (Silva-Corvalán 1997). Coding verbs according to their TMA will make it possible to examine the behavior of specific verb tenses with respect to their morphological and disourse-related properties.

Animacy classifies subjects as animate and inanimate. The significance of this variable is generally not tested in previous work, because null inanimate subjects do not alternate with a pronominal form in Spanish. However, as noted by de Prada (2009), they do alternate with overt lexical forms. Since DP subjects are examined in this dissertation, *Animacy* was considered in the quantitative analysis.

Structural variables related to word order phenomena and the properties of specific syntactic constructions are also examined. They are as follows:

(29) Structural variables:

a. Subject position

b. Finiteness

c. Clause Type

d. Interrogative

Subject Position is included in this study to explore the word order differences that emerge between pronominal and lexical subjects, and among the different pronouns in the paradigm. *Finiteness* distinguishes finite and non-finite clauses, and it is examined in order to accurately describe to what extent overt preverbal subjects are attested with infinitives when compared with finite verb forms. *Clause Type* is designed to test for differences among main, conjoined, embedded and relative clauses. *Interrogative* is included to determine whether grammatical restructuring is more evident in interrogative clauses, which have been claimed to reflect a more advanced stage of change regarding word order phenomena (Lizardi 1993).

In sum, each token was coded for 12 internal variables and 6 external variables. Table 5.2. illustrates all factor groups examined in the statistical analysis and their associated constraints:

Table 5.2. The present study: selected factor groups and associated constraints.		
Dependent variable	***Constraints (factors)***	
SUBJECT FORM	*Null* *Pronominal subject* *DP subject*	
Independent variables		
DISCOURSE-RELATED VARIABLES		
CO-REFERENCE	*Co-referent* *Non-co-referent*	
TOPIC CONTINUATION	*Topic* *Topic shift* *Emphasis*	
DISTANCE FROM PREVIOUS MENTION	*ø 2 4* *1 3 more than 4*	
TMA CONTINUITY	*Same TMA* *different TMA*	
MORPHO-SYNTACTIC AND LEXICO-SEMANTIC VARIABLES		
PERSON OF THE SUBJECT	*1st p sg (yo)* *2ndpsg+spec(tú,usted)* *2ndpsg–spec(tú,usted)* *3rd psg (él, ella)* *3rd psg (uno)*	*1st ppl (nosotros/nosotras)* *2nd ppl (ustedes)* *3rd ppl +specific (ellos/ellas)* *3rd ppl –specific (ellos)*
VERB	*Frequent verb* *Reflexive* *Other*	
TMA	*Present* *Preterite* *Imperfect* *Pluscuamperfect* *Present Perfect*	*Future (synthetic)* *Future (analytic)* *Future perfect* *Conditional*
ANIMACY	*Animate* *Inanimate*	
STRUCTURAL VARIABLES		
FINITENESS	*Finite* *Non-finite*	
SUBJECT POSITION	*Preverbal* *Postverbal*	
CLAUSE TYPE	*Main* *Embedded* *Relative clause*	
INTERROGATIVE	*Interrogative* *Other*	

For instance, a token like the one underlined in (30), taken from example (16) above, would be coded as *pronominal subject*, *co-referent*, *topic*, *ø, different TMA*, *first person singular*, *trabajar*, *imperfect*, *animate*, *finite*, *preverbal*, *main* and *non-interrogative.*

(30) Sample token

Yo tengo mi hermana que vive allá, ella estaba aquí, *ø* se fue hoy, si *ø* me encuentra un trabajito, *yo* me voy, *yo* no quiero volver a la zona franca porque *yo* no sé coser en máquina, entonces *yo* trabajaba en empaque, dobladora (…)

'I have my sister who leaves there, she was here, left today, if she finds me a little job I will leave, I don't want to go back to the factory because I don't know how to sew with the machine, so I worked in the packing department, I was folding.'

(Ramona 25, RW).

The results returned from statistical analysis with Goldvarb X are provided in Chapter Six. Theoretical implications of the quantitative results and qualitative evidence found in the data set are discussed in Chapter Seven.

5.6. Summary

In this chapter I have outlined the research questions that guide this dissertation and the methodology employed to address the issues raised by these questions. Broad inquiries that probe into the description of the Dominican subject expression grammar have been outlined, and specific questions related to null and overt subject patterns, word order phenomena and the theoretical implications of the quantitative study have been formulated. In discussing the methodology for the study, I have described the speaker sample, as well as the transcribing and coding procedure applied to the data set at hand. Selected contexts of variability and factor groups and corresponding constraints have also been laid out. Results from the quantitative analysis are offered in the ensuing chapter.

Chapter Six. Results: The configuration of a restructured subject expression grammar. Null and overt subject distribution and subject position patterns in Dominican Spanish.

6.1. Introduction. Null and overt subject distribution and subject position patterns.

The 6005 tokens that comprise the final data set analyzed in this study were submitted to statistical analysis with the variable rule program Goldvarb X. This program calculates the probabilities of application of the dependent variable in all the contexts coded in the independent variables. As outlined in Chapter Four, statistical analysis with Goldvarb returns two types of measures, in addition to percentages of occurrence and overall significance of the analysis: first, it selects which factor groups are involved in the application of the dependent variable and it provides a ranking of these factor groups; second, it measures the relative strength of each of the constraints within factor groups.

The goal in this study is to examine null and overt subject distribution, as well as to explore subject position patterns. Thus, two sets of analyses were carried out: in the first set of calculations, the dependent variable comprises the factors null, overt pronominal and overt lexical subject, and in the second, it comprises preverbal and postverbal subjects. In each set of analyses, the distribution of pronominal and lexical subjects is compared. Dependent variables in Goldvarb are analyzed as binary (for instance, application and non-application of the null subject value). Thus, for the examination of null and overt subject patterns, null subjects are first compared with pronominal and lexical subjects, and subsequently only pronominal and only lexical subjects are compared with their null counterparts. The results of statistical analysis are presented in the following sections.

This chapter is organized as follows: results for the distribution of lexical and pronominal versus null subjects are presented in section 6.2. Logistic regression analyses were performed to examine the strength of discourse-related, morphosyntactic and structural factor groups. The same logistic regression analyses were carried out counting only pronominal subjects as the non-

application value. The comparison of the results obtained with each of these data sets will allow us to explore the differences that emerge among pronominal and lexical subjects. Section 6.3 focuses on the distribution of lexical and null subjects. The comparison of the findings in section 6.3. with the ones presented in sections 6.2. and section 6.1. will make clear that the distribution of lexical subjects is derived from the discursive properties associated with lexical subjects. The results of analyzing subject position patterns are discussed in section 6.5. Section 6.6. contains a summary.

6.2. Null and overt subject patterns: Overt pronominal and lexical subjects versus null subjects.

When pronominal and lexical subjects are both included in the analysis, the overall rate of overt subject production is 61%. This rate is between 10 and 20 percentage points higher than the rates found in previous work on other Caribbean varieties (Bentivoglio (1987) reported a rate of 40% of overt subjects in the Spanish of Caracas, Venezuela, Cameron (1992) obtained a rate of 45% in Puerto Rican Spanish, and Travis (2007) 48% in Coastal Colombia. Only Cameron (1996) reports a rate close to 60% of overt subject production in his study of Puerto Rican second person subjects. However, Bentivoglio, Travis and Cameron include only pronominal subjects in the envelope of variation. More accurate comparisons with rates of null and overt subject use will be drawn in 6.3, where results of the analysis of pronominal versus null subjects are presented.

Even if lexical and pronominal subjects jointly are compared with null subjects in this section, the distribution of overt subjects in this data set reflects the variation among the different persons of the paradigm found in the previous literature (Cameron 1992, Otheguy et al. 2007). Table 6.1. offers overall rates of null and overt subjects for the nine person distinctions coded in the corpus gathered for this study (cf. 4.3):

Table 6.1. Dominican Spanish. Overall percentages of null and overt subject use: persons of the paradigm.

	Null		*Overt*	
	N	*%*	*N*	*%*
Singular				
1ST PERSON YO	848	39.7	1286	60.3
2ND PERSON [+SPECIFIC] *tú*	30	10.9	244	89.1
2ND PERSON [-SPECIFIC] *tú*	54	21.8	194	78.2
3RD PERSON [+SPECIFIC] *él/ella*	537	28.8	1326	71.2
3RD PERSON [-SPECIFIC] *uno*	37	23.7	119	76.3
Plural				
1ST PERSON PLURAL: *nosotros/nosotras*	354	73.8	126	26.2
2ND PERSON PLURAL: *ustedes*	15	37.5	25	62.5
3RD PERSON PLURAL [+SPECIFIC]: *ellos/ellas*	295	49.1	306	50.9
3RD PERSON PLURAL [-SPECIFIC]: *ellos*	178	87.3	26	12.7
Total	2348	39.1	3652	60.9

As observed in Table 6.1, singular and plural persons differ largely in overall rates of expression. However, leaving aside third person plural [-specific] occurrences, the singular and plural paradigms follow the same pattern: the first person displays the lowest percentage of overt expression, at rates of 60.3 and 26.2 respectively, while the highest percentages are found in the second person (89.1 for [+specific] second person singular and 78.2 for [-specific] occurrences, and 62.5 for second person plural); the third person is ranked after second person subjects in both paradigms as well (71.2 for the singular and 50.9 for the plural). Non-specific instances of third person plural clearly depart from the rest of the singular and plural paradigms, only reaching an overt rate of 12.7%. In spite of this very low rate of overt production, the fact that generic third persons are available with an overt subject in Dominican Spanish constitutes in itself a phenomenon worth noting, since generic subjects are categorically null in other European and Latin American Spanish varieties. As discussed in §4.3.2, these overt subjects have also been documented in Puerto Rican Spanish (Ávila-Jiménez 1996) and in Spanish in contact with English in the United States (Lapidus and Otheguy 2005).

After obtaining raw percentages of null and overt subject expression, data were submitted to logistic regression analysis with Goldvarb. The variable rule program calculates the relative

relevance of the independent variables included in the analysis for the application of one of the values associated with the dependent variable, in this case null or overt subject. In other words, while all the factor groups included in the present analysis have been shown to have a role in subject expression in previous studies, a specific set among them will arise as the most significant when all variables are analyzed jointly.

As is usually the case in variationist work when the relative significance of several factor groups is calculated, a number of these factor groups had to be recoded in order to carry out accurate logistic regression analysis (cf. Tagliamonte 2006:170). The coding schema designed for the present study included factor groups that are relevant only for the analysis of certain subsets of the data; for instance, *Subject Position* is relevant for the comparison of pronominal and lexical subjects, while not relevant for the analyses in which null and overt subject patterns are examined. Futhermore, the more factor groups and the more factors per factor group are included in the analysis, the more likely it is that a given factor does not yield enough tokens for quantitative analysis. This uneven configuration of the data is also the result of gathering spontaneous speech samples, in which some linguistic structures are extremely frequent while others tend to appear in very small numbers. Two procedures are available when a given factor does not yield enough tokens for logistic regression analysis: either that factor is left out of the calculations, or it is merged with another factor that exhibits similar distribution and can be argued on theoretical grounds to be related. Both procedures were applied to the data set used in this dissertation. The re-coding of relevant factor groups is outlined in the following paragraphs.

The data were coded for four different discourse-related factor groups associated with the notion of continuity: *Discourse Function*, *Co-reference*, *TMA of the previous verb* and *Distance from previous mention*. The latter included the factors *0 clauses*, *1 clause*, *2 clauses*, *3 clauses*, *4 clauses*, *more than 4 clauses*, and *new referent*. The distribution of null and overt subjects in each of these categories is illustrated in Table 6.2:

Table 6.2. Dominican Spanish. *Distance from previous mention*: Overall rates of null and overt subjects (N=6005).

	Null		***Overt***	
DISTANCE FROM PREVIOUS MENTION	***N***	***%***	***N***	***%***
0 clauses	1501	52.6	1350	47.4
1 clause	232	32.9	473	67.1
2 clauses	109	33	221	67
3 clauses	54	25.7	156	74.3
4 clauses	38	29.7	90	70.3
More than 4 clauses	165	22.9	557	77.1
New referent	249	23.6	805	76.4

The percentages in Table 6.2 show that roughly, as distance increases, the rate of overt subject expression is higher. The only case in which this parallel increase is not attested is in the rates obtained for *3 clauses* and *4 clauses* of distance. Given the small number of tokens[45] in these two categories, and that *more than 4 clauses* and *new referent* display very similar rates, the factor group *Distance from previous mention* was re-coded into a factor group with four associated factors: *0 clauses, 1 clause, 2 clauses,* and *3 or more clauses*.

The factor group *Discourse function* was also recoded to carry out multivariate analyses. Initial factors designed for this variable were *Topic*, *Shift* and *Emphasis*. However, in the data set, emphatic contexts were very few (N=7) and categorically overt. For this reason, *Emphatic* and *Topic Shift* subjects were merged and *Discourse Function* was transformed into a binary variable with the associated factors *Topic* and *Shift*. *Co-reference* and *TMA of the previous verb* did not require re-coding.

As far as morpho-syntactic factor groups are concerned, the coding schema outlined in Chapter Five included *Person*, *TMA*, *Verb*, and *Animacy*. *Person* was re-coded as follows: the initial nine-factor classification depicted in Table 6.1. was reduced to a six-factor classification: first

[45] Having a small number of tokens for a given factor leads to the *interaction* of its associated factor group with other factor groups. For instance, consider the factor first person within *Person* and the factor *3 clauses* within *Distance from previous mention*. An interaction would occur if no first person subjects at a distance of 3 clauses from its previous mention were attested in the corpus. While Goldvarb is able to develop logistic regression analyses even if interactions are detected, the overall accuracy of the analysis is compromised. For this reason, the usual practice is to eliminate all interactions from the data set before submitting the data to logistic regression analysis (Tagliamonte 2006).

person, second person [+specific], second person [-specific] in the singular, and first and third person in the plural. Second person plural subjects, like in Cameron's (1992) and Otheguy et al's (2007) work, were not submitted to logistic regression analyses due to the small number of occurrences (N=40); third person plural non-specific subjects (N=204) were also left out of the analysis due to their small rate of overt occurrence (12.7%). Furthermore, non-specific second and third person subjects were conflated in a single factor, given the similar rates of overt production they display (78.2% and 76.3 respectively).

Animacy, with the associated factors of animate and inanimate, is not considered in the logistic regression analyses that examine the distribution of pronominal and lexical subjects, as null inanimate subjects only alternate with lexical counterparts.

TMA was classified in the categories illustrated in Table 6.3:

Table 6.3. Dominican Spanish. *TMA* : Overall rates of null and overt subjects. (N=6005)

	Null		***Overt***	
	N	***%***	***N***	***%***
Indicative				
Present	812	33	1647	67
Present perfect	54	34.6	102	65.4
Preterite	706	47	795	53
Imperfect	413	33.3	828	66.7
Past perfect	8	20.5	31	79.5
Conditional	2	22.2	7	77.8
Future analytic	48	37.2	81	62.8
Future synchretic	3	37.5	5	62.5
Subjunctive				
Present	48	44	61	56
Imperfect	43	51.2	41	48.8
Past perfect[46]	2	25	6	75
Infinitive	209	81.3	48	18.7

As observed in Table 6.3, infinitives clearly depart from the rest of the verbal paradigm in that they only bear overt subjects at a rate of 18.7%. Their overall occurrence in the corpus is also small (N=257). For this reason, infinitives are not considered in logistic regression analyses, and the relevance of these constructions for the configuration of the Dominican grammar of subject

[46] No occurrences of present perfect subjunctive were attested in the data set.

expression will be discussed in Chapter Seven, where the theoretical implications of the DS system of null and overt subjects for formal theories of dialectal variation are approached.

Finite verb tenses were re-coded for logistic regression analysis, with two goals in mind: first, to examine whether morphologically ambiguous tenses favor overt subjects, as suggested in previous work (Travis 2005, 2007); second, to explore whether the discursive role of specific verb tenses is also involved in null or overt subject choice (Silva-Corvalán 2001). A first look at the percentages in Table 6.3. suggests that while morphological ambiguity might be involved in subject selection, other constraints are relevant as well: the preterite, a non-ambiguous verb tense, displays the lowest rate of overt subjects among the finite indicative tenses; past perfect indicative, conditional past perfect subjunctive, all three-way ambiguous verb tenses[47], display the highest rates of overt subjects, 79.5% 77.8% and 75% respectively. However, the imperfect indicative and the imperfect subjunctive are also three-way ambiguous verb tenses and their rates of overt subject expression only reach 66.7% and 48.8% respectively. Furthermore, both present and present perfect indicative, and the synchretic and analytic futures are two-way ambiguous verb tenses and their rates of overt subject expression are very close to the rate displayed by the imperfect indicative (67%, 65.4% for the former, and 62.8, 62.5 for the latter). Leaving aside past perfect (indicative and subjunctive) and conditional, which occur very infrequently in the corpus (N=8, N=2, N=2, respectively), the pattern that emerges from Table 6.3 is that generally tenses in the indicative trigger overt subject use in higher rates than tenses in the subjunctive; consider rates of imperfect indicative (66.7%) and imperfect subjunctive (48.8%) and the rates of present indicative (67%) and present subjunctive (56%). Subjunctive tenses occur only in embedded clauses, and these type of clauses have been found to favor null subjects in previous work (Silva-Corvalán 1994, Morales 1997), which suggests that *Clause Type* operates jointly with morphological ambiguity. Also, note

[47] Recall that in three-way ambiguous verb tenses, first, second and third person singular become homophonous in DS due to final –s deletion (cf. past perfect '(yo) había comido', '(tú) habías comido', '(él/ella) había comido', *I/you/he/she had eaten*), while in two-way ambiguous verb tenses only second and third person become homophonous (cf. present '(yo) como', '(tú) comes', '(él/ella) come', *I/you eat, he/she eats*).

that reference tracking in embedded clauses is straightforward when their subject is co-referent with matrix clause subject. Taking as a point of departure the similarities and differences in overall rates, TMA was re-coded as follows: tenses with very few tokens (past perfect indicative and subjunctive and conditional) were left out of the analysis; preterite and imperfect tenses constituted two separate factors, representing a non-ambiguous and three-way ambiguous tenses respectively; present, present perfect and synchretic and analytic futures, which displayed similar percentages and are all two-way ambiguous, were merged into a single factor; finally, subjunctive tenses were conflated into a single factor. This re-coding procedure yields the four-code classification in (1):

(1) *TMA* re-coding:

a. Imperfect.

b. Preterite.

c. Present (includes present, present perfect, future analytic, future syncretic).

d. Subjunctive (includes present subjunctive and imperfect subjunctive).

Subjunctive verb tenses had to be excluded from the logistic regression analyses that examined all coded factor groups jointly, due to interactions. The effects of subjunctive tenses on subject choice were explored, however, in the analyses that examine the relative role of morphological and contextual ambiguity (see §6.2.3. below). Note that this classification will also allow me to test whether the discursive role of specific verb tenses is relevant for null and overt subject choice, as relevant differences have been found by Silva-Corvalán (2001) between the imperfect, which tends to favor overt subjects and the preterite, which favors null subjects.

The factor group *Verb* was also re-coded for logistic regression analyses. The coding schema outlined in Chapter Five included a separate code for all the verbs that featured more than 50 occurrences in the corpus, and 24 different codes were obtained. When percentages of null and overt subjects were calculated, natural classes arose in the corpus. These classes correlated with both semantic and syntactic properties: stative verbs behaved very similarly, with the highest rates of overt expression (‘ser’, *to be*, 77.4%, N=603; ‘ver’, *to see*, 78.2%, N=142, ‘saber’, *to know*,

81.9%, N=171). Note that according to a syntactic classification, 'ser' is copulative while 'ver' and 'saber' are transitive; semantically, all three verbs are stative, which suggests that lexico-semantic constraints override syntactic constraints in this case, as found in previous work (Bentivoglio 1987, Travis 2007, de Prada 2009). However, the rest of the verbs in the corpus seemed to pattern according to syntactic constraints: inergative and unaccusative verbs show very similar percentages of overt use ('vivir', *to live* (N=93), 67.7%, *to talk* (N=56), 64.3%, 'llegar', *to arrive*, 65.3%). Transitive verbs display slightly lower rates of overt subjects ('tener', *to have* (N=375), 63.5%, 'hacer', *to do* (N=105), 54.4%). Interestingly, the lowest rates of overt subjects correspond to ditransitive verbs ('dar', *to give* (N=77), 46.8%, 'preguntar', *to ask*, (N=166), 35.3%). These percentages seem to show that the more argumental structure a verb bears, the less it favors overt subject use. An exception to this trend is the ditransitive 'decir', *to say* (N=377), with a rate of overts of 60.7%. This difference with other ditransitives is attributed to the fact that 'decir' is a speech act verb that can also be used by speakers to express their opinion; these type of verbs in previous work have been found to favor overt subjects (Enríquez 1984, Bentivoglio 1987, Silva-Corvalán 1994, Travis 2007). Ditransitives and 'decir' were left out of logistic regression analyses, the former due to the small number of tokens (N=315), and the latter because it did not pattern with other categories. Finally, reflexive verbs, which were merged together in the initial coding schema and displayed an overt subject rate of 58.7% (N=494), did not cause any interactions, so they were not re-coded.

In sum, *Verb* was re-coded into a six-code factor group according to the natural classes that arose in the corpus:

(2) *Verb* re-coding

a. Stative

b. Transitive

c. Inergative

d. Unaccusative

e. Reflexive

f. Other.

Regarding the factor groups that operationalize structural constraints, i.e., *Subject Position*, *Finiteness*, *Interrogative* and *Clause Type*, re-coding procedures were as follows: *Clause Type* was not re-coded; as explained above, *Subject Position* is excluded from the logistic regression analyses that examine null and overt subject patterns; *Finiteness* was left out of all logistic regression analyses since non-finite clauses will only be examined in Chapter Seven. Regarding *Interrogative*, very significant differences were found in the distribution of overt and null subjects in interrogative and non-interrogative clauses. For instance, while the rate of overt subject use in non-interrogatives is 61%, as mentioned above, this rate is 84% in interrogatives. While interrogative clauses only amount to 190 tokens and they could not be included in logistic regression analyses that examine null and overt subject patterns[48], quantitative analyses of the distribution of null and overt subjects in interrogatives were carried out by way of cross-tabulating *Interrogative* with a number of independent factor groups. The results of these analyses will be presented in 6.2.4. and 6.3.4. below.

As far as language-external variables are concerned, i.e. *Origin, Gender, Age, Social Class,* and *Education*, recall from the discussion in §4.3 that inherent asymmetries were found in the corpus gathered for this study in *Social class* and *Education*. The latter has the associated factors *illiterate*, *some education* (middle or high school) and *more education* (university-educated speakers); while illiterate speakers were only found in the rural area of El Cibao where data were

[48] *Interrogative*, on the contrary, is included in the logistic regression analyses that examine the distribution of preverbal and postverbal subjects (see 6.5. below).

collected, no illiterate speakers were recruited in the city. Conversely, no university-educated speakers were found in the rural community. *Social class* was coded into high-middle class and working class; while both high-middle and working class were interviewed in Santo Domingo, no high-middle class speakers were found in El Cibao. In order to examine the relevance of all external variables for subject expression in DS, three separate analyses that included different external variables were carried out with pronominal and lexical subjects: first, an analysis that included *Origin*, *Age* and *Gender*, in which the external factors that would cause interactions were left out. Second, analyses in which only urban speakers were analyzed and that included the factor groups *Social Class* and *Education*, in addition to *Age* and *Gender*, were performed. Finally, an analysis that included only working class speakers in the urban and rural communities and that examined the role of *Origin*, *Age* and *Gender* was also carried out.

Table 6.4. presents the results for the first of these three logistic regression analyses. According to the re-coding procedures just outlined, the following factor groups are included in the analysis: *Person, Verb, TMA, Distance from previous mention, Co-reference, Discourse function, TMA of the previous verb, Clause Type, Origin, Age* and *Gender*. The dependent variable is null or overt subject. Since Goldvarb requires that the dependent variable is binary, null subject was selected as the application value in these calculations. Non-applications of the dependent variable correspond then to overt pronominal and lexical subjects. Selecting null subject as the application value entails that the factor weights in Table 6.4. represent the favoring or disfavoring of null subject insertion. Percentages of application, located to the right of factor weights, refer to null subjects as well. Table 6.4. and all ensuing tables also include the following information: the total number of subject occurrences (both null and overt) for every factor (to the right of percentages of application), and relevant information of the logistic regression analysis carried out at the top of the table: the corrected mean or overall measure of the application of the rule, in this case null subject;

the log likelihood or measure of goodness of fit of the analysis[49], the overall significance of the analysis, and the total number of tokens included in the calculation:

Table 6.4. Dominican Spanish. Null subjects versus pronominal and lexical subjects: Patterns of variation.

Corrected mean				0.346
Log likelihood				-3211.478
Significance				0.001
Total N				6005
***App*: Null Subject**		**Factor weight**	**%**	**N**
PERSON	1psg *yo*	.44	39.7	2134
	2psg specific *tú*	.19	10.9	274
	2psg, 3rdpsg [-spec] *tú/uno*	.25	22.5	404
	3psg él/ella	.42	28.8	1863
	1ppl nosotros/nosotras	**.88**	73.8	480
	3ppl [+spec] *ellos/ellas*	**.77**	58.8	805
	Range	*69*		
CLAUSE TYPE	Main	.47	36	3555
	Conjoined	**.67**	59.5	557
	Embedded	**.52**	40.2	1641
	Relative	.40	31.2	247
	Range	*27*		
DISTANCE FROM PREVIOUS MENTION	0	**.60**	52.6	2851
	1	.47	32.9	705
	2	.47	33	330
	3	.38	23.9	2114
	Range	*22*		
VERB	Stative	.40	24.3	1241
	Transitive	**.51**	39.3	866
	Inergative	.45	35.5	293
	Unaccusative	.44	34.9	398
	Reflexive	**.53**	41.3	494
	Other	**.57**	47.4	1774
	Range	*17*		
CO-REFERENCE				
	Coreferent	**.58**	51.7	2992
	Non-coreferent	.42	26.6	3008
	Range	*16*		
TMA	Present indicative	.49	33.3	2752
	Preterite	**.58**	47	1501
	Imperfect indicative	.44	33.3	1289
	Range	*14*		
DISCOURSE FUNCTION	Topic	**.53**	43.2	4612
	Shift	.41	25.5	1388
	Range	*12*		
TMA OF PREVIOUS VERB	Same TMA	**.52**	51.7	2992
	Different TMA	.48	26.5	3008
	Range	*4*		
GENDER	Men	[.49]	40.3	3074
	Women	[.51]	37.9	2926
AGE	+50	[.49]	39.1	2881
	-50	[.51]	39.2	3119
ORIGIN	Urban	[.50]	39.2	3799
	Rural	[.49]	39.1	2201

[49] In log likelihood numbers, figures closer to zero represent better models than those further removed from zero (Tagliamonte 2006:265).

In Table 6.4. and all ensuing tables factor weights and percentages for both significant and non-significant factor groups are reported. Following standard variationist notation, factor weights of non-significant factor groups appear in brackets. Tagliamonte (2006) points out that while factor weights of non-significant variables are sometimes not reported, they might provide valuable information for the explanation of patterns of distribution. As will become clear when the results of analyzing null and overt and subject position patterns are compared in section 6.5, this is the case in the present study. Before moving on to present the results of the statistical analyses that include the external factor groups *Social Class* and *Education*, the results depicted in Table 6.4. are discussed in the following section.

6.2.1. Pronominal and lexical subjects versus null subjects. Variables, variable ranking and constraint ranking: Discussion.

As discussed in Chapter Four, statistical analysis with Goldvarb provides, in addition to the list of significant variables, three types of measures that constitute lines of evidence for explaining patterns of variability:

(i) The overall significance of the analysis (0.001 in Table 6.4).

(ii) Constraint rankings within significant factor groups, reflected in factor weights. Generally, values above .5 are taken to favor the application of the dependent variable, while values under .5 are taken to disfavor its application.

(iii) The relative strength of significant factor groups, as expressed by the Range, obtained by subtracting the lowest factor weight from the highest factor weight within a factor group.

In Table 6.4, *Person* (Range=69) arises as the most significant variable in subject expression, as in Otheguy's et al. (2007) analysis of Caribbean newcomers in New York City. Constraint rankings within *Person* are also remarkably similar in the present study and in the one carried out by Otheguy et al (cf. Table 4.10 in Chapter Four), even if in the statistical analyses presented in Table 6.4 not only pronominal, but also lexical subjects are taken into account: second

person singular [+specific] and [-specific] strongly disfavor null subjects, with factor weights of .19 and .25 respectively. First and third person plural, on the other hand, strongly favor null subjects with factor weights of .88 and .77. In Otheguy's et al investigation, in which statistics were carried out with SPSS, Exp(B) values for second person [+specific] and [-specific] were 3.22 and 3.13 respectively, while first person plural and third person plural Exp(B) values were 0.30 and 0.29[50]. Furthermore, first and third person singular also pattern very similarly in both analyses, in that they favor overt subjects, however not as strongly as second person subjects; values for first and third person are Exp(B) 1.04 and Exp(B) 1.09 in Otheguy et al, and .44 for first person and .42 for third person in the present study.

The second ranked variable in Table 6.4 is *Clause Type* (Range=27). Recall from the discussion in Chapter Four that in the studies that examine this variable, the general finding is that embedded clauses favor null subjects, while neither overt nor null subjects are favored in main clauses (Silva-Corvalán 1994, Otheguy et al 2007). The latter result has been interpreted to reflect that factor groups other than *Clause Type* regulate null and overt subject insertion in main clauses. In Table 6.4, main clauses do not favor null subjects, but with a factor weight of .47, which is very close to .5, that would entail that neither overt nor null subjects are favored. Embedded clauses slightly favor null subjects at .52. The factor group *Clause Type* includes, in addition to main and embedded, conjoined and relative clauses. Conjoined clauses favor null subjects at a rate of .67, while relative clauses disfavor them at .40. This result for relative clauses is interesting in light of the syntactic phenomena that have been described in previous work on DS. Specifically interrogative and non-interrogative clefted structures, which are widespread in this dialect, contain relative clauses. This result would suggest then that Dominican vernacular constructions favor overt subject insertion. It is also an interesting result that *Clause Type* is ranked second in the factor group hierarchy in DS. This factor group has been found significant in previous work (Morales 1997, Otheguy et al. 2007), however its effect has been reported to be less significant than the effect of

[50] Recall that in Otheguy et al. (2007) Exp(B) <1.0 indicates that a factor favors nulls and Exp(B)>1.0 indicates that factor favors overts.

other discourse-related and morpho-syntactic variables. The results obtained in the present study might be attributed, on the one hand, to the inclusion of conjoined structures in the analysis, which strongly favor null subjects and, on the other hand, to the fact that relative clauses strongly favor overt subjects in DS.

Distance from previous mention (Range=22) follows *Clause Type* in the factor group hierarchy. This finding reflects the results found in previous work (Cameron 1992, Otheguy et al.2007) in that the effect of a discourse-related factor group overrides the effects on subject expression of morpho-syntactic factor groups such as TMA (Range=14), which is ranked sixth in this analysis. In fact, another discourse-related factor group, *Co-reference* (Range=16), is also ranked higher than TMA. *Discourse function* (Range=12) and *TMA of the previous verb* (Range=4) are the discourse-related factor groups that, while being significant, have a smaller effect on subject expression. Factor weights within *Verb* (Range=17) reflect the trends observed when overall percentages were obtained: transitive and reflexive verbs favor null subjects at .51 and .53 respectively; stative verbs disfavor nulls at .40, and inergative and unaccusative verbs disfavor null subjects with very similar factor weights, .45 and .44. Reflexive verbs were also found to favor null subjects in Otheguy et al. (2007) with an Exp(B) value of 1.98.

Factor weights within *TMA* show that the preterite, a non-ambiguous verb tense, favors null subjects, while the imperfect, a three-way ambiguous tense, disfavors them. The present, with a factor weight of .49, is very close to neither favoring nor disfavoring null subjects, along the lines of the results in Silva-Corvalán (2001), who suggests that it is the discursive role of verb tenses rather than morphological ambiguity that regulates subject choice. However, since only imperfect, preterit and present are considered as factors in this statistical analysis, it is not possible to determine whether discursive or morphological properties are involved in subject selection. A detailed analysis that aims to distinguish the relative strength of morphological ambiguity in subject expression in DS is carried out in 6.2.3.

Language-external factor groups, i.e. *Gender*, *Age*, and *Origin* do not arise as significant. This finding parallels the finding in previous work (Silva-Corvalán 1994, Travis 2007). However, as the results presented in the following section show, other language-external factor groups, such as *Social class*, emerge as significant when they are included in the statistical analysis.

6.2.2. On the role of other language-external variables: *Social class* and *Education*.

In order to examine the role of *Social class* in subject expression, a logistic regression analysis that included this factor group was carried out. Since high-middle class speakers were only found in Santo Domingo, while rural speakers are homogeneous with respect to *Social class*, only urban speakers were included in these calculations. The asymmetries in *Social class* in the data set are illustrated in Table 6.5:

Table 6.5. Dominican Spanish. Asymmetries in language external factor groups: Social class and Education (N=3606).

Santo Domingo (N=21; 3801 tokens)			**El Cibao** (N=13; 2205 tokens)		
SOCIAL CLASS			**SOCIAL CLASS**		
HIGH	LOW		HIGH	LOW	
12	9		------	13	
EDUCATION[51]			**EDUCATION**		
ILLITERATE	SOME SCHOOL	MORE SCHOOL	ILLITERATE	SOME SCHOOL	MORE SCHOOL
-------	10	11	4	9	-------

After carrying out the analysis of *Social class* in the city, working classes in the urban and rural communities were compared in a separate statistical analysis. In Table 6.5. I present the results for urban speakers:

[51] In this table, 'some school' includes speakers that have a middle/high school background, while speakers with a professional school or university background were ascribed to the 'more education' group.

Table 6.6. Dominican Spanish. Social class and Education in Santo Domingo. Patterns of variation for null versus overt (pronominal and lexical) subjects (N=3606).

Corrected mean				0.345
Log likelihood				-1845.073
Significance				0.012
Total N				3606
***App*. Null Subject**		***Factor weight***	***%***	***N***
PERSON	1stp sg *yo*	*.42*	*37.8*	*1144*
	2stp sg [+spec] *tú*	.14	8.4	203
	2stp sg [-spec] *tú*	.24	20.4	270
	3rdp sg *él/ella*	.41	29	1151
	1stp pl *nosotros/nosotras*	**.90**	77.5	342
	3rdp pl *ellos/ellas*	**.81**	63.6	467
	Range	*76*		
CLAUSE TYPE	Main	.47	35.7	2108
	Conjoined	**.67**	61.1	316
	Embedded	**.51**	41.4	1012
	Range	*20*		
DISTANCE FROM	0	**.58**	51.3	1719
PREVIOUS MENTION	1	**.51**	36.2	414
	2	.47	32.2	141
	3	.38	25.2	1265
	Range	*20*		
CO-REFERENCE	Co-referent	**.57**	50.8	1739
	Non-co-referent	.42	28.7	1867
	Range	*15*		
VERB	Stative	.41	24.4	771
	Transitive	.47	37.2	557
	Inergative	.47	35.4	175
	Unaccusative	.45	35.8	257
	Reflexive	**.56**	41.9	346
	Other	**.57**	48.4	963
	Range	*15*		
TMA	Imperfect	.45	32.9	762
	Preterite	**.59**	48.9	939
	Present	.48	31.8	1599
	Range	*14*		
DISCOURSE FUNCTION	Topic	**.53**	42.7	2745
	Shift	.40	28.6	615
	Range	*13*		
SOCIAL CLASS	High-middle class	.47	37.5	1931
	Working class	**.52**	41.4	1675
	Range	*5*		
TMA PREVIOUS	Same TMA	**.52**	41.9	1773
	Different TMA	.48	36.9	1875
	Range	*4*		
AGE	+50	[.48]	38.4	1398
	-50	[.51]	39.9	2208
GENDER	Men	[.49]	40.4	2120
	Women	[.51]	37.8	1486
EDUCATION	University	[.50]	39.1	2467
	High school	[.49]	39.8	1139

The main difference that emerges between the results in Table 6.4. and Table 6.6. is that the external factor group *Social class* (Range=5) is significant in 6.6, while all external factors were not significant in 6.4. Among urban speakers, working class favors null subjects at .52, while high-middle-class disfavors them at .47. *Social class* is ranked over the least significant discourse-related factor group, *TMA of the previous verb* (Range=4), and below all other language-internal factor groups. *Gender*, *Age* and *Origin* are not significant for urban speakers, similarly to what emerged from Table 6.4. Furthermore, *Education* is not selected as significant in Table 6.6.

The top three factor groups in the hierarchy, i.e. *Person* (Range=76), *Clause Type* (Range=20) and *Distance from previous mention* (Range=15) are the same in Tables 6.4 and 6.6, and associated factors also display very similar weights. As far as *Clause Type* is concerned, in these calculations relative clauses had to be left out of the analysis due to interactions; quantitative results that show the behavior of different types of clauses, obtained by cross-tabulating *Clause Type* with *Person* are presented below.

The results of the analysis of working classes in urban and rural communities are presented in Table 6.7. Note that *Origin* is not significant when both communities are compared, which suggests that working classes are homogeneous in Santo Domingo and El Cibao. *Gender* and *Age* are not selected as significant either, as in Tables 6.6 and 6.4. *Person* (Range=67), *Clause Type* (Range=25) and *Distance from previous mention* (Range=24) are the three factor groups at the top of the hierarchy, also similarly to the results presented in Tables 6.6. and 6.4.

Table 6.7. Dominican Spanish. Working class in Santo Domingo and El Cibao: Patterns of variation for null versus overt (pronominal and lexical) subjects (N=3851).

Corrected mean				0.369
Log likelihood				-2111.283
Significance				0.001
Total N				3851
***App*. Null Subject**		***Factor weight***	***%***	***N***
PERSON	1stp sg *yo*	.46	42.8	1513
	2stp sg [+spec] *tú*	.20	11.8	152
	2stp sg [-spec] *tú*	.27	26.3	213
	3rdp sg *él/ella*	.42	28.7	1170
	1stp pl *nosotros/nosotras*	**.87**	73.4	289
	3rdp pl *ellos/ellas*	**.73**	55.2	498
	Range	*67*		
CLAUSE TYPE	Main	.47	37.7	2309
	Conjoined	**.65**	58.6	353
	Embedded	**.52**	40.3	1056
	Relative	.40	33.1	133
	Range	*25*		
DISTANCE FROM	0	**.61**	54.7	1866
PREVIOUS MENTION	1	.45	32.1	471
	2	.44	33.3	198
	3	.37	23.6	1316
	Range	*24*		
CO-REFERENCE	Co-referent	**.58**	53.5	923
	Non-co-referent	.42	26	1380
	Range	*16*		
TMA	Imperfect	.43	34.1	791
	Preterit	**.59**	47.4	897
	Present	.48	35.2	1864
	Range	*16*		
VERB	Stative	.43	26.9	769
	Transitive	**.54**	43.7	563
	Inergative	.43	43.5	211
	Unaccusative	.43	35.1	231
	Reflexive	.49	40.8	240
	Other	**.55**	47.6	1235
	Range	*12*		
DISCOURSE FUNCTION	Topic	**.52**	44.8	2994
	Shift	.41	24.3	857
	Range	*11*		
TMA PREVIOUS	Same TMA	**.52**	42.7	2008
	Different TMA	.48	37.4	1843
	Range	*7*		
ORIGIN	Urban	[.52]	41.7	1650
	Rural	[.49]	39.1	2201
GENDER	Men	[.49]	41.2	1922
	Women	[.51]	39.2	1929
AGE	+50	[.49]	39.6	1716
	-50	[.51]	40.7	2135

In Table 6.7. results differ from previous tables with respect to the strength of *TMA*, which is ranked over *Verb* when only working classes in urban and rural communities are compared. Table 6.8 offers a comparison among the factor weights and percentages within Verb in Tables 6.4, 6.6 and 6.7. The percentages that stand out are highlighted in gray:

Table 6.8. Dominican Spanish. *Verb* Patterns of variability: Urban and Rural speakers and Social Class.

	All speakers (N=6005)			**Urban speakers (N=3606)**			**Working class in Urban and Rural communities (N=3851)**		
Verb	***F. W.***	***%***	***N***	***F.W.***	***%***	***N***	***F.W.***	***%***	***N***
STATIVE	.40	24.3	1241	.41	24.4	771	.43	26.9	769
TRANSITIVE	.51	39.3	866	.47	37.2	557	.54	43.7	563
INERGATIVE	.45	35.5	293	.47	35.4	175	.43	43.5	211
UNACCUSATIVE	.44	34.9	398	.45	35.8	257	.43	35.1	231
REFLEXIVE	.53	41.3	494	.56	41.9	346	.49	40.8	240
OTHER	.57	47.4	1774	.57	48.4	963	.55	47.6	1235
Range	17			15			12		

Factor weights across communities are fairly similar: stative verbs disfavor null subjects in all three communities, with factor weights of .40 for all speakers, .41 for urban speakers, and .43 for urban and rural working classes. However, small differences are attested as well among urban and rural speakers: in urban and rural working classes, inergative and unaccusative verbs are not distinguished from stative verbs, and display identical factor weights, while these two types of verbs display factor weights closer to .5 in urban speakers and when all speakers are considered (.45 and .44 for the latter, and .47 and .45 for the former). As far as transitive verbs are concerned, they favor null subjects both in the analysis that included all speakers (at .51) and in the analysis that considers only working classes in Santo Domingo and El Cibao (at .54), while null subjects are disfavored by transitive verbs in the analysis of urban speakers (.47). Finally, while reflexive verbs favor nulls for all speakers and urban speakers, at .53 and .56 respectively, null subjects are neither favored nor disfavored by working classes at .49.

A different pattern emerges when factor weights for discourse-related factors are compared in the three communities examined so far. The results of this comparison are offered in Table 6.9:

Table 6.9. Dominican Spanish. Patterns of variability, discourse-related factor groups: Urban and Rural speakers and Social Class.

	All speakers (N=6005)			Urban speakers (N=3606)			Working class in urban and rural communities (N=3851)		
Distance	***F.W.***	***%***	***N***	***F.W.***	***%***	***N***	***F.W.***	***%***	***N***
0 CLAUSES	.60	24.3	1241	.58	24.4	771	.61	26.9	769
1 CLAUSE	.47	39.3	866	.51	372	557	.45	437	563
2 CLAUSES	.47	35.5	293	.47	35.4	175	.45	43.5	211
3 OR MORE	.38	34.9	398	.38	35.8	257	.37	35.1	231
Range	*22*			*20*			*24*		
Co-reference									
CO-REFERENT	.58	51.7	2992	.57	50.8	1739	.58	53.5	923
NON-CO-REFERENT	.42	26.6	3008	.42	287	1867	.42	26	1380
Range	*16*			*15*			*16*		
Discourse function									
TOPIC	.53	43.2	4612	.53	42.7	2745	.52	44.8	2994
SHIFT	.41	25.5	1388	.40	28.6	615	.41	24.3	857
Range	*12*			*13*			*11*		
TMA previous									
SAME TMA	.52	51.7	2992	.48	38.4	1398	.52	42.7	2008
DIFFERENT TMA	.48	26.5	3008	.51	39.9	2208	.48	37.4	1843
Range	*4*			*3*			*4*		

Note how factor weights for all three communities are remarkably similar in *Co-reference*, *Discourse function*, *Distance from the previous mention* and *TMA of the previous* verbs. Factor weights for *TMA* are very similar in all three communities as well, as observed in Table 6.10:

Table 6.10. Dominican Spanish. TMA Patterns of variability: Urban and Rural speakers and Social Class.

	Urban and Rural speakers (N=6005)			Urban speakers (N=3606)			Working class in urban and rural communities (N=3851)		
TMA	***F. W.***	***%***	***N***	***F.W.***	***%***	***N***	***F.W.***	***%***	***N***
IMPERFECT	.44	33.3	1289	.45	32.9	762	.43	34.1	791
PRETERIT	.58	47	1501	.59	48.9	939	.59	47.4	897
PRESENT	.49	333	2752	.48	31.8	1599	.48	35.2	1864
Range	14			14			16		

To sum up, what has emerged from the comparisons in Tables 6.8, 6.9 and 6.10 is that all three communities display the same patterns of variability with respect to *TMA* and discourse-related factors. Slight differences among communities are attested in the factor group *Verb*, in

which urban speakers depart from all speakers and working classes in urban and rural communities, also generally disfavoring null subjects at higher rates than the rest of the speakers in the data set.

A comparison of the distribution of the different persons of the paradigm with respect to different types of clauses also follows this trend according to which urban speakers depart from the rest. *Clause Type* has been selected as very significant factor group in the three communities examined. Conjoined clauses strongly favor null subjects in all three communities, and relative clauses have been shown to disfavor null subjects in the analysis of all speakers and working class speakers in urban and rural communities. Relative clauses had to be excluded from the analysis of urban speakers due to interactions. In order to assess the differences among communities with respect to *Clause Type*, this factor group and *Person* were cross-tabulated in the data set that comprises all speakers and in the data sets containing only urban speakers and only working class speakers. Cross-tabulation[52] shows how two specific factors are related, so carrying out this operation with *Clause Type* and *Person* will also allow to observe whether all presons of the paradigm follow the same trends in favoring or disfavoring null subjects in different clause types; this information becomes particularly relevant in the corpus at hand, since strong differences have arisen across the persons of the paradigm. Table 6.11 displays the results of cross-tabulating *Person* and *Clause Type* in the data set that comprises all speakers (N=6005, logistic regression analyses offered in Table 6.4.). Rates of overt subject use are highlighted in gray:

[52] Tagliamonte (2006) stresses the relevance of cross-tabulation for understanding the distribution of a given linguistic phenomenon in a data set. Cross-tabulating is not only fundamental for spotting the interactions that might occur among factor groups, but also for accurately assessing how the phenomenon at hand is distributed in every possible intersection of the factor groups included in the analysis.

Table 6.11. Dominican Spanish. Cross-tabulation of Person and Clause Type. All speakers (N=6005).

		MAIN		CONJOINED		EMBEDDED		RELATIVE	
		null	***overt***	***null***	***overt***	***null***	***overt***	***null***	***overt***
1STP SG *yo*	***N***	481	827	121	83	230	306	16	70
	%	37	63	59	41	43	57	19	81
2NDP SG [+SPEC] *tú*	***N***	14	175	3	3	13	53	0	13
	%	7	93	50	50	20	80	0	100
2NDP SG [-SPEC] *tú/uno*	***N***	37	128	19	18	33	147	2	20
	%	22	78	51	49	18	82	9	91
3RDP SG [+SPEC] *él/ella*	***N***	278	842	96	99	146	349	17	36
	%	25	75	49	51	29	71	36	68
1STP PL *nosotros/as*	***N***	210	76	37	6	91	31	16	13
	%	73	27	86	14	75	25	55	45
3RDP PL [+SPEC] *ellos/ellas*	***N***	162	197	33	14	88	82	12	13
	%	45	55	70	30	52	48	48	52

Overall rates of null and overt subject use in Table 6.11. reflect the great variation among persons of the paradigm that could be observed in Table 6.1. In all persons, relative clauses disfavor null subjects and conjoined clauses favor them. A comparison of main and embedded clauses shows that for all persons except for the second person [-specific], rates of overt subject use are higher in main than in embedded clauses. In Table 6.11. we also observe how the second person singular [+specific] departs from the rest of the persons of the paradigm in rates of overt subjects for all four types of clauses: rates for second person [+specific] and [-specific] are very similar, followed by the first or third person singular; first and third person plural bear the lowest rates of overt subject use in all types of clauses.

The results of cross-tabulating *Person* and *Clause Type* for the data set that comprises only urban speakers (N=3606, logistic regression analyses offered in Table 6.6) are offered in Table 6.12:

Table 6.12. Dominican Spanish. Cross-tabulation of *Person* and *Clause Type*. Only urban speakers (N=3606).

		MAIN		CONJOINED		EMBEDDED		RELATIVE	
		null	***overt***	***null***	***overt***	***null***	***overt***	***null***	***overt***
1STP SG *yo*	***N***	232	454	64	42	128	169	9	46
	%	34	66	60	40	43	57	16	84
2NDP SG [+SPEC] *tú*	***N***	6	133	2	2	9	41	0	0
	%	4	96	50	50	18	82	10	100
2NDP SG [-SPEC] *tú/uno*	***N***	19	80	6	10	28	109	2	16
	%	19	81	38	62	20	80	11	89
3RDP SG [+SPEC] *él/ella*	***N***	181	522	59	61	89	206	9	24
	%	26	74	49	51	30	70	27	73
1STP PL *nosotros/as*	***N***	152	42	26	3	75	22	12	10
	%	78	22	90	10	77	23	55	45
3RDP PL [+SPEC] *ellos*	***N***	156	113	35	5	85	43	21	9
	%	58	42	88	12	66	34	70	30

As far as main clauses are concerned, rates in Table 6.12 are higher than rates in Table 6.11. for all persons in the paradigm. Differences among persons follow the trends observed in Table 6.11: the second persons [+specific] and [-specific] display the highest rates of overt subject use, followed by the third person singular and the first person singular; the lowest rates of overt subjects are found in the first and third person plural. Also like in Table 6.11, the lowest rates of overt subject use arise with conjoined clauses, while relative clauses display the highest rates of overt subject use. In both types of clauses, roughly overt subjects are higher in the data set that only includes urban speakers than in the one in which all speakers are included. Overt subject rates are very similar for all persons in embedded clauses, with the exception of the third person plural, in which urban speakers have a lower rate of overt subject use (34%) than all speakers (48%).

Table 6.13 offers the results of cross-tabulating *Person* and *Clause Type* in the data set that includes working class speakers from Santo Domingo and El Cibao:

Table 6.13. Dominican Spanish. Cross-tabulation of *Person* and *Clause Type*. Working class speakers in Santo Domingo and El Cibao (N=3781).

		main		conjoined		embedded		relative	
		null	***overt***	***null***	***overt***	***null***	***overt***	***null***	***overt***
1STP SG *yo*	N	378	557	83	56	175	214	11	39
	%	40	60	60	40	45	55	22	70
2NDP SG [+SPEC] *tú*	N	11	93	1	2	6	31	0	8
	%	11	89	33	67	16	84	0	100
2NDP SG [-SPEC] *tú*	N	23	69	17	10	16	70	0	8
	%	25	75	63	37	19	81	0	100
3RDP SG [+SPEC] *él/ella*	N	176	524	54	61	96	232	10	17
	%	25	75	47	53	29	71	37	63
1STP PL *nosotros/as*	N	132	50	25	4	46	17	9	6
	%	73	27	86	14	73	27	60	40
3RDP PL [+SPEC] *ellos*	N	148	139	27	13	86	61	14	10
	%	52	48	68	32	59	41	58	42

As illustrated in Table 6.13, rates of overt subject use in main clauses for working class speakers across urban and rural communities are closer to the rates displayed by all speakers and depart from the rates displayed by urban speakers. This trend is roughly followed in conjoined, embedded and relative clauses. Working class speakers also follow the pattern observed in the two previous tables regarding person and clause type differences: the second person singular [+specific] and [-specific] display the highest rates of overt subject use, and overt subjects rates reach the highest numbers in relative clauses and the lowest numbers in conjoined clauses.

In order to assess whether the differences in rates observed in Tables 6.12, 6.13 and 6.14 are statistically significant, chi-squares calculations were performed. Urban speakers were compared with all speakers and with working classes in urban and rural environments. The results of these calculations are presented in Table 6.14:

Table 6.14. Dominican Spanish. Chi-squares of the differences in Clause Type. Urban speakers versus all speakers and working class urban and rural speakers versus urban speakers.

SANTO DOMINGO SPEAKERS VERSUS ALL SPEAKERS		
	Chi-square	*p (significant=p≤0.05)*
Main	7.315	.198
Conjoined	1.406	.843
Embedded	5.863	.320
Relative	9.042	.107
SANTO DOMINGO SPEAKERS VERSUS URBAN AND RURAL WORKING CLASSES		
Main	18.997	.002
Conjoined	3.759	.440
Embedded	19.481	.002
Relative	12.087	.034

As observed in Table 6.14, no statistically significant differences are found in any of the clause types examined when the Santo Domingo and all speakers' data sets are compared. However, significant differences emerge in main, embedded and relative clauses when Santo Domingo speakers are compared with the data set that includes urban and rural working class speakers.

To sum up, the logistic regression and cross-tabulations performed so far show that a myriad of constraints regulate null or overt subject choice in DS. Generally, *Person* and discourse-related factor groups have been shown to be stable across the data sets examined so far. Regarding the effect of language-external factor groups, *Social class* has emerged as significant (to a small degree) in the data set that examines the speech of urban speakers, with working class speakers favoring null subjects. Significant differences in subject expression have been found between urban speakers and urban and rural working class speakers, as attested by the chi-square values provided in Table 6.14. *Verb* and *TMA* have also emerged as significant factor groups in the three communities examined; while differences among communities have been found with respect to *Verb*, *TMA* has been found to be a stable factor group across urban and rural speakers and across high-middle and working classes. However, the re-coding procedures that were followed in order to perform accurate logistic regression analyses have not allowed exploring the relative strength of morphological and contextual ambiguity in subject expression. In order to examine this issue,

subsequent logistic regression analyses in which only the effects of TMA and discourse-related factor groups were carried out. Calculating the effects of a smaller number of factor groups allows including subjunctive verb tenses in the logistic regression analyses. Furthermore, since verb form ambiguity only occurs in the singular persons of the paradigm, in these analyses plural persons have been left out of the data set. Results are presented in the following section.

6.2.3. On the relative role of contextual and morphological ambiguity.

Previous work on Caribbean dialects (Cameron 1992, 1996) has shown that morphological ambiguity does not override contextual ambiguity. This find cast doubts on the accuracy of the *Functional Compensation Hypothesis* (Hochberg 1986). As will become clear when results are presented in the following paragraphs, the Dominican data examined in the present study seem to confirm Cameron's findings regarding the relative strength of morphological and contextual ambiguity. The logistic regression analyses to follow have been carried out with the data set that includes all speakers (logistic regression analyses of all language-internal factor groups provided in Table 6.4. above). The results probing for the effects of TMA in the singular paradigm are provided in Table 6.15:

Table 6.15. Dominican Spanish. Null versus pronominal and lexical subjects: TMA patterns of variation.				
Corrected mean				0.299
Log likelihood				-2780.141
Significance				0.001
Total N				4675
***App*. Null Subject**		***Factor weight***	***%***	***N***
PERSON	1stp sg *yo*	**.58**	39.7	848
	2ndp sg [+spec] *tú*	.22	10.9	274
	2ndp sg person [-spec] *tú*	.40	22.5	404
	3rdp sg *él/ella*	.47	28.8	1863
	Range	*36*		
TMA	Imperfect	.44	25.8	1005
	Preterit	**.60**	39.7	1164
	Present	.47	26.4	2143
	Subjunctive	**.60**	34.7	147
	Range	*16*		
ORIGIN	Urban	.48	30.4	2898
	Rural	**.52**	35.1	1777
	Range	*4*		
AGE	+50	[.50]	33.2	2242
	-50	[.50]	31.3	2433
GENDER	Men	[.50]	31.2	2260
	Women	[.50]	33.1	2415

In Table 6.15. *Person* (Range=36) ranks first in the factor group hierarchy, as in Table 6.4, where all linguistic factor groups were considered. This result suggests that differences among persons of the paradigm are relevant not only for the comparison of singular and plural persons, but also for the comparison of singular persons against each other. *TMA* (Range=16) follows Person in the factor group hierarchy. Within *TMA*, the preterit, a non-ambiguous tense, and the subjunctive tenses, which are three-way morphologically ambiguous, favor null subjects at .60, while the imperfect disfavors them at .44. The present also disfavors null subjects, but with a factor weight of .47, i.e., closer to the .5 reference factor weight that would reflect that null subjects are neither favored nor disfavored. The comparison of constraint rankings of the preterite and the subjunctive tenses do not support the *Functional Compensation Hypothesis* (Hochberg 1986), since both non-ambiguous and three-way ambiguous tenses favor null subjects. The comparison of constraint rankings of the present and the imperfect could be taken as evidence supporting both the *Functional*

Compensation Hypothesis and Silva-Corvalán's (2001) hypothesis on the discursive properties of verb tenses. On the one hand, the imperfect is three-way ambiguous and it disfavors null subjects at a higher factor weight than the present, a two-way ambiguous tense. On the other hand, according to Silva-Corvalán, the imperfect, as a tense that is used in speech for the description of background events, disfavors null subjects more than the present, the tense that she identifies as neither favoring nor disfavoring null subjects.

Interestingly, among language-external factor groups, *Origin* is selected as significant in Table 6.15, with rural speakers favoring null subjects (.52) more than urban speakers (.48). This result is not replicated when only discourse-related factor groups are considered in the analysis; as illustrated in Table 6.16, which suggests that urban and rural speakers are homogeneous with respect to the strength of these factor groups:

Table 6.16. Dominican Spanish. Null versus pronominal and lexical subjects: patterns of variation in discourse-related factor groups.

Corrected mean				0.279
Log likelihood				-2534.701
Significance				0.005
Total N				4675
***App*. Null Subject**		***Factor weight***	***%***	***N***
PERSON	1stp sg *yo*	**.53**	39.7	848
	2ndp sg [+spec] *tú*	.26	10.9	274
	2ndp sg person [-spec] *tú*	.35	22.5	404
	3rdp sg *él/ella*	**.54**	28.8	1863
	Range	*28*		
DISTANCE FROM PREVIOUS MENTION	0 clauses	**.61**	47.8	2339
	1 clause	**.50**	26.6	567
	2 clauses	.48	24.4	254
	3 clauses or more/new	.34	11.6	1515
	Range	*27*		
DISCOURSE FUNCTION	Topic continuation	**.54**	37.2	3772
	Topic shift	.34	11.3	88.7
	Range	*20*		
CO-REFERENCE	Co-referent	**.59**	46.6	2451
	Non-coreferent	.40	16.4	2224
	Range	*19*		
TMA OF PREVIOUS VERB	Same TMA	.50	35.2	2344
	Different TMA	.47	29.2	2331
	Range	*3*		
ORIGIN	Urban	[.49]	47.8	2339
	Rural	[.52]	35.1	1777
AGE	+50	[.50]	33.2	1497
	-50	[.49]	31.3	2433
GENDER	Men	[.50]	31.2	2260
	Women	[.50]	33.1	2415

In Table 6.16. *Person* (Range=28) ranks first in the factor group hierarchy, followed by *Distance of the previous mention* (Range= 27). Note that in this logistic regression analysis *Person* does not depart as much in Range from the factor groups that follow it in the hierarchy, suggesting that discourse-related factors have a similar effect on all singular persons of the paradigm. Furthermore, if we compare these results with the ones reported in Table 6.4, where all linguistic factors were included in the calculations, we can observe that the strength of discourse-related factor groups is roughly the same: Distance from the previous mention and TMA of the previous verb (Range=3) are the strongest and weakest factor groups respectively, while *Discourse function*

and *Co-reference* have a similar effect in terms of strength; the former had a range of 20 in Table 6.4. and a range of 15 in Table 6.16, while the latter had a range of 17 in Table 6.4. and a range of 15 in Table 6.16.

Table 6.17. offers the results of logistic regression analysis when the effects of both *TMA* and discourse-related factors are considered jointly:

Table 6.17. Dominican Spanish. Null versus pronominal and lexical subjects: TMA and discourse-related factors patterns of variation.

Corrected mean				0.299
Log likelihood				-2780.141
Significance				0.027
Total N				4675
***App*. Null Subject**		***Factor weight***	***%***	***N***
PERSON	1stp sg *yo*	**.53**	39.7	2134
	2ndp sg [+spec] *tú*	.26	10.9	274
	2ndp sg person [-spec] *tú*	.36	22.5	404
	3rdp sg *él/ella*	**.54**	28.8	1863
	Range	*28*		
DISTANCE FROM PREVIOUS MENTION	0	**.61**	47.8	2339
	1	.50	26.6	567
	2	.47	24.4	254
	3	.34	11.6	1515
	Range	*27*		
DISCOURSE FUNCTION	Topic	**.54**	37.2	3772
	Shift	.34	11.3	903
	Range	*20*		
CO-REFERENCE	Co-referent	**.58**	46.6	2451
	Non-co-referent	.41	16.4	2224
	Range	*17*		
TMA PREVIOUS	Same TMA	**.58**	35.2	2344
	Different TMA	.41	29.2	2331
	Range	*17*		
TMA	Imperfect	.42	25.8	1005
	Preterit	**.60**	39.7	1164
	Present	.48	26.4	2143
	Subjunctive	**.60**	34.7	147
	Range	*17*		
ORIGIN	Urban	[.49]	30.4	2898
	Rural	[.52]	35.1	1777
AGE	+50	[.50]	33.2	2242
	-50	[.50]	31.3	2433
GENDER	Men	[.50]	31.2	2260
	Women	[.50]	33.1	2415

Person (Range=28) and *Distance from the previous mention* (Range=27) bear in Table 6.17. very similar values to the ones presented in Table 6.16: *Person* is ranked over *Distance*, however their ranges are very close. *Discourse function* (Range=20) follows these two variables in the hierarchy. *Co-reference*, *TMA of the previous verb* and *TMA* are the three variables at the bottom of the hierarchy, all three of them with a Range of 17. This similarity in Ranges suggests that verb ambiguity does not override contextual ambiguity in DS, as found in previous work for other Caribbean varieties (Cameron 1996).

In Tables 6.15, 6.16, and 6.17 *Person* has emerged as a very significant factor group, suggesting that relevant differences emerge among singular persons of the paradigm. For this reason, logistic regression analyses were also carried out including only specific persons of the paradigm in the calculations. The effect on subject choice of only *TMA*, only discourse-related factors and *TMA* and discourse-related factors was calculated for first and second person, on the one hand, and first and third person, on the other[53]. These analyses were performed with two aims: first, to further assess the accuracy of the *Functional Compensation Hypothesis*; if this hypothesis were on the right track, the second and the third person, where morphological ambiguity occurs with both two-way ambiguous and three-way ambiguous verb tenses, should show similar patterns of variation. Second, these calculations make possible to explore whether the discursive properties of first, second and third person singular have an impact on overt or null subject selection. Recall from the discussion in §5.3 that while first and second person are deictic, which makes reference tracking straightforward, the third person is potentially anaphoric.

In Table 6.18. I present the results obtained calculating the effects of TMA on first and second person singular:

[53] Note that the other logical alternative is to perform analyses comparing first and second person singular jointly, on the one hand, and second and third person jointly, on the other. These analyses could not be performed with the available data set due to interactions.

Table 6.18. Dominican Spanish. Null versus pronominal and lexical subjects: First and second person singular *TMA* patterns of variation.				
Corrected mean				0.312
Log likelihood				-1672.819
Significance				0.004
Total N				2812
***App*. Null Subject**		***Factor weight***	***%***	***N***
PERSON	1stp sg *yo*	**.56**	39.7	2134
	2ndp sg [+spec] *tú*	.27	10.9	274
	2ndp sg [-spec] *tú*	.38	22.5	404
	Range	*18*		
TMA	Imperfect	.43	28	553
	Preterit	**.61**	45.5	629
	Present	.47	26.9	1369
	Subjunctive	**.57**	31	87
	Range	*18*		
ORIGIN	Urban	.47	31.3	1697
	Rural	**.54**	39.2	1115
	Range	*7*		
AGE	+50	[.51]	36.2	1337
	-50	[.49]	32.9	1475
GENDER	Men	[.50]	33.2	1327
	Women	[.50]	35.6	1485

In this table, *Person* and *TMA* have the same strength, as observed in their identical Range of 18. Furthermore, constraint rankings within *TMA* confirm the trends in Table 6.15. above: the three-way ambiguous subjunctive tenses and the non-ambiguous preterite both favor null subjects, with factor weights of .57 and .61 respectively. *Origin* (Range=7) also arises as significant, like in Table 6.15, again with rural speakers favoring null subjects (.54) and urban speakers disfavoring them (.47).

This latter result for *Origin* is replicated when first and third persons are analyzed, as illustrated in Table 6.19. Results differ from the ones in 6.18. in that when first and third persons are analyzed, *TMA* (Range=14) ranks over *Person* (Range=8). Constraint rankings within *TMA* replicate the ones in Tables 6.15. and 6.18:

Table 6.19. Dominican Spanish. Null versus pronominal and lexical subjects: First and third person singular *TMA* patterns of variation.				
Corrected mean				0.265
Log likelihood				-1192.362
Significance				0.013
Total N				2019
***App*. Null Subject**		***Factor weight***	***%***	***N***
TMA	Imperfect	.43	26.2	941
	Preterit	**.59**	40.4	1134
	Present	.48	30.3	1650
	Subjunctive	**.57**	38	108
	Range	*14*		
PERSON	1stp sg *yo*	**.56**	39.7	2134
	3rdp sg *él/ella*	.44	28.8	1863
	Range	*8*		
ORIGIN	Urban	.49	33.6	2404
	Rural	**.51**	36.3	1593
	Range	*2*		
GENDER	Men	[.50]	34	1892
	Women	[.50]	35.2	2105
AGE	+50	[.50]	34.8	1951
	-50	[.50]	34.5	2046

The effect of discourse-related factors of the first versus the second person and the first versus the third person were also calculated. Table 6.20. offers the results for the first of these calculations. *Distance from previous mention* is ranked first in the factor group hierarchy (Range=17), followed by *Person* (Range=19). As in the case of the full singular paradigm, language-external factor groups are not selected as significant:

Table 6.20. Dominican Spanish. Null versus pronominal and lexical subjects: patterns of variation in discourse-related factor groups, first and second person singular.

Corrected mean				0.312
Log likelihood				-1607.207
Significance				0.008
Total N				2656
***App*. Null Subject**		***Factor weight***	***%***	***N***
DISTANCE FROM PREVIOUS MENTION	0 clauses	**.60**	47.5	1557
	1 clause	.47	26.7	397
	2 clauses	.33	23	161
	3 clauses or more	.32	12.5	697
	Range	*27*		
PERSON	1stp sg *yo*	**.56**	39.7	2134
	2ndp sg [+spec] *tú*	.26	10.9	274
	2ndp sg [-spec] *tú*	.37	22.9	404
	Range	*19*		
CO-REFERENCE	Co-referent	**.56**	45.8	1664
	Non-coreferent	.41	18	1148
	Range	*15*		
TMA OF PREVIOUS VERB	Same TMA	**.53**	37.5	1381
	Different TMA	.47	31.5	1431
	Range	*6*		
DISCOURSE FUNCTION	Topic	[.50]	35.5	2641
	Shift	[.56]	18.1	171
ORIGIN	Urban	[.48]	31.3	1697
	Rural	[.52]	39.2	1115
AGE	+50	[.51]	36.2	1337
	-50	[.49]	32.9	1475
GENDER	Men	[.50]	33.2	1327
	Women	[.50]	35.6	1486

The results of calculating the effects of discourse-related factors on the first and third persons in the singular are presented in Table 6.21. Note that *Person* does not reach significance in the factor group hierarchy in this analysis. Interestingly, in this analysis the most significant variable is *Discourse function* (Range=23), while in previous analysis it was always overridden by *Distance from previous mention* (cf. Table 6.15) or was not significant (Table 6.20). *Distance*, *Co-reference* and *TMA of the previous verb* follow *Discourse function*, and language-external factor groups are not selected as significant:

Table 6.21. Dominican Spanish. Pronominal and lexical subjects versus null subjects: patterns of variation in discourse-related factor groups, first and third person singular.				
Corrected mean				0.311
Log likelihood				-2263.816
Significance				0.001
Total N				3997
***App*. Null Subject**		***Factor weight***	***%***	***N***
DISCOURSE FUNCTION	Topic	**.55**	40.4	3201
	Shift	.32	11.4	796
	Range	*23*		
DISTANCE FROM PREVIOUS MENTION	0 clauses	**.60**	49.6	2054
	1 clause	.50	28.9	492
	2 clauses	.49	27.6	221
	3 clauses or more	.38	13.3	1230
	Range	*22*		
CO-REFERENCE	Co-referent	**.58**	48.8	2133
	Non-coreferent	.40	18.5	1864
	Range	*18*		
TMA OF PREVIOUS VERB	Same TMA	**.53**	37.9	1986
	Different TMA	.47	31.4	2011
	Range	*6*		
PERSON	1stp sg *yo*	**.51**	28.8	1863
	3rdp sg *él/ella*	.49	39.7	2134
	Range	*2*		
ORIGIN	Urban	.[49]	33.6	2404
	Rural	.[51]	36.3	1593
AGE	+50	[.50]	34	1892
	-50	[.50]	34.5	2046
GENDER	Men	[.50]	34	1892
	Women	[.50]	35.2	2105

The results in Table 6.21 confirm that the differing discursive properties of the third person compared with first and second person have an impact on the factor groups that emerge as significant: specifically, *Discourse function* becomes particularly relevant for third person singular subjects; not being a deictic person, whether the referent is available in the immediate discourse or not becomes a decisive constraint on subject choice.

Results of the effect of both *TMA* and discourse-related factors for the first and second person singular are offered in Table 6.22:

Table 6.22. Dominican Spanish. Null versus pronominal and lexical subjects: patterns of variation in TMA and discourse-related factor groups, first and second person singular.				
Corrected mean				0.290
Log likelihood				-1522.128
Significance				0.007
Total N				2812
***App*. Null Subject**		***Factor weight***	***%***	***N***
DISTANCE FROM PREVIOUS MENTION	0	**.61**	47.8	1557
	1	.47	26.7	397
	2	.42	23	161
	3	.30	12.5	697
	Range	*31*		
TMA	Imperfect	.41	28	553
	Preterit	**.63**	45.5	629
	Present	.47	26.9	1369
	Subjunctive	**.61**	31	87
	Range	*20*		
PERSON	first	**.55**	39.7	2134
	second [+specific]	.29	10.9	274
	second [-specific]	.38	22.5	404
	Range	*18*		
CO-REFERENCE	Co-referent	**.55**	45.8	1664
	Non-co-referent	.42	18	1148
	Range	*13*		
TMA PREVIOUS	Same TMA	**.55**	37.5	1381
	Different TMA	.45	31.5	980
	Range	*10*		
ORIGIN	Urban	48	31.3	1697
	Rural	**.54**	39.2	1115
	Range	*6*		
DISCOURSE FUNCTION	Topic	[.55]	35.5	2641
	Shift	[.50]	18.1	171
GENDER	Male	[.51]	33.2	1327
	Female	[.49]	35.6	1485
AGE	+50	[.50]	33.2	1327
	-50	[.50]	32.9	1475

Note that *Person* (Range=18) is ranked third in this analysis, following *Distance from previous mention* (Range=31) and *TMA* (Range=20). *Co-reference* (Range=13) and *TMA of the previous verb* (Range=10) are ranked below *Person*, and among language-external factor groups *Origin* (Range=6) is selected as significant, with rural speakers favoring null subjects (.54) and urban speakers disfavoring them (.48).

Different patterns emerge when first and third persons are analyzed, as observed in Table 6.23:

Table 6.23. Dominican Spanish. Null versus pronominal and lexical subjects: patterns of variation in *TMA* and discourse-related factor groups, first and third person singular.

Corrected mean				0.303
Log likelihood				-2216.657
Significance				0.018
Total N				3997
***App*. Null Subject**		***Factor weight***	***%***	***N***
TOPIC	Topic	**.54**	40.4	3201
	Shift	.33	11.4	796
	Range	*21*		
TMA	Imperfect	.41	26.2	941
	Preterit	**.59**	40.4	1134
	Present	.48	30.3	1650
	Subjunctive	**.61**	38	62
	Range	*20*		
CO-REFERENCE	Co-referent	**.58**	48.8	2133
	Non-co-referent	.40	18.5	1864
	Range	*18*		
DISTANCE FROM PREVIOUS MENTION	0	**.60**	49.6	205
	1	.49	28.9	492
	2	.47	27.6	221
	3	.45	13.3	1230
	Range	*15*		
TMA PREVIOUS	Same TMA	[.54]	37.9	1986
	Different TMA	[.46]	31.4	2011
PERSON	first person yo	[.49]	39.7	2134
	third person él/ella	[.51]	28.8	1863
ORIGIN	Urban	[.49]	33.6	2404
	Rural	[.51]	36.3	1593
AGE	+50	[.49]	34.8	1951
	-50	[.50]	34.5	2046
GENDER	Men	[.49]	34	1892
	Women	[.50]	35.2	2105

In Table 6.23, similarly to what emerged from Table 6.21. and departing from the results of analyzing the first and second persons, presented in Table 6.22, *Discourse function* (Range=21) is the highest ranked factor group, followed by *TMA* (Range=20), *Co-reference* (Range=18) and *Distance from previous mention* (15). Also like in Table 6.21, *Person* is not selected as significant.

A summary of the similarities and differences between first-second and first-third persons is presented in Table 6.24. Ranges for each factor group appear in parenthesis:

Table 6.24. Dominican Spanish. TMA and discourse-related factor groups: variable ranking in first and second, first and third person singular.

TMA	
1st and 2nd sg	***1st and 3rd sg***
1.Person (18)	1.TMA (14)
2.TMA (18)	2.Person (8)
3.Origin (7)	3.Origin (2)
4.Age	4.Age
5.Gender	5.Gender
DISCOURSE-RELATED FACTOR GROUPS	
1st and 2nd sg	***1st and 3rd sg***
1.Distance (27)	1.Discourse function (23)
2.Person (19)	2.Distance (22)
3.Co-reference (15)	3.Co-reference (18)
4.TMA previous (6)	4.TMA previous (6)
5.Discourse function	5.Person
6.Origin	6.Origin
7.Age	7.Age
8.Gender	8.Gender
TMA AND DISCOURSE-RELATED FACTOR GROUPS	
1st and 2nd sg	***1st and 3rd sg***
1.Distance (31)	1.Discourse function (21)
2.TMA (20)	2.TMA (20)
3.Person (18)	3.Co-reference (18)
4.Co-reference (13)	4.Distance (15)
5.TMA previous (10)	5.TMA previous
6.Origin (6)	6.Person
7.Discourse function	7.Origin
8.Gender	8.Age
9.Age	9.Gender

Factor group hierarchies corresponding to the effects of TMA are similar across first-second persons and first-third analyses in that Origin is selected as significant in both data sets. *Person* is ranked first in the hierarchy with first-second person data set and second in the first-third person data set, which suggests that first and third person pattern similarly while first and second person differ. This result does not support the *Functional Compensation Hypothesis*, according to which second and third persons should be conditioned by the same factors. Factor group hierarchies corresponding to the effect of discourse-related factor groups are homogeneous in both data sets with respect to the effects of language-external factor groups, which are not selected as significant in either calculation. Data sets differ in the strength displayed by *Discourse function* and *Person*: while *Person* is not significant for first and third persons, it is ranked second with first and second

persons; conversely, while *Discourse function* is ranked first with first and third persons, it is not significant with first and second persons. This result further confirms that first and second person do not behave similarly. In addition, it reflects the different discursive properties of the first and second persons versus the third, in that *Discourse function* arises as significant in the data set that includes the third person singular, which is potentially anaphoric and reference tracking is not straightforward. This difference is further confirmed in the logistic regression analyses that calculate the joint effects of *TMA* and discourse-related factor groups; furthermore, these analyses confirm that no significant differences arise between the first and the third person. Finally, *Origin* is selected as significant for the first and second person data set, while none of the language-external factor groups are significant in the first and third person data set.

Note that as in the logistic regression analyses that calculated the effect of all linguistic factor groups, discourse-related constraints have emerged across Tables 6.16-6.24 as very significant for the distribution of null and overt subjects. These findings are in line with previous work, in which discourse-related variables have been claimed to regulate null and overt subject insertion in all Spanish dialects (Bentivoglio 1987, Silva-Corvalán 1994, Otheguy et al. 2007, Travis 2005, 2007, de Prada 2009). As pointed out in § 6.2.1, interrogative clauses differ from non-interrogative clauses in this respect. The factor group *Interrogative* could not be included in logistic regression analyses due to the small number of interrogative clauses features in the corpus gathered for this study (N=190). In order to assess the differences between interrogative and non-interrogative clauses, *Interrogative* was cross-tabulated with a number of factor groups, and chi-squares of the differences between interrogatives and non-interrogatives were calculated. Results are presented in the next section.

6.2.4. Interrogative and non-interrogative clauses: Differences.

As outlined in §6.2.1, the overall rate of overt subject use for non-interrogative clauses in the data set is 61% (N=3488), and null subjects amounted to 39% (N=2322). These percentages

significantly shift for interrogative clauses, which display an overt subject rate of 86% (N=164) and a null subject rate of 14% (N=26).

As the results of cross-tabulations presented in Tables 6.25-6.28 show, the effects of discourse-related factors is not perceived in interrogative clauses. Table 6.25 offers the results of cross-tabulating *Interrogative* and *Distance from previous mention*. The difference between interrogative and non-interrogative clauses is significant, as shown by the *p*-value (=0).

Table 6.25. Dominican Spanish. Cross-tabulation of *Interrogative* and *Distance from previous mention*. (N=6005)

		0 CLAUSES		1 CLAUSE		2 CLAUSES		3 OR MORE CLAUSES	
		N	***%***	***N***	***%***	***N***	***%***	***N***	***%***
NON-INTERROGATIVE	***Null***	1492	53	228	33	108	35	494	26
(N=5815)	***Overt***	1320	47	462	67	205	65	1501	74
INTERROGATIVE	***Null***	9	23	4	27	1	6	12	17
(N=190)	***Overt***	30	77	11	73	16	94	107	83

chi-square=42.81
p = 0

The results of cross-tabulating *Interrogative* and *Co-reference* are presented in Table 6.26. Note how differences between co-referent and non-co-referent subjects are very small in interrogative clauses. Furthermore, the difference between interrogative and non-interrogative is statistically significant, with a *p*-value of 0.0006:

Table 6.26. Dominican Spanish. Cross-tabulation of *Interrogative* and *Co-reference* (N=6005).

		CO-REFERENT		NON-CO-REFERENT	
		N	***%***	***N***	***%***
NON-INTERROGATIVE	***Null***	1537	52	785	27
(N=5815)	***Overt***	1401	48	2087	73
INTERROGATIVE	***Null***	10	19	16	12
(N=190)	***Overt***	44	81	120	88

chi-square=11.652
p=0.0006

Rates of overt subject use are higher in interrogative and non-interrogative clauses in both *Topic* and *Shift* contexts, as illustrated by the cross-tabulation of *Interrogative* and *Discourse Function*. However, this difference does not arise as statistically significant, as observed in Table 6.27:

Table 6.27. Dominican Spanish. Cross-tabulation of *Interrogative* and *Discourse Function*. (N=6005)

		TOPIC		SHIFT	
		N	***%***	***N***	***%***
NON-INTERROG (N=5815)	***Null***	1974	44	348	26
	Overt	2493	56	995	74
INTERROGATIVE (N=190)	***Null***	20	14	6	13
	Overt	125	86	39	87
chi-square=1.738 *p*=.187					

The difference between interrogative and non-interrogative clauses is not statistically significant either when *TMA of the previous verb* is considered:

Table 6.28. Dominican Spanish. Cross-tabulation of *Interrogative* and *TMA of the previous verb*. (N=6005)

		DIFFERENT TMA		SAME TMA	
		N	***%***	***N***	***%***
NON-INTERROG (N=5810)	***Null***	1064	37	1258	42
	Overt	1780	63	1708	58
INTERROGATIVE (N=190)	***Null***	13	12	13	16
	Overt	96	88	68	84
chi-square=3.53 *p*=0.06					

In sum, the effects of discourse-related factor groups are not as strong in interrogative clauses, whereas they have been shown to significantly constrain subject choice in non-interrogative clauses.

Interrogative was also cross-tabulated with *Person*, in order to explore whether significant differences in overt subject rates are observed among persons of the paradigm. With the exception of first person plural, overt subject rates are significantly higher with interrogative clauses, as expressed in the *p*-value=0:

Table 6.29. Dominican Spanish. Cross-tabulation of *Interrogative* and *Person*. (N=6005)					
		NON-INTERROGATIVE		INTERROGATIVE	
		Null	***Overt***	***Null***	***Overt***
1STP SG YO	***N***	839	1263	9	23
	%	40	60	28	72
2NDP SG [+SPEC] TÚ	***N***	27	150	3	94
	%	15	85	3	97
2NDP SG [-SPEC] TÚ	***N***	90	303	1	10
	%	23	77	9	91
3RDP SG ÉL/ELLA	***N***	530	1304	7	22
	%	29	71	24	76
1STP PL NOSOTROS/NOSOTRAS	***N***	351	126	3	0
	%	74	26	100	0
3RDP PL [+SPEC] ELLOS/ELLAS	***N***	294	299	1	7
	%	50	50	12	88
chi-square=741.013 *p*=0					

The cross-tabulation of *Interrogative* and *Clause Type* shows that both main and embedded questions significantly differ from main and embedded non-interrogative clauses:

Table 6.30. Dominican Spanish. Cross-tabulation of *Interrogative* and *Clause Type*.					
		NON-INTERROGATIVE		INTERROGATIVE	
		Null	***Overt***	***Null***	***Overt***
MAIN	***N***	1265	2154	15	121
	%	37	63	11	89
EMBEDDED	***N***	648	947	11	24
	%	41	59	35	76
chi-square=12.926 *p*=0.0003					

Interrogative was also cross-tabulated with the language-external variables *Origin*, *Gender*, *Age* and *Social class*. Results are presented in Tables 6.31-6.34. The results in Table 6.31 show that the differences between urban and rural speakers are statistically significant:

Table 6.31. Dominican Spanish. Cross-tabulation of *Interrogative* and *Origin*. (N=6005)					
		NON-INTERROGATIVE		INTERROGATIVE	
		Null	***Overt***	***Null***	***Overt***
URBAN	***N***	1469	2181	19	130
	%	40	60	13	87
RURAL	***N***	853	1307	7	34
	%	39	61	17	83
chi-square=18.889 *p*=0.00001					

A different picture emerges from the comparison of men and women. Results for the cross-tabulation of *Interrogative* and *Gender* show that differences among men and women are not statistically significant:

Table 6.32. Dominican Spanish. Cross-tabulation of *Interrogative* and *Gender*. (N=6005)					
		NON-INTERROGATIVE		INTERROGATIVE	
		Null	***Overt***	***Null***	***Overt***
MEN	***N***	1229	1757	11	77
	%	41	59	39	61
WOMEN	***N***	1033	1731	15	87
	%	39	61	15	85
chi-square=0.733 *p*=0.392					

The differences between speakers over 50 and below 50 years old are not statistically significant either:

Table 6.33. Dominican Spanish. Cross-tabulation of *Interrogative* and *Age*. (N=6005)					
		NON-INTERROGATIVE		INTERROGATIVE	
		Null	***Overt***	***Null***	***Overt***
+50	***N***	1113	1685	13	70
	%	40	60	16	84
-50	***N***	1209	1803	13	94
	%	40	60	14	86
chi-square=1.986 *p*=0.159					

However, when *Interrogative* and *Social class* are cross-tabulated, the differences between high-middle and working classes are significant:

Table 6.34. Dominican Spanish. Cross-tabulation of *Interrogative* and *Class*. (N=6005)					
		NON-INTERROGATIVE		INTERROGATIVE	
		Null	***Overt***	***Null***	***Overt***
HIGH-MIDDLE CLASS	***N***	857	1378	13	94
	%	38	62	12	88
WORKING CLASS	***N***	1465	2110	13	70
	%	41	59	16	84
chi-square=20.65 *p*=0					

In sum, discourse-related factor groups do not have in interrogative clauses the strong effects they have in non-interrogative clauses. As far as *Person* is concerned, rates of overt subject use are higher in interrogative clauses than in non-interrogative clauses in all persons of the paradigm, except in the firs person plural. Furthermore, interrogative and non-interrogative clauses significantly differ in both main and embedded questions. Regarding language-external factor groups, similarly to what was the case with non-interrogative clauses, *Origin* and *Social class* differences are significant, while *Gender* and *Age* differences are not statistically significant.

Having presented the results of examining the null and overt distribution of lexical and pronominal subjects, I now move on to discuss the findings obtained of examining the patterns exhibited of pronominal versus null subjects. In order to draw accurate comparisons between the two data sets, parallel logistic regression analysis and cross-tabulations were carried out with the pronominal subject corpus.

6.3. Null and overt subject patterns: Overt pronominal subjects versus null subjects.

The overall rate for pronominal subjects in the Dominican data set gathered for the present study is 51.4%. This rate, while closer to the overall rates reported for other Caribbean varieties (45% for Puerto Rican Spanish, Cameron (1992), 48% in Travis (2007) for Coastal Colombian Spanish) than the overall rates of pronominal and lexical subjects considered jointly, still situates DS as the Caribbean variety with the highest percentages of overt subject use. Rates for the

different persons of the paradigm are presented in Table 6.35. In this table, these percentages are compared with the ones reported by Cameron (1992) for the Spanish of San Juan, Puerto Rico:

Table 6.35. Dominican Spanish. Comparison with Person overall rates San Juan, Puerto Rico (Cameron 1992: 233).

SINGULAR		
FIRST PERSON: YO		
	N overt subjects	*%*
San Juan	949	50
Dominican Republic	1259	59.8
SECOND PERSON: TÚ [+SPECIFIC]		
	N overt subjects	*%*
San Juan	145	48
Dominican Republic	235	88.7
SECOND PERSON: TÚ [-SPECIFIC]		
	N overt subjects	*%*
San Juan	188	69
Dominican Republic	193	78.1
THIRD PERSON: ÉL/ELLA		
	N overt subjects	*%*
San Juan	443	39
Dominican Republic	427	44.3
THIRD PERSON: UNO (-SPECIFIC)		
	N overt subjects	*%*
San Juan	39	85
Dominican Republic	119	76.3
PLURAL		
FIRST PERSON: NOSOTROS/NOSOTRAS		
	N overt subjects	*%*
San Juan	220	15
Dominican Republic	105	22.9
SECOND PERSON: USTEDES		
	N overt subjects	*%*
San Juan	-	-
Dominican Republic	24	61.5
THIRD PERSON: ELLOS/ELLAS		
	N overt subjects	*%*
San Juan	138	25
Dominican Republic	103	25.9
THIRD PERSON: ELLOS (-SPECIFIC)		
	N overt subjects	*%*
San Juan	-	-
Dominican Republic	17	8.7

As illustrated in Table 6.35, rates of overt subject use are higher in DS than in Puerto Rican Spanish for every person in the paradigm except third person singular *uno*. Furthermore,

percentages are very close to the ones obtained when both pronominal and lexical subjects were counted. Differences emerge with respect to the third person [+specific], with a 71.2% when pronominal and lexical subjects are examined and 44.3% when only pronominal subjects are counted. Similarly, third person plural [+specific] bears an overt subject rate of 50.9% when pronominal and lexical subjects are considered, and a 25.9% when only pronominal subjects are counted. The second person plural displays very similar rates in both data sets (62.5% and 61.5%), and the third person plural [-specific] has the lowest rates of overt expression in both data sets as well (12.7% and 8.7%).

The results of a logistic regression analysis that include linguistic factor groups in addition to the language-external variables *Origin*, *Gender* and *Age* are presented in Table 6.36. The re-coding of factor groups carried out to allow for logistic regression calculations with Goldvarb is identical to the one presented in §6.2.1 for the analyses in which pronominal and lexical subjects were examined, with the exception of the factor group *Distance from previous mention*. In the analyses of pronominal subjects, constraints within this factor group were reduced to *0 clauses*, *1 clause*, and *2 or more clauses* in order to avoid interactions. The application value for the dependent variable null or overt subject is pronominal subject in all the tables that follow:

Table 6.36. Dominican Spanish. Pronominal subjects versus null subjects: Patterns of variation.

Corrected mean				0.532
Log likelihood				-2657.984
Significance				0.017
Total N				4830
***App:* Pronominal subject**		***Factor weight***	***%***	***N***
PERSON				
	1stp sg *yo*	**.59**	59.8	2107
	2ndp sg [±specific] *tú*	**.82**	81.9	668
	3rdp sg *él/ella*	.44	44.3	964
	1stp pl nosotros/as	.17	22.9	459
	3rdp pl ellos/ellas	.19	20.2	593
	Range	*63*		
CLAUSE TYPE	Main	**.57**	54.3	2801
	Conjoined	.35	32.8	494
	Embedded	.48	50.2	1324
	Relative	**.64**	63.5	211
	Range	*29*		
DISCOURSE FUNCTION	Topic	**.52**	53.2	4265
	Shift	.33	37.3	565
	Range	*19*		
CO-REFERENCE	Co-referent	.43	43.4	2734
	Non-co-referent	**.60**	61.8	2096
	Range	*17*		
DISTANCE FROM	0	.43	42.8	2623
PREVIOUS MENTION	1	**.56**	63.8	640
	2	**.60**	60.8	1567
	Range	*17*		
VERB	Stative	**.60**	65.2	869
	Transitive	**.52**	56.4	780
	Inergative	**.58**	59.4	256
	Unaccusative	**.54**	52.2	291
	'Ir'	.47	46.3	283
	Reflexive	.43	47.6	389
	Other	.43	43.8	1497
	Range	*17*		
TMA	Imperfect	**.57**	58.2	1012
	Preterite	.42	41.6	1209
	Present	**.51**	58.1	2191
	Range	*7*		
TMA OF PREVIOUS VERB	Same TMA	.47	48.5	2467
	Different TMA	**.53**	54.4	2363
	Range	*6*		
ORIGIN	Urban	[.49]	51.1	3042
	Rural	[.51]	51.9	1788
GENDER	Men	[.50]	49.6	2458
	Women	[.50]	53.3	2372
AGE	+50	[.50]	50.5	2277
	-50	[.50]	52.1	2553

The results in Table 6.36 are discussed in the following section.

6.3.1. Overt pronominal subjects versus null subjects: Significant factor groups, factor group ranking and constraint ranking.

In Table 6.36, *Person* (Range=63) is the highest factor group in the hierarchy, followed by *Clause Type* (Range=29), similarly to what was the case with pronominal and lexical subjects. A further similarity in the strength of linguistic factor groups is found in *TMA of the previous verb*, which is the least significant of the discourse-related factor groups in both data sets. Similarities also arise regarding language-external factor groups; *Origin*, *Gender* and *Age* are not selected as significant in either calculation. The rest of factor groups differ in their situation in the hierarchy of significance, as observed in Table 6.37:

Table 6.37. Dominican Spanish. Factor group ranking in lexical and pronominal subjects (N=6005) versus null subjects and pronominal subjects versus null subjects (N=4830).

LEXICAL AND PRONOMINAL SUBJECTS VERSUS NULL SUBJECTS		PRONOMINAL SUBJECTS VERSUS NULL SUBJECTS	
Factor group	***Range***	***Factor group***	***Range***
Person	69	*Person*	63
Clause Type	27	*Clause Type*	29
Distance from previous mention	22	*Discourse function*	19
Verb	17	*Co-reference*	17
Co-reference	16	*Distance from previous mention*	17
TMA	14	*Verb*	17
Discourse function	12	*TMA*	7
TMA of the previous verb	4	*TMA of the previous verb*	6
Age	2	*Origin*	2
Gender	2	*Gender*	0
Origin	1	*Age*	0

In both data sets discourse-related factor groups override morpho-syntactic ones in ranking. However, while *Distance from the previous mention* is the highest ranked discourse-related factor group for pronominal and lexical subjects, when only pronominals are counted *Distance* is ranked below *Discourse function* and *Co-reference*. In fact, except for *TMA of the previous verb*, all discourse-related factor groups override morpho-syntactic constraints in the pronominal data set. This result replicates the findings in previous work (Cameron 1996, Otheguy et al. 2007), and has

usually been interpreted as evidence supporting the claim that contextual ambiguity is more relevant than morphological ambiguity for subject expression in Spanish. While this might well be the case, this result could also be interpreted as a consequence of including only pronominal subjects in the analysis. Lexical subjects are maximally non-ambiguous as compared with pronominal subjects, and hence when the speaker's intention is to clearly distinguish the referent of an utterance from other potential referents, (s)he will employ a lexical subject. On the contrary, ambiguity among potential referents arises when pronominal subjects are used, and therefore the relevance of discourse-related factors strictly related with reference tracking like *Discourse function* or *Co-reference*. For pronominal and lexical subjects, *Distance from previous mention*, a discourse-related factor group not only associated with reference tracking but also with processing or memory constraints, arises as the most relevant factor group. Note also constraint rankings within *Discourse function* in the pronominal data set: overt subjects are favored in *Topic* contexts, while disfavored in *Shift* environments. This might arise at first sight as a counter-intuitive result; however, in light of differences between lexical and pronominal subjects, this is not so; the high rates of overt subject use attested in DS entail that overt pronominals are favored in *Topic* contexts. These pronominals are disfavored in *Shift* contexts probably because in this environment lexical subjects are selected by the speakers.

Morpho-syntactic factor groups are selected as significant in both data sets. Interestingly, *Verb* arises as more significant than *TMA* for subject choice for pronominals and DPs considered jointly and for pronominals only.

The next relevant question is whether DS patterns with other Caribbean varieties in factor group ranking. A comparison with the rankings reported in Otheguy et al. (2007) in their study of Caribbean newcomers in New York City and the ones offered in Orozco and Guy's (2008) investigation of the Spanish of Barranquilla, Colombia, is provided in Table 6.38[54]. *Person* is the highest ranked factor group in all three studies, which suggests that as noted in Otheguy et al.

[54] Comparisons have been established with these two studies because the factor groups included in the analysis approximate the ones selected in the present study.

(2007), Caribbean and Non-Caribbean varieties differ in the strength attributed to this factor group. In Non-Caribbean varieties, even if significant differences also arise between the singular and plural paradigms, discourse-related factor groups such as *Discourse function* usually override *Person*:

Table 6.38. Comparison of factor group rankings: The present study, Otheguy et al. (2007) and Orozco and Guy (2008)

THE PRESENT STUDY: DOMINICAN REPUBLIC.	OTHEGUY ET AL. (2007): CARIBBEAN NEWCOMERS IN NYC.	OROZCO AND GUY (2008): BARRANQUILLA, COLOMBIA.
1.Person	1.Person	1.Person
2.Clause Type	2.Speech connectedness	2.TMA
3.Discourse function	3.TMA	3.Clause Type
4. Co-reference	4.Clause Type	4.Semantic verb type
5.Distance from prev. mention	5.Semantic verb type	5.Speech connectedness
6.Verb	6.Genre	6. Genre
7. TMA	7.Picture	7. Reflexive
8. TMA of the previous verb	8.Reflexive	8. Gender
9.Origin	9.Specific	9.Age
10.Gender		
11.Age		

It can also be observed in Table 6.38 that *Clause Type* is among the top five variables in the hierarchy in all three studies as well. Differences in the strength of this factor group might be the consequence of the coding decisions made in each of these studies: while Otheguy et al. only considered main and embedded clauses, Orozco and Guy distinguished main, embedded, relative and conjoined clauses, along the lines of the present study, which results in a higher ranking of *Clause Type*. Constraint rankings within *Clause Type* in Orozco and Guy's study and this investigation are very similar as well: main clauses favor overt subjects in both studies, at factor weights of .55 in Orozco and Guy and .57 here; conjoined clauses disfavor them at .40 in Orozco and Guy and at .35 in this investigation[55]; relative clauses strongly favor pronominal subjects in both works, with an identical factor weight of .64.

As far as the ranking differences and similarities between morpho-syntactic and discourse-related factor groups is concerned, one of the latter override the former in Otheguy et al. and this dissertation, while *TMA* is ranked over *Speech connectedness*, the only discourse-related factor

[55] Subordinate and conjoined clauses are merged in Orozco and Guy (2008) for logistic regression analyses.

group considered in Orozco and Guy. All three studies converge in finding *Verb* as a significant factor group, even though an all-semantic classification of verbs is carried out in Otheguy et al. and Orozco and Guy and a classification that takes into account semantic and syntactic properties has been performed in this study.

In sum, in spite of shifts in ranking, the results for the comparison of overt pronominal and null subjects obtained in this study are along the lines the findings in previous work on Caribbean varieties: *Person, Clause Type*, one or another discourse-related factor group, *Verb* and *TMA* arise as the strongest factor groups for Caribbean subject expression.

Going back to the comparison of the results obtained in this study for pronominal and lexical versus null subjects and only pronominal versus null subjects, while factor groups rankings differ for the two data sets, as discussed above, constraint rankings within factor groups follow the same trends in both corpora, as illustrated in Table 6.39. Bear in mind for the comparison of factor weights that the application value for the first set of calculations is null subject, whereas the application value for the second set is pronominal subject:

Table 6.39. Dominican Spanish. Constraint rankings in null versus lexical and pronominal subjects (N=6005) and pronominal subjects versus null subjects (N=4830).

Factor group		LEXICAL AND PRONOMINAL VERSUS NULL SUBJECTS **App. Null subject**	PRONOMINAL VERSUS NULL SUBJECTS **App. pronominal subject**
PERSON	1stp sg	.44	.59
	2nd sg [+spec]	.19	.82[56]
	2ndp sg [-spec]	.25	-
	3rdp sg	.42	.44
	1stp pl	.88	.17
	3rdp pl	.77	.19
Range		*69*	*63*
CLAUSE TYPE	Main	.47	.57
	Conjoined	.67	.35
	Embedded	.52	.48
	Relative	.40	.64
Range		*27*	*29*
DISTANCE	0 clauses	.60	.43
	1 clause	.47	.56
	2 clauses	.47	.60
	3 clauses	.38	-
Range		*22*	*17*
CO-REFERENCE	Co-referent	.58	.43
	Non-co-ref.	.42	.60
Range		*16*	*17*
DISC. FUNC.	Topic	.53	.52
	Shift	.41	.33
Range		*12*	*19*
VERB	Stative	.40	.60
	Transitive	.51	.52
	Inergative	.45	.58
	Unaccusative	.44	.54
	Reflexive	.53	.43
Range		*17*	*17*
TMA	Imperfect	.44	.57
	Preterit	.58	.42
	Present	.49	.51
Range		*14*	*7*
TMA PREV.	Same TMA	.52	.47
	Different TMA	.48	.53
Range		*4*	*6*

[56] Recall that second persons [+/-specific] were merged for the logistic regression analyses of pronominals.

Having presented the results of the logistic regression analyses that include all speakers, I now move on to discuss the results of examining the language-external factors *Social class* and *Education*.

6.3.2. Language-external variables: *Social class* and *Education*.

Further re-coding of factor groups was required to carry out logistic regression analysis of the distribution of pronominal subjects in Santo Domingo. *Clause Type* had to excluded from the calculation, and within *Verb*, unaccusative and inergative verbs had to be left out of the analysis as well.

In order to compare the distribution of null and overt subjects with the calculations carried out in 6.3.1, *Clause Type* was cross-tabulated with *Person*. Results are presented in 6.40:

Table 6.40. Urban Dominican Spanish. Cross-tabulation of *Person* and *Clause Type*: Pronominal subjects versus null subjects (N=4830).

		1ST PSG		2ND PSG		3RD PSG		1ST PPL		3RD PPL		TOTAL	
		N	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*
MAIN	*Overt*	447	98	209	98	153	29	35	83	33	29	877	65
	Null	7	2	4	2	369	71	7	17	80	71	467	35
CONJOINED	*Overt*	42	100	12	100	27	44	3	100	2	40	86	70
	Null	0	0	0	0	34	56	0	0	3	60	37	30
EMBEDDED	*Overt*	166	98	148	99	66	32	15	68	14	33	409	69
	Null	3	2	2	1	140	68	7	32	29	67	181	31
RELATIVE	*Overt*	45	98	26	100	8	33	8	80	3	33	90	78
	Null	1	2	0	0	16	67	2	20	6	67	25	22
TOTAL	*Overt*	700	98	395	99	254	31	61	79	52	31	1462	67
	Null	11	2	6	1	559	69	16	21	118	69	710	33

If these results are compared with the ones obtained for lexical and pronominal subjects (cf. Table 6.12.), relevant differences emerge: in both corpora second person singular displays the highest rates of overt subject use in all clause types, followed by the first person singular. However, while the third person singular also displayed high rates of overt subject use when lexical and pronominal subjects were counted together, this is not the picture that emerges from the pronominal data set: the rate of overt subject use for pronominals is 29% for main clauses, 44% for conjoined clauses, 32% for embedded clauses and 33% for relative clauses, versus versus 74%, 51%, 70% and 73% in the lexical and pronominal subjects data set. Conversely, while rates of overt subject use for

the first person plural were low in the pronominal and lexical subjects corpus as compared with the rates obtained for the singular paradigm, this trend is reversed in the pronominal corpus: rates for pronominal and lexical subjects are 22% for main clauses, 10% for conjoined clauses, and 23% and 45% for embedded and relative clauses respectively; rates for pronominal subjects are 83% for matrix clauses, 100% for conjoined ones, and 68% and 80% for embedded and relative clauses respectively. These differences show that among third person singular subjects several occurrences of overt subjects correspond to lexical subjects, while most instances of the first person plural correspond to overt pronominal subjects. This constitutes an interesting result: on the one hand, from the perspective of the discursive properties of the different persons of the paradigm, it suggests that lexical subjects are employed in higher numbers in the person of the paradigm for which reference tracking is less straightforward, i.e., the third person. On the other, from the perspective of the comparison of null and overt subject patterns in lexical versus pronominal subjects, these results suggest that overt subject rates are higher in the pronominal data set for all non-third persons.

The difference between third and non-third persons that has emerged from the cross-tabulation of *Person* and *Clause Type* in Table 6.41 is also attested in the differing factor weights displayed by these persons when logistic regression analyses are performed. Results are offered in Table 6.41:

Table 6.41. Dominican Spanish. *Social class* and *Education* in Santo Domingo. Patterns of variation for overt pronominal subjects versus null subjects.

Corrected mean				0.790
Log likelihood				-508.351
Significance				0.017
Total N				1787
		Factor weight	***%***	***N***
PERSON	1stp sg *yo*	**.93**	98.5	711
	3rdp sg *él/ella*	.12	31.2	813
	1stp pl *nosotros/as*	**.63**	79.2	77
	3rdp pl *ellos/ellas*	.17	30.6	170
	Range	*81*		
DISCOURSE FUNCTION	Topic	**.68**	81.7	1238
	Shift	.15	12.9	549
	Range	*53*		
VERB	Stative	.43	51.3	472
	Transitive	**.64**	78	282
	Inergative	**.68**	75.3	93
	Unaccusative	.48	49.3	148
	Reflexive	.34	57.4	155
	Other	**.51**	63	392
	Range	*34*		
DISTANCE FROM PREVIOUS MENTION	0	**.56**	76.7	703
	1	**.69**	78.5	214
	2	.40	43.1	870
	Range	*29*		
CO-REFERENCE	Co-referent	**.59**	78.5	698
	Non-co-referent	.44	49	1089
	Range	*15*		
SOCIAL CLASS	High-middle class	[.47]	57.3	968
	Working class	[.54]	64.3	819
GENDER	Men	[.47]	60.1	1024
	Women	[.53]	61.2	763
EDUCATION	University	[.52]	59.4	1208
	High school	[.45]	62.9	579
AGE	+50	[.46]	59	746
	-50	[.53]	61.7	1041
TMA	Imperfect	[.54]	62.1	472
	Preterite	[.48]	57.4	460
	Present	[.48]	62.1	791
TMA PREVIOUS	Same TMA	[.49]	59.8	816
	Different TMA	[.51]	61.2	971

A further difference between the results presented in Table 6.41. and the results for the distribution of pronominal and lexical subjects in Santo Domingo speakers offered in Table 6.6. above is that while *Social class* was selected as significant in Table 6.6, it arises as non-significant in Table 6.41. This result suggests that speakers of different social classes are homogeneous with respect to the rates of use of overt pronominal subjects.

Results for the distribution of pronominal subjects across urban and rural working classes are presented in Table 6.42. The main difference among urban speakers and working class speakers in urban and rural communities arises with respect to the strength of discourse-related and morpho-syntactic factors: *Distance from previous mention* overrides *Discourse function* in working classes, while *Discourse function* is the highest ranked discourse-related factor group in urban speakers, immediately followed by *Verb*. Note that similarly to what emerged with respect to *Verb* in the pronominal and lexical subject data set, differences among verb types are not as strong in working class speakers as in urban speakers:

Table 6.42. Dominican Spanish. Working class and Santo Domingo and El Cibao. Patterns of variation for overt pronominal subjects versus null subjects.

Corrected mean				0.525
Log likelihood				-1803.273
Significance				0.007
Total N				3139
		Factor weight	***%***	***N***
PERSON	1stp sg *yo*	**.57**	56.7	1494
	[+spec] *tú*	**.85**	87.8	147
	[-spec] *tú*	**.77**	73.6	212
	3rdp sg *él/ella*	.46	45.5	616
	1stp pl *nosotros/as*	.18	24.3	280
	3rdp pl *ellos/ellas*	.25	26.5	374
	Range	*67*		
CLAUSE TYPE	Main	**.52**	53.1	1856
	Conjoined	.36	34.3	315
	Embedded	.48	50.1	853
	Relative	**.62**	61.7	115
	Range	*26*		
DISTANCE FROM PREVIOUS MENTION	0	.42	41.5	1744
	1	**.58**	65.1	433
	2	**.60**	63.3	180
	Range	*22*		
DISCOURSE FUNCTION	Topic	**.52**	52.2	2805
	Shift	.31	37.7	334
	Range	*21*		
CO-REFERENCE	Co-referent	.43	41.8	1824
	Non-co-referent	**.60**	63	1315
	Range	*17*		
VERB	Stative	**.59**	62.2	548
	Transitive	.49	52.3	516
	Inergative	**.60**	59.7	186
	Unaccusative	**.56**	54.7	179
	Reflexive	.46	46.7	184
	Other	.45	43.8	1047
	Range	*15*		
TMA	Imperfect	**.57**	57.7	639
	Preterite	.43	43.4	751
	Present	**.51**	55.5	1473
	Range	*8*		
TMA OF THE PREVIOUS VERB	Same TMA	.47	47.7	1639
	Different TMA	**.53**	54	1500
	Range	*6*		
ORIGIN	Urban	[.48]	49.1	1351
	Rural	[.51]	51.9	1788
GENDER	Men	[.51]	493	1562
	Women	[.49]	52.1	1577
AGE	+50	[.50]	51.1	1392
	-50	[.50]	50.3	1747

As in section 6.2.3. above, constraint rankings for *Verb*, discourse-related factors and *TMA* are compared in order to determine in which factor groups differences are attested among the three communities examined. Table 6.43 provides these comparisons for *Verb*:

Table 6.43. Dominican Spanish. *Verb* Patterns of variability: Urban and Rural speakers and Social Class. Pronominal versus null subjects

	All speakers (N=4830)			**Urban speakers (N=1787)**			**Working class in urban and rural communities (N=3139)**		
Verb	***F.W.***	***%***	***N***	***F.W.***	***%***	***N***	***F.W.***	***%***	***N***
STATIVE	.60	65.2	869	.43	51.3	472	.58	62.2	548
TRANSITIVE	.52	56.4	780	.64	78	282	.48	52.3	516
INERGATIVE	.58	59.4	256	.68	75.3	93	.59	59.7	186
UNACCUSATIVE	.54	52.2	291	.48	49.3	148	.56	54.7	179
REFLEXIVE	.47	46.3	389	.34	57.4	155	.46	46.7	184
OTHER	.43	43.8	1497	.51	63	392	.44	43.8	1047
Range	*17*			*34*			*15*		

Inergative verbs strongly favor overt pronominal subjects, while reflexive verbs disfavor them in all three communities. Urban speakers depart from all and working class speakers in the effects of stative and unaccusative verbs: all speakers and working classes favor overt subjects with statives, in line with the results obtained in all the logistic regression analyses carried out so far, while urban speakers disfavor them at .43, which constitutes a surprising result. Similarly, while all and working class speakers favor overt subjects with unaccusative verbs, urban speakers disfavor them at .48. In sum, the pattern found in the pronominal and lexical data set, according to which the more argumental structure a verb has the more it disfavors overt subjects does is not so straightforwardly observed in the pronominal corpus.

Table 6.44 offers a comparison of discourse-related factor groups across communities. Recall that with the exception of *Distance from previous mention*, all these factors were very similar in all the social groups examined in the lexical and pronominal data set. When only pronominal subjects are counted, differences across communities are to some extent similar to the differences found in previous calculations: all three communities display very similar factor weights in *TMA of the previous verb*, the lowest ranked discourse-related factor group in Tables 6.36, 6.41 and 6.42.

All three communities pattern with respect to Discourse function as well, with the difference that this factor group is stronger in urban speakers (Range=52) when compared with all speakers (Range=19) and working classes (Range=22). All speakers and working class speakers pattern as well with respect to *Distance from previous mention* and *Co-reference*, while the trends for these two factor groups are reversed in urban speakers:

Table 6.44. Dominican Spanish. Patterns of variability, discourse-related factor groups: Urban and Rural speakers and Social Class: Pronominal versus null subjects.

	All speakers (N=4830)			**Urban speakers (N=1787)**			**Working class in urban and rural communities (N=3139)**		
Distance	***F.W.***	***%***	***N***	***F.W.***	***%***	***N***	***F.W.***	***%***	***N***
0 CLAUSES	.43	42.8	2623	.56	76.7	703	.42	41.5	1744
1 CLAUSE	.56	63.8	640	.69	78.5	214	.58	65.1	433
2 CLAUSES	.60	60.8	1567	.40	43.1	870	.58	63.3	180
3 OR MORE	-	-	-	-	-	-	.61	60.2	782
Range	*17*			*29*			*19*		
Co-reference									
CO-REF	.43	43.4	2734	.58	78.5	698	.43	41.8	1824
NON-CO-REF	.60	61.8	2096	.44	49	1089	.60	63	1315
Range	*17*			*14*			*17*		
Discourse function									
TOPIC	.52	53.2	4265	.68	81.7	1238	.52	52.2	1805
SHIFT	.33	37.3	565	.16	12.9	549	.30	37.7	334
Range	*19*			*52*			*22*		
TMA of the previous verb									
SAME TMA	.47	48.5	2467	.51	59.8	816	.47	47.7	1639
DIFFERENT TMA	.53	54.4	2363	.49	61.2	971	.53	54	1500
Range	*6*			*2*			*6*		

In sum, similarly to what emerged with respect to the lexical and pronominal subject data set, urban speakers depart in patterns of distribution from the other two communities examined. This similarity between the two corpora extends to the patterns of distribution found for TMA, offered in Table 6.45. Note that differences across communities are smaller in this factor group: all three communities display practically identical factor weights in the imperfect tense; all speakers and working class speakers are also identical with respect to the factor weights of preterit and

present, while these two tenses bear very similar factor weights (.49 and .48 respectively) in urban speakers:

Table 6.45. Dominican Spanish. TMA Patterns of variability: Urban and Rural speakers and Social Class. Pronominal versus null subjects.

	Urban and Rural speakers (N=4830)			Urban speakers (N=1787)			Working class in urban and rural communities (N=3139)		
TMA	***F.W.***	***%***	***N***	***F.W.***	***%***	***N***	***F.W.***	***%***	***N***
IMPERFECT	.57	58.2	1012	.54	62.1	472	.57	57.7	639
PRETERIT	.42	41.6	1209	.49	57.4	460	.43	43.4	751
PRESENT	.51	58.1	2191	.48	62.1	791	.51	55.5	1473
Range	*7*			*6*			*8*		

The differences among lexical and pronominal and only pronominal subjects in the fine-grained analysis of the relative role of contextual and morphological ambiguity are discussed in the following section.

6.3.3. On the relative role of contextual and morphological ambiguity.

The effects of TMA on the pronominal singular paradigm are offered in Table 6.46:

Table 6.46. Dominican Spanish. Pronominal subjects versus null subjects, singular paradigm: *TMA* patterns of variation.

Corrected mean				0.626
Log likelihood				-2312.305
Significance				0.001
Total N				3739
		Factor weight	**%**	**N**
PERSON	1stp sg *yo*	.49	59.8	2107
	2ndp sg [+specific] *tú*	**.83**	88.7	265
	2ndp sg [-specific] *tú*	**.68**	77.4	403
	3rdp sg *él/ella*	.33	44.3	964
	Range	*35*		
TMA	Imperfect	**.59**	67.3	791
	Preterite	.41	49.6	917
	Present	**.52**	66.8	1703
	Subjunctive	.38	55.7	115
	Range	*21*		
ORIGIN	Urban	**.51**	61.5	2290
	Rural	.48	56.9	1449
	Range	*3*		
AGE	+50	[.49]	58.1	1777
	-50	[.51]	59.7	1507
GENDER	Men	[.50]	60.2	1772
	Women	[.50]	59.3	1967

The results in Table 6.46 totally parallel the ones obtained for the pronominal and lexical subjects corpus: *Person* is ranked over *TMA*, which suggests that relevant differences arise within the singular paradigm with respect to the effects of morpho-syntactic factor groups, and the language-external *Origin* is selected as significant (cf. Table 6.15 above). As illustrated in Table 6.47, results for discourse-related factor groups also fully parallel the ones obtained for the lexical and pronominal subject data set: factor group rankings are identical, and language-external factor groups are not selected as significant:

Table 6.47. Dominican Spanish. Pronominal subjects versus null subjects, singular paradigm: Discourse- related factors patterns of variation.

Corrected mean				0.621
Log likelihood				-2236.141
Significance				0.002
Total N				3739
		Factor weight	**%**	**N**
PERSON	1p sg *yo*	.49	59.8	3488
	2psg [+specific] *tú*	**.79**	88.7	265
	2p sg [-specific] *tú*	**.70**	77.4	403
	3p sg *él/ella*	.34	44.3	964
	Range	*30*		
DISTANCE	0	.43	48.3	2161
FROM PREVIOUS MENTION	1	**.55**	70.8	518
	2	**.62**	72.2	223
	Range	*19*	79.0	837
DISCOURSE FUNCTION	Topic	**.51**	59.7	3488
	Shift	.33	59.4	251
	Range	*18*		
CO-REFERENCE	Co-referent	.43	49.2	2246
	Non-co-referent	**.60**	75.6	1493
	Range	*17*		
TMA OF PREVIOUS VERB	Same TMA	.47	63.3	1858
	Different TMA	**.53**	56.1	1881
	Range	*6*		
ORIGIN	Urban	[.51]		
	Rural	[.49]		
GENDER	Male	[.49]		
	Female	[.50]		
AGE	+50	[.49]		
	-50	[.50]		

Similarities between data sets do not extend to the analyses in which both *TMA* and discourse-related factors are examined. While the latter were ranked over *TMA* in the lexical and pronominal subject data set, *TMA* emerges as more significant in the pronominal corpus:

Table 6.48. Dominican Spanish. Pronominal subjects versus null subjects, singular paradigm: *TMA* and discourse-related factor groups.

Corrected mean				0.634
Log likelihood				-2154.049
Significance				0.000
Total N				3739
		Factor weight	***%***	***N***
PERSON	1st p sg *yo*	.49	59.8	2107
	2nd p sg [+spec] *tú*	**.80**	88.7	265
	2nd p sg [-spec] *tú*	**.70**	77.4	403
	3rd p sg *él/ella*	.34	44.3	964
	Range	*46*		
TMA	Imperfect	**.60**	67.3	791
	Preterit	.40	49.6	917
	Present	**.52**	66.8	1703
	Subjunctive	.33	55.7	115
	Range	*27*		
DISTANCE FROM PREVIOUS MENTION	0	.42	48.3	2161
	1	**.55**	70.8	518
	2	**.58**	72.2	223
	3	**.64**	79	837
	Range	*22*		
DISCOURSE FUNCTION	Topic	**.51**	59.7	3488
	Shift	.32	59.4	251
	Range	*19*		
CO-REFERENCE	Co-referent	.43	49.2	2246
	Non-co-referent	**.60**	75.6	1493
	Range	*17*		
TMA PREVIOUS	Same TMA	.46	56.1	1881
	Different TMA	**.54**	63.3	1858
	Range	*8*		
ORIGIN	Urban	[.51]	61.5	2290
	Rural	[.48]	56.9	1449
GENDER	Men	[.49]	60.2	1772
	Women	[.51]	59.3	1967
AGE	+50	[.49]	58.1	1777
	-50	[.51]	61.2	1962

Table 6.49. offers the full-fledged comparison of the factor group rankings just discussed:

Table 6.49. Dominican Spanish.Comparison of variable rankings in TMA and discourse-related factors. Pronominal versus null subjects, all singular paradigm.

PRONOMINAL SUBJECTS AND TMA	
PRONOMINAL AND LEXICAL VS NULL SUBJECTS	PRONOMINAL VS NULL SUBJECTS
1. Person (36)	1. Person (35)
2. TMA (16)	2. TMA (21)
3. Origin (4)	3. Origin (3)
4. Age	4. Age
5. Gender	5. Gender
PRONOMINAL SUBJECTS AND DISCOURSE-RELATED FACTOR GROUPS	
PRONOMINAL AND LEXICAL VS NULL SUBJECTS	PRONOMINAL VS NULL SUBJECTS
1. Person (28)	1. Person (30)
2. Distance from previous mention (27)	2. Distance from previous mention (19)
3. Discourse function (20)	3. Discourse function (19)
4. Co-reference (19)	4. Co-reference (17)
5. TMA of the previous verb (3)	5. TMA of the previous verb (6)
6. Origin	6. Origin
7. Age	7. Gender
8. Gender	8. Age
PRONOMINAL SUBJECTS, TMA AND DISCOURSE-RELATED FACTOR GROUPS	
PRONOMINAL AND LEXICAL VS NULL SUBJECTS	PRONOMINAL VS NULL SUBJECTS
1. Person (28)	1. Person (46)
2. Distance from previous mention (27)	2. TMA (27)
3. Discourse function (20)	3. Distance from previous mention (22)
4. Co-reference (17)	4. Discourse function (19)
5. TMA of the previous verb (17)	5. Co-reference (17)
6. TMA (17)	6. TMA of the previous verb (8)
7. Origin	7. Origin
8. Age	8. Gender
9. Gender	9. Age

Patterns of distribution of pronominal subjects were also examined for first and person singular, on the one hand, and first and third person singular, on the other. As illustrated in Table 6.50, factor group rankings are identical in the pronominal and lexical subjects and pronominal subjects data sets. The only difference emerges with respect to the strength of *Person* and *TMA*, which displayed identical ranges in the pronominal and lexical subjects corpus (Range=18), whereas *Person* is a stronger factor group in the pronominal corpus:

Table 6.50. Dominican Spanish. Pronominal versus null subjects: First and second person singular *TMA* patterns of variation.

Corrected mean				0.684
Log likelihood				-1656.690
Significance				0.000
Total N				2775
		Factor weight	***%***	***N***
PERSON	1st p sg *yo*	.44	59.8	265
	2nd p sg [+spec] *tú*	**.78**	88.7	403
	2nd p sg [-spec] *tú*	**.62**	77.4	2107
	Range	*18*		
TMA	Imperfect	**.57**	71.5	544
	Preterit	.39	53.7	618
	Present	**.53**	72.8	1354
	Subjunctive	.44	68.6	86
	Range	*13*		
ORIGIN	Urban	**.53**	68.3	1680
	Rural	.46	60.1	1095
	Range	*7*		
AGE	+50	[.49]	63.2	1315
	-50	[.51]	66.8	1460
GENDER	Men	[.50]	66.2	1306
	Women	[.50]	64.1	1469

Table 6.51. offers the parallel logistic regression analysis for first and third person singular. Once more, the examination of pronominal and lexical subjects and only pronominal subjects yield identical results (cf. Table 6.19 above):

Table 6.51. Dominican Spanish. Pronominal versus null subjects: First and third person singular *TMA* patterns of variation.

Corrected mean				0.564
Log likelihood				-2035.306
Significance				0.007
Total N				3071
		Factor weight	***%***	***N***
TMA	Imperfect	**.60**	66.1	729
	Preterit	.41	48.4	888
	Present	.50	58.9	1216
	Subjunctive	.40	46.8	77
	Range	*20*		
PERSON	1st p sg *yo*	**.55**	59.8	2107
	3rd p sg *él/ella*	.39	44.3	964
	Range	*16*		
ORIGIN	Urban	**.51**	55.2	1802
	Rural	.49	54.5	1269
	Range	*2*		
GENDER	Men	[.49]	54.5	1412
	Women	[.50]	55.3	1659
AGE	+50	[.50]	54.6	1494
	-50	[.50]	55.2	1577

Table 6.52. includes the results of analyzing the effects of discourse-related factor groups on first and second person pronouns. Results parallel the ones obtained for the pronominal and lexical subjects corpus, except with respect to *Discourse function*. This factor group was not selected as significant in 6.2.3. (cf. Table 6.20), while it is the third ranked factor group in Table 6.52. This result reflects the relevance of *Discourse function* with pronominal subjects, also found when all linguistic factors were analyzed together (cf. Table 6.36):

Table 6.52. Dominican Spanish. Pronominal subjects versus null subjects: Patterns of variation in discourse-related factor groups, first and second person singular.

Corrected mean				0.684
Log likelihood				-1590.713
Significance				0.018
Total N				2775
		Factor weight	***%***	***N***
DISTANCE FROM PREVIOUS MENTION	0	.39	52.2	1546
	1	**.53**	73.2	395
	2	**.56**	76.9	160
	3	**.70**	87.1	674
	Range	*31*		
PERSON	1st p sg yo	.44	59.8	2107
	2nd p sg [+spec] tú	**.74**	88.7	265
	2nd p sg [-spec] tú	**.64**	77.4	403
	Range	*30*		
DISCOURSE FUNCTION	Topic	***.51***	64.3	2628
	Shift	.36	78.9	147
	Range	*15*		
CO-REFERENCE	Co-referent	.44	53.9	1654
	Non-co-referent	**.59**	81.5	1121
	Range	*15*		
TMA OF THE PREVIOUS VERB	Same TMA	.47	62	1364
	Different TMA	**.53**	68	1411
	Range	*6*		
ORIGIN	Urban	[.52]	68.3	1680
	Rural	[.47]	60.1	1095
AGE	+50	[.49]	63.2	1315
	-50	[.51]	66.8	1460
GENDER	Men	[.50]	66.2	1306
	Women	[.50]	64.1	1469

The results in Table 6.53 show that for first and third person singular, results are totally parallel in the two data sets examined (cf. Table 6.21. above). The only difference emerges in *Person*, which was not significant for pronominal and lexical subjects, and it emerges as significant in the pronominal subject data set:

Table 6.53. Dominican Spanish. Pronominal subjects versus null subjects: Patterns of variation in discourse-related factor groups, first and third person singular.

Corrected mean				0.684
Log likelihood				-1590.713
Significance				0.018
Total N				2775
		Factor weight	***%***	***N***
DISCOURSE FUNCTION	Topic	**.51**	55.7	2920
	Shift	.28	39.7	151
	Range	*23*		
DISTANCE FROM PREVIOUS MENTION	0	.44	45.7	1877
	1	**.56**	67.9	443
	2	**.55**	67.9	190
	3	**.62**	70.9	561
	Range	*18*		
CO-REFERENCE	Co-referent	.44	46.1	1931
	Non-co-referent	**.60**	69.8	1140
	Range	*16*		
PERSON	1stp sg *yo*	**.54**	59.8	2107
	3rdp sg *él/ella*	.40	44.3	964
	Range	*14*		
TMA OF THE PREVIOUS VERB	Same TMA	.46	50.7	1527
	Different TMA	**.53**	59.1	1544
	Range	*7*		
ORIGIN	Urban	[.50]	55.2	1802
	Rural	[.49]	54.5	1269
GENDER	Men	[.49]	54.5	1412
	Women	[.50]	55.3	1659
AGE	+50	[.50]	54.6	1494
	-50	[.50]	55.2	1577

The joint effects of *TMA* and discourse-related factor groups in the distribution of pronominal subjects were also examined. If the results in Table 6.54. are compared with the same logistic regression analysis carried out with the pronominal and lexical data set (cf. Table 6.22. above), it can be observed that no differences are attested between corpora:

Table 6.54. Dominican Spanish. Pronominal subjects versus null subjects: Patterns of variation in TMA and discourse-related factor groups, first and second person singular.

Corrected mean				0.705
Log likelihood				-1523.289
Significance				0.011
Total N				2775
		Factor weight	***%***	***N***
DISTANCE FROM PREVIOUS MENTION	0	.39	52.2	47.8
	1	**.54**	73.2	395
	2	**.67**	85.1	834
	Range	*28*		
TMA	Imperfect	**.59**	71.5	544
	Preterite	.36	53.7	618
	Present	**.53**	72.8	1354
	Range	*23*		
PERSON	1st sg yo	.44	59.8	2107
	2nd sg tú	**.73**	88.7	265
	2nd sg tú	**.62**	77.4	403
	Range	*18*		
CO-REFERENCE	Co-referent	.45	53.9	1654
	Non-co-referent	.58	81.5	1121
	Range	*13*		
TMA OF THE PREVIOUS VERB	Same TMA	.45	62	1364
	Different TMA	**.54**	68	1411
	Range	*9*		
ORIGIN	Urban	**.52**	68.3	1680
	Rural	.47	60.1	1095
	Range	*5*		
GENDER	Men	[.49]	66.2	1306
	Women	[.50]	64.1	1469
AGE	+50	[.50]	63.2	1315
	-50	[.50]	66.8	1460

The similarities do not extend for the comparison of first and third person singular across data sets. For pronominal subjects, *TMA* is the highest ranked factor group, whereas Discourse function had a stronger effect on pronominal and lexical subjects considered together (cf. Table 6.23 above). Note that this same difference emerged when comparing the effects on *TMA* on both data sets. This result could be interpreted as evidence that *TMA* becomes a more relevant factor group on subject expression in the pronominal data set because pronouns are not as referentially non-ambiguous as lexical subjects, hence the relevance of inflectional morphology in the identification of the subject by the speakers:

Table 6.55. Dominican Spanish. Pronominal versus null subjects: Patterns of variation in TMA and discourse-related factor groups, first and third person singular.

Corrected mean				0.565
Log likelihood				-1901.460
Significance				0.018
Total N				2775
		Factor weight	***%***	***N***
TMA	Imperfect	**.61**	66.1	729
	Preterite	.40	48.4	888
	Present	**.57**	58.9	1216
	Subjunctive	.35	46.8	77
	Range	*26*		
DISCOURSE FUNCTION	Topic	**.51**	55.7	2920
	Shift	.27	39.7	151
	Range	*24*		
DISTANCE FROM PREVIOUS MENTION	0	.44	46.1	1931
	1	**.56**	67.9	443
	2	**.57**	67.9	190
	3	**.63**	70.9	561
	Range	*19*		
PERSON	1st sg *yo*	**.54**	59.8	2107
	3rd sg *él/ella*	.40	443	964
	Range	*14*		
TMA OF THE PREVIOUS VERB	Same TMA	.46	50.7	1527
	Different TMA	**.54**	59.1	1544
	Range	*8*		
GENDER	Men	[.49]	54.5	1412
	Women	[.51]	55.3	1659
ORIGIN	Urban	[.51]	55.2	1802
	Rural	[.50]	54.5	1269
AGE	+50	[.50]	54.6	1494
	-50	[.50]	55.2	1577

The differences and similarities across data sets associated with the relative role of morphological and contextual ambiguity just discussed are summarized in Table 6.56:

Table 6.56. Dominican Spanish.Comparison of variable rankings in TMA and discourse-related factors. Pronominal versus null subjects, all singular paradigm.

Pronominal and lexical vs null subjects	*Pronominal vs null subjects*
PRONOMINAL SUBJECTS AND TMA	
1STP AND 2NDP SG	1STP AND 2NDP SG
1. Person (18)	1. Person (18)
2. TMA (18)	2. TMA (13)
3. Origin (7)	3. Origin (7)
4. Age (2)	4. Age (2)
5. Gender (0)	5. Gender (2)
3RDP AND 1STP SG	3RDP AND 1STP SG
1. TMA (14)	1. TMA (20)
2. Person (8)	2. Person (16)
3. Origin (2)	3. Origin (2)
4. Age (0)	4. Gender (1)
5. Gender (0)	5. Age (0)
PRONOMINAL SUBJECTS AND DISCOURSE-RELATED FACTOR GROUPS	
1STP AND 2NDP SG	1STP AND 2NDP SG
1. Distance (27)	1. Distance from previous mention (31)
2. Person (19)	2. Person (30)
3. Co-reference (15)	3. Discourse function (15)
4. TMA of the previous verb (6)	4. Co-reference (15)
5. Discourse function (6)	5. TMA of the previous verb (6)
6. Origin (4)	6. Origin (5)
7. Age (2)	7. Age (2)
8. Gender (0)	8. Gender (0)
3RDP AND 1STP SG	3RDP AND 1STP SG
1. Discourse function (23)	1. Discourse function (23)
2. Distance from previous mention (22)	2. Distance from previous mention (18)
3. Co-reference (18)	3. Co-reference (16)
4. TMA of the previous verb (6)	4. Person (14)
5. Person (2)	5. TMA of the previous verb (7)
6. Origin (2)	6. Origin (1)
7. Age (0)	7. Gender (1)
8. Gender (0)	8. Age (0)
PRONOMINAL SUBJECTS, TMA AND DISCOURSE-RELATED FACTOR GROUPS	
1STP AND 2NDP SG	1STP AND 2NDP SG
1. Distance from previous mention (32)	1. Distance from previous mention (28)
2. TMA (20)	2. TMA (23)
3. Person (18)	3. Person (18)
4. Co-reference (13)	4. Co-reference (13)
5. TMA of the previous verb (10)	5. TMA of the previous verb (9)
6. Origin (5)	6. Origin (5)
7. Discourse function (5)	7. Gender (1)
8. Gender (2)	8. Age (0)
9. Age (0)	
3RDP AND 1STP SG	3RDP AND 1STP SG
1. Discourse function (21)	1. TMA (26)
2. TMA (20)	2. Discourse function (24)
3. Co-reference (18)	3. Distance from previous mention (19)
4. Distance from previous mention (15)	4. Person (14)
5. TMA of the previous verb (8)	5. TMA of the previous verb (8)
6. Person (2)	6. Gender (2)
7. Origin (2)	7. Origin (0)
8. Age (2)	8. Age (0)
Gender (1)	

As made clear in the discussion carried out in this section, morpho-syntactic and discourse-related factor groups behave very similarly in the pronominal and lexical subjects and the pronominal subjects data sets. The differences that emerge have been interpreted as deriving mainly from the distinct properties displayed by lexical and pronominal subjects. In the following section differences between corpora are explored in interrogative clauses.

6.3.4. Interrogative and non-interrogative clauses: Distribution of pronominal and null subjects.

Interrogative was cross-tabulated with a number of factor groups in order to assess whether significant differences are found between interrogative and non-interrogative clauses. First, cross-tabulations of Interrogative and discourse-related factor groups were performed. Table 6.57. provides the results of cross-tabulating *Interrogative* and *Distance from previous mention*:

Table 6.57. Dominican Spanish. Cross-tabulation of *Interrogative* and *Distance from previous mention*, pronominal subjects (N=4830).

		0 CLAUSES		1 CLAUSE		2 CLAUSES		3 CLAUSES OR MORE	
		N	***%***	***N***	***%***	***N***	***%***	***N***	***%***
NON-INTERROGATIVE	***Null***	1492	58	228	36	108	39	494	42
	Overt	1093	42	398	64	168	61	673	58
INTERROGATIVE	***Null***	9	24	4	29	1	6	12	11
	Overt	29	76	10	71	15	94	96	89

chi-square= 90.578
p=0

As attested by the *p*-value, the difference between interrogative and non-interrogative clauses is statistically significant, reproducing the pattern found for pronominal and lexical subjects (cf. Table 6.25). Similarities in statistical significance in data sets extend to the cross-tabulation of *Interrogative* and *Co-reference*:

Table 6.58. Dominican Spanish. Cross-tabulation of *Interrogative* and *Co-reference*, pronominal subjects (N=4830).

		CO-REFERENT		NON-CO-REFERENT	
		N	***%***	***N***	***%***
NON-INTERROGATIVE	***Null***	1537	57	785	40
	Overt	1148	43	1184	60
INTERROGATIVE	***Null***	10	20	16	13
	Overt	39	80	111	87
chi-square= 30.474 *p*=0					

As illustrated in Table 6.59, results differ across data sets with respect to the cross-tabulation of *Interrogative* and *Topic*, for which the difference between interrogative and non-interrogative clauses is statistically significant for pronominal subjects, whereas it was not for pronominal and lexical subjects (p=.187, cf Table 6.27.):

Table 6.59. Dominican Spanish. Cross-tabulation of *Interrogative* and *Discourse function*, pronominal subjects (N=4830).

		TOPIC		SHIFT	
		N	***%***	***N***	***%***
NON-INTERROGATIVE	***Null***	1974	48	181	34
	Overt	2151	52	348	66
INTERROGATIVE	***Null***	20	14	6	17
	Overt	120	86	30	83
chi-square=4.268 *p*=0.039					

Results pattern across corpora in the cross-tabulation of *Interrogative* and *TMA of the previous verb*; differences between interrogative clauses are not significant in either data set (cf. p=0.006 in Table corpus (p=0.22), while they were significant for the pronominal and lexical subjects data set (p=0.06, Table 6.28):

Table 6.60. Dominican Spanish. Cross-tabulation of *Interrogative* and *TMA of the previous verb*, pronominal subjects (N=4830).

		DIFFERENT TMA		SAME TMA	
		N	***%***	***N***	***%***
NON-INTERROGATIVE	***Null***	1064	47	1258	53
	Overt	1201	53	1131	47
INTERROGATIVE	***Null***	13	13	13	17
	Overt	85	87	65	83

chi-square=1.506
p=0.22

To sum up, similarly to what was the case with pronominal and lexical subjects, within the pronominal data set the effects of discourse-related factor groups significantly constrain subject choice in non-interrogative clauses, whereas their effects are not as strong in interrogative clauses.

The results of cross-tabulating *Interrogative* and *Person* are provided in Table 6.61.

Table 6.61. Dominican Spanish. Cross-tabulation of *Interrogative* and *Person*, pronominal subjects (N=4830).

		NON-INTERROGATIVE		INTERROGATIVE	
		Null	***Overt***	***Null***	***Overt***
1STP SG *yo*	***N***	830	1236	2	23
	%	40	60	28	72
2NDP SG [+SPEC] *tú*	***N***	27	16	3	93
	%	142	84	3	97
2NDP SG [-SPEC] *tú*	***N***	53	184	1	9
	%	22	78	10	90
3RDP SG [+SPEC] *él/ella*	***N***	530	414	7	13
	%	56	44	35	65
1STP PL *nosotros/as*	***N***	351	105	3	0
	%	77	23	100	0
3RDP PL [+SPEC] *ellos/ellas*	***N***	100	294	1	3
	%	25	75	25	75

chi-square=1336.126
p=0

As observed in Table 6.61, the difference between interrogative and non-interrogative clauses is statistically significant. This was also the finding for the pronominal and lexical subjects data set. Similarly, the differences between interrogative and non-interrogative clauses arise as significant when *Interrogative* and *Clause Type* are cross-tabulated:

Table 6.62. Dominican Spanish. Crosstabulation of *Interrogative* and *Clause Type*, pronominal subjects (N=4830).

		MAIN		EMBEDDED	
		N	***%***	***N***	***%***
INTERROGATIVE	***Null***	1265	47	648	51
	Overt	1407	53	635	49
NON-INTERROGATIVE	***Null***	15	12	11	27
	Overt	114	88	30	73

chi-square=6.695
p=0.01

Finally, *Interrogative* was cross-tabulated with the language-external factors *Origin*, *Gender*, *Age* and *Social class*. As illustrated in the results in Tables 6.63-6.66, results replicate in all four cases the findings obtained when Interrogative and language-external factor groups were cross-tabulated with the pronominal and lexical data set. Differences are only significant with *Origin* and *Social class*:

Table 6.63. Dominican Spanish. Cross-tabulation of *Interrogative* and *Origin*, pronominal subjects (N=4830).

		URBAN		RURAL	
		N	***%***	***N***	***%***
NON-INTERROGATIVE	***Null***	1461	51	853	49
	Overt	1435	49	897	51
INTERROGATIVE	***Null***	19	14	7	18
	Overt	119	86	31	82

chi-square=19.075
p=0

Table 6.64. Dominican Spanish. Cross-tabulation of *Interrogative* and *Gender*, pronominal subjects (N=4830).

		MEN		WOMEN	
		N	***%***	***N***	***%***
NON-INTERROGATIVE	***Null***	1129	52	1093	48
	Overt	1149	48	1183	52
INTERROGATIVE	***Null***	11	14	15	16
	Overt	69	86	81	84

chi-square=0.603
p=0.437

Table 6.65. Dominican Spanish. Cross-tabulation of *Interrogative* and *Age*, pronominal subjects (N=4830).

		+50		-50	
		N	***%***	***N***	***%***
NON-INTERROGATIVE	***Null***	1113	51	1209	49
	Overt	1086	49	1246	51
INTERROGATIVE	***Null***	13	17	13	13
	Overt	65	83	85	87

chi-square=0.594
p=0.44

Table 6.66. Dominican Spanish. Cross-tabulation of *Interrogative* and *Social class*, pronominal subject (N=4830).

		HIGH-MIDDLE CASS		WORKING CLASS	
		N	***%***	***N***	***%***
NON-INTERROGATIVE	***Null***	857	49	1465	50
	Overt	883	51	1449	50
INTERROGATIVE	***Null***	69	51	65	83
	Overt	11	49	13	17

chi-square=67.999
p=0

The results that emerged from analyzing the patterns of distribution of lexical versus null subjects are presented in the following section.

6.4. Null and overt subject patterns: Lexical subjects versus null subjects.

The results of examining the distribution of lexical versus null subjects when all linguistic and extra-linguistic factor groups are considered appear in Table 6.67:

Table 6.67. Dominican Spanish. Lexical subjects versus null subjects: Patterns of variation.				
Corrected mean				0.539
Log likelihood				-947.567
Significance				0.020
Total N				2121
App: Lexical subject		*Factor weight*	*%*	*N*
PERSON	3rdp sg *él/ella*	**.64**	62.6	1436
	3rdp pl *ellos/ellas*	.22	30.9	685
	Range	42		
DISCOURSE FUNCTION	Topic	.31	29.5	1099
	Shift	**.70**	77	1022
	Range	*39*		
DISTANCE FROM	0	.35	26	811
PREVIOUS MENTION	1	.46	42.5	147
	2	.42	42.2	87
	3	**.63**	74.4	1046
	Range	*28*		
CLAUSE TYPE	Main	**.55**	57.5	1246
	Conjoined	.29	28.8	212
	Embedded	.48	51.3	587
	Relative	.38	43.4	76
	Range	*23*		
CO-REFERENCE	Co-referent	.38	28.2	865
	Non-co-referent	**.58**	69	1256
	Range	20		
ANIMACY	Animate	.45	42.9	1542
	Inanimate	**.62**	77.5	579
	Range	*17*		
TMA	Imperfect	**.54**	58.9	448
	Preterite	.46	48.3	563
	Present	.50	55.8	963
	Range	*8*		
GENDER	Men	**.53**	54.8	1160
	Women	.46	50.3	961
	Range	*7*		
AGE	+50	**.53**	53.8	1052
	-50	.47	51	1069
	Range	*6*		
ORIGIN	Urban	[.51]	52.2	1382
	Rural	[.49]	52.6	739
TMA OF THE PREVIOUS VERB	Same TMA	[.50]	49.3	1127
	Different TMA	[.50]	55.8	994

6.4.1. Variables, variable ranking and constraint ranking

The highest ranked factor group in Table 6.67 is *Person* (Range=42). Within *Person*, third person singular strongly favors lexical subjects at .64, while third person plural disfavors them at

.22. First and second persons were not considered in this analysis, given the very few occurrences of lexical subjects with these persons. Since first and second persons are deictic in nature and therefore always present in the universe of the discourse by default, it follows that pronominal subjects are favored with these two persons. By the same reasoning, it also follows that third persons, which are potentially anaphoric and therefore their reference is not readily available to speakers strongly favor lexical subjects.

Discourse function (Range=39) is ranked second in the factor group hierarchy. Note that constraint rankings for this factor group are reversed in Table 6.67 with respect to the constraint rankings displayed in the analysis of the pronominal data set: lexical subjects are strongly favored in *Shift* contexts with a factor weight .70, while disfavored in *Topic* contexts at .31. In Table 6.36, pronominal subjects were favored in *Topic* contexts at .52, while disfavored in *Shift* environments at .33. These factor weights neatly complement each other; they show that when a referent is new in the discourse, speakers resource to lexical subjects, while when the referent is available in the previous speech, pronominal subjects are preferred. In a variety like DS in which rates of overt subject expression are very high, pronominal instead of null subjects are chosen in *Topic* contexts.

Next in the factor group ranking is *Distance from previous mention* (Range=28). Constraint rankings within this factor group roughly suggest that the further is the previous mention of a given subject, the more likely is that speakers opt for a lexical subjects.

Clause Type (Range=23) follows *Distance from previous mention* in the factor group ranking. Interesting differences are attested as well among pronominal and lexical subjects: recall that relative clauses strongly favored overt pronominal subjects with a factor weight of .64. This pattern is not replicated by lexical subjects, which disfavor overts at a factor weight of .38. This result, once again, arises as a consequence of the different properties of pronominal and lexical subjects: since relative clauses involve picking up the reference of an element present in the preceding clause (i.e. the antedecedent), it follows that pronominal subjects are preferred in these contexts.

Co-reference (Range=20) follows *Clause Type*. Non-co-referent subjects favor lexical subjects (.58) while co-referent subjects disfavor them (.38).

Animacy (Range=17) and *TMA* (Range=8) are the least significant linguistic factor groups in the hierarchy of significance. Inanimate subjects favor lexical subjects with a factor weight of .62. This result naturally follows from the fact that inanimate subjects do not have a pronominal counterpart in Spanish. Within *TMA*, the imperfect favors overt lexical subjects with a factor weight of .54, and the preterit disfavors them at .46, following the same trends found for the pronominal data set, where imperfects favored overt subjects at .57 and preterites disfavored them at .42. The present does not favor nor disfavors overt subjects in both data sets, with constraint rankings of .51 for pronominal subjects and .50 for lexical subjects.

Regarding language-external factor groups, interesting differences also emerge between pronominal and lexical subjects. No language-external variables were significant when pronominal subjects were examined. On the contrary, *Age* and *Gender* are selected as significant with lexical subjects: speakers over fifty years old favor lexical subjects, whereas younger speakers disfavor them. Men favor lexical subjects as well, while women disfavor them.

Note that the factor group *Verb* was not included in the logistic regression analyses of lexical versus null subjects. In order to examine to some extent whether relevant differences emerge among verb types, *Verb* was cross-tabulated with *Person*; results are presented in Table 6.68. First and second person are included in this crosstabulation:

Table 6.68. Dominican Spanish. Cross-tabulation of Verb and Person, lexical subjects (N=3371).

		1STP SG YO		2NDP SG TÚ		3RDP SG ÉL/ELLA		1STP PL NOSOTROS/AS		3RDP PL ELLOS/ELLAS	
		N	***%***	***N***	***%***	***N***	***%***	***N***	***%***	***N***	***%***
STATIVE	***Lexical***	6	7	4	57	310	73	3	5	49	54
	Null	80	93	3	43	113	27	54	95	42	46
TRANSITIVE	***Lexical***	3	2	0	0	65	54	3	5	15	20
	Null	138	98	5	100	55	46	61	95	61	80
INERGATIVE	***Lexical***	1	3	0	0	25	60	1	5	10	33
	Null	36	97	2	100	17	40	21	95	20	67
UNACCUSATIVE	***Lexical***	1	2	0	0	79	75	1	4	25	56
	Null	62	98	1	100	27	25	26	96	20	44
REFLEXIVE	***Lexical***	1	1	0	0	83	60	4	13	17	41
	Null	83	99	6	100	55	40	26	87	24	59
OTHER	***Lexical***	12	4	0	0	186	53	8	7	70	25
	Null	292	96	12	100	166	47	109	93	215	75

What emerges from the numbers in 6.68 is that it is *Person* and not *Verb* the crucial factor in the distribution of lexical and null subjects. Rates of lexical subject use are higher for the third person than for the first person singular and plural in all verb types.

6.4.2. Null and overt subject patterns in Dominican Spanish: Summary.

The results presented in sections 6.2, 6.3. and 6.4 have aimed to describe the null and overt subject patterns of DS. In light of these findings, subject expression grammar in this dialect has emerged as presenting both similar and different properties from other varieties of Spanish: first, similarly to what has been reported in previous work, discourse-related factors have emerged as very significant for the distribution of null and overt subjects. The differing strengths of specific discourse-related factors has been attributed to the distinct properties of lexical and pronominal subjects, on the one hand, and to the properties ascribed to the different persons of the paradigm. Structural factor groups are also very significant: *Clause Type* is ranked at the top of the factor group hierarchy in the pronominal, lexical, and pronominal and lexical subjects data set. *Animacy* has emerged as a significant factor group for the distribution of lexical subjects. Morpho-syntactic

factors also play a relevant role in subject expression: *Verb* is significant for the pronominal and pronominal-lexical subjects data sets, and *TMA* has an impact on subject expression in all three data sets examined. However, when the relative role of morphological and contextual ambiguity were examined, results showed that generally discourse-related factor groups overrode *TMA*.

Language-external factor groups have been shown to not have the strong effect that linguistic variables have on subject expression. *Gender* and *Age* are generally not selected as significant, except in the lexical subject data set. On the contrary, *Social class* and *Origin* have emerged as significant in some of the analyses performed so far. Specifically, *Origin* is significant when *TMA* and *TMA* and discourse-related factor groups are examined, and *Social class* differences have been proved to be significant for urban Dominican speakers.

6.5. Subject position patterns: On the distribution of preverbal and postverbal subjects.

Results obtained for the analysis of the subject position patterns attested in DS are presented in the following section. As will be clear as findings on the distribution of preverbal and postverbal subjects are presented, differences between null and overt subject patterns and subject position patterns emerge precisely within language-external factor groups.

The decision to explore the distribution of preverbal and postverbal subjects in addition to the examination of null and overt subject patterns stems from one of the main hypotheses put forward in this dissertation (cf. §5.2.3), i.e. that if ongoing language change was taking place in DS, its effects would be more visible in structural properties associated with word order than in null and overt subject patterns.

In the logistic regression analyses that explore the subject position patterns displayed by pronominal and lexical subjects the dependent variable is *Subject Position*. Independent factor groups considered in the analysis include: *Subject form*, which distinguishes overt pronominal and overt lexical subjects;[57] relevant discourse-related factor groups (*Discourse function, Co-reference,*

[57] Null subjects are left out of Goldvarb analyses that explore subject position patterns.

TMA of the previous verb), as it has been noted in theoretical work that reference and focus constraints are relevant for the position of subjects in Spanish (cf. Chapter Two); *Person*, in order to assess whether subject position patters are stable across the singular and plural paradigms; *Interrogative*, since one of the main differences found between Caribbean and Non-Caribbean dialects is the availability of not inverting verb and subject in wh-questions; language-external factor groups (*Origin, Gender, Age, Social class* and *Education*). With the aim of comparing null and overt subject patterns and subject position patterns, Goldvarb analyses were performed for the three communities examined in sections 6.2, and 6.3, i.e. all speakers, only urban speakers, and working classes in urban and rural environments. Results are discussed in the following paragraphs.

Findings associated with subject position patterns in the all-speakers community are offered in Table 6.69:

Table 6.69. Dominican Spanish. Distribution of preverbal (versus postverbal) subjects (N=3344).

Corrected mean				0.951
Log likelihood				-899.873
Significance				0.042
Total N				3344
App: Preverbal subject		***Factor weight***	***%***	***N***
SUBJECT FORM	Pronominal subject	**.70**	97.8	2458
	Lexical subject	.16	75.4	1169
	Range	*54*		
INTERROGATIVE	Non-interrogative	**.51**	90.9	3190
	Interrogative	.26	89.9	154
	Range	*25*		
ORIGIN	Urban	**.55**	91.3	2092
	Rural	.42	87.7	1252
	Range	*13*		
AGE	+50	.44	87.5	1639
	-50	**.56**	92.3	1705
	Range	*12*		
DISCOURSE FUNCTION	Topic	**.53**	95.1	2355
	Shift	.44	77.6	989
	Range	*9*		
PERSON	1stp sg *yo*	[.46]	97	1286
	2ndp sg *tú*	[.52]	95.9	244
	3rd ppl *ellos/ellas*	[.45]	78.9	332
GENDER	Men	[.47]	88.6	1662
	Women	[.53]	91.2	1682
CO-REFERENCE	Co-referent	[.51]	93	1293
	Non-co-referent	[.48]	88	2051
TMA OF THE PREVIOUS VERB	Same TMA	[.50]	89.3	1591
	Different TMA	[.50]	90.5	1753

Note that *Person* does not emerge as a significant factor group in Table 6.69, which suggests that all persons of the paradigm do not differ significantly in subject position patterns.

The highest ranked factor group is *Subject form* (Range=54), with pronominal subjects strongly favoring the preverbal position and lexical subjects disfavoring it. Next in the factor group hierarchy is *Interrogative*: while the preverbal position is favored in non-interrogative clauses with a factor weight of .51, it is disfavored in interrogative clauses with a factor weight of .26. This result confirms the observations made in theoretical work: SVO constitutes the unmarked word order for non-interrogative clauses in all Spanish dialects. As far as interrogative clauses are concerned, in Non-Caribbean Spanish subjects can be null, and when they are overt, subject verb

inversion is mandatory in Non-Caribbean Spanish; on the contrary, a three-way alternation is available to Caribbean speakers: subjects can be null as well, although they tend to arise at low rates, and when they are overt subject-verb inversion is optional. Furthermore, it has been pointed out in theoretical and quantitative work (Suñer 1982, Lizardi 1993) that pronominal subjects are non-inverted in Caribbean Spanish at much higher rates than lexical subjects. This difference between pronominal and lexical subjects might be responsible for the finding in Table 6.69 that interrogative clauses disfavor preverbal subjects.

Perhaps the most interesting result attested in Table 6.69 is that two language-external factor groups, *Origin* and *Age*, are ranked third and fourth in the hierarchy of significance. *Age* has not arisen as significant in any of the analysis carried out to describe Dominican null and overt subject patterns. *Origin* has been selected as significant in some of the Goldvarb analyses reported in previous sections of this chapter; however, it was always a low-ranked factor group. In Table 6.69, *Origin* (Range=13) and *Age* are ranked over the linguistic factor group *Discourse function*. Constraint rankings within *Origin* and *Age* are organized as follows: urban speakers favor preverbal subjects at .55, while rural speakers disfavor them at .42; younger speakers favor preverbal subjects at .56, whereas older speakers disfavor them at .44. If the hypothesis put forward in Chapter Five according to which ongoing language change is related to the extension of the SVO word order, the results in Table 6.69 would suggest that urban and young speakers are leading grammatical restructuring of subject expression.

Verb could not be included in the Goldvarb analysis reported in Table 6.69 due to the presence of interactions. In order to determine to some extent whether verb type is relevant for the choice of a preverbal or postverbal subject, *Subject Form* and *Verb* were cross-tabulated:

Table 6.70. Dominican Spanish. Subject position patterns. Cross-tabulation of Subject Form and Verb (N=3344).

		PRONOMINAL		LEXICAL		TOTAL	
		N	***%***	***N***	***%***	***N***	***%***
STATIVE	*Preverbal*	557	98	330	89	887	94
	Postverbal	10	2	42	11	52	6
TRANSITIVE	*Preverbal*	431	98	75	87	506	96
	Postverbal	9	2	11	13	20	4
INERGATIVE	*Preverbal*	152	100	33	89	185	98
	Postverbal	0	0	4	11	4	2
UNACCUSATIVE	*Preverbal*	149	98	46	43	195	75
	Postverbal	3	2	61	57	64	25
REFLEXIVE	*Preverbal*	181	98	73	70	254	88
	Postverbal	4	2	32	30	36	12
OTHER	*Preverbal*	643	98	205	74	848	91
	Postverbal	13	2	72	26	85	9

chi-square=103.794
p=0

Note the striking similarity of rates across verb types for preverbal subjects. As far as lexical subjects are concerned, rates of preverbal expression approximate the rates of pronominal subjects in stative, transitive and inergative verbs. Unaccusative verbs depart from other verb types in that lexical subjects display the lowest rates of preverbal expression. This result aligns DS with other Spanish dialects, in which unaccusative verbs tend to bear postverbal subjects (cf. Chapter Two). The differences among verb types are statistically significant, as attested by the *p*-value of 0.

Like *Verb*, *Clause Type* could not be included in the Goldvarb analysis presented in Table 6.71. *Subject form* and *Clause Type* were cross-tabulated to explore whether relevant differences in subject position patters emerge across clause types:

Table 6.71. Dominican Spanish. Subject position patterns. Cross-tabulation of Subject Form and Clause Type (N=3344).

		PRONOMINAL		LEXICAL		TOTAL	
		N	***%***	***N***	***%***	***N***	***%***
MAIN	*Preverbal*	1484	98	574	76	2058	90
	Postverbal	37	2	180	24	217	10
CONJOINED	*Preverbal*	157	97	45	71	202	90
	Postverbal	5	3	18	29	23	10
EMBEDDED	*Preverbal*	653	98	241	76	894	91
	Postverbal	12	2	76	24	88	9
RELATIVE	*Preverbal*	133	99	22	61	155	91
	Postverbal	1	1	14	39	15	9

chi-square=15.943
p=0.001

Rates of preverbal expression for pronominal subjects are practically identical across clause types, and very similar for lexical subjects as well, with the exception of relative clauses, which display a lower rate of lexical subjects in preverbal position. Differences among clause types were found statistically significant, as expressed in the *p*-value indicated at the bottom of the table.

The distribution of preverbal and postverbal subjects was also explored in urban speakers. Similarly to what was the case when examining null and overt subject patterns, analyzing the speech of only urban speakers makes possible to include *Social class* in the logistic regression analysis. Results are reported in Table 6.72[58]:

[58] *Person* had to be left out of this Goldvarb run due to interactions with other factor groups. Since *Person* was not selected as significant in previous runs, cross-tabulations of *Person* and other factor groups are not reported.

Table 6.72. Dominican Spanish. Distribution of preverbal and postverbal subjects in Santo Domingo (N=2188).

Corrected mean				0.967
Log likelihood				-471.335
Significance				0.031
Total N				2188
App: Preverbal subject		***Factor weight***	***%***	***N***
SUBJECT FORM	Pronominal subject	**.70**	98.4	1477
	Lexical subject	.14	78.6	711
	Range	*56*		
INTERROGATIVE	Non-interrogative	**.52**	92.1	2062
	Interrogative	.19	89.8	127
	Range	*33*		
AGE	+50	.40	89.1	861
	-50	**.57**	93.8	1328
	Range	*17*		
DISCOURSE FUNCTION	Topic	**.54**	96.5	1574
	Shift	.39	80.3	615
	Range	*15*		
SOCIAL CLASS	High-middle class	**.54**	92.1	1206
	Working class	.45	91.8	983
	Range	*9*		
EDUCATION	University	[.47]	92	1503
	High school and below	[.56]	91.8	686
GENDER	Men	[.47]	91.3	1265
	Women	[.54]	92.9	924
CO-REFERENCE	Co-referent	[.48]	90	1333
	Non-co-referent	[.51]	95	856
TMA OF THE PREVIOUS VERB	Same TMA	[.51]	92.2	1007
	Different TMA	[.49]	91.8	1182

Subject form and *Interrogative* are at the top of the factor group hierarchy like in Table 6.69. *Age* is also selected as significant, once again with younger speakers favoring preverbal subjects at .57 and older speakers disfavoring them at .40. *Social class* is also significant; high-middle classes favor preverbal subjects (.54) whereas working classes disfavor them (.45). Recall that *Social class* also emerged as a significant variable when null and overt subject patterns were examined in sections 6.2. and 6.3, where high-middle class speakers where shown to favor overt subjects more than working class speakers.

Urban and rural working classes are compared in Table 6.73. *Interrogative* could not be included in this Goldvarb run due to interactions with other factor groups. Results for the cross-tabulation of *Interrogative* and *Subject form* are provided below.

Table 6.73. Dominican Spanish. Distribution of preverbal and postverbal subjects in urban and rural working classes (N=2308).

Corrected mean				0.951
Log likelihood				-575.757
Significance				0.046
Total N				2308
App: Preverbal subject		***Factor weight***	***%***	***N***
SUBJECT FORM	Pronominal subject	**.70**	97.7	1591
	Lexical subject	.13	73.2	712
	Range	*57*		
ORIGIN	Urban	**.57**	92.5	964
	Rural	.45	88.5	1344
	Range	*12*		
AGE	+50	.45	88.1	1039
	-50	**.54**	91.9	1269
	Range	*9*		
GENDER	Men	.46	88.9	1132
	Women	**.54**	91.4	1176
	Range	*8*		
DISCOURSE FUNCTION	Topic	[.51]	95.2	1658
	Shift	[.46]	77.2	650
TMA OF THE PREVIOUS VERB	Same TMA	[.48]	89.1	1151
	Different TMA	[.52]	91.2	1157
CO-REFERENCE	Co-referent	[.48]	92.9	924
	Non-co-referent	[.51]	88.4	1384

In Table 6.73, *Origin* (Range=12), *Age* (Range=9) and *Gender* (Range=8) are ranked over discourse-related factor groups, which are not even selected as significant. *Age* behaves in the same way as in Tables 6.69 and 6.72, with younger speakers favoring preverbal subjects. *Origin* patterns with the results in Table 6.69, with urban speakers favoring preverbal subjects and rural speakers disfavoring them. Interestingly, *Gender*, which did not arise as significant when all and urban speakers were examined, is also significant across urban and rural working classes: women favor preverbal subjects (.54) while men disfavor them (.46).

Results from the cross-tabulation of *Interrogative* and *Subject Form* are offered in Table 6.74. Differences between interrogative and non-interrogative clauses are statistically significant. Note also that while rates of preverbal subject expression are very similar for pronominal subjects in interrogative and non-interrogative clauses, differences are observed between these two types of clauses in lexical subjects:

Table 6.74. Dominican Spanish. Subject position patterns. Cross-tabulation of *Subject form* and *Interrogative* (N=2308).

		PRONOMINAL		LEXICAL		TOTAL	
		N	***%***	***N***	***%***	***N***	***%***
NON-INTERROGATIVE	*Preverbal*	1488	98	520	74	2008	90
	Postverbal	35	2	187	26	222	10
INTERROGATIVE	*Preverbal*	67	99	1	20	68	93
	Postverbal	1	1	4	80	5	7
TOTAL	*Preverbal*	1555	98	521	73	2076	90
	Postverbal	36	2	191	27	227	10

chi-square=20.875
p=0

In sum, a core set of factor groups regulate subject position patterns in all three communities examined. *Subject form* is the highest ranked factor group when all speakers, only urban speakers and working class speakers are included in logistic regression analyses. *Interrogative* follows *Subject form* in the former two communities. *Discourse function* has emerged as significant for all and urban speakers. Furthermore, *Social class* is significant for Santo Domingo speakers, whereas *Gender* is significant across working urban and rural working classes. A comparison of factor group rankings in all three data sets analyzed is provided in Table 6.75:

Table 6.75. Dominican Spanish. Subject position patterns. Comparison of factor group rankings for all speakers, urban speakers and working class speakers.

ALL SPEAKERS (N=4830)	SANTO DOMINGO (N=2188)	WORKING CLASS (N=2308)
1. Subject form (54)	1. Subject form (56)	1. Subject form (57)
2. Interrogative (25)	2. Interrogative (33)	2. Origin (12)
3. Origin (13)	3. Age (17)	3. Age (9)
4. Age (12)	4. Disc.function (15)	4. Gender (8)
5. Disc.function (9)	5. Social class (9)	5. Disc.function
6. Person	6. Education	6. TMA previous
7. Co-reference	7. Gender	7. Co-reference
8. TMA previous	8. Co-reference	
	9. TMA previous	

6.6. Null and overt subject distribution and subject position patterns in Dominican Spanish. Summary.

This chapter has focused on the description of the findings related to the distribution of null and overt subjects, on the one hand, and subject position patterns, on the other. As far as linguistic factor groups are concerned, while discourse-related factor groups are relevant for both null and overt subject choice and preverbal or postverbal subject insertion, the rest of the constraints involved in both linguistic phenomena have been shown to differ. Table 6.76 offers a comparison of the language-internal factor groups that have emerged as significant in analyses carried out throughout this chapter:

Table 6.76.Dominican Spanish. Null and overt subject patterns and subject position patterns. Linguistic factor groups.

NULL AND OVERT SUBJECT PATTERNS (PRONOMINAL AND LEXICAL SUBJECTS VERSUS NULL SUBJECTS)		
ALL SPEAKERS	URBAN SPEAKERS	WORKING CLASS SPEAKERS
1. Person (69) 2. Clause Type (27) 3. Distance (22) 4. Verb (17) 5. Co-reference (16) 6. TMA (14) 7. Discourse function (12) 8. TMA previous (4)	1. Person (76) 2. Clause Type (20) 3. Distance (20) 4. Co-reference (15) 5. Verb (15) 6. TMA (14) 7. Discourse function (13) 8. TMA previous (4)	1. Person (67) 2. Clause Type (25) 3. Distance (24) 4. Co-reference (16) 5. TMA (16) 6. Verb (12) 7. Discourse function (11) 8. 8. TMA prev (7)
SUBJECT POSITION PATTERNS		
ALL SPEAKERS	URBAN SPEAKERS	WORKING CLASS SPEAKERS
1. Subject form (54) 2. Interrogative (25) 3. Discourse function (9) 4. Person 5. Co-reference 6. TMA previous	1. Subject form (56) 2. Interrogative (33) 3. Discourse function (15) 4. Co-reference 5. TMA previous	1. Subject form (57) 2. Topic 3. TMA previous 4. Co-reference

Language-external factor groups have also emerged as significant for explaining the distribution of subjects in DS, particularly for subject position patterns. A comparison of these factor groups across communities is provided in Table 6.77:

Table 6.77.Dominican Spanish. Null and overt subject patterns and subject position patterns. Language-external factor groups

NULL AND OVERT SUBJECT PATTERNS (PRONOMINAL AND LEXICAL SUBJECTS VERSUS NULL SUBJECTS)		
ALL SPEAKERS	URBAN SPEAKERS	WORKING CLASS SPEAKERS
1. Gender (2) 2. Age (2) 3. Origin (1)	1. Class (9) 2. Age (3) 3. Gender (2) 4. Education (2)	1. Origin (3) 2. Gender (2) 3. Age (2)
SUBJECT POSITION PATTERNS		
ALL SPEAKERS	URBAN SPEAKERS	WORKING CLASS SPEAKERS
1. Origin (13) 2. Age (12) 3. Gender (6)	1. Age (16) 2. Class (9) 3. Education (7) 4. 4. Gender (7)	1. Origin (12) 2. Age (9) 3. Gender (8)

The interpretation of these results in light of the research questions outlined in Chapter Five, as well as the theoretical implications that can be drawn from these quantitative results are discussed in Chapter Seven.

Chapter Seven. Theoretical implications: Formal and variationist contributions to a theory on subject expression in Dominican Spanish.

7.1. Introduction.

As outlined in Chapter Five, three broad inquiries have articulated this study:

(i) How can we describe the Dominican subject expression system in its synchronic state?

(ii) Does the Dominican subject expression system reflect a situation of ongoing language change? Is language change reflected in competing grammars in the minds of the speakers?

(iii) How can formal grammars account for cross-dialectal variation?

The results presented in Chapter Six have shown that possible solutions for these inquiries center around two related linguistic phenomena: the null and overt subject patterns attested in DS, on the one hand, and the distribution of preverbal and postverbal subjects, on the other. The effects of language change are visible across a number of shifts in the semantic and syntactic properties of subjects.

Take for instance the distribution of subjects in interrogative and non-interrogative clauses. While interrogative clauses strongly favor overt subjects, preverbal subjects are favored in non-interrogative clauses. This suggests that these constitute distinct phenomena, even if preverbal subjects appear in much higher numbers than postverbal subjects overall. Certain morpho-syntactic factor groups, such as *Verb*, have been proved to have an effect on null/overt subject distribution and subject position patterns: recall that while stative verbs strongly favor overt subjects, followed by inergative and unaccusative verbs, transitive and (more strongly) reflexive verbs disfavor them. DP subjects that appear with unaccusative verbs have been shown to favor preverbal subjects.

Differences and correlations between null and overt subject distribution and subject position patterns are also found across language-external factor groups: *Origin* only emerged as significant

in the examination of null and overt subject patterns when *TMA* and discourse-related factor groups were examined, with urban speakers favoring overt subjects in both cases. Urban speakers have also been show to differ from the other communities examined in this dissertation in preverbal and postverbal distribution, in that they favor preverbal subjects. *Age* and *Gender* did not emerge as significant in the analysis of null and overt subject selection, whereas they are significant for the configuration of subject position patterns, as speakers under 50 years old favor preverbal subjects. Men and women do not behave differently when the data set that includes all speakers is submitted to statistical analysis. However, rural men and women differ in subject position distribution, with women favoring preverbal subjects. Finally, *Social class* has emerged as significant for null and overt subject distribution: in Santo Domingo, high-middle class speakers favored overt subjects more than working class speakers; no significant differences were found across working classes in urban and rural environments. Note that this difference only arises when lexical and pronominal subjects are counted together, while urban and rural speakers do not display differences in behavior when only pronominal subjects are included in the calculations. In subject position analyses, high-middle classes were also shown to favor preverbal subjects.

In sum, as predicted by the hypotheses put forward in Chapter Five, ongoing language change is more visible in shifts and differences in subject position patterns than in null and overt subject distribution. Furthermore, urban, high-middle class, young, and female speakers have emerged as the social groups leading language change.

A formal definition of the DS null and overt subject profile is also pursued in this dissertation. As outlined in Chapter Two, a theory of Spanish subject expression needs to address the following issues: What are the mechanisms for null subject licensing? Do they necessarily involve 'rich' agreement morphology across dialects? What are the structural positions of null and overt subjects? What are the structural positions of overt preverbal and postverbal subjects and what is their semantic nature?

In what follows I offer an interpretation of the quantitative results presented in Chapter Six and I explore the theoretical implications that emerge from the analysis of a spontaneous speech corpus. Essentially, I will propose that the Dominican null and overt subject grammar may be accounted for without resorting to competing grammars in the minds of the speakers, and that vernacular Dominican constructions can be accommodated within the Spanish clausal structure. The patterns of distribution of null and overt subjects in DS are discussed in section 7.2, where the results obtained for language-internal and language-external factor groups in the present study will be interpreted in light of current theories on language change. In section 7.3. I approach the formal view on restructured subject grammars, and a formal definition of the DS NSP profile is offered in 7.4. Finally, a summary of the account put forward in this dissertation is offered in section 7.5.

7.2. Null and overt subjects: Patterns of distribution.

The following sections focus on the explanation of the language-internal and language-external variables that regulate subject use in Spanish. The core factor groups that have been found significant in previous work are discussed in light of Dominican facts. Moreover, structural factor groups are examined in some depth. Finally, the patterns that emerge within language-external variables will be discussed.

7.2.1. Rates and factor group rankings: on the role of language-internal factor groups in Dominican subject expression.

In this section I present my interpretation of the results presented in Chapter Six regarding the distribution of null and overt subjects in Dominican Spanish. Core issues for the analysis of restructured subject grammars, among them attested shifts in factor group and constraint rankings, the role of discourse-related and morpho-syntactic factor groups, or the strength of structural, i.e. syntactic constraints will be addressed. These issues were explored in research questions 1-10 outlined in Chapter Five. In the following paragraphs, these research questions and hypotheses are discussed in turn.

(1) Overt and null subjects in Dominican Spanish: Rates.

Research question 1:

What are the rates of overt subject expression in DS?

Hypothesis 1:

Overall rates are expected to be higher in DS than in other Caribbean dialects. On the one hand, DS has been hypothesized to be "at the forefront of linguistic innovation" (Toribio 2000:343) with respect to its Caribbean counterparts. On the other, if ongoing language change is confirmed, rates should have increased in the 20-year span that exists between the data set for this study, gathered in 2008, and the comparable data sets from other Caribbean varieties, which were collected in the early 1990s (Cameron 1992, Lizardi 1993, Ávila-Jiménez 1996).

The hypothesis put forward in (1) has been confirmed by the results obtained in this study: if the speech all 34 speakers included in this study is examined, the rate of overt subject use for pronominal and lexical subjects counted together is 61%. This rate of overt subject expression is fifteen to twenty percentage points higher than the rates reported in the relevant literature for other Caribbean variaties (Bentivoglio 1987 reported a 40% overt rate for first person singular subjects, Cameron 1992 reports a 45% for the Spanish of San Juan, Puerto Rico, and Orozco and Guy 2008 obtain a rate of 48% of expressed subjects in the Spanish of Cali, Colombia). When only pronominal subjects are examined, the rate of overt subject use obtained is 51.4%. While this rate is closer to the ones reported in previous work than the 61% of pronominal and lexical subjects, it still emerges as the highest rate documented up to the date in Caribbean Spanish, to the best of my knowledge. This result is attributed to two different factors: on the one hand, as proposed in previous theoretical work (Toribio 2000a, 2001) DS seems to be more advanced in subject distribution shifts as compared to its Caribbean counterparts (Toribio 2000:343); on the other, the

two decade-span between the corpus gathered for this study and the data sets examined in previous work.

The pattern that emerges with respect to the different persons of the paradigm replicates the findings in previous work: strong differences are attested between the singular and the plural paradigms; while the highest rates of overt subject use correspond to singular persons, particularly to the second person singular [+/-specific], plural persons favor null subjects. It is interesting to note as well that third person non-specific overt subjects are used by Dominican speakers at a rate of 12.7%. While this rate is much lower that the ones reported for other persons of the paradigm, the fact that these expressed subjects are attested also speaks to ongoing language change, as non-specific third person plural subjects are categorically null in other Latin American and European Spanish dialects. Overt third person plural subjects have also been documented in Spanish in contact with English in the United States (Lapidus and Otheguy 2005), and in the speech of other Puerto Rican speakers (Ávila-Jiménez 1996). In Puerto Rico, overt subject rates for these subjects, as reported by Ávila-Jiménez, were 7%.

(2) Language-internal variables in Dominican subject expression: significant factor groups and factor group ranking.

Research question 2:

What are the significant language-internal variables in DS? Do they differ from the variables found significant in other studies?

Hypothesis 2:

Significant factor groups in DS are not expected to depart from the core set of variables found significant in other dialects, as high rates of overt subject use have not been shown to translate into differences in factor group ranking (Tagliamonte 2006). Therefore, discourse-related variables such as *Co-reference* or *Distance from previous mention* should be ranked at the top of the factor group hierarchy in DS as well. Moreover, the morpho-syntactic variable *Person* is expected to strongly

constrain subject expression in DS, as it has been found in previous work that differences in the behavior of specific pronominal elements tear apart Caribbean from Non-Caribbean dialects (Otheguy, Zentella and Livert 2007).

The hypothesis put forward in (2) is borne out by the results obtained in this study: *Person* arises as the strongest factor group that constrains Dominican subject expression in all the logistic regression analyses that explore null and overt subject patterns. A discourse-related constraint, either *Distance from previous mention* in the case of lexical and pronominal subjects or *Topic* when pronominals are anayzed, follows *Person* in the factor group hierarchy[59].

Morpho-syntactic constraints such as *Verb* or *TMA* follow in the ranking. Relevant differences, however, have been found when the speakers of different communities (all speakers, urban speakers and working class speakers across urban and rural environments) have been examined. Lexical and pronominal subjects also depart from each other in significant respects as well:

When lexical and pronominal subjects are counted together all three communities are identical with respect to the factor groups ranked at the top, *Person, Clause Type*, and *Distance from previous mention* (cf. Table 6.6). Moreover, factor group rankings across communities are remarkably similar regarding *Co-reference*, *Discourse function*, and *TMA of the previous verb*. Constraint rankings within *Person* are very similar to what has been found in the literature (Otheguy et al. 2007). *Clause Type* has also emerged as significant in other investigation on Caribbean dialects (Morales 1997, Orozco and Guy 2008, Otheguy et al 2007), however it is not usually ranked at the top of the factor group hierarchy. In the data set collected for this dissertation, relative clauses strongly favor overt subjects, whereas conjoined clauses strongly disfavor them. Main clauses also favor overt subjects, and embedded clauses disfavor them. It should be added that second person singular has the highest rates of overt expression in all clause types.

[59] Differences in the role of *Co-reference* in DS and other Caribbean dialects will be addressed below.

As far as morpho-syntactic factor groups are concerned, the three communities display very similar *TMA* patterns; however, in these calculations it was not possible to assess whether the distribution of null and overt subjects lends support to the *Functional Compensation Hypothesis* (Hochberg 1986), since verb tenses had to be re-coded in order to carry out Goldvarb runs. Differences have been found from community to community in *Verb;* verb classes behave more homogeneously in urban speakers, while in the all-speakers and working class speakers data sets display the following patterns: stative verbs strongly favor overt subjects, and reflexive verbs disfavor them the most; in between are transitive verbs, which favor null subjects, and unergative and unaccusative verbs, which disfavor them.

Regarding language-external factor groups, *Social class* is selected as significant in the statistical analyses that explore subject expression in Santo Domingo (cf.Table 6.24), and shows that high-middle class speakers favor overt subjects more than working class speakers. On the contrary, *Origin* was not significant across working classes in Santo Domingo and El Cibao. Gender was not significant in any of the communities studied.

For the pronominal data set (cf. Table 6.36), *Person* and *Clause Type* belong at the top of the factor group ranking, followed by *Discourse function, Co-reference* and *Distance from previous mention*. Morpho-syntactic factor groups appear in the ranking below *Discourse function*, which suggests that this type of factors do not override discourse-related constraints on subject expression. Only *TMA* from the previous verb ranks below morpho-syntactic constraints. Language-external factor groups are not selected as significant. In terms of factor group rankings, the picture that emerges from the examination of pronominal versus null subjects in DS is very similar to what has been found in previous work: *Person* and discourse-related factors are the strongest factor groups. The next relevant question that arises is whether these results reflect intra-Caribbean syntactic variation. In this respect, Poplack and Tagliamonte (2001), Tagliamonte (2006) point out that in addition to constraint rankings, differences in the strength of factor groups have proven insightful measures of dialectal variation. Note also that in DS *Discourse function* and *Co-reference* are

ranked second and third in the hierarchy; however, their ranges only reach 19 and 17 respectively, whereas *Co-reference* had a range of 35 in Puerto Rican Spanish. This suggests that the effect of *Co-reference* is stronger for San Juan speakers; in DS overt subjects arise in high numbers in both co-referent and non-co-referent contexts, and *Co-reference* is a weaker factor group.

Also note the difference in the discourse-related factor group ranked higher in the lexical-pronominal and the pronominal data sets: *Discourse function* ranks over *Distance from previous mention* in pronominals; this difference has been attributed in §6.3.2. to the distinct semantic nature of lexical and pronominal subjects. The former, as compared to the latter, are maximally non-ambiguous, and thus selected by the speakers when their intention is to convey the identity of the referent; it follows that whether the referent is familiar in the discourse would be a very significant factor in selecting a pronominal or a lexical subject. In fact, *Discourse function* is also the second-ranked factor group in the results corresponding to the third data set examined in this dissertation, which compares lexical and null subjects.

Person is the strongest factor group for the lexical subject corpus, with third person strongly favoring lexical subjects and the first person strongly disfavoring them. Again, this pattern follows from the semantic differences among pronominal and DP subjects, which interacts with the distinct discursive profile of the first and second versus the third person: while first and second person are potentially deictic and therefore always familiar in the universe of the discourse, the third person is potentially anaphoric and not straightforwardly available to the participants in a speech act. Thus maximally non-ambiguous lexical subjects are favored in the third person. By the same token, within *Clause Type* lexical subjects are disfavored in relative clauses, which strongly favored overt pronominal subjects. *Animacy* also emerges as significant for the lexical subjects data set. This is also an expected result, since null inanimate subjects do not have a pronominal counterpart in Spanish. *TMA* effects are very similar to the ones found for the pronominal data sets, and differences among verb types are not remarkable, which suggests that within morpho-syntactic factor groups *Person* has a much stronger effect in lexical subject choice than *Verb*.

Interestingly, *Age* and *Gender*, two language-external factor groups that have arisen consistently as non-significant in previous statistical analyses, are selected as significant for the lexical subjects data set. Moreover, for *Gender*, tendencies are reversed with respect to the findings obtained for the pronominal data set, as illustrated in Table 7.1:

Table 7.1. Dominican Spanish. Gender and Age null/overt and preverbal/postverbal subject patterns.

Null and Overt subject patterns			
	PRONOMINAL AND LEXICAL SUBJECTS		PRONOMINAL SUBJECTS
GENDER		*Factor weight*	*Factor weight*
	MEN	.49	.50
	WOMEN	.51	.50
	Range	2	0
AGE	+50	.49	.50
	-50	.51	.50
	Range	2	0
ORIGIN		*Factor weight*	*Factor weight*
	URBAN	.50	.49
	RURAL	.49	.51
	Range	1	2
Subject position patterns			
GENDER		*Factor weight*	
	MEN	.47	
	WOMEN	.53	
	Range	6	
AGE		*Factor weight*	
	+50	.44	
	-50	.56	
	Range	12	
ORIGIN		*factor weight*	
	URBAN	.55	
	RURAL	.42	
	Range	13	

It is worth to note at this point, along the lines of Ashby and Bentivoglio (1993), that the fact that *Age* and *Gender* are not selected as significant in the logistic regression analyses that examine null and overt subject patterns does not entail that these factor groups do not play a role in Spanish subject expression. What it does entail is that relative to the other factor groups considered in the analysis, their effects are not very strong, but they might arise as significant in statistical analyses that include less factor groups in the calculations, or in analyses performed on a data set

with more tokens. From this discussion stems the relevance of considering not only significant but also non-significant factor groups for the understanding of linguistic phenomena; in the case at hand, as will be clear from the discussion in §7.2.3. below, among language-external factor groups, at least the factor weights within *Gender* with respect to null and overt subject patterns could be interpreted as reflecting differences between men and women in the process of ongoing language change, even if this variable is not selected as significant. Specifically, women seem to be forward in the change, since they favor both overt and preverbal subjects at higher rates than men.

Going back to the discussion related to the differences between the pronominal and lexical and pronominal data sets, while *Social class* differences were attested in the analysis of the former data set, Santo Doming high-middle and working classes behave homogeneously in the pronominal corpus. *Origin* differences are not found significant among working classes either.

Urban speakers, however, depart from the all-speaker and the working class communities in *Verb*: inergative verbs favor overt subjects and reflexive verbs disfavor them across communities; unaccusatives favor overts as well in the all-speaker and working class communities. Overt subjects are strongly favored by stative verbs in the all-speaker and working class communities, while urban speakers disfavor them. As can be observed in Tables 6.24-6.36 in Chapter Six, factor weights for all verb types are rather close to .5, which is usually interpreted as neither favoring nor disfavoring the application of the independent variable. Tagliamonte (2006) stresses that the .5 value should be taken just as a standarized measure of reference, and that factor weights should be interpreted in the context of the factor group they appear in; a .6 difference within a factor group like *Person*, with a range around 70, might not be interpreted as relevant, while it might be relevant within *Verb*, where differences across verb types differ in a maximum of 10 points.

Unlike *Verb*, discourse-related factor groups and *TMA* are stable across communities.

In sum, subject expression in Dominican Spanish has emerged as a complex phenomenon regulated by the effects of a number of constraints operationalized in discourse-related, morpho-syntactic, structural or syntactic and language-external factor groups. Differences are attested across

the three data sets examined in this dissertation: the pronominal and lexical subjects data set, the pronominal subjects data set, and the lexical subjects data sets. Differences and similarities among the three communities investigated have also been outlined. The relative role of morphological and contextual ambiguity, an issue that has been approached by reference to data from other Caribbean varieties, is addressed in the answer to research questions 3-5.

Research question 3:

What is the relative role of discourse-related variables and morpho-syntactic variables?

Hypothesis 3:

Both types of variables are expected to strongly constrain subject expression in DS. However, *Person* is expected to override discourse-related variables in DS, similarly to what has been found for other Caribbean dialects (Otheguy, Zentella and Livert 2007, Orozco and Guy 2008). It is argued in this dissertation that the preeminent role of *Person* in Caribbean Spanish stems from the substantial shifts in the semantic and syntactic properties of certain pronouns.

Hypothesis 3 has been confirmed by the findings outlined in Chapter Six. *Person*, as already pointed out, is the strongest significant factor group for all subject types and across communities. Strong differences arise among singular and plural persons, and within the singular paradigm, in which the first and second persons significantly depart from the third. This distribution of null and overt subjects is claimed to stem from the distinct feature assembly of Caribbean and Non-Caribbean Spanish in this study, following Camacho (2008). The details of the account of pronominal feature composition in DS are discussed in §7.4.2 below. I now move on to discuss the relative role of morphological and contextual ambiguity in the data set at hand.

Research question 4:

What is the role of morphological ambiguity? What is the relative role of morphological versus contextual ambiguity?

Hypothesis 4:

> Morphological ambiguity is not expected to override contextual ambiguity in the corpus under study. The former has been measured in previous variationist work with the factor group *Verb Ambiguity* (Cameron 1996, Travis 2005, 2007), which distinguishes morphologically ambiguous and unambiguous forms. Since morphological ambiguity only arises in specific verb tenses, the variable *Tense* also tests for the role of inflectional morphology in subject licensing. While in early studies on Caribbean dialects it was proposed that the loss of inflectional distinctions was at the source of shifts in overt subject rates (*The Functional Compensation Hypothesis*, Hochberg 1986) subsequent investigations have shown that: first, the relative weight of *Verb Ambiguity* does not significantly depart in Caribbean and Non-Caribbean dialects (Cameron 1996); and second, that morphological ambiguity cannot be assessed accurately without taking into account contextual ambiguity (Cameron 1992, 1993, 1996, Paredes Silva 1993).

Hypothesis 4 is borne out by the patterns found in the data under study. Recall that a study of ambiguity could not be carried out along with the analysis of the whole set of linguistic variables that constrain subject expression in DS, since re-coding procedures have to be followed in order to permit accurate calculations with Goldvarb. With the aim of exploring the relative role of inflectional morphology and reference tracking in subject choice, Goldvarb runs were performed including only the factor groups *Person* and *TMA*, *Person* and discourse-related factors, and *Person*, *TMA* and discourse-related factors. These runs were replicated with lexical and pronominal subjects and pronominal subjects only, and with lexical subjects only. A summary of the results obtained is provided in Table 7.2:

Table 7.2. Dominican Spanish. Effects of TMA and discourse related factors on pronominal and lexical subjects; pronominal subjects; lexical subjects.

PRONOMINAL AND LEXICAL SUBJECTS VERSUS NULL SUBJECTS			
	TMA	DISCOURSE	TMA & DISCOURSE
FULL PARADIGM	-*Person* ranks first. -*TMA* ranks second. -*Origin* is significant.	-*Distance* ranks first. - All factor groups stable across communities.	-*Person* ranks first. -*Distance* ranks second. -*TMA* is significant. -*Origin* is not significant.
1STP AND 2NDP SG	-*Person* and *TMA* rank both first. -*Origin* is significant.	-*Distance* ranks first. - All factor groups stable across communities.	- *Distance* ranks first. - *TMA* ranks second. -*Person* ranks third. - *Origin* is not significant
1STP AND 3RDP SG	-*TMA* ranks first. - *Person* ranks second. -*Origin* is significant.	-Discourse function ranks first. -*Distance* ranks second. -All factor groups stable across communities. -*Person* is not significant.	-Discourse function ranks first. -*TMA* ranks second. -*Person* is not significant. -*Origin* is not significant.
PRONOMINAL VERSUS NULL SUBJECTS			
	TMA	DISCOURSE	TMA & DISCOURSE
FULL PARADIGM	-*Person* ranks first. -*TMA* ranks second. -*Origin* is significant.	-*Person* ranks first. -*Distance* ranks second. -Discourse function ranks third. -*Co-reference* ranks fourth. -*Origin* is not significant.	-*Person* ranks first. -*TMA* ranks second. -*Distance* ranks third. -Discourse function ranks fourth. -All discourse-related factor groups are stable across communities.
1STP AND 2NDP SG	-*Person* ranks first. -*TMA* ranks second. -*Origin* is significant.	-*Distance* ranks first. -*Person* ranks second. -Discourse function ranks third. -*Co-reference* ranks fourth. -*Origin* is not significant.	-*Distance* ranks first. -*TMA* ranks second. -*Person* ranks third. -*Co-reference* ranks fourth. -*Origin* is significant.
1ST AND 3RDP SG	-*TMA* ranks first. -*Person* ranks second. -*Origin* is significant.	-Discourse function ranks first. -*Distance* ranks second. -*Co-reference* ranks third. - All factor groups are stable across communities.	-*TMA* ranks first. -Discourse function ranks second. -*Distance* ranks third. -*Origin* is not significant.
LEXICAL SUBJECTS VERSUS NULL SUBJECTS			
	TMA	DISCOURSE	TMA & DISCOURSE
full paradigm	------	-*Person* ranks first. -*Co-reference* ranks second. -*Gende*r ranks third. -*Origin* and *Age* are stable across communities.	-*Person* ranks first. -Discourse function ranks second. -*Distance* ranks third. -*Co-reference* ranks fourth. -*TMA* ranks fifth -*Gender* is significant.

Table 7.2. includes information on the top three ranked factor groups for each of the analyses presented, as well as information on the significance of language-external factor groups. The picture that emerges is as follows: discourse-related factor groups are stable across data sets,

with factor group rankings only reflecting the differing properties of the first and second versus the third person. When TMA and discourse-related factor groups are analyzed, *Person* arises as the first-ranked factor group for the pronominal and lexical subject data set, whereas it is ranked below discourse-related factors when first and second persons are compared, and is not selected as significant in the analysis of first and third person. This suggests that the first and third person are quite similar to each other, which, as already pointed out in Chapter Six, does not support the *Functional Compensation Hypothesis* (Hochberg 1986). If this hypothesis were on the right track, second and third persons should pattern together, since morphological ambiguity is stronger in these persons than in the first. Similar patterns emerge in the analysis of pronominal subjects. A difference is found in the significance of *Origin*, however. This factor group emerges as significant when *TMA* and discourse-related factor groups are analyzed in the first and second persons; in this calculation, urban speakers are shown to favor overt subjects at higher factor weights than rural speakers. Interestingly, first and second singular are the persons that favor overt subjects the most, and *Origin*, with the trend just discussed, also emerges as significant when only *TMA* is examined.

(3) Language-internal variables in Dominican subject expression: Constraint rankings within *Person*.

Research question 5:

a. What are the constraint rankings within the factor group *Person*?

b. Person and the discursive status of the different persons in the paradigm: Are there differences between first and second persons versus the third?

Hypothesis 5:

a. Dialectal variation is not necessarily detected on the specific factor groups that are involved in subject expression, but rather in the relative ordering of these factor groups across varieties, and even more so, in the constraint rankings within factor groups (Tagliamonte 2006). In fact, constraint rankings within *Person* depart significantly in Caribbean and Non-Caribbean

varieties: plural pronominal subjects have been shown to disfavor overt occurrences in both Mainland and Caribbean varieties (Cameron 1992, Travis 2005, 2007, de Prada 2009). As far as singular pronominal subjects are concerned, while in Mainland dialects the first person (Silva-Corvalán 1982, 1994, de Prada 2009) favors overt subjects the most, in Caribbean dialects these expressed subjects are most favored with second person singular specific and non-specific subjects. In light of these results, second person singular specific-and non-specific is expected to favor overt subjects the most in this study. Plural subjects are expected to conform to the patterns found in other Spanish dialects in not favoring overt subject insertion (Cameron 1992, Otheguy, Zentella and Livert 2007.

b. Further differences are hypothesized to arise in the singular paradigm between first and second persons versus the third person singular. First and second persons are deictic, whereas the third person is potentially anaphoric. This fact entails differences in reference tracking. For this reason, third person singular pronouns are expected to favor overt subjects less than first and second pronouns.

In other words, while the hypothesis in (3a) that second person singular will favor overt subject use the most in DS is based on the shifts in the referential meaning of pronouns attested in Caribbean Spanish, the hypothesis in (3b) that first and second person will pattern differently from the third stems from the differing discursive properties of pronominals. Further support for this claim is found in the fact that in other restructured varieties, such as Brazilian Portuguese (cf.§3.3), third person singular has been shown to allow for zero subjects in contexts where first and second expressed pronouns are no longer available (Modesto 2000, 2009).

As far as Hypothesis 5a is concerned, the results discussed in §6.2.3. and §6.3.3. show that DS behaves very similarly to other Caribbean dialects (Otheguy et al. 2007, Cameron 1992, 1996):

second person [+/-specific] favors overt subjects the most, followed by first and third person singular. Plural persons pattern with other Spanish varieties in disfavoring overt subjects. An exception to this trend is the second person plural, with rates closer to the ones displayed in the singular paradigm (cf. Table 6.1.).

Regarding Hypothesis 5b, *Person* is the strongest factor group for the lexical subject corpus, with third person strongly favoring lexical subjects and the first person strongly disfavoring them. Again, this pattern follows from the semantic differences among pronominal and DP subjects, which interacts with the distinct discursive profile of the first and second versus the third person: while first and second persons are potentially deictic and therefore always familiar in the universe of the discourse, the third person is potentially anaphoric and not straightforwardly available to the participants in a speech act. Thus maximally non-ambiguous lexical subjects are favored in the third person. By the same token, within *Clause Type* lexical subjects are disfavored in relative clauses, which strongly favored overt pronominal subjects.

Further differences between the first-second persons are the third are observed within *Clause Type*. Table 7.3. offers a summary of the overt subject rates of first, second and third person subjects in main, conjoined, embedded and relative clauses:

Table 7.3. Dominican Spanish. Rates of overt subject use across clause types. First, second, and third person singular subjects.

	MAIN	CONJOINED	EMBEDDED	RELATIVE
Pronominal subjects	***% overt***	***% overt***	***% overt***	***% overt***
THIRD PERSON SINGULAR	29	44	32	33
SECOND PERSON SINGULAR	98	100	99	100
FIRST PERSON SINGULAR	98	100	98	98
Pronominal and lexical subjects				
THIRD PERSON SINGULAR	74	51	70	73
SECOND PERSON SINGULAR	96	50	82	100
FIRST PERSON SINGULAR	66	40	57	84

Note that third person subjects display lower rates of overt subject expression than first and second person subjects in the two data sets represented in Table 7.3. Differences are stronger, however, in the case of pronominal subjects. Once more, this result is attributed to the different

properties of the third person singular: where reference tracking is not provided by discursive constraints, overt pronominals are dismissed in favor of maximally non-ambiguous lexical subjects.

So far I have discussed how the factor groups that have been found significant in previous work are shaped in the DS grammar. In the next section I approach the syntactic factor groups examined in this dissertation. These factor groups have been studied on occasion in previous work (Morales 1997) with revealing results for the characterization of subject expression in the Caribbean.

7.2.2. Structural constraints on Dominican subject expression.

Questions related to the role of structural variables in DS were covered in research questions 6-10 in Chapter Five. Answers to these questions are approached in the following paragraphs.

(4) Language-internal variables in DS: structural properties.

Research question 6:

Are variables related to structural properties such as *Subject Position*, *Clause Type*, *Interrogative* or *Syntactic Verb Type* significant in DS?

Hypothesis 6:

Variables that operationalize syntactic properties are expected to arise as significant in the corpus examined. Moreover, the hypothesis put forward in this dissertation is that grammatical restructuring will not be as visible in null and overt patterns, as zero and expressed subject insertion is guided in DS by the same set of factors that are relevant across dialects. On the contrary, shifts in structural properties will inform us on the effects of grammatical restructuring.

Hypothesis 6 is confirmed. For the explanation of null and overt subject patterns, *Clause Type* has emerged as a very strong factor group for all types of subjects. *Verb*, which has been discussed among morpho-syntactic related variables, can also be considered a structural variable as the lexico-semantic information of specific verbs triggers syntactic operations and is relevant for

subject position choices. Both semantic and syntactic factor groups have an effect on null or overt subject selection by the speakers.

Furthermore, very relevant differences have been found between interrogative and non-interrogative clauses. *Interrogative* was ranked second in the analyses that explore subject position patterns, only overridden by *Subject form*. Interrogative clauses are not frequent in the corpus under study (N=190), and for this reason they could not be included in the logistic regression analyses that examined null and overt subject patterns. *Interrogative* was subsequently cross-tabulated with a number of factor groups, and what emerged from this comparisons is as follows: discourse-related factor groups do not seem to have a strong effect on interrogative clauses; as far as *Person* is concerned, overt subject rates are higher for all persons in interrogative clauses, and second person singular [+specific] subjects are almost categorically overt in these clauses.

Subject form has arisen as the strongest factor group in the analyses in which *Subject position* was selected as the dependent variable. Furthermore, language-external factor groups are ranked at the top of the hierarchy of factor groups that constraint preverbal and postverbal subject choice.

Research question 7:

What differences emerge between pronominal and DP subjects with respect to *Subject Position*?

Hypothesis 7:

Significant differences are expected to arise between pronominal and DP subjects. Specifically, overt preverbal pronominal subjects are expected to be strongly favored across constructions, whereas substantial variation is predicted with respect to DP subjects.

Origin and *Age* follow *Subject form* and *Interrogative* in the hierarchy of factor groups that constrain preverbal and postverbal subject patterns. *Gender*, as has been usually the case so far, is not selected as significant. Urban speakers and speakers under 50 years old favor preverbal subjects

more than rural speakers and Dominicans over 50 years old. Note that *Age* did not emerge as a significant factor group for null and overt subject patterns, while *Origin* was significant in the analyses that explored the relative role of TMA and discourse-related factor groups in null or overt subject choice.

Discourse function is ranked below language-external factors, with *Topic* subjects strongly favoring preverbal subjects. This result resonates with the findings in theoretical work, in which SVO has been claimed to be the default word order in Spanish. From this fact also stems that *Co-reference* does not reach significance: preverbal subjects are highly favored by the speakers, so much so that *Person*, the strongest factor in null or overt subject choice, is not significant either for subject position patterns.

Research question 8:

What differences emerge among verb types with respect to *Subject Position*?

Hypothesis 8:

Overt subjects have been argued to be widespread in Dominican Spanish (Toribio 2000, Camacho 2008). In the data set at hand, transitive and unergative verbs are predicted to strongly favor overt subjects, and unaccusative verbs are not expected to favor postverbal subjects.

The results outlined in §6.5. related to the cross-tabulation of *Subject form* and *Verb* suggest that pronominal subjects of all persons behave similarly in preverbal and postverbal subject patterns, following Non-Caribbean Spanish trends in this respect. Differences are attested in lexical subjects, which favor postverbal subjects with unaccusative verbs. This result rejects *Hypothesis 8*, according to which differences across lexico-syntactic verb types were expected to have blurred in DS as a result of ongoing language change.

Research question 9[60]:

What differences emerge between main and embedded clauses?

Hypothesis 9:

I predict for relevant differences to emerge between main and embedded clauses. Specifically, object relative clauses are expected to favor overt preverbal subjects, in line with the results in Morales (1997). The source of this tendency will be argued to be in the development of (vernacular) clefting structures in DS, which involve preverbal subject insertion (cf 1.3.1).

The results obtained from cross-tabulating *Subject form* and *Clause Type* have shown that pronominal subjects are identical across clause types in their subject position patterns. Pronominals are almost categorically preverbal in main, embedded, relative and conjoined clauses. Differences emerged in the patterns displayed by lexical subjects, which are disfavored with relative clauses. As discussed in §6.5, this behavior is attributed to the distinct semantic properties of pronominal and lexical subjects; specifically, the former appear frequently in relative clauses, since these clauses involve reference tracking of elements in the matrix clause.

Research question 10:

a. What differences emerge between interrogative and non-interrogative constructions?

b. Are preverbal pronominal subjects categorical in any given construction?

Hypothesis 10:

a. Interrogative constructions are predicted to strongly favor overt preverbal subjects as compared with their non-interrogative counterparts. This hypothesis stems from the findings in previous work with respect to the

[60] The distribution of null and overt subjects in non-finite clauses was also explored in research question 9 in Chapter Five. The contribution of the personal infinitives gathered in the corpus under study to our understanding of the properties of these structures are presented in §7.3.2. below.

long-standing availability of non-inverted *wh*-interrogatives in Caribbean Spanish: Navarro-Tomás (1948) documented them in Puerto Rico in the late 1920s, and de Granda (1987) reported their existence in DS in the 19th century already.

b. Pronominal subjects are expected to be almost categorical in *wh*-interrogatives, in line with the results in Lizardi (1993). Furthermore, differences are predicted to emerge among different pronouns of the paradigm: first and second person subjects, which also feature the highest rates of overt use in Caribbean Spanish (Cameron 1992, 1996), are expected to favor preverbal subject insertion the most.

The hypothesis in 10a is only partly confirmed by the results obtained. As explained elsewhere, interrogative clauses significantly depart from non-interrogative clauses in null and overt subject patterns. The effect of discourse-related factor groups is not visible in interrogative clauses, which very strongly favor overt subjects (overt rate=84%, N=190); this difference extends to both main and embedded clauses, with overt subject rates in interrogatives of 87% and 83%. This trend, however, is reversed when subject position is examined. Non-interrogative clauses favor preverbal subjects, while interrogatives favor null subjects. This pattern is interpreted as resulting from the constraints that regulate subject expression in *wh*-interrogatives (Lizardi 1993). A three-way alternation is available to Caribbean speakers, who in addition to the null and verb-subject inverted clauses attested in Non-Caribbean Spanish, may select a non-inverted question. Differences in (non-)inversion patterns are attributed by Lizardi, among others (cf. Torrego 1998 and Ordóñez and Olarrea 2006), to three constraints:

First, the nature of the *wh*-word is relevant (complex wh-words tolerate non-inversion while simpler ones do not in Non-Caribbean Spanish). Second, [+/-argumental] distinction; non-inversion is allowed in Non-Caribbean Spanish with non-argumental wh-words, while ungrammatical with argumental *wh*-words. In Caribbean Spanish, non-argumental questions display overt subjects at

higher rates than argumental questions. Third, differences arise between pronominal and lexical subjects. While non-inversion is almost categorical with pronominal subjects, particularly second person singular, variation is reported for lexical subjects.

Finally, in the corpus under study *Origin* has emerged as significant in the distribution of non-interrogative clauses, in which urban speakers produce overt subjects at higher rates than rural speakers. Santo Domingo and El Cibao speakers, on the contrary, are stable with respect to their rates of overt subjects in interrogative clauses. This result reflects the fact that non-inverted interrogatives have been part of the Caribbean grammar of subject expression for a long period of time; argumental and non-argumental pronominal wh-questions with SV word order were documented in the Dominican Republic in the 19th century by de Granda (1987), and in the early 1920s in Puerto Rico by Navarro-Tomás (1948). The comparison of corpora provided by Lizardi (1993) shows that overt pronominal subjects have increased their rates in both argumental and non-argumental questions in the last few decades, and that the SV word order has extended, to a lesser degree, in *wh*-questions with lexical subjects. Thus Dominican speakers of different socio-demographic profiles are stable in the distribution of subjects in interrogatives. *Gender* and *Age*, like *Origin*, were not statistically significant. *Social class* emerged as significant, situating urban ahead of rural speakers in overt subject rates.

In sum, urban and high-middle class speakers emerge as the social groups ahead in the ongoing language change, which is to be found in non-interrogative clauses and in the shifts in the behavior of lexical subjects with respect to subject position patterns. Pronominal subject distribution has been shown to be stable across communities.

As far as Hypothesis 10b is concerned, overt pronominal subject rates are very high in interrogative clauses (overt rate=86%, N=140), and almost categorical with [+/-specific] second person singular subjects (overt rates= 97% and 90% respectively). The patterns found in the examination of subject position are as follows: preverbal subjects are almost categorical in non-interrogative and interrogative clauses (overt rates=99% and 98% respectively), whereas variation is

attested with lexical subjects (overt rates of preverbal subjects in interrogative clauses=20%, overt subject rate in non-interrogative clauses=74%). No *Person* differences have been found significant in subject position patterns.

Aside from these differences, subject position patterns replicate the results obtained when examining null and overt subject patterns, situating again urban, high-middle class speakers ahead in the change: the effect of discourse-related factor groups is not visible with pronominal subjects; within *Clause Type*, interrogatives display higher rates of overt use across clauses; finally, regarding language-external factor groups, *Origin* arises as significant, with urban speakers displaying higher rates of preverbal subjects; *Social class* is also significant, and high-middle class speakers obtain the highest rate of expressed subjects; *Gender* and *Age* are not significant. These factor groups are discussed in detail in the following section.

7.2.3. Language-external factor groups and language change.

(5) Language-external variables in DS subject expression: *Age*.

Research question 11:

Are there apparent time differences? Do these differences reflect real time ongoing language change?

Hypothesis 11:

Significant differences across age groups in rates and variables constraining subject use are not expected to arise in the data set examined, whereas apparent time differences are predicted to emerge in the corpus with respect to the strength of structural variables such as *Subject Position*. This hypothesis stems from the results reported in previous work: while clear *Age* differences have been found in the study of other restructured varieties such as Brazilian Portuguese (Duarte 2000, cf.3.3), this variable is typically returned as a non-significant for subject expression investigations that examine the distribution of null and overt subjects (Bentivoglio 1987, Cameron 1992) in Caribbean dialects. An exception to this trend is Ávila-

Jiménez (1996), who found age differences significant in her corpus of Puerto Rican Spanish. A cut-off point of 50 years old was found in her data, with speakers below 50 showing significantly higher rates of overt subject use. *Age* differences have also emerged in the studies that examined structural word order properties rather than null and overt subject insertion. A case in point is Lizardi (1993). In this investigation, 50 years old is taken as the cut-off age point as well, and significant differences emerge between younger and older speakers with respect to the preverbal positioning of both pronominal and DP subjects.

Hypothesis 11 is borne out in subject position patterns. Speakers under 50 years old favor preverbal subjects at .56, speakers over 50 years old disfavor them at .46. Urban, high-middle class speakers have also been shown to favor preverbal subjects. This parallelism in results suggest that apparent time differences in fact reflect ongoing language change, and that these social groups are leading the process. On the contrary, *Age* was not significant for the configuration of null and overt subject patterns. Ranges for this factor group go from 0 to 2 depending on the data set being analyzed. An exception is found in the lexical subjects data set, where *Age* was significant, with speakers over 50 years old favoring overt lexical subjects. This result has been interpreted above as a result of the higher number of overt pronominal subjects being concentrated in younger speakers.

The results obtained for the pronominal and lexical subjects considered jointly and pronominal data sets parallel the findings in previous work. Language-external variables have been found consistently non-significant in the dialects studied up to date, with notable exceptions (Lizardi 1993, Ávila-Jiménez 1996). In both studies 50 years old arose as the significant cut-off age in Puerto Rican Spanish. Ávila-Jiménez reported higher rates of overt subject use for speakers under 50, and Lizardi found higher rates of preverbal subjects for the same age group. The comparison of these results suggests that expressed subjects and preverbal position are the two innovative variants in the speech of Puerto Ricans. In the Dominican data set gathered for this corpus, *Age* does not emerge as significant for null and overt subject distribution. This lack of

patterning could be interpreted in two different directions. On the one hand, it could suggest that the extension of overt pronominalization and preverbal subject insertion constitute two separate linguistic phenomena. On the other, it could be the case that selecting a different cut-off age point for the analyses would yield different results. Examining Dominican Spanish subject expression across diverse age groups is a task I leave for future research.

(6) Language-external variables in DS subject expression: *Gender*.

Research question 12:

Are there differences between men and women? How can these differences be interpreted for the characterization of ongoing change?

Hypothesis 12:

Similarly to what is the case with *Age*, *Gender* has not been found as a significant variable for the distribution of null and overt subjects, and it has only been found relevant for the description of structural phenomena such as word order patterns of pronominal and DP subjects. Lizardi (1993) reports that women, along with younger speakers, show the most advanced trends in preverbal subject insertion: on the one hand, they almost categorically favor preverbal insertion of pronominal subjects; on the other, they display significantly higher rates of preverbal DP subject insertion. This leads her to suggest that Caribbean ongoing change in subject expression constitutes an instance of 'change from below' (cf. Labov 1991, 1994 and §4.2.3), as women usually lead this type of shift. In the present study, *Gender* differences are only expected to arise in the cross-tabulation of *Subject Position* with pronominal and DP subjects.

The results obtained for *Gender* in this dissertation suggest that ongoing change is to be found in subject position differences across groups. This factor group has only emerged as significant in the examination of subject position patterns in working classes across urban and rural environments. Working class women favor preverbal subjects at higher rates (=.54) than men

(=.46). Since grammatical restructuring is carried on by shifts in subject position patterns, women emerge as the social group, along with young, urban, high-middle class speakers, that are ahead in the change.

The fact that *Gender* is not significant in null and overt subject patterns could also be attributed to the nature of the ongoing grammaticalization. Lizardi (1993) notes that in internally motivated language change, if the process of grammaticalization is not new in a given variety, differences across social groups might not be very large. Like what is the case for other language-external factor groups, only statistical analyses carried out with a larger number of tokens, or analyses in which the strength of only a few variables are included could reveal whether *Gender* has a stronger role in subject expression than the one reflected by the present results.

(7) Language-external variables in DS subject expression: *Social class* and *Education*.

Research question 13:

Are there differences among social classes and/or educational levels?

Hypothesis 13:

No significant differences at the level of factor groups and constraint rankings are predicted in this study's data set with respect to *Social Class* and *Education*, in spite of potential differences in overt subject use overall rates. Lizardi (1993) and Ávila-Jiménez (1996), find the variable *Occupation* significant: as discussed in §4.2.3, in the latter the speaker group that included professional and skilled workers featured higher rates of overt subject expression than the speaker group conformed by unskilled workers; she reports very similar results with respect to *Education*, with more educated speakers producing displaying the highest percentages of expressed subjects. Similar results were returned from Lizardi's analysis of interrogative constructions. Generally, however, *Social Class*, *Education* and related variables such as *Occupation* are not returned significant in previous analyses.

Hypothesis 13 is not supported by the results obtained in this study. *Social class* differences have emerged as significant in the examination of null and overt subject patterns and in exploring the distribution of preverbal and postverbal subjects. High-middle class speakers favor overt preverbal subjects, and working class speakers disfavor both variants. *Education* has not emerged as significant in any of the calculations performed on the data set at hand. *Occupation* was not one the factor groups included in the present investigation. However, since it is usually the case that high-middle class speakers carry out skilled and professional work in the Dominican Republic, these results could be interpreted as indirectly supporting Lizardi's (1993) results.

(8) Language-external variables in DS subject expression: *Urban* and *Rural*.

Research question 14:

a. Are there differences between Urban and Rural speakers?

b. Do urban and rural Dominicans conform two distinct speech communities as far as subject expression is concerned?

Hypothesis 14:

The differences between urban and rural Caribbean speech have only very rarely explored in previous work. To the best of my knowledge, only Ávila-Jiménez (1996) gathered a subset of her data in rural areas of Puerto Rico. She did not find significant differences between urban San Juan and rural speakers. In the rest of contemporary investigations, data sets on Caribbean subject expression patterns have been gathered in urban settings (Bentivoglio 1987, Cameron 1992, Flores-Ferrán 2002, 2004, Guy and Orozco 2008). On the contrary, early descriptions of Caribbean varieties, carried out within traditional dialectology frameworks (Navarro-Tomás 1949, Henríquez Ureña 1940), typically report the speech of rural areas. In recent work, early rural descriptions and contemporary urban corpora are compared in order to address progress in language change. While relevant differences between urban and rural speech are attested in morpho-phonological

phenomena, for instance (Bullock and Toribio 2008), no qualitative differences have been reported with respect to patterns of null and overt subject insertion. Taking into account this partial, indirect evidence, the null hypothesis in this study is that urban and rural speakers share an underlying grammar of subject expression.

Hypothesis 14 is only confirmed to some extent by the results reported in Chapter Six. On the one hand, urban and rural Dominicans have been shown to differ with respect with null and overt subject patterns as well as with respect to the distribution of preverbal and postverbal subjects, with urban speakers favoring overt preverbal subjects. This has been interpreted elsewhere as suggesting that Santo Domingo speakers are ahead in ongoing language change. On the other hand, Santo Domingo and El Cibao speakers have been shown to be rather homogeneous as far as discourse-related factor groups are concerned. Since the constraints operationalized by these factor groups have been found to be remarkably stable across Spanish dialects, this result is expected in the present study. The theoretical implications urban/rural differences are developed in the following section.

7.3. Grammatical restructuring and formal grammars.

The view sustained in this dissertation is that the restructured Dominican subject expression grammar should be explained by formal grammars. In generative terms, this implies that parametric space should be granted for the Dominican Spanish NSP profile. Two issues are approached in §7.3.1. and §7.3.2: first, whether cues on the parametric configuration of a language-specific grammar may be drawn from quantitative evidence. Second, the relevance of qualitative evidence for linguistic theorizing is discussed.

7.3.1. On the quantitative expression of parameters.

(9) On the theoretical implications of quantitative evidence.

Research question 15:

Are parametric connections expressed quantitatively?

Hypothesis 15:

> In recent work (Adger and Smith 2003, 2005; and Cornips and Corrigan 2005) the analysis of quantitative data is deemed relevant not only as far as it can lend support to structural analysis; rather, the potential of bridging quantitative and formal paradigms is conceived as an instrumental tool for the explanation of syntactic variation. Following this line of research, the hypothesis put forward in the present study is that quantitative analysis may reveal unattested correlations in the DS subject expression grammar, which in turn may contribute to an accurate description of the specific instantiation of the *Null Subject Parameter* represented by this variety.

The investigation of the form that parameters might take in natural languages is a central concern in generative syntactic theory in its two most recent incarnations, the Government and Binding Model (Chomsky 1981, 1986) and the Minimalist Program (Chomsky 1995, 2001, 2004). The GB theory was founded in two main ideas: the correlation between morphology, and syntax and the notion of parameter clustering. The first idea is largely based on facts related to the *Null Subject Parameter* (NSP) in Germanic versus Romance languages (Rizzi 1982). The availability of null subjects was attributed to the 'rich' morphological system of Romance languages. However, the loose concept of 'rich' agreement has resisted a formal definition ever since (Roberts 1993, Rohrbacher 1999, Holmberg 2005, among others). The notion of parameter clustering entails that a whole compartment of a language-specific grammar is created around the possibility of having null referential subjects.

These two ideas articulate a powerful theory that reconciles synchronic and diachronic concerns, since parameter setting is at the core of both language acquisition and language change processes. However, the extant cross-linguistic research carried out in the eighties and nineties revealed that subject expression grammars were more complex than the one defined in the early

formulations of the NSP, and that a number of mixed patterns were available cross-linguistically. Variation ranges from Null Expletive languages such as German or Icelandic to Partial Null Subject languages like Finnish or Hebrew.

In addition, quantitative evidence on the parametric shifts attested in acquisition and diachronic processes in language-specific grammars has also been central to test the accuracy of parameter clustering theories.

The developmental studies carried out within GB focused on the evidence for the relationship between the acquisition of inflectional paradigms and parameter settings in language development processes. Results showed that the acquisition of agreement morphology does not necessarily entail the setting of all the grammatical properties associated with the NSP. 'Optionality' was identified as one of the essential elements in the developmental grammars that correspond to transitional stages of first and second language acquisition and in the ultimate attainment, end-state grammars of non-native speakers, as shown by the extant research on the issue carried out in second language acquisition studies (Bley-Vroman 1989, Eubank 1994, Epstein, Flynn and Martohardjono 1996, Schwartz 1998, Schwartz and Sprouse 1994, 1997, Liceras 1998, 2007, Zobl and Liceras 2005, Liceras et al. 2006, Sorace 1993, 1999), among many others. Language acquisition research suggests: first, that 'single' developmental grammars yield systems that incorporate 'optionality' as a stabilized, rather than transitional, feature; and second, that 'optionality' must be accounted for within the limits of formal grammars, particularly in a model that conceives diachronic change as crucially driven by language acquisition.

Furthermore, 'optionality', defined as the coexistence within and individual grammar of two or more variants of a given construction that make use of the same lexical resources and converge the same meaning, is well attested in native grammatical systems as well, as explained by Biberauer and Roberts (2006).

Similarly, the evidence from language change processes suggests that on the one hand, the co-existence of two constructions corresponding to opposite settings of the NSP is a usual scenario

in diachrony. It also shows that this situation yields a number of possible mixed NSP patterns. Relevant facts from French, Mainland and Island Scandinavian languages, Old Spanish and Brazilian Portuguese will be discussed in 7.3.3. below.

In sum, parametric connections or lack thereof are well attested in quantitative studies on NSP shifts in acquisition and diachronic processes. This evidence, added to the evidence on cross-linguistic and cross-dialectal variation triggered the theoretical move towards the Minimalist Program (Chomsky 1995, 2001, 2004). In this program, language variation is located in the feature composition of lexical items, on the one hand, and in the interface between syntax and the phonological and semantic-pragmatic modules, on the other. This theoretical framework allows deconstructing the NSP on a number of (micro)parameters. The setting of these distinct parameters to a specific value may derive a number of attested mixed NSP patterns[61]. Specifically, Sheehan (2006) claims that three micro-parameters articulate the NSP:

(10) *The Null Subject Parameter:*

(i) *The rich agreement parameter*

T lacks/bears a [uD] feature.

(ii) *The PF-interpretability parameter*

T lacks/bears a *, where * requires the specifier of I to be spelled out at PF.

(iii) *The Weak/strong nominative Case parameter*

Nominative Case feature [uI] lacks/bears *, where * requires movement to Spec, TP.

With the mimimalist view on parametrization, the similarities and differences that arise within closely related grammatical systems might be accounted for in a principled way. For instance, Brazilian Portuguese can be defined as follows: it has a 'negative' setting for the NSP, that is to say, it lacks a [uD] feature on T, meaning that the deletion of referential null subjects under non-distinctness is blocked. However, BP lacks a PF requirement for Spec, TP and so it licenses

[61] For an extensive discussion on this issue, see Roberts and Holmberg (2009).

null locatives and expletives. BP also allows movement from finite clauses via a peripheral Topic position. This gives rise to null bound embedded subjects. As such, the null subjects licensed in BP are also the result of deletion. It can thus be maintained that BP has an EPP requirement of the type 'Merge XP', as has been argued for Null Subject Romance. Moreover, in BP, nominative Case is strong, as it is in English, EP, and French, triggering movement of the subject to Spec, TP. Spec TP is only empty at PF where it contains a deleted expletive or locative or the deleted copy of a moved DP. Note that grammars in competition in the minds of the speakers (Kroch 1989, 1994) have not been postulated to describe the mixed NSP pattern displayed by BP, or the differences that may arise between Brazilian and European Portuguese.

The study of closely related varieties had traditionally been carried out within the field of variationist sociolinguistics. These studies have contributed to the understanding of subject licensing grammars in Spanish in two different directions: on the one hand, they allow for a principled description of syntactic variation patterns, as just discussed; on the other, they have also converge in specific patterns described by theoretical linguistics.

Relevant examples are the results obtained by Cameron (1992) for the distribution of *–s* deletion and null/overt subject patterns, which speak to the non-correspondence of morphology and syntax. Cameron found that the speakers who consistently delete final *–s* (male speakers in his study) do not insert overt pronominal subjects at higher rates than speakers who do not delete final *–s* as frequently. Also relevant in this respect is the correlation that holds, to some extent, between the extension of overt subjects and the grammaticalization of preverbal subjects.

The formal definition of the Dominican subject expression grammar will the approached in section §7.3.4. For now it will suffice to say that in fact unattested correlations among linguistic phenomena may be found in quantitative results; in addition, proposed correlations may be tested against spontaneous speech results. I now move on to discuss whether the differences in methodological approaches usually followed in variationist and formal linguistics have consequences on they way subject expression facts are pictured.

Traditionally, generative and variationist approaches to language variation have been considered complementary at the most. However, already within the GB framework Kato and Taralho (1984) claimed that quantitative results on the shifts attested in the Brazilian pronominal paradigm could be theoretically insightful. The bulk of research that explores the potential of quantitative results has grown in the last two decades with the transition to the Minimalist Program, as shown by the work included in Adger and Smith (2005).

Recently, Henry (2005) has discussed not only the potential but also the challenges of bridging the generative and variationist paradigms. The author discusses the validity of grammaticality judgment tasks in the investigation of non-standard varieties, and concludes that while corpus data inform the description of unattested patterns in the language, the exploration of certain syntactic phenomena cannot be carried out only through reference to data gathered in interviews, since specific structures are rare in speech. From a formal point of view, the frequency of syntactic structures is not necessarily involved in the formulation of theories. Negative evidence, unavailable in speech, is also a fundamental tool in syntactic theorizing. For these reasons, and taking as point of departure the patterns detected in a large corpus of Belfast English, Henry elaborated syntactic well-formedness questionnaires and distributed them among speakers in order to study subject-verb agreement.

Another example of the potential of interdisciplinary work can be observed in the study of Brazilian Portuguese. Synchronic (Kato and Taralho 1984) and diachronic quantitative work (Duarte 2000) have found and described the relevant variation facts for the explanation of the Brazilian NSP profile. Subsequent formal work has isolated the specific contexts in which synchronic BP still allows for null subjects. These contexts are third person null subjects in embedded clauses. The availability of these structures has lead Modesto (2000, 2009) to propose that BP is a Topic-prominent language, while Sheehan (2006) and Holmberg (2009) interpret the facts as picturing BP as a Partial Null Subject language.

Only quantitative data have been gathered for the present study. Therefore, the theoretical implications that may be drawn from the data set might be considered incomplete from a formal perspective. However, the opposite view is taken in this dissertation: gathering a large corpus of Dominican speech has allowed us to observe the specific properties of two linguistic phenomena, null and overt subject distribution and subject position patterns. In addition, a number of the constructions explored in theoretical studies were attested in the interviews with Dominican speakers. Among these constructions, personal infinitives, interrogative structures, relative clauses and overt expletive subjects stand out. Note that this view does not exclude the potential of gathering traditional syntactic data for the fine-grained description of language-specific grammars. In the following section I address the theoretical implications related to the existence in DS of the constructions that did not reach a number of tokens high enough to perform Goldvarb analyses.

7.3.2. Qualitative evidence and the description of grammatical systems.

(11) On the role of qualitative evidence.

Research question 16:

What does it entail that specific constructions are either not attested or marginal in other dialects while possible in DS?

Hypothesis 16:

It is argued in this dissertation that the examination of specific structures, even if they do not arise in high numbers in the corpus, is also relevant for the explanation of the Dominican subject grammar. Recall for instance from the discussion of interrogative structures in §4.2.2.3.1 that clefted questions, even if they are widespread in Caribbean dialects when compared to their Non-Caribbean counterparts, do not arise in spontaneous speech as often as *wh*-interrogatives (in Lizardi's 1993 data set, which comprised 1912 tokens, only 130 clefted questions were reported). However, their existence is relevant for the explanation of Caribbean subject grammars, as they are vernacular constructions that feature the overt

preverbal subjects, i.e, the word order pattern associated with grammatical restructuring in these varieties.

Four structures stand out in the corpus gathered for the present study:

(i) Overt expletives,

(ii) Personal infinitives,

(iii) Clefts, and

(iv) Clefted questions.

These four constructions share the SVO word order preferred in Caribbean Spanish, and are distinct from their Non-Caribbean counterparts in a number of respects. They appear in small numbers in the data set under study; however, their distribution might be relevant for the formal description of the Dominican subject grammar. Note for instance that declarative and interrogative clefts contain a relative clause, and that relative clauses have been found to strongly favor overt subjects in the quantitative analysis.

Different cues on the configuration of the Dominican grammar are provided by personal infinitives and overt expletive constructions. In both structures the preverbal subject position is filled, following the Caribbean trend and departing from Non-Caribbean Spanish, in which these overt subjects either are not attested or bear different properties. In what follows I discuss them in turn.

7.3.2.1. Overt expletive constructions.

Martínez Sanz and Toribio (2008) offer an account of Dominican overt expletives that significantly departs from the proposals in previous work (Uriagereka 1995, Silva-Villar 1998). These authors propose that unlike French and German expletives, Romance expletives are accorded a discourse role, and consequently merged in Spec, CP. Evidence for the Spec, CP hypothesis is found in the availability of Transitive Expletive Constructions (cf. also §2.3.4.):

(12) Ello yo no sé por qué mi padre no me llamó la torda o la papagaya

'I don't know why I was not called either the thrush or the parrot by my father.'

(17th century Spanish, Silva-Villar 1998:257).

Furthermore, Romance expletives always appear in the leftmost position of the sentence:

(13) *Xa sei que el chove.

'Of course I know it rains.'

(Galician, Silva-Villar 1998:255).

Recently, Hinzelin and Kaiser (2007) have supported this hypothesis, proposing that they are akin to expletives in other Romance languages. They find evidence for this claim in the availability of TECs in Old Spanish (cf. §2.3.4.). Martínez and Toribio (2008) find a number of flaws in this account. These authors show that overt expletive structures are documented in a number of contexts:

(14) Metereological predicates:

Ello/ø hace frío sí.

'It's cold yes.'

(15) Impersonal passives:

a. Ya para mayo ello/ø están las matas llenas y ello/ø se ven esos mangos maduritos.

' Already in may the trees are full and those mangoes look ripe.'

b. Ponte el uniforme de los nueve días porque ello/ø se han muerto muchos en estos días.

'Put on your 'nine days' uniform because many have died these days.'

(16) Unaccusatives:

a. Parece que por aquí ello/ø no pasan motoconchos por la mañana.

'It seems that around here motorcycle-taxis don't pass by in the morning.'

b. Yo llevo es tres horas esperando y ello/ø casi no ha pasado ni un vehículo.

'I have been waiting for three hours and almost no vehicle has passed.'

(17) Embedded contexts:

a. A mí no me molesta lavar cuando ello hay agua porque yo termino rápido.

'It doesn't bother me to wash clothes when there is water because I finish quickly.'

b. Yo creo que ello tiene que ser noche de luna llena para ello aparecer lechuza.

'I think that it has to be a full moon night for an owl to show up.'

(18) TECs are ungrammatical in DS:

*Ello yo no sé por qué mi papa me puso ese nombre.

'I don't know why my dad gave me that name.'

Interestingly, Mainland and Caribbean traditional Spanish grammars document a referential pronoun *ello*. This pronoun seems to be in decline, as is not found in the corpus gathered by the authors:

(19) Lo que/Eso/Ello que tú quieras te lo doy.

'What/That/ what you want I give it to you.'

In light of these facts, the proposal in Martínez Sanz and Toribio is that two types of *ello* are available in in DS, expletive and referential pronoun; referential *ello* belongs in Spec, CP, while expletive *ello* belongs in the TP field. Building up on the account in Martínez Sanz (2007), it is put forward that *ello* is merged in Spec, TP to check EPP on T; in DS, the EPP feature cannot be checked by agreement; consequently, expletive and personal pronouns emerge in preverbal position. Pragmatic force may be imparted by adverbials:

(20) a. Q: Yo creo que hay suficiente arroz para Radhamés.

'I think there is enough rice for Radhamés.'

A: No, no hay suficiente / No, ello no hay suficiente / No, ello no hay suficiente no.

No, no is enough / No, exp no is enough / No, exp no is enough no

'No, there is not enough.'

The account in Martínez Sanz (2007), heavily relied on the non-availability of verb movement in Dominican Spanish, which in turn was attributed to the impoverishment of agreement morphology. However, the existence of languages like French, in which verb movement and overt subjects co-exist, poses a challenge for this theory. In the present study, the hypothesis in Martínez Sanz (2007) is improved by proposing that the Spanish clausal structure does have a number of projections available within TP for hosting different types of subjects, following Cardinaletti and Starke (1999), Cardinaletti (2004). The existence of these positions is independently justified in other Spanish dialects, and Dominican word order patterns further confirm it. Specifically, preverbal positions will be postulated within the TP field for hosting overt pronominal and lexical vernacular subjects. This theory, in turn, makes verb movement available. The distribution of overt pronominal subjects in personal infinitives further supports this claim.

7.3.2.2. Personal infinitives.

Personal infinitives in Non-Caribbean and Caribbean Spanish present a very similar distribution (Fernández Lagunilla and Anula 1994, Rigau 1995, Torrego 1998). They can only occur in adjunct or subject positions. The Case of the subject is nominative in these structures; as illustrated in (22), when the personal infinitive is introduced by a preposition, the latter does not license the Case on subject. Subjects can only be preverbal if they bear contrastive focus (23):

(21) Para corregirlo yo, tenía que trabajar toda la noche.
for correct.INFIN-it I.NOM had-to.1SG work all the night
'In order to correct it, I had to work all night.'

(22) *Para mí corregirlo, tenía que trabajar toda la noche.
for acc.I correct.INFIN-it had-to.1sg work.INFIN all the night
'In order to correct it, I had to work all night.'

(23) Al ELLA tener que pagar, y no los otros, María gritó
On she.NOM have-to.INFIN pay and not the others Maria shouted
'Since she had to pay and not the others, Maria shouted'
(Uriagereka 1995:216)

In Caribbean Spanish, however, the preverbal position does not convey contrastive focus (Suñer 1986, Lipski 1991, Toribio 1993b). Furthermore, full DP subjects are also available in the preverbal position. In the corpus under study, personal infinitives are only attested with singular pronominal subject, as in (24):

(24) a. Para tú llegar al sitio donde se bañaba uno…
For you to-arrive to-the place where CL.bath.PAST one
'For you to get to the place where you could swim…'

b. Para uno hacer cualquier cosa..
For one to-do anything…
'For you to do anything…'

c. Para ella ser chula bien…
For her to-be cool alright
'For her to be cool…'

d. Entonces al día siguiente de yo casarme…

Then the day next to I to-get married

'Then, the day after I got married…'

Subjects of infinitives are null in the data set at a rate of 82%. The proposal put forward here is that overt subjects of infinitives occupy the same position in the structure as overt pronominal subjects. Furthermore, overt infinitive subjects are also documented in constructions that involve a position akin to Spec, CP, such as Topicalization:

(25) Sí, yo para graduarme…

Yes, I for to-graduate

'Yes, myself, for me to graduate…'

Examples like (25) show that several subject positions within the TP and CP fields are also available in DS. Similar patterns of variation are attested in clefted structures, which are discussed in the following section.

7.3.2.3. Clefts and Focus: Affirmative and interrogative clauses.

Of great interest for the study carried out in this dissertation is to probe into the strategies that DS has available for the focusing of subjects, since the expression of Focus in Spanish triggers syntactic phenomena such as shifts in subject position patterns, on the one hand, and the use of clefted structures, on the other.

The grammar of Spanish offers two basic options for focusing constituents, namely prosodic and syntactic focus, illustrated in (26) and (27) respectively, which are borrowed from Camacho (2008:1). Focused constituents are underlined:

(26) a. Fue Marta la que trajo los regalos. ['true' cleft structure]
was Marta cl.acc that brought the gifts.

b. Quien trajo los regalos fue Marta. [pseudocleft structure]
who brought the gifts was Marta.

c. Marta fue quien trajo los regalos. [inverted pseudocleft structure]
Marta was who brought the gifts
'It was Marta who brought the gifts.'

(27) Los regalos los trajo Marta.
'Marta brought gifts.'

It would be reasonable to think that in a variety like DS, in which overt subjects are widespread and tend to bear a Topic reading, prosodic focus constructions should be rather infrequent, more so given that prosodic focus in Spanish entails inserting a subject in postverbal position. The extent to which these constructions are available in DS has not been explored in previous work; however, it has been pointed out (Toribio 2000, 2001) that clefting is a very productive strategy in this dialect, even in interrogative constructions, as in (28). In these clefted questions subjects are often preverbal, although there is some degree of variability, as illustrated in (28b):

(28) a. ¿Qué es lo que *yo* tengo que hacer?
what is CL.ACC that I have.1PSG that do
'What is it that I have to do?'

b. ¿Cómo es que se llaman *esas maticas*, Maria Amelia?
How is that CL call.3PPL those little.plants Maria Amelia
'How is it that those little plants are called, Maria Amelia?'

Suñer (1986) already pointed out the availability of vernacular clefted questions with preverbal subjects in Venezuelan Spanish. She noticed the great frequency with which information

questions were asked by using the formula *wh*-word-*ser-(lo)-que wh*-word-*to-be-(lo*)-that. This form parallels the French *qu'est-ce que…?* formula for *wh*-questions, and can be viewed as a way of avoiding subject-verb inversion. The next question that arises is whether this is a strategy for marking either narrow or contrastive focus, or a strategy simply to formulate *wh*-interrogatives avoiding subject-verb inversion. It could even be suggested that it might the case that these two possibilities do not exclude each other. This latter suggestion stems from the distribution of clefted structures in affirmative clauses. The discussion in 7.34. below will show that such a pattern is displayed by French, according to Hamlaoui (2007). While clefted subjects convey narrow Focus in French, clefted objects convey contrastive Focus.

Affirmative sentences in which subjects are focalized through clefting are widespread in the data set gathered for this study. Inverted pseudoclefts, in which the subject appears in a preverbal position (29a,b) alternate in the speech of most speakers with the pseudoclefts that bear a postverbal subject (29c,d); a small number of speakers feature also the 'true' clefts in (29e), which also require a postverbal subject. Note that in all these constructions, the tendency in DS is to omit the accusative clitic that appears in these constructions in Non-Caribbean Spanish:

(29) a. Yo la vi en el Banco Popular, y yo digo, yo como que conozco a esta vieja… ella me miró a mí, [ella fue que me dijo mi nombre.]
'I saw her at the bank, and I thought, I think I know this old lady…
it was her who looked at me, it was her who called my name.'
(Néstor, 56, UH)

b. Entonces como yo era pitcher, yo dije en mi mente, [esa es que dio a mi hermana.]
'And then, since I was the pitcher, I said in my mind, that is the one who hit my sister earlier.' (Ramona, 28, RW)

c. Yo no me siento mal, al revés, yo me siento bien, porque [fui yo que le dije]
'I don't feel bad, on the contrary, I feel good, because it was me who told him.' (Bruno, 52, UW)

d. Cuando decían El Jefe nadie se equivocaba, [era Trujillo que llamaba.]
'When they said El Jefe nobody got it wrong, it was Trujillo who was calling.' (Federico, 81, UH)

e. Espectacular me pasó con una artista, que teníamos una producción en la cual [la que se fajó fui yo.]
'Spectacular is what happened to me with an artist, that we had a production together in which the one that did all the hard work was me.'
(Atilia, 50, UW)

The quantitative study carried out in this dissertation is not specifically defined to determine whether the strategies available in DS for the focusing of subjects are part of a system of focus assignment that reflects a significant restructuring of the Dominican grammar, or whether the discursive constraints that regulate the organization of the informational content in syntactic structures are applied in DS as well. However, the distribution of subjects in clefted constructions as attested in the corpus under study could contribute to the understanding of the Dominican Focus grammar; at the very least it could inform syntactic investigations on relevant research avenues.

7.4. On the formal definition of the Dominican subject expression grammar.

Informally, DS could be defined as a language that allows for both null referential and overt pronominal and expletive subjects. Overt subjects are almost categorically preverbal with the first and second persons singular, while null and overt variation is witnessed in the third person singular. This definition is very similar to the one offered for Brazilian Portuguese by Sheehan (2006) and Modesto (2000, 2009). However, as languages generally are complex and rich systems, (micro)dialectal variation is expected to arise between BP and DS. This question is addressed in

what follows. Also relevant for the definition of the Dominican grammar are the data from language change processes in Spanish. The data offered in the context of the ensuing discussion will show that the sort of microvariation attested in diachronic Romance derives also from the same core parameters of variation afforded by formal grammars.

Finally, the co-existence of overt expletives and null subjects has been found in non-typologically related languages, such as the Swedish dialect studied by Platzack (1987) discussed in section 7.4.2. While overt expletive syntax bears different properties in Scandinavian and Romance languages, the phenomena shared by these languages all offer additional evidence for split projections across the extended functional field. Leaving aside intervening factors like the V2 property of Scandinavian languages, mixed patterns in these languages are very similar to the ones found in Romance, which confirms that cross-linguistic and cross-dialectal variation are governed by the same set of constraints.

The language-specific issues that a theory of DS subject expression grammar needs to deal with are outlined in (30):

(30) On the nature of dialectal variation: The formal definition of the Dominican subject grammar.

Research question 17:

a. What is the specific role of rich agreement in subject licensing in DS? What mechanisms are available for Nominative Case assignment?

b. What is the nature in terms of semantic properties and feature assembly of Dominican overt subjects? Are they distinct from prototypical, NSL Spanish overt subjects?

c. Does widespread use of overt subject affect the discourse-pragmatic functions already conveyed through overt pronominalization in other Spanish dialects? If this is the case, are there innovative strategies in DS for the expression of discourse-related phenomena such as Focus?

Hypothesis 17:

a. As discussed in § 2.2.2. extensive research has shown that mechanisms other than 'rich' agreement are available cross-linguistically for the licensing of null subjects. Taking this finding as a point of departure, it will be argued, with Toribio (2000) and Sheehan (2006) that Case might be assigned configurationally in DS. Furthermore, it will be proposed, following Sheehan (2006), that subjects in Caribbean and Non-Caribbean Spanish belong in the TP domain, which entails that EPP features are licensed via subject movement to Spec, TP.

b. Following Camacho (2008), it will be proposed that Dominican overt pronominal subjects differ from their Non-Caribbean counterparts with respect to their feature composition. Consequently, the quantitative and qualitative evidence offered in this study lends support to theories that propose that different subject positions are available within TP in the Spanish clausal structure for the landing of subjects (Suñer 2003, Cardinaletti 2004, Camacho 2008). DS subjects are merged lower in the structure, which would explain their limited distribution, as well as the specific properties of Caribbean structures that require subject-verb adjacency.

c. The results found in variationist work have repeatedly shown that very similar discourse-related variables constrain overt and null subject insertion in both Non-Caribbean and Caribbean dialects. In theoretical work, however, a number of Caribbean-specific constructions related with the expression of discourse-related phenomena have been attested. Cases in point are clefted structures, which can convey Focus in Spanish and are widespread in Caribbean dialects. Carrying out a thorough investigation of the Focus

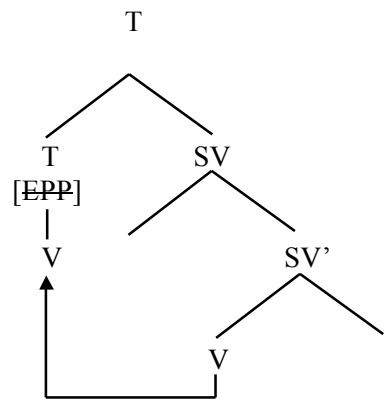
structures available to Dominican speakers is not among the main goals of this dissertation. However, from the observation of the clefted structures featured in the corpus at hand it will be suggested that the availability of vernacular Focus constructions in which the preferred SV word order is maintained is in fact related with the process of restructuring of the DS subject grammar.

In the following three subsections I address these questions in turn.

7.4.1. Case assignment in Dominican Spanish.

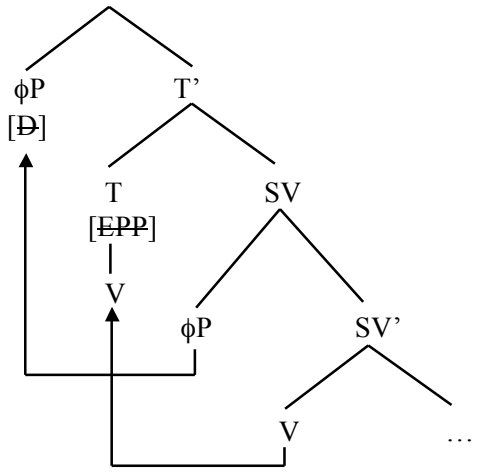
Recall from the discussion in Chapter Two that for *Hypothesis A* NSP theories (Alexiadou and Anagnostopoulou 1998, among others) agreement is interpretable and thus [+pronominal] in Null subject Romance; the inflectional endings of the verb incorporate interpretable [D], [phi] and [Case] features, which entails that the subject may stay in situ and the specifier position of T is not projected. As a consequence, these theories can dispense with the empty category *pro*. Overt subjects, when they occur, occupy Spec, CP, in consonance with the information structure associated with overt subjects in Spanish. Since Spec, TP is not projected, it stems from *Hypothesis A* theories that the default word order in Spanish is VSO, and not SVO as usually proposed. Overt subjects are attributed the same properties as Clitic Left-Dislocated constructions (CLDD). EPP licensing in Null Subject Romance as conceived by these theories is depicted in (31), which repeats (42) in Chapter Two. Nominative Case is assigned under Government:

(31)



The existence of the empty category *pro*, on the contrary, is maintained in *Hypothesis B* theories. Agreement is uninterpretable and *pro* is conceived as an ordinary pronoun, however without PF features. As such, it moves to Spec, TP to pick up the referential [+D] feature. Thus referential subjects belong in Spec, TP and the default word order in Spanish is SVO. Note that this claim does not entail that overt subjects with specific information structure meaning could not be merged in Spec, CP. What it does explicitly entail is that while it might be the case that some overt subjects are in Spec, CP it cannot be the case that all of them are.

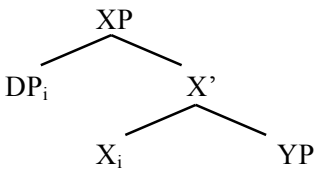
(32)



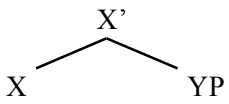
Case-assigment is carried out through Spec-head agreement in the structure in (32). The claim that mechanisms other than Government are available in Spanish was already put forward within the GB framework by Koopman and Sportiche (1991):

(33) Case assignment mechanisms (Koopman and Sportiche 1991):

a. Spec-Head agreement Case



b. Government Case



Sheehan (2006) translates Koopman and Sportiche's account in minimalist terms as follows: in some languages nominative Case is specified as [strong], which entails that it requires a local, Spec, Head configuration with the agreeing head T in order for the unintepretable feature [Nominative] to delete. This would be the case in a language like English, for instance, in which subjects surface necessarily in Spec, TP. In some other languages, Non-Caribbean Spanish being a case in point, nominative Case is [weak], and a DP specified as [nominative] can remain *in situ*, as long as the EPP is otherwise satisfied.

In other words, in Non-Caribbean Spanish both Spec-head agreement and Government agreement are available; only where another element satisfies the EPP, subjects are free to remain in-situ to check Case and agreement. This is clearly the case in instances of XPVSO word order:

(34) a. ¿Qué pasó ayer?
'What happened yesterday?'

b. Ayer ganó Juan la lotería.
yesterday won Juan the lottery

b'. #Ganó Juan la lotería.
won Juan the lottery
'Juan won the lottery yesterday'.

Since elements other than the subject can satisfy the EPP in Spanish, Sheehan (2006) proposes that Spec, TP, rather than the D-feature in Chomsky (1995), has the properties of an edge feature, as in Chomsky (2001). Edge features manifest the requirement for the specifier of a head to be filled. Sheehan's account goes as follows: if the head T bears a (Topic or Focus) discourse feature, the EPP associates itself with this discourse feature, triggering movement of a Topic or Focus element to Spec, TP. If there is no discourse feature in T, the EPP associates itself with phi-features, and attracts the subject to Spec, TP. This contrast, attested both with transitive and intransitive verbs, is illustrated in (35)-(36):

(35) a. ¿Qué pasó?

b. Juan ganó la lotería.

(36) a. ¿Qué pasó ayer?

b. Ayer Juan ganó la lotería.

In (35b) there is no Topic, and the whole sentence receives wide focus as new information; the EPP is satisfied by the subject, which agrees in phi-features with T and has its Case feature valued. In (36b) a Topic is present. The EPP is therefore construed as an edge feature and triggers movement of the adjunct ayer to Spec, TP. Agreement and Case checking between the subject and T proceeds as in (35b), the only difference being that no subject movement takes place, and thus the subject remains *in situ* in Spec, vP, and Case is valued via Agree.

Facts on *wh*-interrogatives also support *Hypothesis B* and the Spec-head agreement theory. Two alternative accounts have been put forward in the relevant literature to explain subject-verb inversion in Non-Caribbean Spanish: either the verb moves to C (Torrego 1984, Rizzi 1996), or the *wh*-phrase moves to Spec, TP (Zubizarreta 2008, Sheehan 2006). According the latter theory, and similarly to what just discussed regarding VS word orders in declaratives, the subject can remain in situ in Spec, vP as long as another element lands in Spec, TP and is able to satisfy the EPP. Case and Agreement are checked locally by the subject in Spec, vP, under Agree. Furthermore, the same

generalizations hold in embedded questions and in matrix contexts, suggesting that the inversion does not involve I-to-C movement.

Non-inversion, however, is possible with the complex wh-word *por qué* 'why' (Gallego 2006:19):

(37) ¿Por qué Juan quiere salir antes que los demás?

Why Juan wants to-go out before that the others

'Why does Juan want to go out before the others?'

(38) ¿Cómo Juan hizo eso?

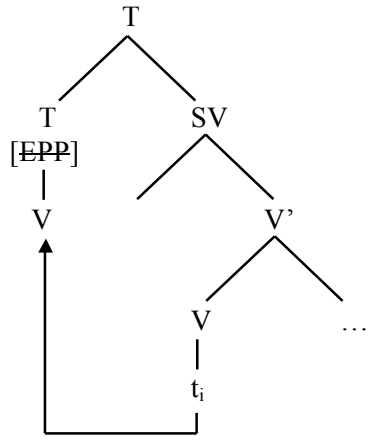
How Juan did that

'How did Juan do that?'

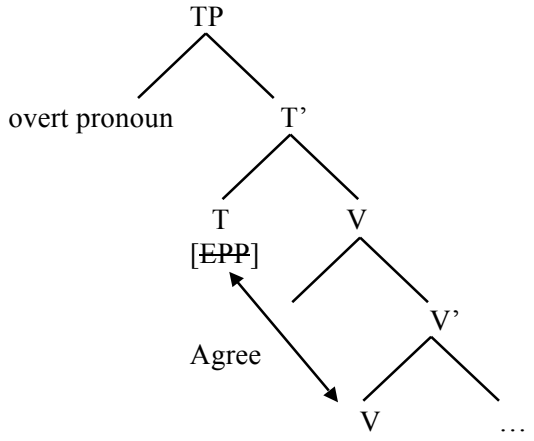
Sheehan takes these facts as evidence that the low *wh*-words can satisfy the EPP in Spanish. The high adverbs, as they are merged above TP, raise to Spec, CP. As a result, inversion only obtains optionally, subject to the information structure of the constituents.

Caribbean Spanish facts related to word order in interrogative and non-finite constructions also support the claim that Spec-head agreement is instantiated in Spanish. In my previous work (Martínez Sanz 2007) Spec-head agreement was proposed to derive the word order patterns of Dominican overt pronominal and overt expletive constructions. The differences between Non-Caribbean and Caribbean Spanish are represented in (39) (cf. also §2.3.4.):

(39) Non-Caribbean Spanish



Dominican Spanish



Suñer (1986) and Toribio (1993b) had proposed that Spec-head agreement is responsible for nominative Case assignment in Dominican infinitives as well.

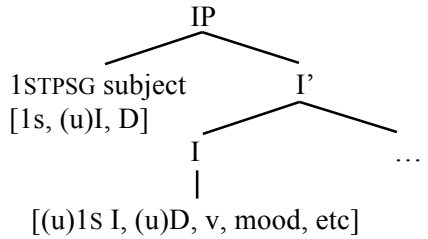
Suñer (1986), Toribio (1993b). In these structures, Case is assigned in structural terms in the absence of Agreement. The possibilities for linking Case assignment mechanisms with the properties of Dominican pronominal subjects are explored in the following section.

7.4.2. On the nature of null and overt pronominal subjects in Dominican Spanish.

The properties of subjects in Spanish have been the object of painstaking research in the last three decades within the generative framework. In current theoretical work, two main theories have been put forward to disentangle what are the properties of Spanish clausal structure.

In *Hypothesis A* theories (Alexiadou and Anagnostopoulou 1998) the licensing of null Romance subjects is straightforwardly derived, as inflectional morphology is attributed pronominal nature. In *Hypothesis B* theories (Sheehan 2006) a mechanism of pronominal deletion needs to be proposed. Following Roberts (2003), Sheehan claims that pronouns are deleted under 'feature non-distinctness'. In (40) the features of the pronoun are a subset of the features contained in Inflection. As a result, the pronoun is deleted under non-distinction:

(40) Feature non-distinctness (Sheehan 2006:224)



The theory depicted in (40) also derives SVO as the default word order in consistent NSLs. Two phenomena remain to be explained by this theory: on the one hand, the cases in which null non-thematic pronouns are compulsory, and the contexts in which overt pronominal referential subjects are equally categorical in Romance NSLs, on the other. The first case is instantiated by null expletive structures, and the second by Focus constructions.

As far as NSL expletive constructions are concerned, Sheehan proposes expletive deletion in Romance is also subject to feature non-distinctness as formalized in (46), i.e., null expletives are deleted under non-distinctness with T, just like null referential pronouns. This is so because the PF-features of expletives are a subset of those in T, which with it agrees in person only. Focus constructions bear an additional interpretable [+focus] feature that triggers overt insertion.

The next question that arises is how Dominican overt subjects are licensed in the theory represented in (40). Since unexpressed subjects are allowed in DS, the null hypothesis is that zero Dominican subjects are licensed like their Non-Caribbean counterparts. Overt subjects, on the contrary, display differing properties. This question has been resolved in previous work by appealing to grammars in competition in the minds of the speakers (Kroch 1989, Toribio 2000, Martínez Sanz 2007). In recent work this view has been challenged in light of language development and language change facts:

From a language change point of view, it is a common scenario to find different structures available for speakers to convey the same meaning; however, it is not clear whether this necessarily entails these optional phenomena correspond to differing mental grammars. As suggested in Chapter Two, it might be the case that these competing structures correspond to a single, and for a given period of time stable, grammar. From the point of view of language development, it is not infrequent either that this sort of 'optionality' becomes a core property of interlanguage or language contact grammars. Furthermore, recent work has shown that the deficiencies reported in morphological inflection in L2 and bilingual speech does not entail the erosion of syntactic properties. As discussed in §1.4, evidence for this claim is found in the properties of these grammars' clausal structure, which align them with a complete syntactic system with intact functional projections (Bullock and Toribio 2006).

In light of these facts and drawing on previous work (Cardinaletti 2004, Camacho 2008), I propose that to account for the DS null and overt subject patterns the clausal structure in (40) should be extended to include the different types of subjects available to Dominican speakers. With

Cardinaletti (2004), I will suggest that (40) does not exhaust yet the distribution of subjects cross-linguistically.

Cardinaletti suggests that the four properties attributed to preverbal subjects (being the subject of predication, checking the EPP feature, checking phi-features, checking nominative Case) should be deconstructed across discrete functional projections, each of which realizes a feature or a set of features. She proposes this to be the clausal structure of both NSLs and non-NSLs. Thus the preverbal subject field would contain the following projections:

(i) SubjP. This position hosts the subject of predication feature, and as such it typically hosts strong subjects. Preverbal dative and locative PPs (cf. §7.3 above), which do not agree with the verb and are not assigned nominative Case, also land in Spec, SubjP. Therefore, their movement cannot be motivated by the need for checking phi-features and Case. Moreover, Subj^{o} cannot be the locus of phi-features and Case. Subj^{o} must contain some feature that attracts this rather heterogeneous set of elements, and this is the subject-of-predication feature.

(ii) AgrSP, located below SubjP. Phi-features of nominative DPs are checked in this position, which hosts the grammatical subject. This results on nominative Case on the subject DP and subject-verb agreement. AgrSP is is the position of weak subjects, and of non-referential subjects. Referential subjects, which can be either weak or strong, can be merged in Spec, SubjP or in Spec, AgrSP. Topic non-nominative XPs also belong in this position, which independently argues for the existence of this functional projection.

(iii) EPP, which hosts a distinct functional projection.

(iv) DP and pronoun subjects occupy different (middlefield TP) positions, as evidenced by data from Hebrew (Shlonsky 2000:342):

(41) a. ‘Eyn Rina madaberet rusit.
not Rina speaks Russian

b. ‘Eyn hi madaberet rusit.

not she speaks Russian

c. ‘Eyn ze kase la-daber rusit.

not it difficult to-speak Russian.

(42) a. *’Eyn Rina gveret Levi.

not Rina Mrs Levi

b. ‘Eyn hi gveret Levi.

not she Mrs. Levi

‘She is not Mrs. Levi.’

Cardinaletti and Starke (1999) establish a tripartite typology of pronouns in natural languages: strong, weak and clitic pronouns. Both strong and weak pronouns may appear as subjects in a given language (the Italian pronouns *lui* and *egli*, strong and weak respectively, being a case in point), and they are syntactically, semantically and prosodically distinct. Basically, weak pronouns are ‘defective’, in a sense that needs to be specified, as compared to their strong counterparts.

Suñer (2003) specifically suggests that overt pronouns in Caribbean varieties such as DS are defective as compared to Non-Caribbean Spanish pronouns. She characterizes DS pronouns as weak, and submits that as a consequence they are limited in their distribution: while Non-Caribben strong subject pronouns like *él*/*ella* can be separated from the verb by a parenthetical, the Dominican overt expletive pronoun *ello* cannot.

(43) a. Juan/Él, a mi parecer, es muy simpático. (Suñer 2003:5)

Juan/he, according to me is very nice

‘Juan/he according to me is very nice.’

b. Ello (*a mi parecer) no sería malo estudiar

it according to me not would be bad to study

‘It (according to me) wouldn’t hurt to study.’

This suggests, according to Suñer, that *ello* is structurally closer to the verb than strong pronouns and lexical NPs, and therefore it is merged lower in the structure, as represented in (44):

(44) $[_{YP}$ {Juan/él} $[_{XP}$ parenthetical $[_{ZP}$ {ello} V]]]

Camacho (2006) finds further evidence to support the claim that Dominican pronouns are merged lower than strong pronouns in subject adjacency effects attested elsewhere in the Spanish grammar: in addition to the adjacency effects seen in (44) with the Dominican pronoun *ello*, preverbal negative quantifier subjects in both DS and Non-Caribbean Spanish must also be adjacent to the verb (46a-b), by contrast to lexical subjects, postverbal quantifiers, and CLLDed items (cf.46a-c):

(45) a. Nadie pudo avanzar tres metros
no-one could advance three meters
'No one could advance three meters.'

b. *Nadie casi pudo avanzar tres metros.
no-one almost could advance three meters
*Ninguno de ellos casi pudo avanzar tres metros
none of them almost could advance three meters

(46) a. La tortuga casi pudo avanzar tres metros
the turtle almost could advance three meters
'The turtle could almost advance three meters.'

b. Casi no pudo avanzar tres metros nadie
almost not could advance three meters no-one
'No one could almost advance three meters.'

c. A ninguno de los culpables, casi lo descubrieron.
to none of the guilty almost cl discovered
'None of the guilty, were almost discovered.'

Thus the preverbal subject field according to Cardinaletti (2004) would look as in (47):

(47) $[_{SUBP}\ [_{AGRSP}\ *\ [_{TP}\ldots\ [\ldots\ [_{VP}\ldots\ ]]]]]$

The derivation of sentences containing preverbal subjects looks as follows: weak subjects such as the Italian null subject *pro* only check nominative Case and phi-features and occur in Spec, AgrSP. Strong subjects such as DPs and Italian strong pronouns like *lui* continue their derivation to Spec, SP since they also check the subject-of-predication feature:

(48) a. $[_{SUBP}\ [_{AGRSP}\ pro_i\ Vfin\ [\ldots\ [_{VP}\ldots\ t_i\ ]]]]$

b. $[_{SUBP}\ Gianni_i/lui_i\ /egli_i\ [_{AGRSP}\ t_i\ Vfin\ [\ldots\ [_{VP}\ldots\ ti]]]]$

Both subject positions are below the lowest Comp projection, FinP (Rizzi 1997) and thus occur in the Infl domain. The articulation of the clause would look as in (49):

(49) (ForceP TopP* FocusP Fin P SubjP AgrSP TP… VP

Comp domain Infl domain verb domain

functional domain lexical domain

The landing sites for weak and strong pronouns are depicted in (50):

(50) strong (and some weak) weak

$[_{SUBP}\ \{Gianni/lui/egli\}$ $[_{XP}$ parenthetical $[_{AGRSP}\ \{pro/tu_{weak}\}$ Vfin […]]]]

Camacho (2008) builds up on Cardinaletti (2004) in his account of DS. He proposes that they key feature specifications that have shifted in Caribbean pronominal systems are related to the feature [±referentiality], on the one hand, and to the feature related to the nature of a given pronoun, on the other. Bearing a [weak] feature implies a requirement for the pronoun to be attached to a head, and a syntactic requirement that this head must be Inflection.

Subject pronouns in consistent NSLs such as Null Subject Spanish are strong, and consequently specified as [+referential]; they bear no [weak] feature, as represented in (51) for a null subject pronoun and an overt third person singular pronoun:

(51) Null Subject Spanish

a. *pro* {α person, α number α ref }

b. *ella* {3rd person, sg, +ref }

If an overt strong pronoun entered the derivation in Non-Caribbean Spanish, two [+referential] features would be present in the computation: a [+referential] feature in Inflection and a [+referential] feature in the pronominal; since the operation Agree must be established between [+] and [–] versions of a feature, it would not be possible in this context. For this reason, the pronoun appears in its null version and the derivation contains a single [+referential] feature. For an overt pronominal to be felicitous in Null Subject Spanish, it would need to carry additional features, related with the information structure of the clause, such as contrast or emphasis.

Overt pronouns in BP and DS, on the contrary, are specified as [-referential] and [weak]. Bearing a [-referential] feature makes the pronoun a potential semantic variable, distinct from a [+referential] item, which is potentially anaphoric. However, [+anaphoric] pronouns are also available in DS, since null subject Non-Caribbean Spanish-like structures are available in this variety as well. This leads Camacho to propose that in DS there are two independent lexical entries for pronouns, one with the [-referential] feature and one without it, as represented in (52):

(52) Dominican Spanish

a. *pro* {α person, α number -ref }

b. *ella* {3rd person, sg, weak, +ref }

c. *ella* {3rd person, sg, weak, -ref }

Postulating two different lexical entries for pronouns is theoretically justified on Cardinaletti and Starke's (1999) tripartite classification of pronouns. Furthermore, it is empirically supported by Spanish facts: note that the availability of two lexical entries for pronouns is also found in Non-Caribbean Spanish; specifically, third person plural pronouns can be specified either as [+referential], in specific readings, or as [-referential], in arbitrary interpretations that convey an impersonal meaning. Relevant examples and the feature specification of the pronouns are represented in (53)-(54):

(53) a. (Ellos) trajeron el vino a la fiesta.
they brought.3PPL.SPECIFIC the wine to the party
'They brought the wine to the party.'

b. *-n* {3rd person, pl, +ref }

(54)

a. Dicen que beber mucha agua es bueno para la salud.
they say. 3PPL.ARBITRARY that to drink a lot of water is good for the health
'They say that drinking a lot of water is good for your health.'

b. *-n* {3rd person, pl, -ref }

According to Camacho, in a variety like DS once the pronominal loses its exclusively [+referential] interpretation and a separate, [-referential] lexical is created, overt pronouns are able to be merged in Spec, TP and satisfy the EPP like in English-type languages. He further argues that once the weak pronominal is able to check the EPP in T, the verb's inflection loses its pronominal nature.

Camacho also notes that there are relevant interpretative differences among overt pronouns that bear a [+referential] feature, which makes them potentially anaphoric, and [-referential] pronouns. Montalbetti (1984) first formulated these differences in the *Overt Pronoun Constraint*. This *Constraint* states that whenever there is a contrast between overt and null pronouns, the overt pronoun cannot be bound as a variable:

(55) a. Todo estudiante$_i$ cree que *pro*$_i$ es inteligente. (Non-Caribbean Spanish)
every student thinks that *pro* is intelligent
'Every student$_i$ thinks that he$_i$ is intelligent.'

b. Todo estudiante$_i$ cree que él$_{*i/j}$ es inteligente.
every student thinks that he is intelligent
'Every student$_i$ thinks that he$_j$ is intelligent.'

In the varieties in which pronouns are not exclusively linked to [+referential] readings, such as BP and DS, we would expect for the Avoid Pronoun Constraint to not apply. This is in fact the case; in embedded clauses, pronouns that are coreferential with matrix subjects can be expressed overtly in BP, as noted by Duarte (2000) and exemplified in (56a), and in DS, as pointed out by Toribio (2000); these examples are also found in the corpus gathered for this study, as illustrated in (56b):

(56) a. Você, no Canadá, você pode ser o que você quiser. (Duarte 2000:191).

b. Tio Papi$_i$ murió tan feliz que *pro*$_i$ no sabe que él$_i$ murió.

Tio Papi died.3PSG so happy that *pro* not know.3PSG that he died

'Tio Papi died so peacefully that he doesn't even know that he's dead.'

The proposal in Camacho (2008) has the significant advantage that it dispenses with postulating two grammars in competition in the minds of DS and BP speakers. The co-existence of null and overt subject structures is accounted for by availing of two separate lexical entries for at least a subset of the pronouns in the paradigm, a possibility that is already attested in the grammar of Spanish for third person plural pronouns, and in other Romance languages like Italian, in which distinct weak and strong pronouns show up as phonetically distinct (*lui* versus *esso*).

The question of what is the source of this sort of diachronic shifts is approached in the following section.

7.4.3. On syntactic change.

Bullock and Toribio (2006) explore to which extent the shifts attested in bilingual systems parallel the processes found in diachrony. Their investigation of two bilingual systems shows that change in the morphological specification of grammatical features need not signal any fundamental loss of formal syntactic properties. While a bulk of recent work has focused on the linguistic domains more vulnerable to change (the peripheral interface components PF and LF, Paradis 2000), Bullock and Toribio investigate how morpho-syntactic change fits within a restrictive syntactic model.

The dissociation between the exponents of inflection and the availability of higher functional categories has been consistently found in previous work on acquisition and bilingualism (cf. Haznedar 2001 and references therein). Most interestingly for the purposes of the present investigation, according to Bullock and Toribio (2006) this dissociation does not constitute a case of 'defective' or incomplete acquisition, but a case of 'incomplete replication'. Their theory is constructed by reference to interview data gathered in Frenchville, Pennsylvania and samples from the diary of a Mexican migrant worker; target-deviant morpho-syntax in these two data sets is exemplified in (57)-(58):

(57) Alterations in verbal morphology: French.

a. Ma femme ne parler pas; sa femme ne parler pas (cf.*parle*).

'My wife doesn't speak; his wife doesn't speak.'

b. Ça prend pas trop de temps pour le perdre si tu ne parler pas. Je parler avec lui de temps en temps… (cf. *parles, parle*).

'It doesn't take a long time to lose if you don't speak it. I speak with him from time to time'.

c. Eux allaient à l'école parler français i pourraient pas parler anglais (cf. *parlant*).

'They went to school speaking French; they couldn't speak English.'

(58) Alterations in verbal morphology: Spanish.

a. Si la cierras se quebra (cf. *quiebra*).

'If you close it it breaks…'

b. Sabes salemos todos negríos de la labor (cf. *salimos*).

'We come out all black from the field.'

c. Es nativo de C. donde vivamos nosotros (cf.*vivimos*).

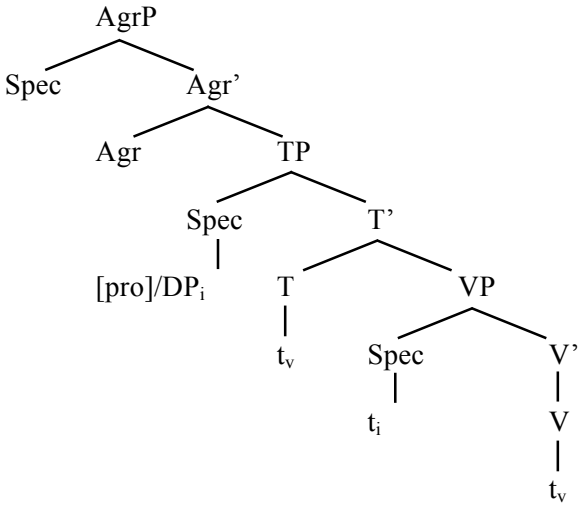
'He is a native of C. where we live.'

The Spanish data feature the presence and projection of uninterpretable features and functional projections that license postverbal and null referential subjects, as in (59):

(59) a. Pero no estaban saliendo los aviones por la tempestad que *pro* había.

Estaba lloviznando bien recío. *pro* llegamos a T. a las doce de la noche.

b.



In similar fashion, there is evidence of the articulation of structure of the left periphery (60). This is also attested in the data set gathered for this dissertation (cf.61a'). The structure of these constructions is represented in (60b).

(60) a. Todos mis gastos A.C. pago por ellos, yo ni dinero traíba. [sic]

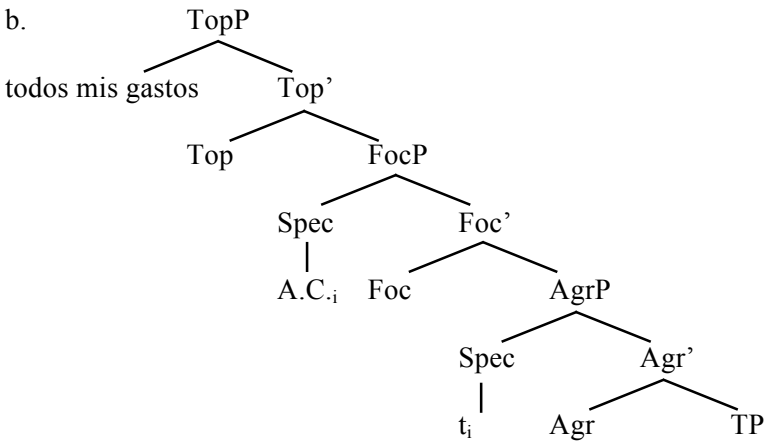
'All my bills, A.C. paid for them, I didn't bring any money.'

(Bullock and Toribio 2006:6).

a'. Las rodillas se las desbarató

'Her knees, she severely injured them.

(Maria Amelia, 56, UFH).



The authors point out that the inflectional deficits that emerge in bilingual speech are very similar to rather frequent 'impoverishment' processes of language change, such as semantic extension. A process along the lines of semantic extension is attributed in Bullock and Toribio to account for the over-extension of native-language syntactic options into new pragmatic contexts. Specifically, bilinguals tend to reduce the syntactic options available for expressing pragmatic differences in a language, and they do so by keeping the structure that is the most congruent across languages (Muysken 2000). This process if usually referred as *convergence*. Consequently a child acquiring French in Frenchville may reinterpret the preference for certain structures in specific pragmatic contexts as the syntactic rule. The former pragmatic contrast, as a result, is no longer available in the bilingual's grammar. Similarly, the child acquiring Spanish from the input provided by the Mexican diary writer could be led to reanalyze the grammar of this language as one with compulsory overt subject pronouns and fixed SVO word order. In other words, a new grammar could be acquired on the basis of the available variable input. Most importantly, all the variants available in the input where syntactically were formed. Processes of semantic extension or reinterpretation of word orders associated with semantic or pragmatic contrasts have also been proposed for the distribution of personal infinitives over time.

7.4.4. Personal infinitives in diachrony.

Lipski (1991) studies non-finite structures with nominative subjects. His comparison of these constructions in a number of Caribbean, vestigial, and non-standard varieties allows him to

probe into the historical origin of this construction. Lipski's hypothesis is that infinitives with preverbal subjects might have arisen through a process of ambiguity resolution. Specifically, the preposition *para*, 'for' is usually chosen in Spanish to introduce infinitival clauses, not only when the following constituent is nominative, but also when it bears oblique Case:

(61) Es difícil para mí hacer eso. (Lipski 1991:205)
is difficult for me.CL.DAT do.INFIN that
'It is difficult for me to do that.'

The sentence in (61) involves a pivotal ambiguity, which would trigger, in a subsequent stage, the use of the infinitive with "objective subject Case" (Lipski 1991:206), like in the Louisiana Isleño example in (62):

(62) Para ti tené un bote tienes que ser sosedano americano.
for you.CL.DAT have.INFIN a boat have-to be citizen American
'For you to have a boat you have to be an American citizen.'
(Lipski 1991:202).

When examples like (62) arise, a grammatical tension is set between the requirement for oblique Case of the preposition, and the reanalysis of the objective pronoun as nominative. Since, according to Lipski nominative is the unmarked Case in Spanish, the tension is resolved in favor of the nominative and the reanalysis can be extended to other prepositions:

(63) a. replace PRO by [NP, +lexical]; or, equivalently:
b. adjoin [NP, +lexical] to any Infinitive.

The explanation in Lipski (1991) does not rely on a process related to the weakening of agreement morphology in Caribbean Spanish, *contra* Suñer (1986). This allows him to derive synchronic personal infinitives from language change driven by maximally unmarked processes. In addition, this account allows for the preverbal merging of both pronominal and DP subjects. Recall that preverbal lexical subjects are not attested in infinitive constructions in the data set under study. However, the fact a preverbal position is independently postulated for the insertion of preverbal

lexical material in *wh*-interrogatives, their availability is predicted by the theory put forward in the present investigation. Only a fine-grained syntactic investigation of the distribution of lexical subjects with infinitives could confirm whether this structural possibility is exploited in DS. In the absence of those data, further evidence for the claim that shifts in the DS subject expression grammar can be accommodated within the Spanish clausal structure can be found on the one hand, in restructuring patterns in other DS structures, and on the other, in the diachronic shifts undergone by other typologically and non-typologically related languages. I start by discussing the latter type of evidence with Scandinavian data (Platzack 1987).

7.4.5. Null subjects in Mainland and Insular Scandinavian.

In Platzack's (1987) view, Icelandic (Insular Scandinavian) is a [+null subject] and Swedish (Mainland Scandinavian) is a [-null subject] language. He bases this claim in the differences attested between both languges:

(i) Icelandic allows for null referential subjects and null expletives. In Swedish, expletive subjects are overt. Icelandic, however, allows for overt expletive structures as well; examples and discussion are provided below.

(64) a. Í dag hafa (*pad) komid margir málvínsindamenn hingad.
(Icelandic)

b. I dag har *(det) kommit många lingvister hit.
(Swedish)
today have there come many linguists here
'Today there have arrived many linguists.'

(ii) Icelandic has number agreement between the finite verb and the subject. Swedish does not have agreement:

(65) a. Hann heggur. Beir höggva. (Icelandic)

b. Hann hugger. De hugger. (Swedish)
'He hews. They hew.'

(iii) Heavy subject proposing, which shifts the subject to a position adjoined to VP, is possible in Icelandic but not in Swedish:

(66) a. Pad munu kaupa bessa bók margir stúdentar. (Icelandic)

b. *Det vill köppa denna bok många studenter. (Swedish)

'Many students will buy this book.'

(iv) Icelandic allows that-trace filter violations, whereas such extractions are not possible in Swedish:

(67) a. Hver heldur pú (ad) hafi framid glæpinn? (Icelandic)

b. Vem tror du (*att) har begat brottet? (Swedish)

'Who do you believe has committed the crime?'

(v) In Icelandic, non-nominative DPs may behave syntactically and semantically like a subject:

(68) a. Hana vantar peninga. (Icelandic)

her (A) lacks money (A)

b. Hon saknar pengar (Swedish)

she (N) lacks money

'She lacks money.'

(vi) In Icelandic, the fronting of elements in subjectlss subordinate clauses or Stylistic Fronting is allowed:

(69) a. Barna er konan, sem kosin var/var kosin forseti (Icelandic)

b. Där är kvinnan som (*vald blev)/blev vald till president (Swedish)

There is the-woman that chosen was/ was chosen (to) president

'There is the woman that was elected president.'

Platzack attributes these differences to the [+pronominal] specification of agreement in Icelandic and its [-pronominal] specification in Swedish, in line with *Hypothesis A* theories like

Alexiadou and Anagnostopoulou (1998). Thus the differences between Icelandic and Swedish illustrated in (70), would be assigned the structure in (71):

(70) a. badan var (*pad) skammt til bæja. (Icelandic)

b. Därifrån var *(det) ikke langt til gårdene. (Swedish)

from-there was there short-way to the-farms

'There is a short way to the farms from here.'

(71) a. [CP XP [CP [C var] [NP t] IP]] (Icelandic)
[+pronoun]

b. [CP XP [CP [C var] [NP det IP]] (Swedish)

Icelandic, however, may display overt expletive subjects in the familiar Transitive Expletive Constructions (TECs):

(72) Pad hafa komid margir málvísindamenn hingad.

There have come many linguists here.

Platzack proposes that Icelandic expletives occupy a position akin to Spec, CP. The Icelandic facts show like for the TP field, which is articulated in a number of functional projections, extended clausal structure is evidenced within CP as well. Overt Scandinavian expletives occur in CP because of the V2 property shared by all Germanic and Scandinavian languages. Thus in Platzack's theory, cross-linguistically the [±pronominal] feature may be a property of Inflection or a property of C.

This difference is not only attested cross-linguistically, but also within the same language-specific grammar. Constructions depending on the presence of [+pronoun] in C occur together in some Mainland dialects, and they were syntactically productive in earlier stages of Swedish. The former case is instantiated by the null subject Swedish dialect Älvdalsmålet. This dialect bears verb-subject agreement, allows for tensed clauses with zero subjects, and displays stylistic fronting, as exemplified in (73a-b):

(73) a. Upo åsem ova riä slaid åv.

on the-ridge has already been-cut off

'The harvesting has already been done on the ridge.'

b. Han…so gamble uar.

he that older was

'He…who was older'

In sum, Scandinavian facts instantiate the type of syntactic variation described for DS. Moreover, Dominican and Scandinavian subject patterns provide support for the cartographic approach to syntactic structure.

Now that the variable pattern of null subject expression of DS as it emerges in the corpus under study has been discussed and it has been placed in the context of other dialectal and diachronic variation processes, one last issue remains to be approached: given that overt pronominalization and preverbal subject insertion conveys specific semantic-pragmatic information in Non-Caribbean Spanish, a relevant question to pose is whether in a variety like DS, in which overt preverbal subjects reach very high rates of use, alternative strategies for conveying this information have been developed. This question is explored in the following section.

7.4.6. On the pragmatic functions of overt pronominalization in Caribbean and Non-Caribbean Spanish.

As explained in Chapter Four, the main finding in cross-dialectal variationist work is that discourse-related factor groups strongly constrain subject expression in all the dialects studied up to date. Dominican Spanish has been proved to be no exception in this respect in this study: *Distance from previous mention* is among the top three variables in the logistic regression analyses that measure null and overt subject distribution. If discourse-related constraints are compared to each other, *TMA of the previous verb* is on other side of the spectrum, with much weaker effects on null and overt subject patterns. *Co-reference* and *Discourse function* have also emerged as significant in most calculations.

This could be interpreted as a in intriguing result in the study of DS, since in several other respects this variety has been found to depart from its Caribbean and Non-Caribbean counterparts. Note however that the discourse-related factor groups investigated in previous variationist work are associated with the notion of *continuity* in discourse; roughly, *continuity* in verb tense, subject form, and reference favors null subjects, while non-continuity disfavors them. The effects of continuity are attested as well in the data set under examination. Nonetheless, differences emerge among subject types: *Discourse function* is a very significant constraint for the distribution of lexical subjects and third person singular pronominal subjects, whereas *Distance from previous mention* is more relevant for first and second person subjects.

Co-reference is a mid-ranked factor group across the sets of results presented in Chapter Six. Interestingly, this factor group has emerged as the strongest constraint in Spanish subject expression in previous work (Cameron 1992, 1996, Flores-Ferrán 2004). As already discussed in §7.2, Cameron (1996) reports that in the Spanish of San Juan, Puerto Rico overt subjects are favored with non-co-referent subjects at a factor weight of .65, and disfavored with co-referent subjects at .35 (*Range*=30). In DS, as outlined in §6.3.2, co-referent and non-co-referent overt pronominal subjects are significantly closer in factor weights, .43 and .60 respectively (*Range*=17). Poplack and Tagliamonte (1991), Tagliamonte (2006) stress the relevance of *Range* numbers for the comparison of factor groups across dialects. The fact that *Co-reference* is a stronger factor group in Puerto Rican than in Dominican Spanish would reflect then that, as a result of rates of overt subject use being higher in DS, co-referent subjects do not disfavor expressed subjects as strongly.

Furthermore, the finding that *Distance from previous mention* is very relevant for in DS subject use, as discussed also in §7.2, could also be interpreted as pointing to variation in the strength and nature of the discourse-related factor groups that constrain null and overt subject distribution in the Caribbean. Specifically, *Distance from previous mention* operationalizes potential constraints on subject insertion that range from reference tracking to memory and processing issues.

Discourse-related factor groups that operationalize purely linguistic-internal constraints, so to speak, do not arise as significant in DS.

Discourse-related factor groups outside the realm of continuity have not been explored as thoroughly in previous variationist work. Among them *Person* and *Clause Type* stand out. Both of these factor groups are usually discussed in the literature among morpho-syntactic constrains on subject expression. However, note that since variation in discursive behavior is attested across the singular paradigm, *Person* could also be discussed in the context of discourse-related constraints. Similarly, since the impact of the configuration of main clauses extends to subordinate and relative clauses, *Clause Type* distinctions extend beyond the sentence boundary.

The effect of the different nature of the deictic first and second persons versus the potentially anaphoric third person has been detected in the data set under study (cf.6.2.3 and 6.3.3.). These properties should also be taken into account for the explanation of the Dominican subject grammar. In fact, recall that overt pronominal subjects are strongly favored with first and second person subjects, while overt lexical subjects are favored with the third. This result has been interpreted here as reflecting the differences between first-second person and the third person in the possibilities of recovering the reference of subjects in the discourse. Differences in lexical and pronominal subject use have also emerged in the examination of different clause types; relative clauses have been shown to favor overt pronominal subjects and disfavor overt lexical subjects. Since relative clauses involve an antecedent in the previous discourse, the speaker does not need to resource to maximally non-ambiguous lexical subjects.

Perhaps more interesting from the perspective of grammatical restructuring are the shifts attested in DS with respect to subject positions. Recall from the discussion in §2.3.4 that subject position in Spanish is regulated, among other factors, by information structure constraints. Both syntactic and prosodic strategies are available in Spanish for encoding old and new information in the discourse.

As SVO is the default word order in Spanish, preverbal subjects typically encode Topic constituents, while postverbal subjects introduce new, focused information. Matters have been shown to be more complex than this, however. Preverbal subjects can bear contrastive focus if proper intonation is added.

Moreover, wide focus, where the whole sentence is focused, might be conveyed by SVO and VSO word orders, although VSO is the unmarked option. Finally, narrow focus, in which a constituent is focused, also triggers postverbal subjects; in this case, both VSO and VOS word orders are possible. In addition, syntactic and prosodic constraints interact for narrow focus assignment (cf. §2.3.4. and the ensuing discussion). Table 7.4. summarizes the word order possibilities of Spanish:

Table 7.4. Focus and word order in Spanish.

FOCUS	PREVERBAL	POSTVERBAL	PROSODY
CONTRASTIVE	✓	X	✓
WIDE	✓	✓ *	X
NARROW	X	✓ **	✓

*VSO is the wide focus unmarked word order.
** both VSO and VOS are possible.

Contrastive focus in DS will not be addressed in depth in this study, as a prosodic analysis of subjects in the data set was not among the goals of the investigation. As far as wide focus is concerned, it may be obtained via SVO word order in Spanish, and therefore does not pose a problem for the DS subject expression system. The relevant question is then how is narrow focus assigned in a Spanish dialect that shows a strong preference for preverbal subjects. The proposal put forward in this dissertation is that Dominican speakers resort to the syntactic strategies available in Spanish for narrow-focusing subjects, by availing of specific (vernacular) clefted structures that allow maintaining the preferred SVO word order.

Zubizarreta (1998) put forward a theory on narrow and wide focus assignment in Romance in which word order effects are derived from constraints at the prosody/syntax interface. The core

elements of Zubizarreta's analysis of narrow focus, presented in §2.3.4, are provided again below for ease of discussion:

(i) All natural languages are subject to the *Focus Prominence Rule*:

(74) The Focus Prominence Rule

Given to sister categories C_i (marked [+Focus] and C_j (marked [-Focus], C_i is more prominent than C_j. (Zubizarreta 1998:21)

(74) formally states that focused constituents are more intonationally prominent than unfocused constituents. Intonational prominence is assigned to focused constituents via a *Nuclear Stress Rule*. In Zubizarreta's analysis prosody corresponds directly to information structure in Spanish, which entails that nuclear stress encodes narrow focus; nuclear stress will fall on the 'lowest' constituent in the structure:

(ii)

(75) The Nuclear Stress Rule (NSR)

Constituent-driven NSR. Given two sister categories Ci and Cj, the one lower in the asymmetric c-command ordering is more prominent.

(Zubizarreta 1998:19).

Differences in word order in Romance and Germanic languages are attributed by Zubizarreta to a 'prosodic invisibility' constraint that applies in the latter but not in the former: in Germanic languages and French, defocalized constituents are invisible to the NSR, whereas in Spanish and Italian they are not. For this reason preverbal nuclear stress can be assigned in English and French, but not in Spanish or Italian. The examples in (77a-d), all answers to the question 'Who ate the cake?', show the differences in metrical (in)visibility between French and Spanish. Metrically invisible constituents are in italics:

(76) a. Marie ate the cake.

b. Marie a mangé le gateau.

Marie has eaten the cake

'Marie has eaten the cake.'

c. *María ha comido el pastel.

María has eaten the cake

'María has eaten the cake.'

d. Ha comido el pastel María.

Has eaten the cake María

'María has eaten the cake.'

If there is a mismatch between the NSR and the FPR, the alternative solution in Spanish according to Zubizarreta is what she labels P(rosodic)-movement:

(iii)

(77) P-movement. This operation occurs in the phonological component and triggers movement of the object over the subject so the NSR can apply and assign nuclear stress, and narrow focus, to the low subject.

P-movement is responsible for the fact that Spanish VOS is triggered only when the subject receives narrow focus. Therefore, in VOS word orders, the object cannot bear contrastive focus.

French, which Zubizarreta aligns with Germanic languages in her theory, might provide a hint on how narrow focus assignment is carried out in DS. In French, while a sentence like (77b) may convey narrow focus on the subject, the most natural answer to a *wh*-question is a clefted structure, as in (78) (Belletti 2005, Hamlaoui 2007, Eric Mathieu, p.c.):

(78) a. Qui est-ce que a mangé un biscuit?

who is-it that has eaten a cookie

'Who ate a cookie?'

b. C'est Ella (qui a mange un biscuit).

it-is Ella that has eaten a cookie

'Ella ate a cookie.'

In (78), the clefted XP, i.e. the lower element in asymmetric c-command bears main stress, in accordance with (75). Zubizarreta explained structures like (78) by claiming that metrical invisibility is optional in French. Maintaining the view that metrical invisibility generally applies in French, however, entails proposing *ad hoc* solutions for clefted constructions like (78). This is the position taken for instance in Belletti (2005), who submits that the focused constituent appears in a Focus-dedicated position in the left-periphery of VP. On the contrary, Hamlaoui (2007) views clefted structures as a strategy that allows subjects to appear in the position where main stress is assigned, that is, the right-periphery of the clause.

Hamlaoui points out that in terms of stress, French is usually described as a rightward oriented language. To formulate a rule for stress assignment in French, she builds up on the ideas in (i) and (ii):

(i) Selkirk's (1984) prosodic hierarchy: intonation phrase (IntP) > phonological phrase (φ) > prosodic word (ω).

(ii) Szendröi's (2003) stress rule, formulated for English and Italian, according to which main stress is directly assigned in the prosodic structure[62]:

(79) Stress-focus correspondence principle:

62 In Szendröi's approach, prosodic and syntactic structure are associated through the mapping rules in (i)-(ii):

(i) Syntax-phonology mapping rules for phrases:
Align the right edge of a syntactic phrase with the right edge of a phonological phrase.

(ii) Syntax-phonology mapping rules for clauses:
a. Align all the left edges of the largest extended projection of the verb with the left-edge of an intonation phrase.
b. Align all the left-edges of the intonation phrase with the left-edges of the largest extended projection of the verb.
c. Align all the right-edges of the largest extended projection of the verb with the right-edge of an intonation phrase.
d. Align all the right-edges of the intonation phrase with the right-edges of the largest extended projection of the verb.

The focus of a clause is a(ny) constituent containing the main stress of the intonation phrase.

(Szendröi 2003:47)

Hamlaoui proposes the rule in (80) for French:

(80) French stress rule

a. Assign a Strong label to the rightmost phonological word in the phonological phrase. Assign Weak to the other phonological words.

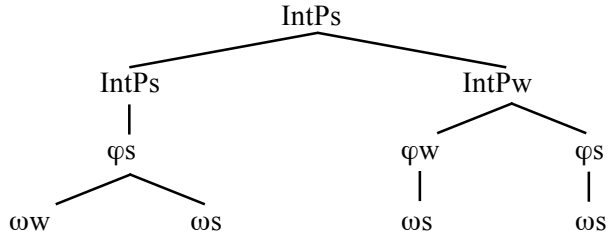
b. Assign a Strong label to the rightmost phonological phrase in the intonation phrase. Assign Weak to the other phonological phrases.

c. Assign a Strong label to the intonation phrase.

(Hamlaoui 2007:3)

In (81), nodes are assigned Weak (w) and Strong (s) labels; the node that is only dominated by Strong labels is the one that bears the main stress. The theory in (81) derives that main stress is assigned to the rightmost constituent in the intonation phrase in French in clefted structures:

(81)



[$_{TP}$ [$_{TP}$ C'est [$_{VP}$ [$_{DP}$[Ella]]] [$_{CP}$ qui a mange [$_{DP}$ un biscuit]]]

Clefted structures offer two potential focus positions, as they are made of two clauses and thus two intonation phrases. These focus positions display different properties, however: Although the cleft sentence offers two potential focus positions, as it is made of two clauses and therefore two intonation phrases, these positions are not equivalent. In (81), the intonation phrase aligned with the relative clause adjoins to the intonation phrase associated with the main clause. The stress rule in (80) would normally apply to both intonation phrases; however, the main stress in (81) falls on the right-periphery of the main clause, since the adjoined intonation phrase receives a Weak label and

its extrametrical at the level of the highest intonation phrase. Nonetheless, within the intonation phrase to which it belongs, the phonological phrase [$_{\varphi}$ un biscuit] receives the main stress.

In sum, French narrow-focuses subjects by resorting to a construction that allows it to not relocate the main stress. Therefore, according to Hamlaoui (2000) and *contra* Zubizarreta (1998), French departs from English in narrow focus strategies.

Interestingly, Hamlaoui points out that object and adjunct narrow-focusing is not carried out as in (82). Rather, a cleft sentences is a non-felicitious answer to object and adjunct *wh*-questions in French:

(82) a. Qu'est-ce que Juan a mangé?
what is-it that Juan has eaten?
'What did Juan eat?'

b. ? C'est un biscuit (que Juan a mangé)
it-is a cookie that Juan has eaten

c. (Juan a mangé) un biscuit.
'Juan has eaten a cookie.'

(83) a. Quand est-ce qu'Ella a mangé un biscuit?
when is-it that Ella has eaten a cookie
'When did Ella ate a cookie?'

b. ? C'est hier (qu'Ella a mangé un biscuit).
it-is yesterday that Ella has eaten a cookie

c. (Ella a mangé un biscuit) hier.
'Ella has eaten a cookie yesterday.'

In (82) and (83) the constituents appearing in right-peripheral positions do not need to appear in a cleft construction in order to receive narrow focus. In this context, the cleft sentences in (82b) and (83b) constitute the marked variant, as they are computationally more costly than its canonical counterpart. Structures such as (82b) and (83b), however, are felicitous in contrastive

contexts. As exemplified in (84b) and (85b), when objects and adjuncts bear contrastive focus, they naturally appear in cleft sentences. Canonical sentences, on the other hand, are not felicitous in these contexts:

(84) a. Juan a mangé un prune.
Juan has eaten a plum
'Juan ate a plum.'

b. Non, c'est un biscuit que Juan a mangé.
no it-is a cookie that Juan has eaten

c. ? Non, (Juan a mangé) un biscuit.
no Juan has eaten a cookie.

(85) a. Ella a mangé un biscuit ce matin.
Ella has eaten a cookie this morning
'Ella ate a cookie this morning.'

b. No, c'est hier qu'Ella a mangé un biscuit.
no, it-is yesterday that Ella has eaten a cookie

c. ? Non, (Ella a mangé un biscuit) hier.
no Ella has eaten a cookie yesterday

Hamlaoui further notes that clefted sentences with narrow-focused subjects the relative clause is optional and might be elided (cf. *Qui est-ce que a mangé un biscuit? C'est* Ella ~~*(qui a mange un biscuit))*~~. In contrastive clefted sentences, on the contrary, the material appearing in the post-focal part of the cleft sometimes has to be stressed; i.e. not only one constituent but two need to occupy a strong position with respect to stress:

(86) a. Pourquoi vous intéressez-vous tellement à Paris?
'Why are you so interested in Paris?'

b. C'est à Paris/lá que j'ai rencontré ma femme.

it-is in Paris there that I-have met my wife

'I met my wife in Paris/there.'

À Paris, even if constitutes given information and is discourse-linked, is the focus of this sentence. The post-focal material in the right periphery constitutes new information and thus also needs to be stressed.

The proposal put forward in this study is that clefted structures are also displayed in DS as an alternative strategy for conveying narrow focus while keeping the SVO preferred word order. Dominican clefted structures have been discussed in §2.3.4. and in §7.3.2.3. above. In these discussions it was pointed out that while general issues related to focus assignment in DS have been approached in previous work (Toribio 2002, Camacho 2008b), the question of how subjects get narrow-focused in this variety has not received much attention in the relevant literature. Subjects receive narrow focus in Spanish in cleft, pseudocleft and inverted pseudocleft structures in Spanish (cf (29) above). Camacho (2008b) notes that an additional strategy for narrow focus is available in Caribbean Spanish. This structure, usually labeled focalized-*ser* construction or Caribbean focus construction, is exemplified in (87):

(87) Trajo los regalos fue <u>Marta</u>.

brought.3PSG the gifts was.3PSG Marta

'It was Marta who brought the gifts.'

Camacho notes, with Bosque (1999), that the copular verb in (87) cannot modify preverbal subjects because generally it cannot modify any category higher than vP. He further points out that *ser* focalizing structures are not attested in DS for the focusing of subjects, probably because postverbal subjects are dispreferred in this variety. This is confirmed in the data set gathered for this dissertation, in which Caribbean focus constructions are not attested with subject constituents. The preferred structure for the focusing of subjects by the speakers is the inverted pseudocleft construction, which features a preverbal subject also in Non-Caribbean Spanish (cf. (25c) <u>*Marta*</u> *fue*

quien trajo los regalos, 'Marta brought the gifts'). Dominican inverted pseudoclefts were exemplified in (29) above. Example (29a) above is repeated below as (88). When the context in which this utterance was produced is added to this example it becomes clear that narrow focus is conveyed via an inverted pseudocleft. Note also that in (88a) the relative *que* is chosen over Non-Caribbean *quien/el que*, which is usually associated with more formal registers (cf. *Era Trujillo que llamaba* vs *Era Trujillo quien/el que llamaba*). In Non-Caribbean Spanish, relative pronouns are substituted by *que* in informal speech, while complex relative pronouns are usually associated with more formal registers.[63] In the data set under study no complex relatives are found in inverted pseudocleft structures, which suggests that *que* is grammaticalized in Dominican vernacular clefts.

(88) Nosotros por cierto estábamos pescando, y vino una lancha de la marina.
Entonces de cerca preguntaron: "¿Es el ingeniero López? El Jefe lo quiere ver".
El Jefe era Trujillo. Cuando decían El Jefe nadie se equivocaba, *era Trujillo que llamaba*.
'By the way were fishing, and a boat from the navy came over. And then from up close they asked: Are you López the engineer? The Boss wants to see you. The Boss was Trujillo. When they said El Jefe nobody took it wrong, it was Trujillo who was calling.'
(Federico, 81, UHM)

Lexical subjects are also featured in vernacular clefted structures, as illustrated in (89):

(89) Porque su mamá se murió estando ellos ahí; ella quedó chiquita de 9 años, entonces fue su hermana que la crió.
'Because her mom died while they were there; she was 9 years old, so it was her sister who raised her.'

[63] Consider for instance the following Non-Caribbean Spanish examples:
(i) La chica [*que* me encontré ayer].
The girl that I run into yesterday
(ii) La chica [*con la que* me encontré ayer].
The girl with the.FEM that I run into yesterday
'The girl who I run into yesterday.'

To conclude, DS speakers resort to strategies provided by the Spanish grammar for the expression of focus, while maintaining SVO word order. These structures support the view sustained in this dissertation, i.e. that Non-Caribbean and Dominican vernacular structures can be accommodated within the clausal structure of Spanish. In this theory, no competing grammars in the minds of the speakers need to be postulated. With this evidence in mind, a formal definition of the Dominican NSP profile is provided in the following section.

7.5. (Micro) parametric variation in Romance: On the NSP profile of a restructured subject grammar.

NSP variation across grammatical systems has been found in recent minimalist theory in three different sites:

(i) Feature assembly in the TP domain;

(ii) The nature of (null and overt) pronouns; and

(iii) Projection of lack thereof of the Spec, TP position.

As far as (i) and (iii) are concerned, the facts presented throughout this chapter have supported the view featured in *Hypothesis B* theories (Holmberg 2005, Sheehan 2006). Inflectional morphology is uninterpretable in Spanish, and the subject moves to Spec, TP, as in English-type languages, to check the EPP feature and to pick up the [+referential] feature in T. Regarding the nature of null and overt pronouns, the latter have been proposed to bear different properties in Caribbean and Non-Caribbean Spanish, following Ordóñez and Olarrea (2006) and Camacho (2008a). As a consequence, they occupy different positions within a split TP field (Cardinaletti 2004). Vernacular Dominican pronouns are weak in the sense of Cardinaletti and Starke (1999), and as such are merged in a low, Spec, TP position (AgrSP in Cardinaletti's 2004 notation). The higher position (SubjP) where strong pronominal subjects are merged is also available in DS. The existence of both positions is independently motivated by Non-Caribbean Spanish facts (cf.§7.4.2), and further confirmed by Dominican facts on word order.

The three axes of variation in (i)-(iii) are captured by minimalist NSP theories by deconstructing the traditional NSP (Rizzi 1982) in three micro-parameters (Sheehan 2006), as illustrated in (90), which repeats (10) above:

(90) *The Null Subject Parameter:*

(i) *The rich agreement parameter*

T lacks/bears a [uD] feature.

(ii) *The PF-interpretability parameter*

T lacks/bears a *, where * requires the specifier of I to be spelled out at PF.

(iii) *The Weak/strong nominative Case parameter*

Nominative Case feature [uI] lacks/bears *, where * requires movement to Spec, TP.

DS could be formally defined as follows: this language has a positive setting for (i), as attested by the fact that null subjects are available in the Dominican grammatical system. Second, it has a negative setting for (ii), since null expletive constructions are also available to speakers. Finally, it has a positive setting for (iii), since subject movement to Spec, TP has been shown to apply in DS.

If this NSP configuration is compared with the one displayed by Brazilian Portuguese (BP), the following (micro)parametric differences emerge: according to Sheehan (2006), BP has a 'negative' setting for (i). Main and embedded referential null subjects are licensed in this variety through movement from finite clauses via a peripheral Topic position (cf.§3.3.). However, DS and BP present identical settings for (ii) and (iii): BP also lacks a PF requirement for Spec, TP and so licenses null locatives and expletives. Moreover, in BP nominative Case is strong, triggering movement of the subject to Spec, TP. Spec TP is only empty at PF where it contains a deleted expletive or locative or the deleted copy of a moved DP.

Person differences in DS and BP have been found to display the same patterns: second and first person singular subjects very strongly favor overt subjects, while not as strongly favored with

third person singular subjects. Differences are attested, however, in the different types of null subjects available in each of these varieties, as summarized in Table 7.5:

Table 7.5. Null subjects in Dominican Spanish and Brazilian Portuguese.

	DOMINICAN SPANISH	BRAZILIAN PORTUGUESE
NON-THEMATIC SUBJECTS	✓	✓
REFERENTIAL SUBJECTS IN MAIN CLAUSES	✓	X*
REFERENTIAL SUBJS. IN EMBEDDED CLAUSES	✓	✓

*Only available via Topicalization

In BP, null embedded subjects have been shown to be available only in specific circumstances, i.e. when bound by a pronominal subject in the previous clause (cf.§3.3). In this study, in light of the available data, I have assumed as the null hypothesis that Dominican zero subjects are licensed like their Non-Caribbean counterparts. Only a fine-grained syntactic analysis which draws not only on quantitative but also on qualitative evidence could confirm whether this is the case or not. This is a task I have left for future research.

7.6. Summary.

In this chapter I have discussed the quantitative results presented in Chapter Six, as well as the qualitative evidence that can be found in the data set under study. Taking into account both types of evidence, a theory on the instantiation of the NSP represented by DS has been put forward. DS has been shown to share some of its properties with Non-Caribbean Spanish, while also departing from other Spanish dialects in relevant respects. DS has emerged in this comparison as a variety that evolves within the limits afforded by the Spanish clausal structure as defined in formal grammars.

Chapter Eight. The cohesive approach to cross-dialectal variation and language change: contributions and suggestions for further research.

This dissertation has explored subject expression in DS. In spite of being one of the most thoroughly researched features of the Spanish grammar in both the generative and the variationist paradigms, some facts related to subject licensing and overt/null subject insertion in Spanish still await explanation. Among them are the specific role of rich inflection in subject identification, or the possible NSP profiles that emerge in language change processes. I have argued that DS, a variety that has been proposed to display a restructured subject expression grammar, arises as a suitable venue for examining these issues. I have further argued that a cohesive approach to syntactic variation offers the tools for the description and explanation of the mixed systems derived from grammatical restructuring. Generative and variationist theories have traditionally differed in their approach to linguistic facts. As a result, in mainstream linguistic research the received knowledge is that these approaches are complementary at best, if not irreconcilable. However, the potential of building a cohesive approach to language variation that integrates the insights from both the generative and the variationist traditions has been developed in recent work (Adger and Smith 2005, Cornips and Corrigan 2005a,b,c). This dissertation expands on this bulk of research in that theoretical implications are drawn from a large corpus of data gathered following standard variationist methodology. In doing this, the present investigation has contributed to the explanation of synchronic and diachronic Romance subject expression patterns, and more generally to our understanding of the nature and limits of language variation and change. The contributions and limitations of this study are discussed in the following paragraphs.

8.1. Language variation and language change: The Dominican grammar of subject expression.

The variationist work carried out in the last three decades has consistently shown that subject expression is a variable phenomenon across dialects of Spanish, whether they are

undergoing language change or not. Specifically, discourse-related, morpho-syntactic and structural constraints have been found to regulate null and overt subject patterns. Among them, discourse-related constraints have arisen as remarkably stable across dialects. Factor groups that operationalize these constraints emerge as highly significant in all dialects studied up to date (Enríquez 1984, Bentivoglio 1987, Cameron 1992, 1996, Travis 2005, 2007, de Prada 2009, among others). In previous work it has been suggested that differences between Caribbean and Non-Caribbean subject expression are to be found in morpho-syntactic (Otheguy, Zentella and Livert 2007) or structural (i.e.) syntactic constraints (Morales 1997). These claims have been supported by the results obtained in this study: DS has been shown to pattern with other Spanish dialects in that discourse-related factor groups strongly determine null and overt subject patterns; moreover, it patterns with Caribbean dialects in that the morpho-syntactic factor group *Person* represents the strongest constraint on subject expression. Perhaps the most interesting results obtained in this study have emerged in the examination of syntactic factor groups: *Clause Type* has emerged as a very significant factor group in DS; specifically, conjoined clauses strongly disfavor overt subjects, whereas relative clauses strongly favor them. This result has been interpreted in this study as reflecting grammatical restructuring, since vernacular Dominican constructions documented in the previous literature, such as clefted affirmative and interrogative structures, involve relative clauses. In addition, grammatical restructuring is also attested in subject position patterns: while pronominal subjects are almost categorically preverbal, variation in subject position patterns emerges in lexical subjects. Interestingly, language-external factor groups, which typically have not emerged as significant in previous work on null and overt subject insertion, have been shown to be significant for the distribution of preverbal and postverbal subjects. Specifically, young, urban, high-middle class, female speakers arise as the groups leading grammatical restructuring in DS. This pattern corresponds to what has been defined in the variationist literature (Labov 1994) as a language-internally motivated change with no social stigma associated to any of the variants implicated and that can be introduced by any social class. These results speak to the potential of exploring syntactic

factor groups in grammatical restructuring processes of subject expression, which have been explored only occasionally in previous work.

This study also has some limitations:

First, the data set gathered for this study contained inherent asymmetries with respect to the language-external factor groups *Social class* and *Education*. Specifically, high-middle class speakers belong in urban environments only, while illiterate speakers were only recruited in rural areas. Due to these asymmetries and with the aim of examining the role of *Social class* and *Education* in DS, speakers were organized in three communities: all speakers, urban speakers only, and working class speakers in rural and urban environments. Note that this organization of speaker groups could yield results in which all three communities display similar results, since a number of speakers were part of more than one community; for instance, urban working class speakers belong in the all speaker and working class communities. Even if this is the case, significant differences across communities that pointed to grammatical restructuring in progress were found.

Second, subject expression in Spanish is constrained by several factors. Multivariate analysis offers the unique advantage of making possible to analyze the relative role of all these factors simultaneously. However, in this study, and in order to perform accurate multivariate analyses, a number of linguistic factor groups had to be recoded or excluded from specific statistical calculations. A case in point are interrogative constructions, which have been shown to strongly depart from affirmative structures and to have a relevant role in diachronic shifts in subject distribution in Caribbean Spanish. Finally, specific constructions associated with grammatical restructuring, such as clefted interrogatives or lexical subjects in interrogative constructions, do not arise frequently in spontaneous speech. It might well be the case that all the possible variants associated with a given construction have not arisen, or do so in very small numbers, in the data set examined. For this reason, fine-grained syntactic investigations need to be carried out in order to determine whether the structures attested in spontaneous speech exhaust the possibilities afforded by the Dominican grammar for the expression of subjecthood. Furthermore, from a syntactic-

theoretical point of view, the fact that the speakers do not frequently produce specific structures does not entail that they would judge them as grammatical or that these structures are not part of the speakers' grammars. Negative evidence is also an instrumental tool of theoretical-syntactic analyses. For this reason, some of the theoretical implications drawn from quantitative results in this study would need to be further confirmed by an investigation based on a traditional syntactic methodology. The fine-grained analysis of the structures that do not arise in high numbers in the data set gathered for this is a matter I leave for further research.

In sum, this dissertation has offered an accurate description of the Dominican variable null and overt subject pattern, and has shown that grammatical restructuring is to be found in the syntactic constraints that regulate null and overt subject insertion, and particularly, preverbal and postverbal subject distribution. Drawing theoretical implications from quantitative data has proven very fruitful for the description of the Dominican system, while specific issues call for in-depth examination in further research. In the following section I summarize the main contributions of this study to our understanding of the limits afforded by formal grammars for dialectal variation.

8.2. Dialectal variation and formal grammars: the issue of 'optionality'.

The study of closely related grammatical systems has emerged as suitable venue for exploring the limits afforded by formal grammars in recent generative work (Tortora 2003, Adger and Smith 2005, among others). This renewed interest in dialectal variation stems from two main sources: on the one hand, the conception of language variation advocated for in the Minimalist Program (Chomksy 1995, 2001, 2004), in which parameters are defined as devices that trigger mostly local consequences in the speaker's grammar. On the other, already within the Government and Binding framework (Chomksy 1981, 1986) it was argued that differences within the dialects of a language can mirror cross-linguistic differences (Henry 1995, 1997), which suggests that dialectal variation should be constrained within the parametric space defined by formal grammars.

The sort of dialectal variation described in Henry (1995, 1997) is also attested in DS as described by Toribio (2000a, 2001), since in this variety both Non-Caribbean Spanish null subject

structures and vernacular overt subject constructions are documented. In other words, structures corresponding to the two possible values of the NSP are available to Dominican speakers, an unexpected result within traditional, Government and Binding views on parametrization. Within the Minimalist Program accounting for a system of this sort has been suggested to be possible in pre-theoretical terms (Kayne 2005), since the deconstruction of the traditional NSP in smaller-scale (micro)parameters predicts the existence of mixed null and overt subject systems.

An account of the DS subject grammar needs to explain the apparent 'optionality' of null subject structures in this variety. In previous work (Toribio 2000a, Martínez Sanz 2007) this issue was approached by postulating competing grammars in the minds of Dominican speakers, following Kroch (1989, 1994). However, Kroch's view on language change has been questioned recently in both theoretical (Koopman and van der Wurff 2000, Biberauer and Roberts 2006) and empirical terms (Sorace 2000, 2005). In short, in these investigations it is shown that 'optionality', defined as the co-existence within an individual grammar of two or more variants of a given construction that make use of the same lexical resources and convey the same meaning is well attested in grammatical systems. This sort of semantically-vacous optionality is attested in native grammatical systems, it is an essential part of interlanguage grammars, and constitutes a rather frequent scenario in language change processes. Thus a formal theory of grammar should include an explanatory account of optionality.

In this dissertation, the explanation of the Dominican optionality has been suggested to lie in the properties of the Spanish clausal structure. Building up on Cardinaletti (2004), Ordóñez and Olarrea (2006) and Camacho (2008) it has been proposed that the extended TP and CP fields contain the positions to host the different types of subjects attested in DS. The claim that weak and strong pronominals exist in DS is based on the evidence on word order phenomena presented in Suñer (2003) and Camacho (2008). The fact that lexical subjects also arise in preverbal position in DS suggests that a middlefield, Spec TP position should also be available for merging vernacular lexical subjects. Furthermore, following Bullock and Toribio (2006) it has been proposed that the

available DS evidence (for instance on the availability on Non-Caribbean Topicalization constructions, cf.7.4.3) suggests that the impoverishment of the morphological exponents of inflection, instantiated in the loss of final *–s* in DS, does not entail that higher functional categories are not available in the clausal structure of a given variety. Bullock and Toribio put forward their proposal by reference to Romance bilingual data, and note that the inflectional deficits that emerge in bilingual speech parallel the 'impoverishment' phenomena attested in monolingual language change processes. The authors suggest that morpho-syntactic change may emerge as a result of the reinterpretation in language development of specific syntactic options that are associated with different pragmatic meanings in a language. In bilingual contexts, when faced with this variable input, language acquirers will set on the variant that is most congruent across the two languages. This process has been labeled *convergence*. Taking this proposal as a point of departure, it has been suggested in this study that instances of a somewhat similar process are attested in the monolingual restructured grammar of DS. A case in point would be vernacular Focus constructions, which involve clefted structures with overt preverbal subjects. Focus may be expressed in Non-Caribbean Spanish by prosodic and syntactic means. Syntactic strategies for the expression of Focus include the postverbal positioning of subjects and the use of clefted structures. The evidence found in the data set gathered for this dissertation suggest that Dominican speakers often resort to clefted structures for the expression of narrow focus. In other words, it has been suggested that while it might be the case that the expression of narrow Focus via a postverbal subject is available in the Dominican grammar, the speakers resort to the well-formed structures that allow for maintaining their preferred SVO word order. However, putting forward a proposal of subject-Focus assignment in Caribbean Spanish would require the examination of both the prosodic and syntactic strategies available to speakers, a task I have left for future research.

Appendix A. Demographic Information. All Speakers.

SPEAKER (PSEUDO)NAME & NUMBER	CODE	ORIGIN (U/R and area)	AGE	SEX	LITERACY	OCCUPATION	SOCIAL CLASS
1. FEDERICO	A	SD-Zona Universitaria **	81	male	university**	Engineer	H
2. MARIA AMELIA	B	SD-Evaristo Morales**	60	female	college**	Home-maker	H
3. NÉSTOR	C	SD-Naco**	56	male	university**	Engineer	H
4. DANILO	D	SD-La Esperilla**	52	male	university**	Engineer	H
5. JOSÉ	E	SD-Evaristo Morales**	48	male	university**	Physician	H
6. EVANGELITA	F	SD-Naco**	72	female	high school (some)**	Home-maker	H
7. MARTÍN	G	SD-Evaristo Morales**	26	male	university**	Dentist	H
8. CHAGO	H	SD-Villa Consuelo*	39	male	university*	Pharmacy clerk	W
9. ELISER	I	SD- Sabana Perdida*	35	male	high school (some)*	Driver	W
10. BILEIKA	J	SD-Gualey*	29	female	college*	Secretary (finances)	W
11. MAGDALIS	K	SD-Cristo Rey*	34	female	high school**	Secretary	W
12. MAURO	L	SD-Herrera*	25	male	high school	Construction worker	W
13. VALERIA	M	SD-Evaristo Morales**	19	male	high school**	University student	H
14. MARIGABI	N	SD-Evaristo Morales**	19	male	high school**	University student	H
15. OMAR	O	SD-Evaristo Morales**	19	male	high school**	University student	H
16. JOSÉ GABRIEL	P	SD-Evaristo Morales**	34	male	university**	Business owner	H
17. VICDALI	Q	SD-Cristo Rey	44	female	university*	Artist/Small business owner	W
18. ATILIA	R	SD-Herrera*	50	female	high school*	Local television host	W
19. TATICA	S	SD-Ensanche Piantini**	51	female	high school**	Home maker	H
20. MELVYN	T	SD-Villa C.*	31	male	university*	Unemployed	W
21. BRUNO	U	SD-Villa Consuelo	52	male	high school* (some)	Unemployed	W
22. JULIO	a	La Frontera-Partido	40	male	high school* (some)	Musician	W
23. RAMPHIS	b	La Frontera-Loma de Cabrera	40	male	high school (some)*	Civil defense brigade worker	W
24. CAMILA	c	La Frontera-Los Compos	28	male	illiterate	Cook/farmhand	W
25. AMBIORI	d	La Frontera-Los Compos	28	female	high school (some)*	Cook/House Keeper	W
26. MIGUELITO	e	La Frontera-Los Compos	49	male	middle school (some)*	Farmhand	W
27. CARMELA	f	La Frontera-El Pino	60	female	middle school (some)*	Home maker	W
28. BARBARITO	g	La Frontera-El Pino	61	male	middle school (some)*	Farmhand	W
29. QUISQUEYA	h	La Frontera-Los Compos	42	female	illiterate	Cook/Laundress/Farm hand	W

30. MAMÁ CANDITA	i	La Frontera-Partido	72	female	illiterate	Cook/Farmhand	W
31. OSVALDO	j	La Frontera-Los Compos	45	male	illiterate	Farmhand	W
32. VALDERRAMA	k	La Frontera-Los Compos	73	male	high school* (some)	Farmhand	W
33. RAMONA	l	La Frontera-Los Compos	28	female	high school (some)*	Unemployed	W
34. GLORIA	m	La Frontera-Partido	83	female	middle school (some)*	Home maker	W

*=public school, public university
**=private school, private university

*=working class neighbourhood
**=middle/high class neigbourhood

References

Abney, S. (1987): *The English Noun Phrase in its Sentential Aspect*. PhD dissertation. MIT.

Adger, D. (1996): Economy and Optionality: the interpretation of subjects in Italian. *Probus* 8(2), 117-135.

Adger, D. and Smith, J. (2003): *Variation and the Minimalist Program*. Ms. Queen Mary, University of London

Adger, D. and Smith, J. (2005): Variation and the Minimalist Program. In: Cornips, L. and Corrigan K. (eds.): *Biolinguistic and Sociolinguistic Accounts of Syntactic Variation: Doubts about Dichotomy. Current Issues in Linguistic Theory*. Amsterdam/Philadelphia: John Benjamins.

Alexiadou, A. and Anagnostopoulou, E. (1998): Parametrizing AGR: Word order, V-movement and EPP checking. *Natural Language and Linguistic Theory* 16, 491-539.

Amaral, P. and Schwenter, S. (2005): Contrast and the (Non-) occurrence of subject pronouns. In: Eddington, D. (ed.), *Selected Proceedings of the 7th Hispanic Linguistics Symposium*, 116-27. Somerville, MA: Cascadilla Proceedings Project.

Ambar, M. (1992): *Para uma Sintaxe da Inversão Sujeito-Verbo em Português*. Lisboa: Edições Colibri.

Ashby, W. and Bentivoglio, P. (1993): Preferred argument structure in spoken French and Spanish. *Language variation and change* 5, 61-76.

Ávila-Jiménez, B. (1995): A sociolinguistic analysis of a change in progress: Pronominal overtness in Puerto Rican Spanish. *Cornell Working Papers in Linguistics* 13, 25-47.

Ávila-Jiménez, B. (1996): Subject pronoun expression in Puerto Rican Spanish: A sociolinguistic, morphological and discourse analysis. PhD dissertation. Cornell University, Ithaca, New York.

Ávila-Shah, B. (2000): Discourse connectedness in Caribbean Spanish. In: Roca, A. (ed.): *Research on Spanish in the US: Linguistic issues and challenges*, 238-251. Somerville: Cascadilla Press.

Baker (2005): *The Macroparameter in a Microparametric world*. Ms, Rutgers University.

Balasch, S. (2008): La conectividad discursiva en el discurso interactivo. In: Bruhn de Garavito, J. and Valenzuela, E. (eds.): *Selected proceedings of the 10th Hispanic Linguistics Symposium*, 300-311. Somerville, MA: Cascadilla Proceedings Project.

Barbosa, P. (1995): *Null subjects*. PhD dissertation: MIT.

Barbosa, P. (2001): On inversion in *wh*-questions in Romance. In: Hulk, A. and Pollock, J-Y. (eds.): *Romance Inversion*, 20-90. New York: Oxford Press.

Barbosa, P. (2009): Two kinds of subjects *pro*. *Studia Linguistica* 63, 2-58.

Barrenechea, A.M. and Alonso, A. (1977): Los pronombres personales sujetos en el español hablado en Buenos Aires. In: Lope Blanch, J.M. (ed.): *Estudios sobre el español hablado en las principales ciudades de América*, 333-349. Mexico City: Universidad Nacional Autónoma de México.

Bayley, R. and Pease-Álvarez, L. (1996): Null and expressed pronoun variation in Mexican-descent children's Spanish. In: Arnold, J., Blake, R. and Davidson, B. (eds.): *Sociolinguistic*

variation: Data, theory and analysis, 85-99. Stanford: Center for the Study of Language and Information.

Bayley, R. and Pease-Álvarez, L. (1997): Null pronoun variation in Mexican-descent children's narrative discourse. *Language variation and change* 9, 349-371.

Beard (1991): Decompositional Composition: The Illusion of Bracketing Paradox and Opacity. Natural Language and Linguistic Theory 9, 195-229.

Belletti, A. (1998): Aspects of the Low IP Area. In: Rizzi, L. (ed.): *The Structure of IP and CP. The Cartography of Syntactic Structures* Vol. 2. Oxford University Press.

Belletti, A. (2001): Inversion as focalization. In: Hulk, A. and Pollock, J-Y. (eds.): *Subject inversion in Romance and the theory of Universal Grammar*, 60-91. New York: Oxford University Press.

Belletti, A. (2004): Aspects of the low IP area. In Rizzi, L. (ed.): *The structure of CP and IP. The cartography of syntactic structures* Vol. 2, 16-51. New York: Oxford University Press.

Belletti, A. (2005): Extended doubling and the VP periphery. *Probus* 17, 1-35.

Belletti and Rizzi (1988): Psych verbs and Theta-theory. *Natural Language and Linguistic Theory* 6, 291-352.

Benedicto, E. (1993): *The (IP) Functional projections and the Identification of pro*. Ms. Amherst: University of Massachusetts.

Bentivoglio, P. (1983): Topic continuity and discontinuity in discourse: A study of spoken Latin American Spanish. In: Givón, T. (ed.): *Topic continuity in discourse: A quantitative cross-language study*, 255-311. Philadelphia: John Benjamins.

Bentivoglio, P. (1987): *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.

Bentivoglio, P. (1988): La posición del sujeto en el español de Caracas: un análisis de los factores lingüísticos y extralingüísticos. In: Hammond, R.M. y Resnick, M.C. (eds.): *Studies in Caribbean Spanish dialectology*, 13-23. Washington: Georgetown University Press.

Bentivoglio, P. (1994): Spanish Preferred Argument Structure across Time and Space: *DELTA* 10, 277-293.

Benveniste, É. (1946): Structure des relations de personne dans le verbe. In: *Bulletin de la Société de Linguistique de Paris*, XLIII, 1, 1-12.

Berlinck, R.A. (1989): A construção V SN no português do Brasil: uma visão diacrônica do fenômeno da ordem. In: Tarallo, F. : *Fotografias sociolingüísticas*, 95-112. Campinas: Unicamp.

Berlinck, R.A. (1996): The Portuguese Dative. In: van Belle, W. and van Langendonck, W. (eds.): *The Dative*. Amsterdan/ Philadelphia: John Benjamins .

Barrenechea, A.M. and Alonso, A. (1977): Los pronombres personales sujetos en el español hablado en Buenos Aires. In: Blanch, L. (ed.): *Estudios sobre el español hablado en las ciudades principales de América*, 333-349. México: Universidad Autónoma de México.

Biberauer, T. and Richards, M. (2005): Keeping movement motivated: optionality without the tears. In: Chalcraft, F. and Sipetzis, E. (eds.): *Cambridge Occasional Papers in Linguistics (COPiL)* 2, 27-45.

Biberauer, T. and Roberts, I. (2006): The loss of "head-final" orders and remnant fronting in Late Middle English: causes and consequences. In: Hartmann, J. and Molnárfi, L. (eds.):

Comparative Studies in Germanic Syntax: from A(frikaans) to Z(ürich German), 263-297. Amsterdam: Benjamins.

Bickerton, D. (1981): *Roots of Language*. Ann Arbor: Karoma.

Bley-Vroman, R. (1989): The logical problem of second language learning. In: Gass, S. and Schachter, J. (eds.): *Linguistic perspectives on second language acquisition*. Cambridge: Cambridge University Press.

Bock, K. and Griffin, Z.M. (2000): The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General* Vol. 129, 177-192.

Borer, H. (1984): *Parametric syntax. Case studies in Semitic and Romance languages*. Dordrecht: Foris.

Bosque, I. (1999): On focus *vs*. wh-movement: The case of Caribbean Spanish. *Sophia Linguistica: Working papers in linguistics* 44-45, 1-32.

Bullock and Toribio (2009): Reconsidering Dominican Spanish: Data from the rural Cibao. *Revista Internacional de Lingüística Iberoamericana* 14, 49-73.

Bullock, B. and Toribio, A.J. (2007): Kreyol incursions into Dominican Spanish: The percept of Haitianized speech among Dominicans. In: Niño-Murcia, M. and Rothman, J. (eds.): *Linguistic identity and bilingualism*, 175-198. John Benjamins Publishers.

Burzio, L. (1986): *Italian syntax: a Government-Binding approach*. Dordrecht: Kluwer.

Cabrera-Puche, M.J. (2008): *Null subject patterns in language contact: The case of Dominican Spanish*. PhD dissertation. Ann Arbor: University Microfilm International.

Camacho, J. (2006a): In situ Focus in Caribbean Spanish: Towards a Unified Account of Focus. In: Sagarra, N. and Toribio, A.J. (eds): *Selected Proceedings of the 9th Hispanic Linguistics Symposium,* 13-23.

Camacho, J. (2006b): Do subjects have a place in Spanish? In: Nishida, C. and Montreuil, J. (eds.): *New perspectives in Romance linguistics*, 51-66. Amsterdam: John Benjamins.

Camacho, J. (2008): Syntactic Variation: The case of Spanish and Portuguese Subjects. To appear in: *Studies in Hispanic and Lusophone Linguistics* 2.

Cameron, R. (1992): *Pronominal and null subject variation in Spanish: Constraints, dialects and functional compensation*. PhD dissertation. Philadelphia: University of Pennsylvania.

Cameron, R. (1993): Ambiguous agreement, functional compensation, and nonspecific tú in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change* 5, 305-334.

Cameron, R. (1994): Switch reference, verb class and priming in a variable syntax. *Papers from the Regional Meeting of the Chicago Linguistic Society: Parasession on variation in linguistic theory* 30(2), 27-45.

Cameron, R. (1995): The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics* 6/7, 1-27.

Cameron, R. (1996): A community-based test of a linguistic hypothesis. *Language in Society* 25, 61-111.

Cameron, R. and Flores-Ferrán, N. (2004): Preservation of subject expression across regional dialects of Spanish. *Spanish in Context* 1(1), 41-65.

Cardinaletti, A. (2004): Toward a cartography of subject positions. In: Rizzi, L. (ed.): *The structure of CD and IP. The cartography of syntactic structures*, 115-65. Oxford: Oxford University Press.

Cardinaletti, A. and Starke, M. (1999): The typology of structural deficiency: A case study of the three classes of pronouns. In: van Riemsdijk, H. (ed.): *Clitics in the languages of Europe*, 145-235. Berlin: Mouton de Gruyter.

Casielles-Suárez, E. (2004): *The Syntax-Information Structure Interface*. New York and London: Routledge.

Cedergren, H. (1973): *The interplay of social and linguistic factors in Panama*. PhD dissertation. Cornell University.

Cedergren, H. (1983): Sociolingüística. In: López Morales, H. (ed.): *Introducción a la lingüística actual*, 147-165. Madrid: Playor.

Chafe, W. (1994): *Discourse, consciousness, and time. The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Cheshire, J. (1987): Syntactic variation, the linguistic variable, and sociolinguistic theory. *Linguistics* 25, 257-282.

Chomsky, N. (1957): *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965): *Aspects of the Theory of Syntax*. Cambridge MA: The MIT Press.

Chomsky, N. (1981): *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1986): *Knowledge of language: its nature, origin, and use*. Praeger: New York.

Chomsky, N. (1995): *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N. (2000): Minimalist inquiries: the framework. In: Martin, R., Michaels, D. and Urigareka, J. (eds.): *Step by step: essays on minimalist syntax in honor of Howard Lasnik*, 89-156. Cambrige, MA: MIT Press.

Chomsky, N. (2001): Derivation by phase. In: Kenstowicz, M. (ed.): *Ken Hale: a life in language*, 1-53. Cambridge, MA: MIT Press.

Chomsky, N. (2004): Beyond explanatory adequacy. In: Belletti, A. (ed.): Structures *and beyond: the cartography of syntactic structures* Vol. 3, 104-13. Oxford: Oxford University Press.

Cifuentes, H. (1980): Presencia y ausencia del pronombre personal sujeto en el habla culta de Santiago de Chile. *Homenaje a Ambrosio Rabanales: Boletín de Filología de la Universidad de Chile* 31, 743-152.

Cifuentes, R. (1981). 'Presencia y ausencia del pronombre personal en el habla culta de Santiago de Chile'. Homenaje a Ambrosio Rabanales. *Boletín de Filología de la Universidad de Chile*, XXXI.

Cinque, G. (1990): Types of A'-Dependencies. Cambridge MA: The MIT Press.

Cinque, G. (1999): Adverbs and functional heads. A crosslinguistic perspective. New York: Oxford University Press.

Cinque, G. (2002): Functional Structure in DP an IP. *The Cartography of Syntactic Structures* Vol. 1. Oxford: OUP.

Cinque, G. (2006): Restructuring and Functional Heads. *The Cartogtaphy of Syntactic Structures* Vol. 4. Oxford: OUP.

Clahsen, H., Marcus, G., Bartke, S. and Wiese, R. (1996): Compounding and inflection in German child language. In: Booij, G. and van Marle, J. (eds.): *Yearbook of Morphology*, 115–142. Dordrecht: Kluwer.

Clark, R. and Roberts, I. (1993): A computational approach to language, learnability and language change. *Linguistic Inquiry* 24: 299-345.

Contreras, H. (1989): On spanish empty N′ and N. In: Kirschner, C. and Decesaris, J. (eds.): *Studies in Romance Linguistics*. Amsterdam: John Benjamins.

Contreras, H. (1991): On the position of subjects. In: Rothstein, S. (ed.): *Perspectives on Phrase Structure: Heads and Licensing, Syntax and Semantics* 25, 63-80. New York: Academic Press.

Cornips, L. (1998): Syntactic variation, parameters and their social distribution. *Language Variation and Change* 10, 1-21.

Cornips, L. and Corrigan, K. (2005a): Syntax and variation: reconciling the biological and the social. Amsterdam: John Benjamins.

Cornips, L. and Corrigan, K. (2005b): Convergence and divergence in grammar: In: Auer, P., Hinskens, F. and Kerswill, P. (eds.): *Dialect Change: Convergence and Divergence in European Languages*, 96-134. Cambridge: Cambridge University Press.

Cornips, L. and Corrigan, K. (2005c): Toward an integrated approach to syntactic variation: A retrospective and prospective synopsis. In: Cornips, L. and Corrigan, K.P. (eds.): *Syntax and Variation: Reconciling the Biological and the Social*, 1-30. Amsterdam: John Benjamins.

Costa, J. (2004): *Subject positions and interfaces: the case of European Portuguese*. Berlin: Mouton de Gruyter.

Cyrino, S., Duarte, E. and Kato, M. (2000): Visible subjects and invisible clitics in Brazilian Portuguese. In: Kato, M. and Negrão, E. (eds.): *Brazilian Portuguese and the Null Subject Parameter*, 55-73. Frankfurt and Madrid: Vervuert-Iberoamericana.

Davis, J.C. (1971): Tú, ¿Qué Tú Tienes? *Hispania* 54, 331-332.

de Granda, G. (1987): Dos rasgos dialectales del español dominicano en el siglo XVIII. *Lingüística Española Actual* 9, 235-241.

de Prada, A. (2009): *Subject expression in Minorcan Spanish: Consequences of contact with Catalan*. PhD dissertation, The Pennsylvania State University.

Duarte, M.E. (1992): A perda da ordem V(erbo) S(ujeito) em interrogativas *qu-* no português do Brasil. *D.E.L.T.A.* 8, 37-52.

Duarte, M.E. (1993): Do pronome nulo ao pronome pleno. In: Roberts, I. and Kato, M. (eds.): *Português Brasilero: Uma viagem diacrônica (Homenagem a Fernando Tarallo)*, 107-28. Campinas, SP, Editora da UNICAMP.

Duarte, M.E. (1995): *A perda do princípio 'evite o pronome' no português brasileiro*. PhD dissertation. Campinas: Unicamp.

Duarte, M.E. (2000): The loss of the 'Avoid Pronoun' principle in Brazilian Portuguese. In: Kato, M. and Negrão, E. (eds.): *Brazilian Portuguese and the null subject parameter*, 17-36. Frankfurt: Vervuert.

Enríquez. E. (1984): *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas.

Epstein, S., Flynn, S., and Martohardjono, G. (1996): Second language acquisition: theoretical and experimental issues in contemporary research. *Brain and Behavioral Sciences* 19, 677-758.

Esgueva, M. y Cantarero, M. (1981): El habla de la ciudad de Madrid. Materiales para su estudio: Madrid: CSIC.

Eubank, L. (1994): Optionality and the initial state in L2 development. In: Hoekstra, T. and Schwartz, B.D. (eds.): *Language Acqusition Studies in Generative Grammar*, 369–88. Amsterdam/Philadelphia: Benjamins.

Fernández Lagunilla, M. and Anula, A. (1994): Procesos de filtrado de rasgos categoriales en la sintaxis. Los infinitivos y la legitimación del caso nominativo. In: Demonte, V. (ed.): *Gramática del Español*. Florida: El Colegio de Mexico.

Ferreira, M. (2000): *Argumentos nulos em português brasileiro*. MA thesis, University of Campinas.

Ferreira, M. (2004): Hyperraising and Null Subjects in Brazilian Portuguese. In: Castro, A., Hacquard, V., Salanova, A. and Ferreira, M. (eds.): *Romance. Collected Papers on Romance Syntax,* 57-85. Cambridge: MITWPL.

Figueiredo Silva, M.C. (1996): *A Posição Sujeito no Português Brasileiro: Frases finitas e infinitivas*. Campinas: Editora da Unicamp.

Flores-Ferrán, N. (2002): *Subject personal pronouns in Spanish narratives of Puerto Ricans in New York City: A sociolinguistic perspective*. München: Lincolm Europa.

Flores-Ferrán, N. (2004): Spanish subject personal pronoun use in New York City Puerto Ricans: Can we rest the case of English contact? In: *Language Variation and Change* 16, 49-73.

Flores-Ferrán (2007): A bend in the road: Subject personal pronoun expression in Spanish after 30 years of sociolinguistic research. *Language and Linguistics Compass* 4, 1-24.

Flores-Ferrán, N., and Toro, J. (2000): The persistence of dialect features under conditions of contact and leveling. In: *Southwest Journal of Linguistics* 19, 31-42.

Fukui, N. (1993): Parameters and Optionality. *Linguistic Inquiry* 24, 399-420.

Gallego, A. (2006): Phase Effects in Iberian Romance. *Cuadernos de Lingüística del Instituto Universitario Ortega y Gasset*, XIII, 59-68.

García Mayo, P., A. Lázaro Ibarrola and J. M. Liceras. 2006. Agreement in the interlanguage of Basque/Spanish bilinguals: A minimalist farewell to pro. ITL: International Journal of Applied Linguistics 151: 83- 97.

Gianollo, C., Guardiano, C. and Longobardi, G. (2009): Three fundamental issues in parametric linguistics. In: Biberauer, T. (ed.): *The limits of syntactic variation*. Amsterdam: Benjamins, 109-142.

Gilligan, G. (1987): *A cross linguistic approach to the pro-drop parameter*. PhD dissertation. University of Southern California.

Goodall, G. (1993): SPEC of IP and SPEC of CP in Spanish *wh*-questions. In: Mithun, M., Perissinotto, G. and Raposo, E. (eds.): *Linguistic Perspectives on the Romance Languages*, 199–209. Amsterdam/Philadelphia: John Benjamins.

Goodall, G. (2001): The EPP in Spanish. In: Davies, W. and Dubinsky, S. (eds.): *Objects and other subjects: grammatical functions, functional categories, and configurationality*, 193-223. Dordrecht: Kluwer.

Grimshaw, J. (1992): Argument Structure. *Linguistic Inquiry Monograph* 18. Cambridge: MIT Press.

Grimshaw, J. and Samek-Ludovici, V. (1995): Optimal Subjects. In: Beckman, J., Walsh-Dickey, L. and Urbanczyj, S. (eds.): *University of Massachusetts Occasional Papers* 18, 589-606.

Gutiérrez-Bravo, R. (2002): *Structural Markedness and Syntactic Structure*. PhD. dissertation, Santa Cruz: University of California.

Hamlaoui, F. (2007): French cleft sentences and the syntax-phonology inter- face. In: *Proceedings of the 2007 annual conference of the Canadian Linguistic Association*.

Hammond, R.M. (1979): Restricciones sintácticas y/o semánticas en la elisión de /s/ en el español de cubano. *Boletín de la Academia Puertorriqueña de la Lengua Española* 7.2, 41-57.

Haznedar, B. (2001): The Acquisition of the IP System in Child L2 Acquisition. *Second Language Acquisition* 23, 1-39.

Heap, D. (1990): Questions with preverbal subject in dominican spanish: Morphological variation and syntactic restructuring. *Journal of the Atlantic Provinces Linguistic Association/Revue de l'Association de Linguistique des Provinces Altantique* 14, 59-67.

Henríquez Ureña, P. (1940): *El español en Santo Domingo*. Buenos Aires: Casa Editora Coni.

Henry, A. (1995): *Belfast English and Standard English: Dialect Variation and Parameter Setting*. Oxford University Press.

Henry, A. (1997): Viewing change in progress: The Loss of V2 in Hiberno-English imperatives. In: van Kemenade, A. and Vicent, N. (eds.): *Parameters of morphosyntactic change*, 273-296. Cambridge: Cambridge University Press.

Henry, A. (2002): Variation and syntactic theory. In: Chambers, J., Trudgill, P. and Schilling-Estes, N. (eds.): *The Handbook of Language Variation and Change*, 267-282. Malden: Blackwell.

Hinzelin, M. and Kaiser, G. (2007): El pronombre ello en el léxico del español dominicano. In: Mihatsch, W. and Sokol, M. (eds.): *Language contact and language change in the Caribbean and beyond*, 171-188. Frankfurt: Peter Lang.

Hochberg, J. (1986): Functional compensation for /-s/ deletion in Puerto Rican Spanish. In: *Language* 62, 609-621.

Holmberg, A. (2005): Is there a little *pro*? Evidence from Finnish. *Linguistic Inquiry* 36, 533-64.

Holmberg, A. (2009): Three partial null-subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi. In: *Studia Linguistica* 63(1), 59-97. Wiley-Blackwell Publishing.

Holmberg, A. and Nikanne, U. (2002): Expletives, subjects and topics in Finnish. In: Svenonius, P. (ed.): *Subjects, expletives and the EPP*, 71-105. New York: Oxford University Press.

Holmberg, A. and Sheehan, M. (2009): Control into finite clauses in partial null-subject languages. In: Biberauer, T., Holmberg, A., Roberts, I., Sheehan, M. (eds.): *Parametric variation. Null subjects in minimalist theory*. Cambridge: Cambridge University Press.

Hornstein, N. (1999): Movement and control. *Linguistic Inquiry* 30, 69-96.

Hornstein, N. (2003): On control. In: Hendrick, R. (ed.): *Minimalist syntax*, 6-81. Oxford: Blackwell.

Huang, J. (1984): On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15, 531-74.

Hulk, A. and Pollock, J-Y. (2001): Subject positions in Romance and the theory of Universal Grammar. In: Hulk, A. and Pollock, J-Y. (eds.): *Subject inversion in Romance and the theory of Universal Grammar*, 3-19. Oxford: Oxford University Press.

Hyams (1992): *Language acquisition and the theory of parameters*. Dordrecht: Reidel.

Jackendoff, R. (1972): *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Jaeggli, O. and Safir, K. (1989): *The null subject parameter*. Dordrecht: Foris.

Jiménez Sabater, M. (1975): *Más datos sobre el español de Santo Domingo*. Santo Domingo: Instituto Tecnológico de Santo Domingo.

Kato, M. (1999): Strong and weak pronominals in the null subject parameter. *Probus* 11, 1-37.

Kato, M. and E. Negrão (2000): *The null subject parameter in Brazilian Portuguese*. Frankfurt: Vervuert-Iberoamericana.

Kato, M. and Taralho (1983): Anything YOU can do in Brazilian Portuguese. In: Jaeggli, O., Silva-Corvalán, C. (eds.): *Studies in Romance Linguistics*. Dordrecht: Foris, 343-58.

Kayne, R. (1994): *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Kayne, R. (2005): Some notes on comparative syntax, with special reference to England and French. In: Cinque, G. and Kayne, R. (eds.): The *Oxford handbook of comparative syntax*, 3-69. Oxford: Oxford University Press.

Kayne, R. and Pollock, J-Y. (2001): New thoughts on stylistic inversion. In: Hulk, A. and Pollock, J-Y. (eds.): *Subjects inversion in Romance and the theory of Universal Grammar*, 107-62. New York: Oxford University Press.

Koopman, H. and Sportiche, D. (1991): The position of Subjects. *Lingua* 85, 211-58.

Koopman, W.F. and van der Wurff, W. (2000): Two word order patterns in the history of English: stability, variation and change. In: Sornicola, R., Poppe, E. and Shisha-Halevy, A. (eds.): *Stability, Variation and Change of Word-order Patterns over Time*, 259-283. Amsterdam: Benjamins.

Kroch, A. (1989): Reflexes of grammar in patterns of language change. *Language Variation and Change* 1, 199-244.

Kroch, A. (1994): Morphosyntactic variation. In: Beals, K., Knippen, R., Melnar, L., Suzuki, H. and Zeinfeld, E. (eds.): *Proceedings of the 30th Regional Meeting of the Chicago Linguistic Society*, 180-201. Chicago: University of Chicago Press.

Laberge and G. Sankoff (1979): Anything You Can Do. In: Givón, T. (ed.): *Syntax and Semantics 12. Discourse and Syntax*. New York: Academic Press, 419-41.

Labov, W. (1966): The Social Stratification of English in New York City. Washington, D.C.: Center of Applied Linguistics.

Labov, W. (1969): The logic of non-standard English. In: Alatis, J. (ed.): *Georgetown monographs on languages and linguistics* 22, 1-44. Washington, DC: Georgetown University Press.

Labov, W. (1972): Sociolinguistic Patterns. Philadelphia: University of Pennsylvania Press.

Labov, W. (1980): Locating language in time, space and society. *Quantitative analyses of linguistic structure* Vol. 1. New York: Academic Press.

Labov, W. (1982): Bulding on empirical foundations. In: Lehmann, W. and Malkiel, Y. (eds.): *Perspectives on historical linguistics*, 17-92. Amsterdam and Philadelphia: John Benjamins.

Labov, W. (1984): Field methods of the project on linguistic change and variation. In: Baugh, J. and Sherzer, J. (eds.): *Language in use: readings in sociolinguistics*, 28-54. Englewood Cliffs, NJ: Prentice-Hall.

Labov, W. (1991): The three dialects of English. In: Eckert, P. (ed.): *Quantitative Analyses of Sound Change*, 1-44. New York: Academic Press.

Labov, W. (1994): *Principles of linguistic change Vol. I, Internal factors*. Cambridge and Oxford: Blackwell Publishers.

Landau (2000): *Elements of Control. Structure and meaning in infinitival constructions*. *Studies in Natural Language and Linguistic Theory 51*. Dordrecht: Kluwer Academic Publishers.

Landau (2004): The Scale of Finiteness and the Calculus of Control. *Natural Language and Linguistic Theory* 22, 811-877.

Lantolf, J. (1980): Evolutive Change in Syntax: Interrogative Word Order in Puerto Rican Spanish. In: Nuessel, F.H. (ed.): *Contemporary Studies in Romance Languages, Proceedings of the Eighth Annual Symposium on Romance Languages*, 202-19. Bloomington: Indiana University Linguistics Club.

Lapidus, N. and Otheguy, R. (2005): Overt nonspecific ellos in Spanish in New York. In: *Spanish in Context* 2, 157-174.

Lapidus, N. and Otheguy, R. (2009): Shifting sensitivity to continuity of reference: Subject pronoun use in Spanish in New York City. In: Lacorte, M. and Leeman, J. (eds.): *Español en Estados Unidos y otros contextos de contacto. Sociolingüística, ideología y pedagogía*, 111-136. Frankfurt/Madrid: Vervuert Iberoamericana.

Larrázabal Blanco, C. (1975): *Los negros y la esclavitud en Santo Domingo*. Santo Domingo: Julio D. Postigo e hijos, Editores.

Lavandera, B. (1978): Where does the sociolinguistic variable stop? *Language in Society* 7, 171-83.

Li, C. and Thompson, S. (1976): Subject and topic: a new typology of language. In: Li, C. (ed.): *Subject and topic*, 457-89. London/New York: Academic Press.

Liceras, J.M. (1998): On the specific nature of non-native grammars: the whys, whens, wheres and hows. In: Fernández-González, J. and de Santiago Guervós, J. (eds.): *Issues in Second Language Acquisition and Learning* 6, 58-96.

Liceras, J.M. (2007): A 'linguistic approach' to the idiosynchratic nature of second language acquisition: monosyllabic place-holders and morpheme orders. *Bilingualism: Language and Cognition* 10(1), 39-41.

Lightfoot, D. (1979): *Principles of diachronic syntax*. Cambridge: Cambridge University Press.

Lightfoot, D. (1991): *How to set parameters: arguments from language change*. Cambridge, MA: MIT Press.

Lightfoot, D. (1998): *The Development of Language: Acquisition, Change and Evolution*. Oxford: Blackwell.

Lightfoot, D. (2002): *Syntactic effects of morphological change*. Oxford: Oxford University Press.

Lipski, J. (1977): Preposed Subjects in Questions. *Hispania* 60, 61-7.

Lipski, J. (1985): Creole Spanish and Vestigial Spanish: Evolutionary Parallels. *Linguistics* 23, 963-84.

Lipski, J. (1991): In search of the Spanish personal infinitive. In: Wanner, D. and Kibbie, D. (eds.): *New analyses in Romance linguistics, papers from the XVIII Linguistic Symposium on Romance Languages*), 201-220. Amsterdam: John Benjamins.

Lipski, J. (1994a): *Latin American Spanish*. New York: Longman Linguistics Library.

Lipski, J. (1994b): The evolution of null subjects in Philippine Creole Spanish. *Mid-America Linguistics Conference Papers, Volume II*, 387-401. Lawrence, Kansas: University of Kansas Linguistics Department.

Lipski, J. (1999): Null subjects in Romance-derived creoles: routes of evolution. *Society for Pidgin and Creole Linguistics annual meeting*. Los Angeles, January 8, 1999.

Lipski, J. (2004): El español de América: los contactos bilingües. In: Cano, R. (ed.): *Historia de la lengua española*, 1117-1138. Barcelona: Ariel.

Lizardi, C. (1993). *Subject position in Puerto Rican wh-questions: Syntactic, sociolinguistic and discourse factors*. PhD dissertation. New York: Cornell University.

Lizardi , C. and Suñer, M. (1992): Dialectal variation in an argumentation/non-argumental asymmetry in Spanish. In: Amastae, J., Goodal, G., Montalbetti, M. and Phinney, M. (eds.): *Contemporary Research in Romance Linguistics. Papers from the XXII Linguistic Symposium on Romance* 2:83,84

Lopes Rossi, M. (1993): Estudo diacrônico sobre as interrogativas do português do Brasil. In: Roberts, I. and Kato, M. (eds.): *Português Brasileiro: uma viagem diacrónica*, 307-42. Campinas, SP: Ed. Da UNICAMP.

Lopez , L. (2008): *A derivational syntax of information structure. Oxford Studies in Theoretical Linguistics 23*. Oxford: Oxford University Press.

López Morales, H. (1983): Estratificación sociolectal frente a diglosia en el Caribe hispánico. *Lingüistica Española Actual* 5, 205-22.

Lunn, P. (2002): 'Tout se tient' in Dominican Spanish. In: Lee, J., Geeslin, K. and Clements, J.C. (eds.): *Structure, meaning, and acquisition in Spanish*, 65-72. Somerville, MA: Cascadilla.

Manzini, R. and Savoia, L. (2002): Parameters of subject inflection in Italian dialects. In: Svenonius, P. (ed.): *Subjects, expletives and the EPP*, 157-200. Oxford: Oxford University Press.

Martínez-Sanz, C. (2007): *Dialectal microvariation and the null subject parameter: The case of Dominican Spanish*. Ms, University of Ottawa.

Martínez-Sanz, C. and Toribio, A.J. (2008): On the nature of expletive *ello* in Dominican Spanish. Paper presented at the Hispanic Linguistics Symposium, Université Laval, Quebec.

McWhorter, J. (2001): *The missing Spanish creoles: Recovering the birth of plantation contact languages*. Berkeley: University of California Press.

Medina-Rivera, A. (1991): Interaction of (s) and subject expression in the Spanish of Choluteca and El Paraíso, Honduras. Paper presented at NWAV XX, Washington DC.

Meechan, M and Foley, M. (1994): On resolving disagreement: linguistic theory and variation - there's bridges. *Language Variation and Change* 6, 1. 63-85.

Megenney (1984): Elementos subsaháricos en el español dominicano. In: Alba. O. (ed): *El español del Caribe*. Santiago de los Caballeros: Universidad Católica Madre y Maestra, 183-201.

Megenney (1990): *África en Santo Domingo. La herencia lingüística*. Santo Domingo: Museo del Hombre Dominicano.

Meisel, J. M. (1994): *Bilingual First Language Acquisition: French and German Grammatical Development*. Amsterdam: John Benjamins.

Miró Vera, R. and de Pineda, A. (1982). Determinación sociolingüística de la presencia/ausencia del pronombre personal sujeto. In: Teresa Palet Plaja, M. (ed.): *Sociolingüística andaluza 5: Habla de Sevilla y hablas americanas,* 37-44. Sevilla: Servicio de publicaciones de la Universidad de Sevilla.

Modesto, M. (2000a): *On the identification of null arguments*. PhD dissertation: University of Southern California.

Modesto, M. (2000b): Null subjects without 'rich' agreement. In: Kato, M. and Negrão, E. (eds.): *The Null Subject Parameter in Brazilian Portuguese*, 147-74. Frankfurt: Vervuert-Iberoamericana.

Modesto, M. (2009): Topic prominence and null subjects. In: Biberauer, T. (ed.): *The limits of syntactic variation*. Amsterdam/Philadelphia: John Benjamins.

Montalbetti, M. (1984): *After Binding: On the Interpretation of Pronouns*. PhD dissertation. MIT.

Montes Miró, R.G. (1986): Factores discursivos en el análisis de los pronombres personales sujeto en español. *Morphé* 1, 45-71.

Morales, A. (1980): La expresión de sujeto pronominal primera persona en el español de Puerto Rico. *Boletín de la Academia de Lengua Española* 8.3, 91-102.

Morales, A. (1982): La perspectiva dinámica oracional en el español de Puerto Rico. In: Alba, O. (ed.): *El español del Caribe: Ponencias del VI simposio de dialectología*, 205-219. Santo Domingo: Universidad Católica Madre y Maestra.

Morales, A. (1986): *Gramáticas en contacto: análisis sintácticos sobre el español de Puerto Rico*. Madrid: Playor.

Morales, A. (1988): Infinitivo con sujeto expreso en el español de Puerto Rico. In: Hammond, R.M. and Resnick, R. (eds.): *Studies in Caribbean Spanish dialectology*.

Morales, A. (1989): Hacia un universal sintáctico del español del Caribe: el orden SVO. *Anuario de Lingüística Hispánica* 5, 139-152.

Morales, A. (1992): Acerca de la topicalidad de objeto en algunos dialectos del español. *Revista de Filología Española* 72.3, 671-685.

Morales, A. (1997). La hipótesis functional y la aparición del sujeto no nominal: El español de Puerto Rico. *Hispania* 80, 153-65.

Moya Pons, F. (1996): Dominican national identity: A historical perspective. *Punto 7 Review: A Journal of Marginal Discourse* 3(1), 14–25.

Muysken, P. (2000): *Bilingual Speech. A typology of code-mixing*. Cambridge: Cambridge University Press.

Muysken, P. (2005): A modular approach to sociolinguistic variation in syntax. The gerund in Ecuadorian Spanish. In: Cornips, L. and Corrigan, K. (eds.): *Biolinguistic and Sociolinguistic Accounts of Syntactic Variation: Doubts about Dichotomy. Current Issues in Linguistic Theory*. Amsterdam/Philadelphia: John Benjamins.

Navarro-Tomás, T. (1948): *El español en Puerto Rico*. Río Piedras: Editorial Universitaria.

Negrão, E. (1997): Asymmetries in the distribution of overt and empty catogories in Brazilian Portuguese. In: Black, J. and Motapanyane, V. (eds.): Clitics, pronouns and movement, 217-35. Amsterdam: Benjamins.

Negrão, E. (2000): Wh-Extractions and Relative Clauses in Brazilian Portuguese. *DELTA* 16, 141-163.

Nicolis, M. (2005): *On pro-drop*. PhD dissertation: Università degli Studi di Siena.

Nicolis, M. (2008): The null subject parameter and correlating properties: the case of creole languages. In: Biberauer, T. (ed.): *The limits of syntactic variation*, 271-94. Amsterdam: Banjamins.

Núñez Cedeño, R. (1983): La pérdida de transposición de sujeto en interrogativas pronominales del español del Caribe. *Thesaurus* 38, 1-24.

Olarrea, A. (1996): *Pre- and Postverbal Subject Positions in Spanish*. PhD dissertation. University of Washington.

Ordóñez, F. (1997): *Word order and clause structure in Spanish and other Romance languages.* PhD dissertation. CUNY Graduate Center.

Ordóñez, F. (1998): Post-verbal asymmetries in Spanish. *Natural Language and Linguistic Theory* 16, 313-46.

Ordóñez, F. (1999): Focus and subject inversion in Romance. In: Gutiérrez-Rexach, J. and Martínez-Gil, F (eds.): *Advances in Hispanic Linguistics*, 502–18. Somerville, MA: Cascadilla Press.

Ordóñez, F. (2000): *The Clausal Structure of Spanish: A Comparative Study*. Outstanding Dissertations in Linguistics. New York: Garland.

Ordóñez, F. and Olarrea, A. (2006): Microvariation in caribbean/non caribbean spanish interrogatives. *Probus: International Journal of Latin and Romance Lin*guistics 18, 59-96.

Ordóñez, F. and Treviño, E. (1999): Left dislocated subjects and the pro-drop parameter: a case study of Spanish. *Lingua* 107, 39-68.

Orozco, R. (2007): The impact of linguistic constraints on the expression of futurity in the Spanish of New York Colombians. In: Potowski, K. and Cameron, R. (eds.): *Policy, social and linguistic inquiries*. Philadelphia: John Benjamins, 312-327.

Orozco, R. and Guy, G. (2008): El uso variable de los pronombres sujetos: ¿qué pasa en la costa Caribe colombiana? In: Westmoreland, M. and Thomas, J.A. (eds.): *Selected proceedings of the 4th workshop on Spanish sociolinguistics*, 70-80. Sommerville,Mass.: Cascadilla Proceedings Project.

Ortega-Santos, I. (2008): *Projecting Subjects in Spanish and English*. PhD dissertation. University of Maryland, College Park.

Ortiz López, L. (2007): La negación en la frontera domínico-haitiana: variantes y usos (socio)lingüísticos. In: Potowski, K. and Cameron, R. (eds.): *Spanish in contact: Educational, social, and linguistic inquiries*, 211-236. John Benjamins: John Benjamins Publishing.

Ortiz López, L. (2009): Pronombres de sujeto en el español (L2 vs L1) del Caribe. In: Jennifer, L. and Manel, L.: *El Español en Estados Unidos y otros contextos de contacto*, 85-110. Madrid/Frankfurt: Iberoamericana/Vervuert.

Otheguy, R. (1973): The Spanish Caribbean: A creole perspective. In: Bailey, C-J. and Shuy, R. (eds.): *New Ways of Analizing Variation in English*. Washington: Georgetown University Press, 323-339.

Otheguy, R., Zentella, A.C., and Livert, D. (2007): Language and dialect contact in Spanish in New York: Towards the formation of a speech community. *Language* 83(4), 770-802.

Paradis, J., Nicoladis, E. and Genesee, F. (2000): Early emergence of structural constraints on code-mixing: Evidence from French-English bilingual children. *Bilingualism: Language and Cognition* 3(3), 245-261.

Paredes Silva, V.L. (1993): Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change* 5, 35-49.

Patín Maceo, M.A. (1947): *Dominicanismos*. Ciudad Trujillo: Librería Dominicana.

Pérez Guerra, I. (1989): Un caso de prestigio encubierto en el español dominicano: La vocalización cibaena. In: *Actas del III Congreso internacional del Español de América*, 1185-1191.

Pérez-Leroux, A.T. (1999): Innovación sintáctica en el español del Caribe y los principios de la gramática universal. In: Ortiz López, L. (ed.): *El Caribe hispánico: perspectivas lingüísticas actuales, Homenaje a Manuel Alvarez Nazario*, 99-118. Frankfurt an Main/Madrid: Vervuert/Iberoamericana.

Perlmutter, D. (1970): *Deep and surface structure constraints in syntax*. PhD dissertation: MIT.

Pesetsky, D. and Torrego, E. (2001): T-to-C movement: causes and consequences. In: Kenstowicz, M. (ed.): *Ken Hale: a life in language*, 355-426. Cambridge, MA: MIT Press.

Pierce, A. (1992): Language Acquisition and Syntatic Theory: A Comparative Study of French and English Child Grammars. Dordrecht, Kluwer.

Pires, A. (2000): Pro, Movement and Binding in Portuguese. In: Camps, J. and Wiltshire, C. (eds.): *Romance Syntax, Semantics and their L2 Acquisition. Selected Papers from the 30th Linguistic Symposium on Romance Languages*. Gainesville, Florida.

Platzack, C. (1987): The Scandinavian languages and the null-subject parameter. *Natural Language and Linguistic Theory* 5, 377-401.

Platzack, C. (2003): Agreement and null subjects. *Nordlyd. Tromsø University Working Papers on Language and Linguistics* 31(2), 326-355.

Platzack. C. (2004): Agreement and the person phrase hypothesis. *Working Papers in Scandinavian Syntax* 73, 83-112.

Pollock, J-Y. (1989): Verb movement, Universal Grammar and the structure of IP. *Linguistic Inquiry* 20, 365-424.

Poplack, S. (1979): *Function and process in a variable Phonology*. PhD dissertation, University of Pennsylvania.

Poplack, S. and Meechan, M. (1995): Patterns of language mixture: Nominal structure in Wolof-French and Fongbe-French bilingual discourse. In: Muysken, P. and Milroy, L (eds.): *One speaker, two languages*, 199-232. Cambridge: Cambridge University Press.

Poplack, S. and Tagliamonte, S. (1991): African-American English in the diaspora: Evidence from old-line Nova Scotians. *Language Variation and Change* 3(3), 301-339.

Pollock, J-Y. (1989): Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry* 20(3), 365-424.

Poletto, C. and J-Y. Pollock (2004): On the Left Periphery of Some Romance Wh-Questions. In: Rizzi, L. (ed.): *The structure of CP and IP: The Cartography of Syntactic Structures*, 251-296. New York/Oxford: Oxford University Press.

Pratas, F. (2002): *O Sistema Pronominal do Caboverdiano (variante de Santiago): Questões de Gramática*. Dissertação de Mestrado. FCSH, Universidade Nova de Lisboa.

Ranson, D.L. (1991): Person marking in the wake of /s/ deletion in Andalusian Spanish. *Language Variation and Change* 3(2), 133-152.

Raposo, E. (1987): Case Theory and Infl-to-Comp: The Inflected Infinitive in European Portuguese. *Linguistic Inquiry* 18, 85-109.

Raposo, E. (1996): *Towards a Unification of Topic Constructions*: UCSB.

Ribeiro, I. (1994): *A sintaxe da ordem no português arcaico: I efeito V2*. PhD dissertation. Campinas, SP: UNICAMP.

Rigau, G. (1995): The properties of the temporal infinitive contructions in Catalan and Spanish. *Probus* 7.

Rivero, M.L. (1980): On Left-Dislocation and Topicalization in Spanish. *Linguistic Inquiry* 2, 3363-93.

Rizzi, L. (1982): *Issues in Italian syntax*. Dordrecht: Foris.

Rizzi, L. (1986): Null objects in Italian and the theory of *pro*. *Linguistic Inquiry* 17, 501-57.

Rizzi, L. (1987): Three remarks on null subjects. Talk given at the Workshop on Dialectology, GLOW X. University of Venice.

Rizzi, L. (1996): Residual verb second and the wh-criterion. In: Belletti, A. and Rizzi, L. (eds.): *Parameters and Functional heads: Essays in Comparative Syntax*, 63-90. Oxford: Oxford University Press.

Rizzi, L. (1997): On the fine structure of the left periphery. In: Haegeman, L. (ed.): *Elements of grammar*, 281-337. Dordrecht: Kluwer.

Rizzi, L. (1998): Remarks on Early Null Subjects. In: Greenhill, A., Hughes, M., Littlefield, H. and Walsh, H. (eds.): *Proceedings of the 22nd Annual Boston University Conference on Language Development*, 14-39. Sommerville, Mass.: Cascadilla.

Rizzi, L. (2000): Early null subjects and root null subjects. In: Rizzi, L. and Friedemann, M.A. (eds.): *The Acquisition of Syntax*.

Rizzi, L. (2004): Locality and left periphery. In Belletti, A. (ed.): *Structures and beyond: Cartography of syntactic structures*. New York: Oxford University Press, 104-131.

Rizzi, L. and Roberts, I. (1989): Complex inversion in French. *Probus* 1, 1-30

Roberge, Y. (1990): *The Syntactic Recoverability of Null Arguments*. Montréal : McGill-Queen's University Press.

Roberts, I. (1993): *Verbs and diachronic syntax: a comparative history of English and French*. Dordrecht: Kluwer.

Roberts, I. (1999): Verb movement and markedness. In: deGraff, M. (ed.): *Language creation and change*, 287-328. Cambridge, Mass: MIT Press.

Roberts, I. (2001): Language change and learnability. In: Bertolo, S. (ed.): *Parametric linguistics and learnability*, 81-125. Cambridge: Cambridge University Press.

Roberts, I. (2004): The C-system in Brythonic Celtic languages, V2 and the EPP. In: Rizzi, L. (ed.): *The structure of CP and IP. The cartography of syntactic structures* Vol.2, 297-328. Oxford: Oxford University Press.

Roberts, I. (2007): *Diachronic syntax*. Oxford: Oxford University Press.

Roberts, I. and Holmberg, A. (2009): Introduction: Parameters in minimalist theory. In: Biberauer, T., Holmberg, A., Roberts, I. and Sheehan, M. (eds.): *Parametric variation: Null subjects in minimalist theory*. Cambridge: Cambridge University Press.

Roberts, I. and Roussou, A. (1997): *Interface Interpretation*. Ms. Bangor: University of Stuttgart and University of Wales.

Roberts, I. and Roussou, A. (2001): The EPP as a condition on the tense dependency. In: ed. Svenonius, P. (ed.): *Subjects, expletives and the EPP*, 125-157. Oxford: OUP.

Roberts, I. and Roussou, A. (2003): *Syntactic Change. A Minimalist Approach to Grammaticalization*. Cambridge: Cambrigde University Press.

Robertson, D. and Sorace, A. (1999): Losing the V2 constraint. In: Klein, E. and Martohardjono, G. (eds.): *The development of second language grammars: A generative approach*. Amsterdam: John Benjamins.

Rodrigues, C. (2002): Morphology and null subjects in Brazillian Portuguese. In: Lightfoot, D. (ed.): *Syntactic effects of morphological change*, 160-78. Oxford: Oxford University Press.

Rodrigues, C. (2004): *Impoverished morphology and A-movement out of Case-domains*. PhD dissertation. University of Maryland.

Rohrbacher, B. (1994): *The Germanic VO languages and the full paradigm: A theory of V to I raising*. PhD dissertation. Amherst: University of Massachusetts.

Rohrbacher, B. (1999): *Morphology-driven syntax*. Amsterdam: Benjamins.

Romaine, S. (1984): Relative Clauses in Child Language, Pidgins and Creoles. *Australian Journal of Linguistics* 4, 257-81.

Ross, J.R. (1982): Pronoun deletion process in German. Paper presented at the annual meeting of the Liguistic Society of America. San Diego, California.

Sankoff, D. (1975): *Varbrul 2*. Unpublished program and documentation.

Sankoff, D. (1988): Sociolinguistics and syntactic variation. In: Newmeyer, F.J. (ed.): *Linguistics: the Cambridge survey*, 140-61. Cambridge: Cambridge University Press.

Sankoff, D., Tagliamonte, S. and Smith, E. (2006). Goldvarb X. Department of Linguistics, University of Toronto, Toronto, Canada. http://individual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm.

Sankoff, G. (2002): Linguistic outcomes of language contact. In: Chambers, J.K., Trudgill, P. and Schilling-Estes, N. (eds.): *The Handbook of Language Variation and Change*. Oxford: Blackwell, 638-668.

Schwartz, B.D. (1998): On Two Hypotheses of 'Tranfer' in L2A: Minimal Trees and Absolute L1 Influence. In: Flynn, S., Martohardjono, G. and O'Neil, W. (eds): *The Generative Study of Second Language Acquisition*, 35-60. New Jersey: Erlbaum.

Schwartz, B.D. and Sprouse, R.A. (1994): Word order and nominative case in non-native language acquisition: a longitudinal study of (L1 Turkish) German interlanguage. In: Hoekstra, T. and Schwartz, B.D. (eds.): *Language Acquisition Studies in Generative Grammar*, 317-368. Amsterdam: John Benjamins.

Schwartz, B.D. and Sprouse, R.A. (1997): Transfer: A Tradition in Transition. Presentation at the American Association of Applied Linguistics, March 9, 1997.

Sedano, M. (1988): Yo vivo es en caracas: Un cambio lingüístico. In: Hammond, M.R. and Resnick, M.C. (eds.): *Studies in the Caribbean Spanish Dialectology*, 115-123. Washington: Georgetown University Press.

Sedano, M. (1990): *Hendidas y otras construcciones con ser en el habla de Caracas*. Caracas: UCV.

Sedano, M. (1995): A la que yo escribí fue a María vs. yo escribí fue a María. El uso de estas dos estructuras en el español de Caracas. *Boletín de Lingüística* 9, 51-80.

Sedano, M. (2001): Estructuras focalizadoras en español y portugués. *Filología e Lingüística Portuguesa* 4, 53-68.

Selkirk, E. (1984): *Phonology and Syntax: the relation between sound and structure*. Cambridge, Mass: MIT Press.

Sheehan, M. (2006): *The EPP and null subjects in Romance*. PhD dissertation. Newcastle University.

Sigurdsson, H.Á. (1993): Argument drop in Old Icelandic. *Lingua* 89, 247-80.

Silva-Corvalán, C. (1982): Subject expression and placement in Mexican-American Spanish. In: Amastae, J. and Elías-Olivares, L. (eds.): *Spanish in the United States: Sociolinguistic aspects*, 93-120. Cambridge/New York: Cambrige University Press.

Silva-Corvalán, C. (1984): Topicalización y pragmática en español. *Revista Española de Lingüística* 14, 1-19.

Silva-Corvalán, C. (1989): *Sociolingüística. Teoría y análisis*. Madrid: Alambra.

Silva-Corvalán (1992): El español actual en Estados Unidos. In: Hernández Alonso, C. *(ed.): Historia y presente del español de América*, 827-856. Valladolid, Spain: Junta de Castilla y León.

Silva-Corvalán, C. (1994): *Language contact and change: Spanish in Los Angeles*. Oxford: Oxford University Press.

Silva-Corvalán, C. (1997): Variación sintáctica en el discurso oral: problemas metodológicos. In: Moreno Fernández, F. (ed.): *Trabajos de sociolingüística hispánica*, 115-135. Universidad de Alcalá: Nuevo Siglo.

Silva-Corvalán, C. (2001): *Sociolingüística y pragmática del español*. Washington, DC: Georgetown University Press.

Silva-Corvalán, C. (2003): Otra mirada a la expresión del sujeto como variable sintáctica. In: Moreno, F., Gimeno Menéndez, F., Samper, J.A., Gutiérrez Araus, M.L., Vaquero, M. and Hernández, C. (eds.): *Lengua, variación y contexto* Vol.II, 849-860. Madrid: Arco Libros.

Silva-Corvalán, C. (2007): Spanish-English bilingualism: Theoretical issues, empirical analyses. *Southwest Journal of Linguistics* 26, 1-19.

Silva-Villar, L. (1998): Subject position and the roles of CP. In: Schwegler, A., Tranel, B. and Uribe-Etxebarría, M. (eds.): *Romance linguistics: Theoretical perspectives*, 247-270. Amsterdam: John Benjamins.

Sitaridou, I. (2002): *The Synchrony and Diachrony of Romance Infinitives with Nominative Subjects*. PhD dissertation, University of Manchester.

Sorace, A. (1999): Syntactic optionality in non-native grammars. *Second Language Research* 16(2), 93-102.

Sorace, A. (2000): Optionality in Second Language Acquisition. Thematic issue of *Second Language Research* 16(2).

Sorace, A. (2005): Syntactic optionality at interfaces. In: Cornips, L. and Corrigan, K. (eds): *Syntax and Variation: Reconciling the Biological and the Social*, 46-111. Amsterdam: John Benjamins.

Sportiche, D. (1995). Subject clitics in French and Romance Complex inversion and clitic doubling. In: *Studies in Comparative Romance Syntax*, eds. Kyle Johnson and Ian Roberts. Dordrecht: Kluwer.

Suárez-Budenbender, E. (2009): *Perceptions of Dominican Spanish and Dominican self-perception in the Puerto Rican diaspora*. PhD dissertation, The Pennsylvania State University.

Suñer, M. (1982): *Syntax and semantics of Spanish presentational sentence-types*. Washington DC: Georgetown University Press.

Suñer, M. (1986): Lexical Subjects of Infinitives in Caribbean Spanish. In Jaeggli, O. and Silva-Corvalán, C. (eds.): *Studies in Romance Linguistics*, 189-204. Dordrecht: Foris.

Suñer, M. (2003): The Lexical Preverbal Subject in a Romance Null Subject Language: Where are Thou? In: Núñez Cedeño, R. (ed.): *A Romance Perspective on Language Knowledge and Use. Selected Papers from the 31st Linguistic Symposium on Romance Languages (LSRL)*, 341-358. Amsterdam: John Benjamins.

Suñer, M. and Lizardi, C. (1995). Dialectal variation in an argumental/non-argumental asymmetry in Spanish. In: Amastae, J., Goodall, G., Montalbetti, M. and Phinney, M. (eds.): *Contemporary Research in Romance Linguistics*, 187 ff.

Szendröi, K. (2003): A stress-based approach to the syntax of Hungarian focus. *The Linguistic Review* 20(1), 37-78.

Tagliamonte, S.A. (2002): Comparative sociolinguistics. In: Chambers, J., Trudgill, P. and Schilling-Estes, N. (eds): *Handbook of language variation and change*, 729-63. Malden and Oxford: Blackwell Publishers.

Tagliamonte, S.A. (2006): *Analyzing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Taraldsen, T. (1978): *On the NIC, vacuous application and the that-trace filter*. Unpublished Ms: MIT.

Taralho, F. (1986): *Tempos lingüisticos – Itinerário histórico da lingua portuguesa*. São Pauso, SP: Ática.

Terrell, T. (1979): *The interaction of phonological and grammatical constraints on aspiration and deletion*. Ms, University of Irvine.

Toribio, A.J. (1993a): *Parametric Variation in the Licencing of Nominals*. PhD dissertation, Cornell University.

Toribio, A.J. (1993b): Lexical subjects in finite and non-finite clauses. *Cornell Working Papers in Linguistics*. Ithaca, New York: Cornell University.

Toribio, A.J. (1996): Dialectal variation in the licensing of null referential and expletive pronouns. In: Parodi, C., Quicoil, C. and Saltarelli. M. (eds.): *Aspects of Romance linguistics*, 409-432. Washington, DC: Georgetown University Press.

Toribio, A.J. (2000a): Setting parametric limits on dialectal variation in spanish. *Lingua: International Review of General Linguistics* 110, 315-341.

Toribio, A.J. (2000b): Language variation and the linguistic enactment of identity among Dominicans. *Linguistics* 38, 1133-1159.

Toribio, A.J. (2001): Minimalist ideas on parametric variation. In: Hirotani, M., Coetzee, A., Hall, N. and Kim, J. (eds.): *North East Linguistic Society* 30, 627-638. Amherst, MA: University of Massachusetts.

Toribio, A.J. (2002): Focus on clefts in Dominican Spanish. In: Lee, J.F., Geeslin, K.L. and Clements, J.C. (eds): *Structure, meaning, and acquisition in Spanish*, 130-146. Somerville: Cascadilla Press.

Toribio, A.J. (2006): The Revitalization of Hispanic Linguistics. *Hispanic Issues On Line* 1, 133-136.

Torrego, E. (1984): On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15, 103-29.

Torrego, E. (1998): Nominative subjects and *pro*-drop Infl. *Syntax* I.2, 206-19.

Torres Morais, M.A. (1994): *Do Português Clássico ao Português Moderno: um estudo da cliticização e do movimento do verbo*. Phd dissertation. Campinas, SP: UNICAMP.

Tortora, C. (2003): *The Syntax of Italian Dialects*. New York: Oxford University Press.

Traugott, E. and Heine, B. (1991): Introduction. In: Traugott, E.C. and Heine, B. (eds.): *Approaches to grammaticalization* Vol. 1, 1-14. Amsterdam: John Benjamins.

Travis, C. (2005): The yo-yo effect: Priming in subject expression in Colombian Spanish. In: Gess, R.S. and Rubin, E.J. (eds.): *Selected papers from the 34th Symposium on Romance Languages (LSRL)*, 329-349. Amsterdam: John Benjamins.

Travis, C. (2007): Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language variation and change* 19, 101-135.

Uriagereka, J. (1995): Aspects of the syntax of clitic placement in Western Romance. *Linguistic Inquiry* 26, 79-123-

Vainikka, A. (1999): *Cross-linguistic structure building*. Series of lectures 'Structure Building in Acquisition' at the University of Durham.

Vainikka, A. and Levy, Y. (1999): Empty subjects in Finnish and Hebrew. *Natural Language and Linguistic Theory* 17, 613-71.

Vallduví, E. (1990): *The informational component*. PhD dissertation. University of Pennylvania.

Vance, B. (1989): *Null subjects and syntactic change in Medieval French*. PhD dissertation. Cornell University.

Van Riemsdijk, H. (1990): Funtional prepositions. In: Pinkster, H. and Gene, I. (eds): *Unity in Diversity*, 229-241. Dordrecht: Foris.

Weber, E.G., and Bentivoglio, P. (1991): Verbs of Cognition in Spoken Spanish: A Discourse Profile. In: Fleischman, S. and Waugh, L.R. (eds.): *Discourse Pragmatics and the verb: The evidence from Romance*, 194-213. London/New York: Routledge.

Weinrich, U., Labov, W. and Herzog, M. (1968): Empirical foundations for a theory of language change. In: Lehman, W. and Malkiel, Y. (eds.): *Directions for historical linguistics*, 97-195. Austin: University of Texas Press.

Wilson, J. and Henry, A. (1998): Parameter setting within a socially realistic linguistics. *Language in Society* 27, 1-21.

Winford, D. (1996): The problem of syntactic variation. In: Arnold, J., Blake, R., Davidson, B., Schwenter, S. and Solomon, J. (eds.): *Sociolinguistic Variation: Data, Theory and Analysis*, 177-192. Stanford: CSLI.

Wolfram, W. (1999): Principles of donor dialect attribution. Paper presented at *Methods X*, Memorial University of Newfoundland, Saint John's, Newfoundland, Canada.

Wolfram, W. (2001): From definition to policy: The ideological struggle of African-American English. In: Alatis, J.E. and Tan, A. (eds.): *Round table on language and linguistics 1999*, 292-313. Washington, DC: Georgetown University Press.

Yoon, J. and Bonet-Farran, N. (1991): The ambivalent nature of Spanish infinitives. In: Wanner, D. and Kibbee, D.A. (eds.): *New analyses in Romance languages: selected papers from the XVIII Linguistic Symposium on Romance Languages.* Amsterdam: John Benjamins.

Young, R. and Bayley, R. (1996): VARBRUL analysis for second language acquisition research. In: Bayley, R. and Preston, D. (eds.): *Second language acquisition and linguistic variation*, 253-306. Amsterdam: John Benjamins.

Zanuttini, R. (1997): *Negation and Clausal Structure: A Comparative Study of Romance Languages*. Oxford University Press.

Zeijlstra, H. (2009): Dislocation Effects, Uninterpretable Features, Functional Heads, and Parametric Variation: Consequences of Conflicting Interface Conditions. In: Grohmann, K. (ed.): *Interhases. Phase-Theoretic Investigations of Linguistic Interfaces*. Oxford: Oxford University Press.

Zobl, H. and Liceras, J.M. (2005): Accounting for Optionality in Nonnative Grammars: Parametric Change in Diachrony and L2 Development as Instances of Internalized Diglossia. In: Dekydtspotter, L., Sprouse, R.A. and Liljestrand, A. (eds.): *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference* (GASLA 2004), 283-291. MA: Cascadilla Press.

Zubizarreta, M.L. (1998): *Prosody, focus and word order*. Cambridge, MA: MIT Press.

Zubizarreta, M.L. (1999): Funciones informativas: tema y foco. In: Bosque, I. and Demonte, V. (eds.): *Gramática descriptiva de la lengua española* Vol. 3, 4215-4244. Madrid: Real Academia Española y Espasa Calpe.

Zubizarreta, M.L. (2000): The Clitic Projection in Questions, *Cat WPL* 7, 253-277.